



Cite this: *RSC Adv.*, 2019, 9, 8714

# Hybrid multidimensional data acquisition and data processing strategy for comprehensive characterization of known, unknown and isomeric compounds from the compound Dan Zhi Tablet by UPLC-TWIMS-QTOFMS†

Taofang Cheng,<sup>‡a</sup> Ji Ye,<sup>‡b</sup> Huiliang Li,<sup>b</sup> Hongyuan Dong,<sup>b</sup> Ning Xie,<sup>c</sup> Nan Mi,<sup>b</sup> Zhen Zhang,<sup>b</sup> Jingtao Zou,<sup>d</sup> Huizi Jin <sup>\*a</sup> and Weidong Zhang<sup>\*ab</sup>

The compound Dan Zhi Tablet (DZT), a reputable traditional Chinese medicine prescription, is widely used for the treatment of ischemic stroke in clinic. However, its systematic chemical constituents have rarely been elucidated, which hampers its quality evaluation, the study of bioactive constituents and the mechanism of action interpretation. In this study, we developed a combination of multidimensional data acquisition and data processing strategy with the aim to globally and comprehensively identify the chemical constituents in DZT based on UPLC-TWIMS-QTOFMS. First, multidimensional acquisition modes ( $MS^E$ , Fast DDA and  $HDMS^E$ ) were performed on UPLC-TWIMS-QTOFMS. Second, targeted characterizations of the known compounds and their analogues present in DZT were carried out on the basis of the corresponding commercial standards or Mass2Motifs. Third, untargeted identification of unknown compounds in DZT was performed by extracting shared Mass2Motifs from the raw fragmentation spectra. Finally, the coeluting isomers were characterized using a precursor and/or product ion mobility. Consequently, 202 compounds were detected from DZT: 29 of them were unambiguously identified by comparison with reference compounds, 29 unknown compounds were discovered in specific medicinal materials, and ten pairs of coeluting isomers, which could not be distinguished using conventional  $MS^E$  or Fast-DDA, were resolved using  $HDMS^E$  only. This strategy was successfully used for the rapid and global identification of complex compounds including known, unknown and coeluting isomeric compounds in DZT and provided helpful chemical information for further quality control, pharmacology and active mechanism research on DZT.

Received 9th December 2018  
 Accepted 1st March 2019

DOI: 10.1039/c8ra10100k

[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## Introduction

Compound formulae are the main usage form of traditional Chinese medicine (TCM) and have been used for the treatment of complex diseases in China in clinical practice for thousands of years.<sup>1–4</sup> Typically, a TCM formula, which consists of several types of medicinal herbs or animals, can amplify the therapeutic efficacies of each ingredient, leading to enhanced therapeutic efficacy or reduced adverse effects. Considering the

chemical diversity and complexity, the development of a rapid and reliable analytical method for the identification of the constituents of TCM formulae is of great significance to enhance quality control in clinical applications.

In this regard, ultra-performance liquid chromatography coupled with traveling wave ion mobility quadrupole time-of-flight mass spectrometry (UPLC-TWIMS-QTOFMS)<sup>5</sup> is a promising tool well-suited for the high-throughput identification of the constituents from TCM formulae due to its support for multidimensional acquisition modes such as data-independent acquisition ( $MS^E$ ),<sup>6–8</sup> fast data-directed analysis (Fast-DDA)<sup>8,9</sup> and high-definition  $MS^E$  ( $HDMS^E$ ).<sup>10–14</sup> The  $MS^E$  mode is applied to increase the coverage of observable chemical compounds; the Fast DDA mode can obtain cleaner precursor-product links and greatly ease data processing for the identification of compounds, especially unknown compounds. Considering that there are some compounds that cannot be absolutely separated and eluted at the same retention time by UPLC, the ion mobility mode ( $HDMS^E$ ) is applied to separate the coeluting compounds,

<sup>a</sup>School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: [kimhz@sjtu.edu.cn](mailto:kimhz@sjtu.edu.cn); [wdzhangy@hotmail.com](mailto:wdzhangy@hotmail.com); Fax: +86-21-34205989; Tel: +86-21-34205989

<sup>b</sup>School of Pharmacy, Second Military Medical University, Shanghai, 200433, China

<sup>c</sup>State Key Laboratory of Innovative Natural Medicine and TCM Injections, Jiangxi Qingfeng Pharmaceutical Co., Ltd., Ganzhou 341000, China

<sup>d</sup>Tonghua Huaxia Pharmaceutical Co., Ltd., Tonghua, 134100, China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8ra10100k

‡ These authors contributed equally to this work.



especially coeluting isomers for a second time according to their different charge states, shapes and sizes. However, the identification of the chemical constituents in TCM formulae is still a great challenge due to the presence of unknown compounds. Although many data processing strategies such as a key ion filter,<sup>15</sup> mass spectral trees similarity filter,<sup>16</sup> diagnostic ion filter,<sup>17</sup> and mass defect filter<sup>18</sup> can improve the efficiency of identification and decrease the rate of false positives, these strategies were established to analyze the targeted components according to prior knowledge<sup>19,20</sup> and they do not reflect the complexity of the chemical constituents in TCM formulae. It is well-known that fragmentation spectra are replete with abundant information pertaining to chemical substructures and have recurring fragment ions and losses. Therefore, Mass2Motifs containing co-occurring MS/MS fragment ions and neutral losses can be extracted and obtained from the raw fragmentation spectra to discover unknown compounds and putatively support *de novo* structure elucidation.<sup>21–23</sup> This method can be used to identify unknown compounds without prior knowledge about the fragments of interest.

The compound Dan Zhi Tablet (DZT) is a reputable traditional Chinese medicine (TCM) formula that consists of *Astragalus membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao (AM), *Salvia miltiorrhiza* Bge. (SM), *Pheretima aspergillum* (E. Perrier) (PA), *Hirudo nipponica* Whitman (HN) and *Ligusticum Chuanxiong* Hort. (LC) with a ratio of 15 : 12 : 10 : 3 : 6. DZT is capable of tonifying or regulating Qi and activating blood or resolving blood stasis and is widely utilized for the treatment of cerebrovascular diseases such as ischemic stroke. The data from clinical randomized controlled trials have illustrated that the treatment of DZT for stroke recovery is efficient and safe.<sup>24</sup> To the best of our knowledge, to date, systematic and global chemical information of DZT remains unknown. In our previous work, four major groups of absorbed prototype constituents (tanshinones, flavonoids, phthalides and furan sulfonic acids) and their metabolites were detected in the plasma and cerebrospinal fluid of rats after the oral administration of DZT by UPLC-QTOF-MS.<sup>25</sup> However, some important active ingredients such as the salvianolic acids<sup>26–30</sup> in SM having antiplatelet activity and astragalosides<sup>31</sup> in AM showing neuroprotection and other unknown constituents as well as coeluting isomers are still unclear, which seriously hampers the control of quality and the interpretation of the underlying pharmacology and mechanism of DZT.

In the present work, a combination of multidimensional data acquisition and data processing strategy was proposed to comprehensively characterize the multi-components including the known compounds, unknown compounds and coeluting isomers of DZT using UPLC-TWIMS-QTOFMS (Fig. 1). First, data acquisition was conducted using multidimensional modes including MS<sup>E</sup>, Fast DDA and HDMS<sup>E</sup>. Second, targeted identification of known compounds and their unknown analogues in DZT was performed based on standard references or Mass2Motifs. Third, untargeted discovery and characterization of unknown components in DZT were performed according to Mass2Motifs. Finally, the coeluting isomers in DZT were

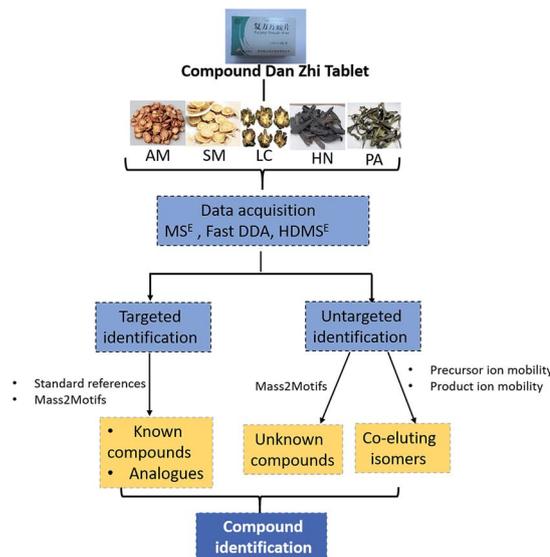


Fig. 1 The flowchart for the analysis of complex chemical compounds from DZT.

distinguished using a precursor and/or product ion mobility. With the established method, a total of 202 compounds were detected from DZT, and twenty-nine unknown compounds of them were discovered in specific medicinal materials. Moreover, ten pairs of coeluting isomers, which could not be distinguished using conventional MS<sup>E</sup> or Fast-DDA, were resolved using HDMS<sup>E</sup> only. This is the first research on the chemical analysis of DZT, which will be beneficial to its quality control and is essential to deeply understand the mechanism of DZT for the treatment of ischemic stroke.

## Experimental

### Materials and reagents

Twenty-four authentic standards of adenosine (12), caffeic acid (17b), danshensu (17c), protocatechuic aldehyde (21), *trans*-5-caffeoylquinic acid (23b), ferulic acid (30), lithospermic acid (45b), salvianolic acid B (48a), baicalin (50a), ononin (50b), salvianolic acid A (61a), salvianolic acid C (70), astragaloside IV (77b), astragaloside III (79b), astragaloside II (83d), farrerol (84b), isoastragaloside II (85b), wogonin (86a), astragaloside I (92b), cyclosiiversigenin (97b), dihydrotanshinone I (102a), tanshinone I (112), levistolide A (121) and tanshinone IIA (123) were all purchased from the National Institute for the Control of Pharmaceutical and Biological Products (Beijing, China). Five reference standards of calycosin 7-O- $\beta$ -D-glucoside (28b), rosmarinic acid (44), calycosin (62a), formononetin (80) and cryptotanshinone (130) were obtained from Shanghai Yuanye Bio-Technology Co., Ltd. (Shanghai, China). The structures of these reference standards (purity > 98%) are shown in Fig. 2. They were stored at 4 °C before use.

Acetonitrile, formic acid and water of LC-MS grade were purchased from Fisher Scientific Company (USA). Methanol of analytical grade was obtained from Runjie Tech Co. (Shanghai, China).



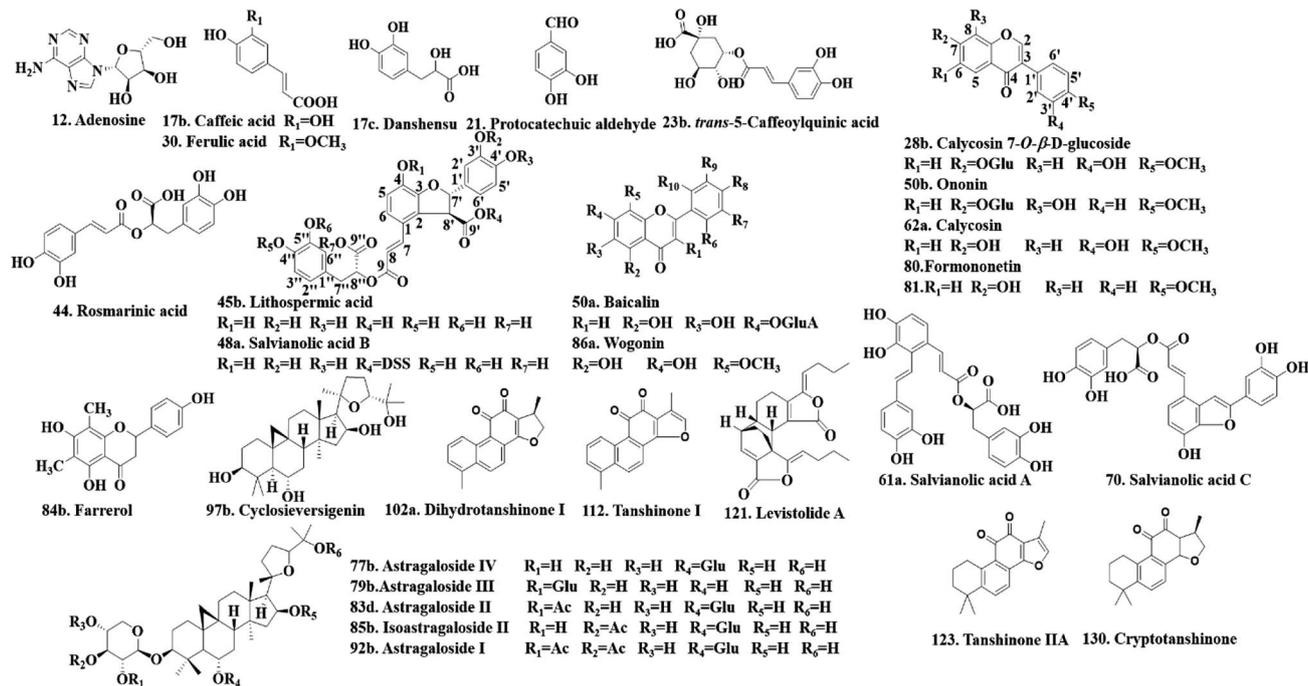


Fig. 2 The chemical structures of the 29 reference standards.

DZT and its five ingredients, *i.e.*, LC, SM, PA, AM, and HN were provided by Tonghua Baishan Pharmaceutical Co., Ltd. (Jilin, China).

### Standard solution preparation

The abovementioned twenty-nine reference compounds were prepared by dissolving each authentic compound in methanol at a concentration of  $1 \text{ mg mL}^{-1}$ . A certain amount of the aforementioned reference compound stock solution was mixed and diluted with methanol to obtain the reference compound mixture solution (about  $30 \text{ }\mu\text{g mL}^{-1}$  for each compound). This solution was centrifuged at 13 000 rpm for 10 min and stored at  $4 \text{ }^\circ\text{C}$  before UPLC-QTOFMS analysis.

### DZT sample preparation

DZT powder (3 g) was extracted with methanol three times ( $3 \times 30 \text{ mL}$ , 1 hour each) using ultrasonic extraction. The extracted solutions were merged and diluted 10-fold with methanol. Then, a certain amount of the diluted solution was centrifuged at 13 000 rpm for 10 min to obtain the sample solutions.

### Multidimensional data acquisition

The analysis of the DZT samples was conducted on a Waters Acquity UPLC system (Waters, Milford, MA, USA) consisting of a binary pump and an autosampler. UPLC separation was performed on an Agilent ZORBAX Eclipse Plus C18 column ( $2.1 \text{ mm ID} \times 150 \text{ mm}$ ,  $1.8 \text{ }\mu\text{m}$ , Agilent Technologies, Santa Clara, California, USA). The mobile phase contained 0.1% formic acid (vol/vol) in water (A) and acetonitrile (B). A gradient elution program was set as follows: 0.0–2.0 min, 1% B; 2.0–4.0 min, 1–8% B; 4.0–10.0 min, 8–28% B; 10.0–12.0 min, 28–30% B; 12.0–13.0 min, 30–

36% B, and 13.0–24.0 min, 36–95% B. The elution rate was set at  $0.4 \text{ mL min}^{-1}$ . The column and autosampler temperatures were maintained at  $35 \text{ }^\circ\text{C}$  and  $10 \text{ }^\circ\text{C}$ , respectively.

Mass detection was performed on SYNAPT G2-Si HDMS (Waters Corp., Manchester, UK) equipped with an ESI interface. MS data acquisition was separately performed in positive and negative ion modes. The  $\text{MS}^E$  parameters were set as follows: capillary voltage, 2.5 kV in positive mode and 2.0 kV in negative mode; cone voltage, 40 eV; source offset voltage, 80 eV; source temperature,  $100 \text{ }^\circ\text{C}$ ; desolvation temperature,  $400 \text{ }^\circ\text{C}$ ; cone gas flow rate,  $50 \text{ L h}^{-1}$ ; desolvation gas flow rate,  $800 \text{ L h}^{-1}$ ; collision energy, 6 eV for low energy function and 6–90 eV ramp for high energy function; mass range, 50–1500 Da; and scan time, 0.2 s. Data were calibrated using an external reference (LockSpray<sup>TM</sup>) by the constant infusion of a leucine-enkephalin solution ( $m/z$  556.2771 in positive ion mode) at a flow rate of  $5 \text{ }\mu\text{L min}^{-1}$ . The Fast DDA parameters were set as follows: dual-dynamic collision energies, 10–40 V for low-mass collision energy and 40–90 V for high-mass collision energy; automatic switching to the MS/MS mode when the intensity of TIC rose above 3000 intensity per s and automatic switching off when 0.6 s elapsed; deisotope peak detection mode,  $\pm 3.0 \text{ Da}$  for the tolerance window and 7.0 Da for the peak extract window. The other parameters for Fast DDA were the same as those for  $\text{MS}^E$ . The MassLynx V4.1 software was used for data acquisition, and UNIFI 1.8 was used for the processing and visualization of the multidimensional mass spectrometry data.

## Results and discussion

Fig. 3 displays the representative base peak chromatograms (BPCs) of DZT both in positive and negative ion modes. In the



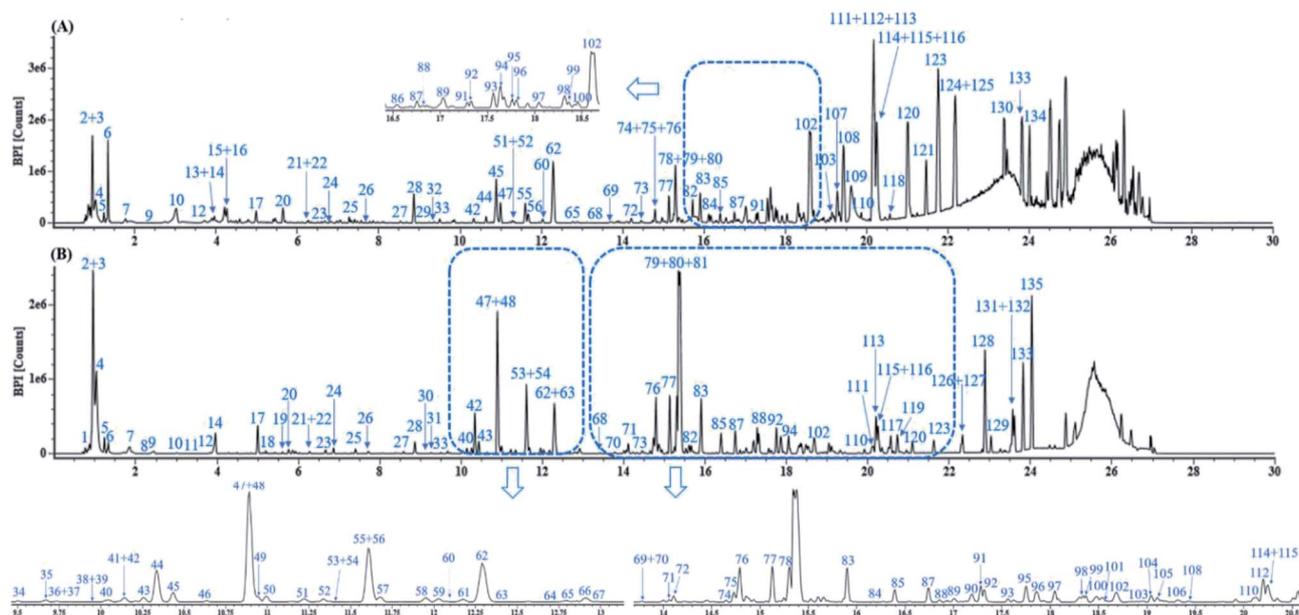


Fig. 3 The representative base peak chromatograms (BPCs) of DZT both in positive (A) and negative ion modes (B).

present work, 202 compounds were detected in DZT, of which 29 compounds were positively identified by the analysis of the corresponding reference standards and 173 compounds were putatively identified based on previous literatures and the shared Mass2Motifs extracted from the raw MS/MS spectrum of DZT or ion mobility technology. All the detected Mass2Motifs in DZT are shown in Table 1. Moreover, 29 unknown compounds, namely, two salvianolic acids (peaks 39 and 41), four astragalosides (peaks 67a, 69a, 83c and 84c), two furan sulfonic acids (peaks 62b and 73a), seven carbohydrate derivatives (peaks 10, 11, 13, 14a, 15a, 16 and 46), and fourteen lysoglycerophosphocholines (peaks 97d, 98a, 98b, 103c, 103d, 106b, 108b, 110, 111, 113, 115b, 117, 119 and 127) were detected in DZT. Furthermore, 10 pairs of coeluting isomers were tentatively assigned using a precursor and/or product ion mobility. Their structures and detailed mass spectra information, including peak number, retention times ( $R_t$ ), formulas, accurate molecular ions, mass measurement errors (within a  $\pm 5$  ppm window), fragment ions, drift time ( $d_t$ ) values, compound names and original sources, are shown in Fig. S1 and Table S1.†

### Targeted characterization of known compounds and their unknown analogues from DZT

**Identification of salvianolic acids in DZT.** Salvianolic acids are the major components in DZT, and they can be divided into monomers and polymers (dimers, trimers and tetramers) according to the number of phenyls in their structures. The fragment ions of the monomers of salvianolic acids were found by the loss of  $\text{CO}_2$ ,  $\text{CO}$  and  $\text{H}_2\text{O}$ ; the fragment ions of the polymers of salvianolic acids were produced by the successive or simultaneous loss of  $\text{C}_9\text{H}_6\text{O}_3$  (162.0317 Da),  $\text{C}_9\text{H}_8\text{O}_4$  (180.0423 Da) and  $\text{C}_9\text{H}_{10}\text{O}_5$  (198.0528 Da), corresponding to caffeoyl, caffeic acid (CA) and danshensu (DSS). Herein, compound 48a ( $R_t = 10.91$  min) was taken as an example to elucidate the

detailed fragmentation pathway of the tetramers. As shown in Fig. 4A, it forms the  $[\text{M} - \text{H}]^-$  ion at  $m/z$  717.1460 and then yields the fragment ions at  $m/z$  519.0956, 339.0520, 321.0414, 295.0613 and 277.0499, corresponding to  $[\text{M} - \text{H} - \text{DSS}]^-$ ,  $[\text{M} - \text{H} - \text{DSS} - \text{CA}]^-$ ,  $[\text{M} - \text{H} - 2\text{DSS}]^-$ ,  $[\text{M} - \text{H} - \text{DSS} - \text{CA} - \text{CO}_2]^-$  and  $[\text{M} - \text{H} - \text{DSS} - \text{CA} - \text{CO}_2 - \text{H}_2\text{O}]^-$ . Compound 48a was unambiguously identified as salvianolic acid B by comparison with its reference standard. Peak 37c ( $m/z$  699.1552) with the molecular formula  $\text{C}_{33}\text{H}_{32}\text{O}_{17}$  showed its fragment ions at  $m/z$  655.1628, 519.0926, 457.1106, 321.0393, 295.0598 and 277.0505, corresponding to  $[\text{M} - \text{H} - \text{CO}_2]^-$ ,  $[\text{M} - \text{H} - \text{CA}]^-$ ,  $[\text{M} - \text{H} - \text{CO}_2 - \text{DSS}]^-$ ,  $[\text{M} - \text{H} - \text{CA} - \text{DSS}]^-$ ,  $[\text{M} - \text{H} - \text{CO}_2 - \text{DSS} - (\text{Glu} - \text{H}_2\text{O})]^-$  and  $[\text{M} - \text{H} - \text{CO}_2 - \text{DSS} - \text{Glu}]^-$ . Peak 37c was assigned as lithospermate-9'-*O*-glucoside by comparison with a previous result.<sup>32</sup> Three isomers (peaks 38a, 39 and 41) showed  $[\text{M} - \text{H}]^-$  at  $m/z$  879.1994 with the molecular formula  $\text{C}_{42}\text{H}_{40}\text{O}_{21}$  and presented masses 180 Da higher than that of lithospermate-9'-*O*-glucoside, suggesting the presence of a caffeic acid group in their structures. Among them, peak 38a was tentatively identified as salvinoside.<sup>32</sup> Although the fragmentation patterns did not allow the elucidation of the structures of peaks 39 and 41, they were hypothesized as two isomers of salvinoside with a caffeic acid group at different positions. Peaks 39 and 41 were potentially new compounds.

A total of 35 salvianolic acids were identified, of which five compounds (danshensu, lithospermic acid, salvianolic acid A, salvianolic acid B and rosmarinic acid) were unambiguously identified by comparing their retention times, accurate molecular masses and MS/MS fragments with those of the standard references; moreover, the remaining 30 compounds were tentatively identified by comparing their exact molecular masses, fragmentation information, retention times and drift times with previously reported values.<sup>32-34</sup>

**Identification of astragalosides in DZT.** Most of the astragalosides contain the same tetracyclic triterpenoids



Table 1 Mass2Motifs of DZT in negative ion mode<sup>a</sup>

No.	Mass2Motifs	Characterization
1	Frag 191.0553 (C <sub>7</sub> H <sub>11</sub> O <sub>6</sub> ), 173.0450 (C <sub>7</sub> H <sub>9</sub> O <sub>5</sub> )	Fragments related to quinic acid
2	Frag 341.1082 (C <sub>12</sub> H <sub>21</sub> O <sub>11</sub> ), 281.0878 (C <sub>10</sub> H <sub>17</sub> O <sub>9</sub> ), 221.0660 (C <sub>8</sub> H <sub>13</sub> O <sub>7</sub> ), 179.0552 (C <sub>6</sub> H <sub>11</sub> O <sub>6</sub> ), 161.0445 (C <sub>6</sub> H <sub>9</sub> O <sub>5</sub> ), 119.0342 (C <sub>4</sub> H <sub>7</sub> O <sub>4</sub> )	Fragments related to disaccharide
3	Frag 545.1732 (C <sub>20</sub> H <sub>33</sub> O <sub>17</sub> ), 383.1190 (C <sub>14</sub> H <sub>23</sub> O <sub>12</sub> ), 221.0660 (C <sub>8</sub> H <sub>13</sub> O <sub>7</sub> ), 179.0552 (C <sub>6</sub> H <sub>11</sub> O <sub>6</sub> ), 161.0445 (C <sub>6</sub> H <sub>9</sub> O <sub>5</sub> ), 119.0342 (C <sub>4</sub> H <sub>7</sub> O <sub>4</sub> )	Fragments related to polysaccharides
4	NL 150.0528 (C <sub>5</sub> H <sub>10</sub> O <sub>5</sub> ), 132.0423 (C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> )	Pentose (xylose/arabinose/ribose) related loss
5	NL 180.0634 (C <sub>6</sub> H <sub>12</sub> O <sub>6</sub> ), 162.0528 (C <sub>6</sub> H <sub>10</sub> O <sub>5</sub> )	Hexose (glucose/galactose) related loss
6	NL 164.0685 (C <sub>6</sub> H <sub>12</sub> O <sub>5</sub> ), 146.0603 (C <sub>6</sub> H <sub>10</sub> O <sub>4</sub> )	Rhamnose-related loss
7	NL 17.0267 (NH <sub>3</sub> )	Amino acid-related loss
8	NL 42.0102 (C <sub>2</sub> H <sub>2</sub> O)	Acetyl group
9	NL 100.0162 (C <sub>4</sub> H <sub>4</sub> O <sub>3</sub> )	Isopropyl group
10	NL 15.0235 (CH <sub>3</sub> )	Methyl group
11	NL 18.0106 (H <sub>2</sub> O)	Loss of water
12	NL 27.9949 (CO)	Loss of C=O
13	NL 43.9898 (CO <sub>2</sub> )	Carboxylic acid (CO <sub>2</sub> ) group
14	Frag 179.0340 (C <sub>9</sub> H <sub>7</sub> O <sub>4</sub> ), 161.0231 (C <sub>9</sub> H <sub>5</sub> O <sub>3</sub> ), 135.0446 (C <sub>8</sub> H <sub>7</sub> O <sub>2</sub> )	Fragments related to caffeic acid
15	NL 180.0423 (C <sub>9</sub> H <sub>8</sub> O <sub>4</sub> )	Caffeic acid/salvianolic acid-related loss
16	NL 162.0317 (C <sub>9</sub> H <sub>6</sub> O <sub>3</sub> )	Caffeoyl group/salvianolic acid-related loss
17	Frag 197.0459 (C <sub>9</sub> H <sub>9</sub> O <sub>5</sub> ), 179.0340 (C <sub>9</sub> H <sub>7</sub> O <sub>4</sub> ), 135.0439 (C <sub>8</sub> H <sub>7</sub> O <sub>2</sub> )	Fragments related to danshensu
18	NL 198.0528 (C <sub>9</sub> H <sub>10</sub> O <sub>5</sub> )	Danshensu/salvianolic acid-related loss
19	Frag 519.0927 (C <sub>27</sub> H <sub>19</sub> O <sub>11</sub> ), 339.0505 (C <sub>18</sub> H <sub>11</sub> O <sub>7</sub> ), 321.0396 (C <sub>18</sub> H <sub>9</sub> O <sub>6</sub> ), 295.0610 (C <sub>17</sub> H <sub>11</sub> O <sub>5</sub> ), 197.0449 (C <sub>9</sub> H <sub>9</sub> O <sub>5</sub> ), 179.0335 (C <sub>9</sub> H <sub>7</sub> O <sub>4</sub> ), 161.0231 (C <sub>9</sub> H <sub>5</sub> O <sub>3</sub> )	Fragments related to salvianolic acids
20	Frag 353.0875 (C <sub>16</sub> H <sub>17</sub> O <sub>9</sub> ), 191.0553 (C <sub>7</sub> H <sub>11</sub> O <sub>6</sub> ), 179.0340 (C <sub>9</sub> H <sub>7</sub> O <sub>4</sub> ), 173.0445 (C <sub>7</sub> H <sub>9</sub> O <sub>5</sub> ), 135.0439 (C <sub>8</sub> H <sub>7</sub> O <sub>2</sub> )	Fragments related to caffeoylquinic acids
21	Frag 515.1190 (C <sub>25</sub> H <sub>23</sub> O <sub>12</sub> ), 353.0875 (C <sub>16</sub> H <sub>17</sub> O <sub>9</sub> )	Fragments related to dicaffeoylquinic acids
22 <sup>a</sup>	Frag 455.3525 (C <sub>30</sub> H <sub>47</sub> O <sub>3</sub> ), 437.3420 (C <sub>30</sub> H <sub>45</sub> O <sub>2</sub> ), 419.3314 (C <sub>30</sub> H <sub>43</sub> O)	Fragments related to astragaloside
23	NL 180.0634 (C <sub>6</sub> H <sub>12</sub> O <sub>6</sub> ), 162.0538 (C <sub>6</sub> H <sub>10</sub> O <sub>5</sub> ), 146.0579 (C <sub>6</sub> H <sub>10</sub> O <sub>4</sub> ), 132.0430 (C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> )	Astragaloside-related loss
24 <sup>a</sup>	Frag 184.0731 (C <sub>5</sub> H <sub>15</sub> NO <sub>4</sub> P), 124.9999 (C <sub>2</sub> H <sub>6</sub> O <sub>4</sub> P), 104.1068 (C <sub>5</sub> H <sub>14</sub> NO), 86.0961 (C <sub>5</sub> H <sub>12</sub> N)	Fragments related to lysoglycerophosphocholines
25	Frag 242.0791 (C <sub>7</sub> H <sub>17</sub> NO <sub>6</sub> P), 224.0693 (C <sub>7</sub> H <sub>15</sub> NO <sub>5</sub> P), 168.0426 (C <sub>4</sub> H <sub>11</sub> NO <sub>4</sub> P), 78.9585 (PO <sub>3</sub> )	
26	NL 60.0220 (CH <sub>3</sub> +CO <sub>2</sub> H)	Lysoglycerophosphocholine-related loss
27	Frag 96.9595 (HSO <sub>4</sub> )	Sulphate group
28	Frag 79.9562 (SO <sub>3</sub> )	Sulfate group
29 <sup>a</sup>	NL 18.0106 (H <sub>2</sub> O), 27.9949 (CO), 28.0310 (C <sub>2</sub> H <sub>4</sub> ), 42.0470 (C <sub>3</sub> H <sub>6</sub> ), 56.0630 (C <sub>4</sub> H <sub>8</sub> )	Loss of phthalides



Table 1 (Contd.)

No.	Mass2Motifs	Characterization	
30	NL	15.0235 (CH <sub>3</sub> ), 18.0106 (H <sub>2</sub> O), 29.0391 (C <sub>2</sub> H <sub>5</sub> ), 58.0780 (C <sub>4</sub> H <sub>10</sub> ), 71.0860 (C <sub>5</sub> H <sub>11</sub> ), 86.1096 (C <sub>6</sub> H <sub>14</sub> )	Loss of furan sulfonic acids
	Frag	79.9568 (SO <sub>3</sub> )	Fragment related to furan sulfonic acids
31	NL	157.1592 (C <sub>10</sub> H <sub>21</sub> O), 141.1643 (C <sub>10</sub> H <sub>21</sub> ), 86.1096 (C <sub>6</sub> H <sub>14</sub> ), 72.0939 (C <sub>5</sub> H <sub>12</sub> )	Loss of alkyl sulfuric acids
	Frag	96.9595 (HSO <sub>4</sub> ), 79.9562 (SO <sub>3</sub> ), 165.0217 (C <sub>5</sub> H <sub>9</sub> O <sub>4</sub> S)	Fragments related to alkyl sulfuric acids

<sup>a</sup> a: Mass2Motifs in positive ion mode; Frag: fragment; NL: neutral loss.

Astragalosides from Radix Astragali were found in DZT. Most of the astragalosides contain the same tetracyclic triterpenoid aglycone and one or several sugar moieties such as xylose (Xyl), arabinose (Ara), rhamnose (Rha), glucose (Glu) and/or galactose (Gal), which can link to the aglycone at C-3, C-6 and/or C-25 via a glucosidic bond. The astragalosides showed a good MS response both in positive and negative ion modes. In the positive ion mode, the astragalosides generated rich characteristic fragment ions at  $m/z$  473.3631 (C<sub>30</sub>H<sub>49</sub>O<sub>4</sub>), 455.3525 (C<sub>30</sub>H<sub>47</sub>O<sub>3</sub>), 437.3420 (C<sub>30</sub>H<sub>45</sub>O<sub>2</sub>), and 419.3314 (C<sub>30</sub>H<sub>43</sub>O), corresponding to the aglycone and successive losses of water molecules; thus, these ions can help us identify as many astragalosides as possible.<sup>35–37</sup> In the negative ion mode, the astragalosides exhibited a precursor ion of [M + HCOOH–H]<sup>–</sup> and rich product ions with respect to the addition of formic acid in the mobile phase, which can be convenient for analyzing the MS/MS fragmentation pathways. As shown in Fig. 4B, astragaloside III (peak 79b,  $R_t$  = 15.26 min) yields the dominant [M + HCOOH–H]<sup>–</sup> at  $m/z$  829.4602 in the negative ion mode and generates several structurally relevant daughter ions at  $m/z$  783.4540, 621.4002 and 489.3618, corresponding to [M – H]<sup>–</sup>, [M–Glu]<sup>–</sup> and [M–H–Glu–Xyl]<sup>–</sup>. Peak 83c showed [M + HCOOH–H]<sup>–</sup> at  $m/z$  957.5070 with the molecular formula of C<sub>47</sub>H<sub>76</sub>O<sub>17</sub> and yielded fragment ions at  $m/z$  749.4493, 603.3890, 585.3736 and 471.3472, corresponding to [M–H–(Glu–H<sub>2</sub>O)]<sup>–</sup>, [M–H–(Glu–H<sub>2</sub>O)–(Rha–H<sub>2</sub>O)]<sup>–</sup>, [M–H–(Glu–H<sub>2</sub>O)–Rha]<sup>–</sup> and [M–H–(Glu–H<sub>2</sub>O)–(Rha–H<sub>2</sub>O)–(Xyl–H<sub>2</sub>O)]<sup>–</sup>. Two isomers (peaks 69a and peak 84c) showed similar fragmentation patterns to that of peak 83c. However, their parent ions [M + HCOOH–H]<sup>–</sup> were  $m/z$  1119.5613 (with molecular formula of C<sub>53</sub>H<sub>86</sub>O<sub>22</sub>) with masses 162 Da greater than that of the compound corresponding to peak 83c, suggesting a glucose or galactose moiety. Peak 67a ( $m/z$  1397.6616) with the molecular formula of C<sub>65</sub>H<sub>106</sub>O<sub>32</sub> presented 324 Da (C<sub>12</sub>H<sub>20</sub>O<sub>10</sub>) greater than peak 69a, suggesting that it might correspond to a sugar chain with two glucoses or galactoses. However, only the fragmentation experiments did not allow for the elucidation of the structures of peaks 67a, 69a, 83c and 84c.

In this work, 21 astragalosides along with one sapogenin were positively or tentatively identified from DZT, and four of

them (peaks 67a, 69a, 83c and 84c from AM) were potentially new compounds.

**Identification of sulfur-containing compounds.** Fourteen sulfur-containing compounds, *i.e.*, peaks 22, 43, 49, 61b, 62b, 66b, 71a, 73a, 79a, 89a, 91, 97a, 100 and 101 were found in DZT. The majority of them except peak 22 were discovered based on the shared fragment ions [SO<sub>3</sub>]<sup>–</sup> at  $m/z$  79.9562. These compounds exhibited a high MS response in the negative ion mode and could be divided into three classes including seven furan sulfonic acids, six alkyl sulfuric acids and one pteridine according to their structural skeletons.

For furan sulfonic acids, the neutral losses were CH<sub>3</sub> (15.0235 Da), H<sub>2</sub>O (18.0106 Da), C<sub>2</sub>H<sub>5</sub> (29.0390 Da), C<sub>4</sub>H<sub>10</sub> (58.0780 Da), C<sub>5</sub>H<sub>11</sub> (71.0860 Da) and C<sub>6</sub>H<sub>14</sub> (86.1090 Da). As shown in Fig. 4C, peak 79a displays its molecular ion [M – H]<sup>–</sup> at  $m/z$  259.1008 and mainly generates product ions at  $m/z$  244.0765,  $m/z$  188.0139,  $m/z$  172.9904,  $m/z$  165.1273,  $m/z$  109.1284 and  $m/z$  79.9564, corresponding to [M–H–CH<sub>3</sub>]<sup>–</sup>, [M–H–C<sub>5</sub>H<sub>11</sub>]<sup>–</sup>, [M–H–C<sub>6</sub>H<sub>14</sub>]<sup>–</sup>, [M–H–CH<sub>3</sub>–SO<sub>3</sub>]<sup>–</sup>, [M–H–C<sub>5</sub>H<sub>11</sub>–SO<sub>3</sub>]<sup>–</sup> and [SO<sub>3</sub>]<sup>–</sup>. Peak 79a can be positively identified as 5-ethyl-2-hexylfuran-3-sulfonic acid by comparing its data with the MS/MS data and the NMR spectra of pure compounds (purity > 98%) isolated in our laboratory. Comparing the MS/MS spectrum with a previously reported result,<sup>38</sup> peaks 61b ( $m/z$  273.0791), 66b ( $m/z$  275.0947) and 71a ( $m/z$  245.0836) were tentatively assigned as 5-ethyl-2-hexanoylfuran-3-sulfonic acid, 5-ethyl-2-(1-hydroxyhexyl) furan-3-sulfonic acid and 5-ethyl-2-pentylfuran-3-sulfonic acid, respectively. Another three isomers (peaks 49, 62b and 73a) of 5-ethyl-2-(1-hydroxyhexyl) furan-3-sulfonic acid were also detected at  $m/z$  275.0948, eluting at 10.95 min, 12.37 min and 14.37 min, respectively. However, their structures could not be confirmed using only the MS/MS data. To the best of our knowledge, two furan sulfonic acids (peaks 62b and 73a) from *Pheretima aspergillum* (E. Perrier) were reported for the first time.

For alkyl sulfuric acids, the fragment ion was  $m/z$  96.9596, corresponding to [HSO<sub>4</sub>]<sup>–</sup>. Two isomers (peaks 89a and 91) with the molecular formula C<sub>10</sub>H<sub>22</sub>O<sub>4</sub>S showed their molecular ions [M – H]<sup>–</sup> at  $m/z$  237.1170 and 237.1156 and then yielded high intensity fragment ions at  $m/z$  96.9596 [HSO<sub>4</sub>]<sup>–</sup> and 79.9570 [SO<sub>3</sub>]<sup>–</sup> by the loss of C<sub>10</sub>H<sub>21</sub> (decyl) and C<sub>10</sub>H<sub>22</sub>O (decyl ether),



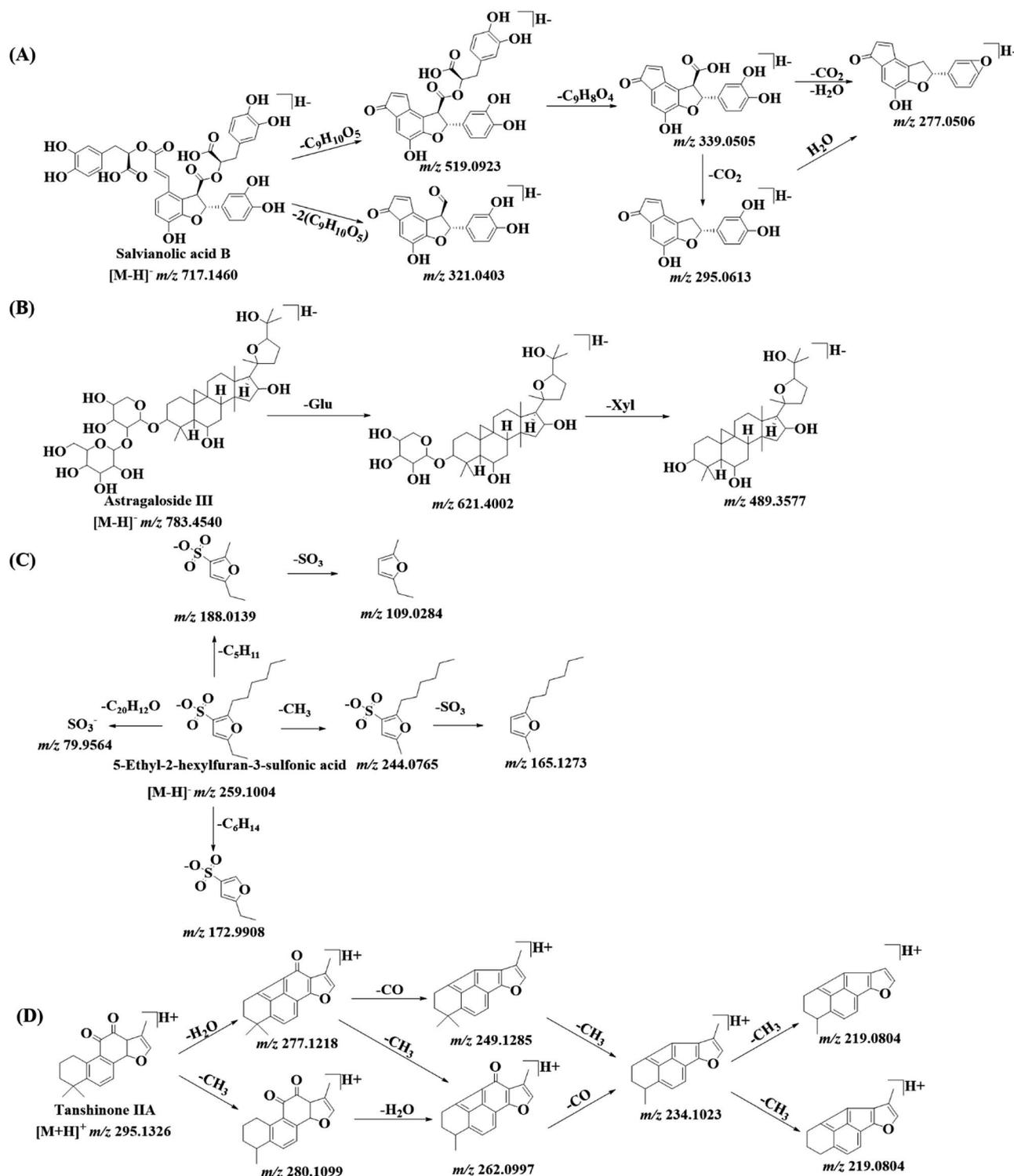


Fig. 4 The fragmentation patterns of salviolic acid B (A), astragaloside III (B), 5-ethyl-2-hexylfuran-3-sulfonic acid (C) and tanshinone IIA (D) in DZT.

respectively. In addition, peak 91 presented an ion at  $m/z$  165.0217, corresponding to the loss of 72.0939 Da ( $C_5H_{12}$ ). Therefore, peak 89a could be tentatively deduced as decyl hydrogen sulfate, and peak 91 might be 5-decanyl hydrogen sulfate. Three isomers (peaks 97a, 100 and 101) were detected at  $m/z$  251.1313 ( $C_{11}H_{24}O_4S$ ) and eluted at 18.02 min, 18.45 min

and 18.58 min, respectively. Peak 100 presented the same fragment ions as that of peak 89a and might be tentatively identified as undecyl sulfuric acid. Peaks 97a and 101 also showed the same fragment ions at  $m/z$  165.0217 as peak 91, corresponding to the loss of 86 Da ( $C_6H_{14}$ ), and they might be putatively assigned as two isomers of undecyl sulfuric acid. Peak



43 ( $m/z$  207.0702  $[M - H]^-$ ) matched the molecular formula  $C_8H_{16}O_4S$  and showed its fragment ion at  $m/z$  79.9562. Peak 43 was a potentially unknown compound. In addition, peak 22 was tentatively assigned as hirudonucleodisulfide B according to a previous literature report.<sup>39</sup>

**Identification of tanshinones in DZT.** In this study, tanshinones could generate  $[M + H]^+$  ions with remarkably higher abundance in the positive ion mode. The characteristic fragment ions of tanshinones were traced by the simultaneous or successive loss of  $CH_3$ ,  $H_2O$ ,  $CO$  or  $CO_2$ . As shown in Fig. 4D, tanshinone IIA (peak 123,  $R_t = 21.76$  min) exhibits the  $[M + H]^+$  ion at  $m/z$  295.1326, and the product ions in its MS/MS spectra are detected at  $m/z$  280.1099, 277.1218, 262.0997, 249.1285, 234.1023, and 219.0804, corresponding to  $[M + H - CH_3]^+$ ,  $[M + H - H_2O]^+$ ,  $[M + H - CH_3 - H_2O]^+$ ,  $[M + H - H_2O - CO]^+$ ,  $[M + H - H_2O - CO - CH_3]^+$ , and  $[M + H - H_2O - CO - 2CH_3]^+$ .

In total, 29 tanshinones in DZT were observed, of which four tanshinones (peaks 102a, 112, 123 and 130) were unequivocally identified by comparing their retention times and MS/MS spectra data with those of authenticated compounds; the remaining 25 compounds were putatively identified by comparing the MS/MS spectral data with previously reported values.<sup>32,40–43</sup>

**Identification of phthalides in DZT.** Phthalides are the main active constituents in the rhizome of *Ligusticum chuanxiong*, and they can be divided into two types including phthalide monomers and phthalide dimers.<sup>44–47</sup> Phthalide monomers can be further classified into two types, *i.e.*, alkyl phthalide and hydroxyl phthalide compounds considering the different types, numbers and positions of the substituents of the mother skeleton. The alkyl phthalide compounds comprise a saturated *n*-butyl and unsaturated butenyl substituent at the C-3 position. Thus, the characteristic fragment ions of the alkyl phthalides were observed by the losses of  $H_2O$ ,  $CO$ ,  $C_2H_4$  and  $C_3H_6$ . It is worth noting that the losses of  $C_4H_8$  can be used to distinguish alkyl phthalides as saturated *n*-butyl and unsaturated butenyl constituents. The characteristic fragment ions of hydroxyl phthalides were found by the loss of  $nH_2O$  according to their number of hydroxyls. The main fragmentation pathway of phthalide dimers found in the DZT sample was first depolymerized into a monomer ion at  $m/z$  191.1077, which further yielded fragment ions similar to that of the alkyl phthalides or hydroxyl phthalides. Senkyunolide A (peak 93, 17.56 min) and senkyunolide F (peak 27, 8.66 min) were used as examples for elucidating the fragmentation pathways of alkyl phthalides and hydroxyl phthalides, respectively. Senkyunolide A exhibited the molecular ion  $[M + H]^+$  at  $m/z$  193.1225 and yielded product ions at  $m/z$  175.1113, 147.1171, 137.0601 and 119.0854, corresponding to  $[M + H - H_2O]^+$ ,  $[M + H - H_2O - CO]^+$ ,  $[M + H - C_4H_8]^+$  and  $[M + H - H_2O - C_4H_8]^+$  (Fig. S2a†). Senkyunolide F displayed  $[M + H]^+$  at  $m/z$  207.1015 and generated fragment ions at  $m/z$  189.0906, 179.1061 and 161.0969, corresponding to  $[M + H - H_2O]^+$ ,  $[M + H - CO]^+$  and  $[M + H - H_2O - CO]^+$  (Fig. S2b†).

As a result, 16 phthalides were putatively identified, including three alkyl phthalides (6,7-epoxyligustilide, 3-butylidene phthalide, senkyunolide A), 11 hydroxyl phthalides (senkyunolide F, 3-*n*-butyl-3-hydroxy-4,5,6,7-tetra-hydro-6,7-

dihydroxy phthalide, senkyunolide N, senkyunolide J, senkyunolide I/senkyunolide H, (–)-3-butyl-7-hydroxyphthalide, 3-butyl-4-hydroxyphthalide, senkyunolide K, senkyunolide G, senkyunolide E, senkyunolide M/senkyunolide Q) and two dimers of phthalide (senkyunolide P and levistolide A).

**Identification of flavonoids in DZT.** In this study, flavonoids produced a wealth of fragment ions in the positive ion mode. The main characteristic fragment ions of the isoflavones were found by consecutively or simultaneously losing  $CH_3$ ,  $CH_3OH$ ,  $CO$  and  $CO_2$ .<sup>48–50</sup> Calycosin 7-*O*- $\beta$ -*D*-glucoside (peak 28b, 8.86 min) was used as an example to illustrate the fragmentation pathway for flavonoids. As shown in Fig. S2c,† calycosin 7-*O*- $\beta$ -*D*-glucoside displays its precursor ion  $[M + H]^+$  at  $m/z$  447.1285 and then yields product ions at  $m/z$  285.0776, 270.0523 and 253.0498, corresponding to  $[M + H - Glu]^+$ ,  $[M + H - CH_3]^+$  and  $[M + H - CH_3OH]^+$  with an error within 5 ppm.

Consequently, 13 flavonoids and their derivatives (six isoflavones, two isoflavans, three flavones, one pterocarpan and one flavonone) were assigned by comparison with previous results.<sup>48–50</sup>

**Identification of carbohydrates in DZT.** Twelve carbohydrates and their derivatives in DZT were discovered based on the shared fragment ions at  $m/z$  341.1089, 221.0661, 179.0552 and 119.0355, corresponding to typical  $C_2$ ,  $^{0,4}A_2$ ,  $C_1$  and  $^{0,2}A_1$ , which were consistent with previously reported results.<sup>34</sup> Verbascose (peak 6a,  $R_t = 1.33$  min) was taken as an example to elucidate the characteristic fragmentation pathway of carbohydrates (Fig. S3†); it showed the precursor ion  $[M - H]^-$  at  $m/z$  827.2687 and yielded product ions at  $m/z$  665.2152, 545.1726, 503.1617, 383.1188, 341.1080, 221.0658, 179.0551, and 119.0342, corresponding to  $C_4$ ,  $^{0,4}A_4$ ,  $C_3$ ,  $^{0,4}A_3$ ,  $C_2$ ,  $^{0,4}A_2$ ,  $C_1$ , and  $^{0,2}A_1$ . Based on previous literature reports,<sup>34,51</sup> peaks 3a, 3b, 4 and 5 were tentatively identified as sucrose, raffinose, stachyose hydrate and isomaltotetraose, respectively.

Three acetylated disaccharide (AD) isomers, namely, AD-383A (peak 10,  $m/z$  383.1189), AD-383B (peak 11,  $m/z$  383.1189) and AD-383C (peak 13,  $m/z$  383.1187) with the molecular formula  $C_{14}H_{24}O_{12}$  eluted at 3.13 min, 3.41 min and 3.84 min, respectively. They presented values 42 Da ( $C_2H_2O$ ) greater than that of sucrose, suggesting their acetylation. However, the fragmentation pathways did not allow for the elucidation of the position of the acetyl moiety. Three isopropyl derivative disaccharide (ID) isomers, namely, ID-441A (peak 14a), ID-441B (peak 15a) and ID-441C (peak 16) with the molecular formula  $C_{16}H_{26}O_{14}$  showed  $[M - H]^-$  ions at  $m/z$  441.1245, eluting at 3.92 min, 4.11 min and 4.29 min, respectively. The three isomers presented values 100 Da ( $C_4H_4O_3$ ) greater than that of sucrose, suggesting isopropylation. Although the position of the isopropyl group could not be confirmed, peaks 14a, 15a and 16 might be derived from sucrose after isopropylation. Peak 46 with the molecular formula  $C_{23}H_{36}O_{12}$  eluted at 10.65 min and displayed fragment ions at  $m/z$  341.1084, 221.0685, and 161.0456, showing a very similar fragmentation pathway to that of sucrose and thus suggesting the presence of a sucrose group. The abundant fragment ions at  $m/z$  191.0560 suggested the presence of a quinic acid group in compound 46. However, the enormous number of structural possibilities did not provide



putative identification for peak 46. In this study, seven compounds including peaks 10, 11, 13, 14a, 15a, 16 and 46 from LC were reported in DZT for the first time.

**Identification of other compounds.** The other eighteen compounds were detected from DZT, including gluconic acid, nicotinic acid, L-pyroglutamic acid, hypoxanthine, succinic acid, uridine, adenosine, 2-hydroxypurine nucleoside, guanosine, phenylalanine, (9E)-9-octadecenedioic acid, eicosapentaenoic acid, 8,12,15-octadecatrienoic acid, (8E, 12Z, 15Z), 2-pentadecenoic acid,  $\beta$ -ursolic acid, arachidonic acid and linoleic acid. Arachidonic acid, linoleic acid, uridine and adenosine were easily identified by comparison with their reference standards. The other 13 compounds were putatively identified by matching the MS/MS fragmentation ions with the previously reported data.<sup>34,52,53</sup>

### Untargeted identification of unknown compounds from DZT

Lysoglycerophosphocholines (Lyso-GPCs) are a representative class of biochemical intermediates comprising

a glycerophosphocholine backbone with a free glycerol hydroxyl at either the *sn*-1 or *sn*-2 position.<sup>54</sup> Nineteen lyso-GPCs were discovered based on co-occurring Mass2Motifs from their fragmentation spectra. In the positive ion mode, all the spectra showed the same Mass2Motifs including the neutral loss of 18.0106 Da ( $\text{H}_2\text{O}$ ) and fragment ions at  $m/z$  184.0731 ( $[\text{P-Ch}]^+$ ), 104.107 ( $[\text{P-Ch-HPO}_3]^+$ ), and 86.096 ( $[(\text{P-Ch})\text{-HPO}_3\text{-H}_2\text{O}]^+$ ) (Fig. 5A). The ion at  $m/z$  184.0731 indicated that phosphocholine as the polar head group of phospholipids was present in the structures.<sup>54</sup> In the negative ion mode, all the spectra shared the same Mass2Motifs including fragment ions ( $m/z$  224.0690 and 168.0424) and a neutral loss of 60.0220 Da (Fig. 5B). Due to the use of formic acid as the elution buffer, 13 compounds were detected as a single intense solvent adduct ion  $[\text{M} + \text{HCOOH-H}]^-$  in the negative ion mode; they easily eliminated  $\text{CH}_3$  to yield  $[\text{M} - \text{CH}_3]^-$  ions by first undergoing facile elimination of  $\text{HCO}_2$ , and their neutral losses were 60.0220 Da ( $\text{HCO}_2 + \text{CH}_3$ ) [49–51]. Based on the Mass2Motifs, element composition and previously reported results,<sup>54–56</sup> these lyso-GPCs can be divided

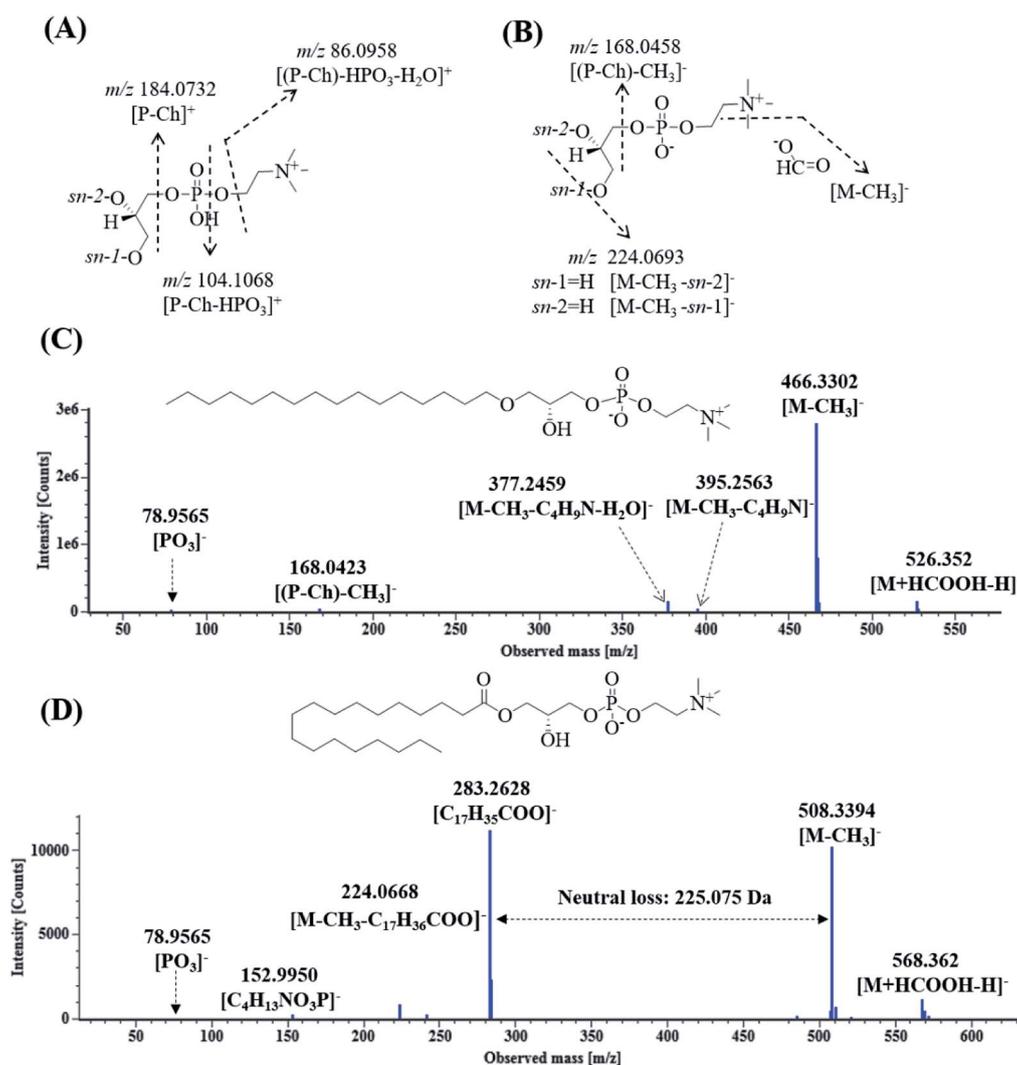


Fig. 5 The proposed fragmentation pathways of lysoglycerophosphocholines in DZT. Fragment pattern in the positive ion mode (A), fragment pattern in the negative ion mode (B); fragmentation pathway of lysophosphatidylcholine O-16 : 0/0 : 0 (C) and fragmentation pathway of 1-stearoyl-lysophosphatidylcholine (D).



into two groups: lysophosphatidylcholines (lysoPCs) and lyso platelet activating factors (lyso-PAFs).

Eleven compounds having peaks 97d, 98a, 98b, 103c, 103d, 110, 115a, 117, 119, 122 and 127 showed neutral losses of 225.0750 Da and yielded abundant fatty acyl-related fragment ions at  $m/z$  301.2165, 301.2164, 277.2165, 303.2324, 279.2320, 281.3487, 281.3488, 307.2630, 307.2635, 283.2635 and 309.2794, respectively. These compounds were classified as lysoPCs, the substituents of which were long-chain fatty acyls at the *sn*-1 or *sn*-2 position. In addition, eight compounds showing peaks 103b, 105a, 106b, 108b, 111, 113, 115b and 126 were grouped into lyso-PAFs, the substituents of which were long-chain ethers at the *sn*-1 or *sn*-2 position. Peaks 115a and 122 were used as examples to elucidate the fragmentation pathway of lyso-PAFs and lyso PCs. As shown in Fig. 5C, peak 115a exhibited the molecular ion  $[M + HCOOH-H]^-$  at  $m/z$  526.3520, and yielded ions at  $m/z$  395.2563, 377.2459, 168.0423 and 78.9565, corresponding to  $[M - CH_3]^-$ ,  $[M-CH_3-C_4H_9N]^-$ ,  $[M-CH_3-C_4H_9N-H_2O]^-$ ,  $[(P-Ch)-CH_3]^-$  and  $[PO_3]^-$ , respectively. As shown in Fig. 5D, peak 122 displayed its molecular ion  $[M + HCOOH-H]^-$  at  $m/z$  568.3620 and generated fragment ions at  $m/z$  508.3394, 283.2628, 224.0668, 152.9950 and 78.9565, corresponding to  $[M - CH_3]^-$ ,  $[C_{17}H_{35}COO]^-$ ,  $[M-CH_3-C_{17}H_{36}COO]^-$ ,  $[C_4H_{13}NO_3P]^-$  and  $[PO_3]^-$ . Peaks 115 and 122 were tentatively identified as lysophosphatidylcholine O-16 : 0/0 : 0 and 1-stearoyl-lysophosphatidylcholine,<sup>55</sup> respectively.

Nineteen compounds were subsequently verified through the LipidMaps (<http://www.lipidmaps.org>) or LipidBank (<http://lipidbank.jp/>) database search. It should be noted that 14 compounds (peaks 97d, 98a, 98b, 103c, 103d, 106b, 108b, 110, 111, 113, 115b, 117, 119 and 127) in DZT were first reported from *Hirudo nipponica* Whitman.

### Isomeric analysis of chemical compounds in DZT

**Characterization of isomers with precursor ions at  $m/z$  717.1456 using precursor ion mobility.** As ion mobility increased the overall peak capacity and specificity of constituent analysis over those of conventional high-resolution mass spectrometry approaches, seven isomers with precursor ions at  $m/z$  717.1456  $\pm$  5 ppm were simultaneously detected for the first time using only the HDMS<sup>E</sup> method. As shown in Fig. 6A, peaks 40, 48, 51 and 53 display the precursor ions  $[M - H]^-$  at  $m/z$  717.1456  $\pm$  5 ppm, eluting at 10.05 min, 10.95 min, 11.19 min and 11.37 min, respectively. Three pairs of coeluting isomers including the mobility peaks 40a/40b, 48a/48b and 51a/51b in the ion mobility mode were derived from peaks 40, 48 and 51, respectively (Fig. 6B). Peak 40a and peak 40b were two coeluting isomers with different drift times ( $d_t$ ) at 5.26 ms and 5.86 ms, respectively. Peaks 40a and 40b showed their product ions at  $m/z$  519.0932, 339.0508, 321.0401 and 295.0608, corresponding to  $[M-H-DSS]^-$ ,  $[M-H-DSS-CA]^-$ ,  $[M-H-2DSS]^-$  and  $[M-H-DSS-CA-CO_2]^-$ . Peaks 40a and 40b were tentatively assigned as salvianolic acid E/isomer by comparing their retention times and MS/MS data with the previously reported results.<sup>57</sup> Peaks 48a and 48b were two isomers with different drift times at 5.30 ms and 5.85 ms, respectively. Peak 48a was positively identified as

salvianolic acid B by comparison with its corresponding standard. The MS/MS spectrum of peak 48b was different from that of salvianolic acid B (Fig. 6C). Peak 48b was putatively assigned as an isomer of salvianolic acid B. Peak 51a ( $d_t = 5.08$  ms) and peak 51b ( $d_t = 5.55$  ms) were two coeluting isomers and showed similar product ions at  $m/z$  519.0927 and 321.0400, corresponding to  $[M-H-DSS]^-$  and  $[M-H-2DDS]^-$ , respectively. Peaks 51a and 51b were tentatively identified as salvianolic acid L/isomer by comparing the retention time and MS/MS data with previously reported results.<sup>57</sup> Peak 53 ( $d_t = 5.56$  ms), showing a very similar fragmentation pattern to that of peak 48b, was also tentatively identified as an isomer of salvianolic acid B (Fig. 6C). However, the exact structures of peaks 40a, 40b, 48b, 51a, 51b and 53 could not be elucidated by the MS/MS data alone.

**Characterization of isomeric caffeoylquinic acids using precursor and/or product ion mobility.** Seventeen caffeoylquinic acids (CQAs), formed between quinic acid and one to two caffeoyl residues, were detected in DZT. These CQAs could be divided into two categories including mono-substituted quinic acid (mono-CQA) and di-substituted quinic acid (di-CQA) based on the numbers of caffeoyl residues. CQAs usually exist in the form of positional and geometrical (*cis-trans*) isomers due to the caffeoyl substituents at the different substituted positions of quinic acid and the double bond of caffeic acid substituents, respectively. These isomers could be differentiated according to their different fragmentation patterns and drift times in the HDMS<sup>E</sup> mode.

(1) *Characterization of mono-CQA isomers with precursor ions at  $m/z$  353.0880 using precursor ion mobility.* Seven mono-CQAs exhibited deprotonated molecular ions  $[M - H]^-$  at  $m/z$  353.0880  $\pm$  5 ppm, which could yield fragment ions at  $m/z$  191.0560, 179.0351, 173.0457 and 135.0450, corresponding to  $[M-H-(CA + H_2O)]^-$ ,  $[M-H-quinic\ acid]^-$ ,  $[M-H-(CA + H_2O)-H_2O]^-$  and  $[M-H-quinic\ acid-CO_2]^-$ . The product ion at  $m/z$  191.0570 was the base peak, indicating that the caffeoyl group was attached to 1-OH, 3-OH or 5-OH. However, the intensities of the product ions at  $m/z$  179.0351 were clearly different (higher for 3-OH and lower for 1- or 5-OH). The product ion at  $m/z$  173.0457 was the base peak, indicating that the caffeoyl group was linked to 4-OH. As shown in Fig. 7A, peaks 18, 20, 23 and 24 in the MS<sup>E</sup> and fast DDA analysis modes display the same precursor ions  $m/z$  353.0880, eluting at 5.25 min, 5.77 min, 6.67 min and 6.86 min, respectively. Fig. 7B shows the overlaid mobility profile of peaks 18, 20, 23 and 24. It was clearly observed that 3 pairs of coeluting isomers including mobility peaks 18a/18b, 23a/23b and 24a/24b in the HDMS<sup>E</sup> mode were derived from peaks 18, 23 and 24, respectively. The corresponding fragmentation spectra for mobility peaks 18a, 18b, 20, 23a, 23b, 24a and 24b are shown in Fig. 7C. On the basis of the comparison of their retention times, drift times and fragmentation data with those of authentic standards, peaks 23b and 24b were unambiguously identified as *trans*-5-caffeoylquinic acid and *trans*-4-caffeoylquinic, respectively. Peak 20 with a single mobility peak, showing its base peak ion at  $m/z$  191.0571 and abundant signal intensity of the ion at  $m/z$  173.0457, was tentatively characterized as *cis/trans*-3-



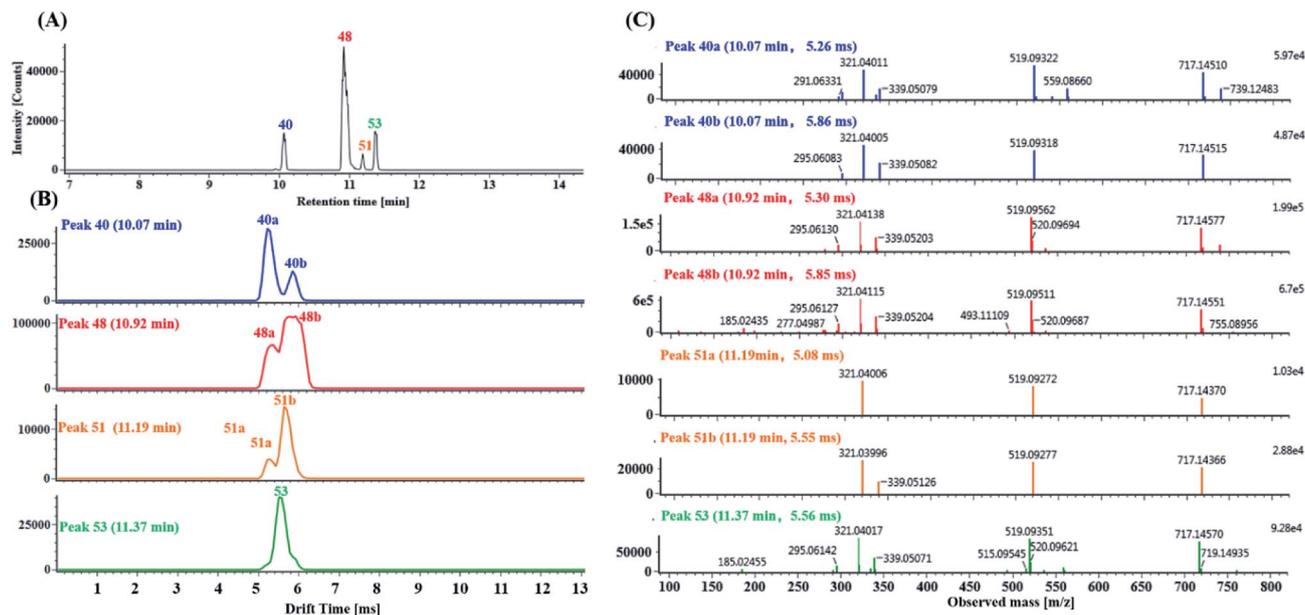


Fig. 6 Characterization of isomers with precursor ions at  $m/z$  717.1456 ppm in DZT: (A) extracted ion chromatography (EIC) of the precursor ions at  $m/z$  717.1456  $\pm$  5 ppm, including peaks 40 (blue trace), 48 (red trace), 51 (orange trace) and 53 (green trace); (B) mobility profiles of the  $m/z$  717.1456 precursor ions of peaks 40, 48, 51 and 53; in addition, among them, peaks 40, 48 and 51 further dissociated into mobility peaks 40a/40b, 48a/48b and 51a/51b, respectively. (C) MS/MS spectrum of the precursor ions at  $m/z$  717.1456 for mobility peaks 40a, 40b, 48a, 48b, 51a, 51b and 53.

caffeoylquinic acid by comparing its retention time and fragmentation data with previously reported data.<sup>58,59</sup> Compared with the data from the literature, peaks 23a and 24a showed similar fragmentation patterns and drift times to those of peaks 23b and 24b, and they were tentatively proposed to be *cis*-5-caffeoylquinic acid and *cis*-4-caffeoylquinic acid, respectively.

From a previous literature, we inferred that 1-caffeoylquinic and 5-caffeoylquinic were undistinguished by the same fragmentation, but the greater hydrophobicity of 5-caffeoylquinic and availability of its reference standard ensured that these two compounds could be differentiated by their different retention times on a reversed-phase column. Moreover, peaks 18a and

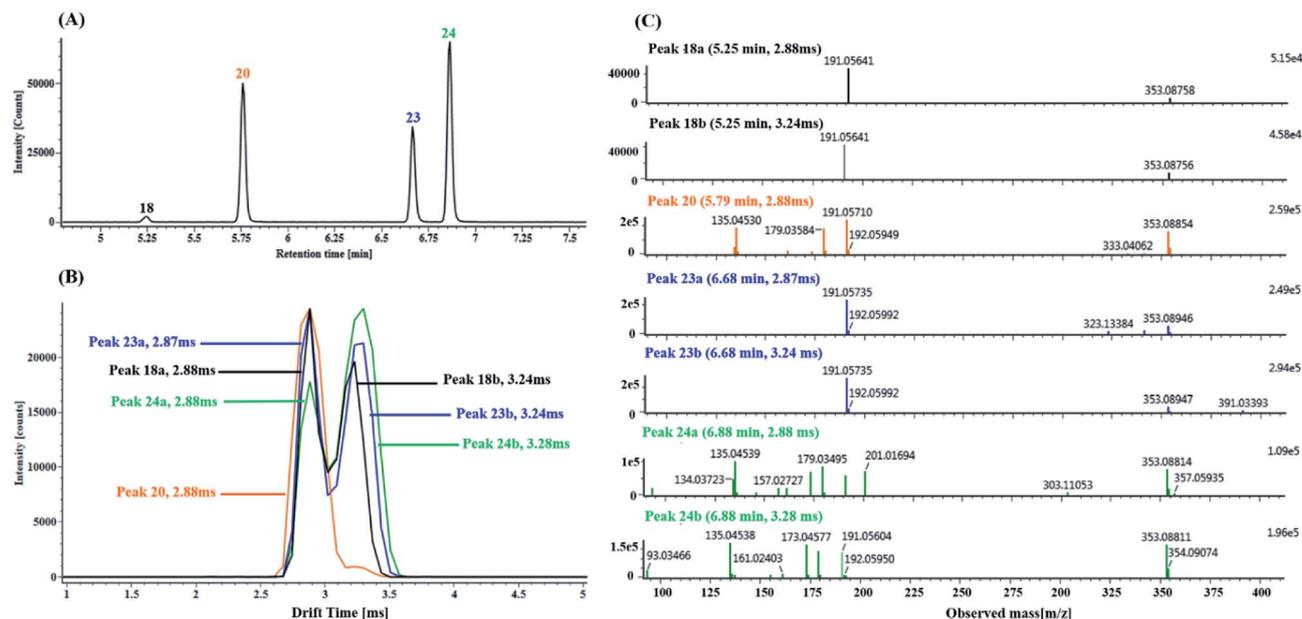
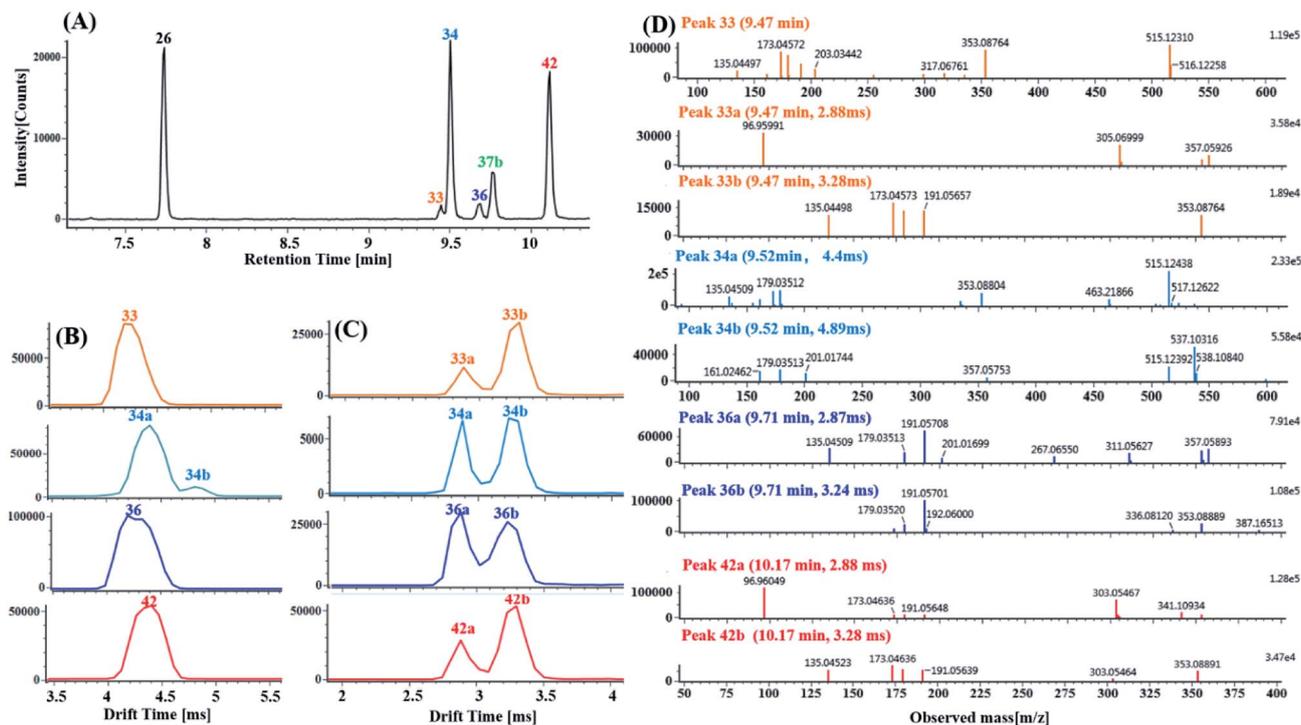


Fig. 7 Characterization of the CQA isomers with the precursor ions at  $m/z$  353.0880 in DZT using precursor ion mobility: (A) extracted ion chromatography (EIC) of the  $m/z$  353.0880  $\pm$  5 ppm precursor ions including peaks 18 (blank trace), 20 (orange trace), 23 (blue trace) and 24 (green trace). (B) Overlaid mobility profiles of the  $m/z$  353.0880 precursor ions of peaks 18, 20, 23 and 24; in addition, peaks 18, 23 and 24 among them further dissociated into valley-to-valley peaks, including mobility peaks 18a/18b, 23a/23b and 24a/24b, respectively. (C) MS/MS spectrum of the precursor ions at  $m/z$  353.0880 for mobility peaks 18a, 18b, 20, 23a, 23b, 24a and 24b.





**Fig. 8** Characterization of the di-CQA isomers with the product ions at  $m/z$  353.0880  $\pm$  5 ppm derived from the precursor ions at  $m/z$  515.1190  $\pm$  5 ppm in DZT using product ion mobility: (A) extracted ion chromatography (EIC) of the  $m/z$  353.0880 product ions including peaks 26 (blank), 33 (orange), 34 (azure), 36 (blue), 37b (green) and 42 (red). (B) Mobility profiles of the  $m/z$  515.1190 precursor ions of peaks 33, 34, 36 and 42. (C) Mobility profiles of the  $m/z$  353.0880 product ions of peaks 33a/33b, 34a/34b, 36a/36b and 42a/42b. (Peaks 33, 34, 36 and 42 further dissociated into valley-to-valley mobility peaks, including 33a/33b, 34a/34b, 36a/36b and 42a/42b, respectively.) (D) MS/MS spectrum of mobility peaks 33a, 33b, 34a, 34b, 36a, 36b, 42a and 42b.

18b showed similar drift times to those of *cis*-5-caffeoylquinic and *trans*-5-caffeoylquinic, respectively. Thus, peaks 18a and 18b were deduced and tentatively identified as *cis*-1-caffeoylquinic acid and *trans*-1-caffeoylquinic acid, respectively.

(2) *Characterization of di-CQA isomers with product ions at  $m/z$  353.0880 derived from precursor ions at  $m/z$  515.1190 using product ion mobility.* Ten di-CQAs exhibited precursor ions  $[M - H]^-$  at  $m/z$  515.1190  $\pm$  5 ppm and generated fragment ions at 353.0880, 191.0570, 179.0351, 173.0457 and 135.0450, corresponding to  $[M - H - (CA - H_2O)]^-$ ,  $[M - H - 2(CA - H_2O)]^-$ ,  $[M - H - (CA - H_2O) - \text{quinic acid}]^-$ ,  $[M - H - 2(CA - H_2O) - H_2O]^-$  and  $[M - H - (CA - H_2O) - \text{quinic acid} - CO_2]^-$ . As shown in Fig. 8A, peaks 26, 33, 34, 36, 37b and 42 in the MS<sup>E</sup> and Fast DDA modes show their precursor ions  $m/z$  515.1190  $\pm$  5 ppm, eluting at 7.76 min, 9.47 min, 9.52 min, 9.71 min, 9.78 min and 10.14 min, respectively. Fig. 8B and C display the corresponding mobility spectra of the precursor ions at  $m/z$  515.1190 and product ions at  $m/z$  353.0880 of peaks 26, 33, 34, 36, 37b and 42. Interestingly, four pairs of coeluting isomers including the mobility peaks 33a/33b, 34a/34b, 36a/36b and 42a/42b were derived from peaks 33, 34, 36 and 42, respectively. As shown in Fig. 8D, peak 33b is easily identified as *trans*-1,4-dicaffeoylquinic acid with the occurrence of product ions at  $m/z$  317.0676, 299.0650, 255.0667, and 203.0331 based on the different experimental drift times and previously reported results.<sup>60</sup> Peak 33a with only a fragment ion at  $m/z$  353.0875 could be tentatively assigned as a dicaffeoylquinic acid isomer. Peaks 26 and 37b were proposed

to be *trans*-1,3-dicaffeoylquinic acid and *cis*-1,3-dicaffeoylquinic acid, respectively, based on their different drift times and similar fragment ions including a detectable ion  $m/z$  335.0772, base peak  $m/z$  191.0570 and strong intensity ion  $m/z$  179.0351 (>50% base peak). Peaks 36a and 36b, having similar mobility behaviors and fragmentation behaviors but no detectable ion  $m/z$  335.0772 as peak 26 and 37b, were tentatively identified as *cis*-3,5-dicaffeoylquinic acid and *trans*-3,5-dicaffeoylquinic acid, respectively. Based on the occurrence of different drift times, the base peak  $m/z$  173.0457, strong ion  $m/z$  179.0351 and weak ion  $m/z$  335.0772, peaks 34a and 34b were tentatively identified as *cis*-3,4-dicaffeoylquinic acid and *trans*-3,4-dicaffeoylquinic acid, respectively. Comparing their similar mobility behaviors and fragmentation patterns to those of peaks 34a and 34b, peaks 42a and 42b were putatively characterized as *cis*-4,5-dicaffeoylquinic acid and *trans*-4,5-dicaffeoylquinic acid, respectively.

## Conclusions

In the present study, the combination of multidimensional acquisition methods with a data processing strategy was established for the comprehensive identification of 202 constituents in DZT. This is the first report about the component analysis of DZT. Based on the Mass2Motifs extracted from fragmentation spectra, 29 potentially unknown compounds were identified or highlighted for the first time. By using ion



mobility analysis, ten pairs of coeluting isomers, which were not discovered or resolved by the conventional MS<sup>E</sup> or Fast DDA mode, were distinguished using the HDMS<sup>E</sup> mode. This study revealed the suitability of the proposed approach to discover and identify the known, unknown and isomeric compounds of DZT. Moreover, this work can be a practical and powerful paradigm for meeting the ongoing demands for the partial exploration of molecular dark matter from TCM formulae or other natural products. However, a major drawback is that the accurate structures of many compounds cannot be confirmed using MS data alone. Therefore, further studies will be required to elucidate these complex compounds using additional characterization techniques such as nuclear magnetic resonance (NMR).

## Conflicts of interest

The authors declare no conflict of interest.

## Acknowledgements

The work was supported by Professor of Chang Jiang Scholars Program, NSFC (81520108030, 21472238), Shanghai Engineering Research Center for the Preparation of Bioactive Natural Products (16DZ2280200), the Scientific Foundation of Shanghai China (13401900103, 13401900101) and the National Key Research and Development Program of China (2017YFC1700200).

## References

- R. Liu, R. S. Runyon, Y. Wang, S. G. Oliver, T. Fan and W. Zhang, *Science*, 2015, **347**, S40–S42.
- L. Wang, G. Zhou, P. Liu, J. Song, Y. Liang, X. Yan, F. Xu, B. Wang, J. Mao, Z. Shen, S. Chen and Z. Chen, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4826–4831.
- W. Lam, S. Liu, Z. Jiang and Y. Cheng, *Science*, 2015, **347**, S43–S44.
- H. Sheridan, B. Kopp, L. Krenn, D. Guo and J. Sendker, *Science*, 2015, **350**, S64–S66.
- H. Zhang, D. Zheng, H. Li, H. Wang, H. Tan and H. Xu, *Anal. Chim. Acta*, 2016, **912**, 85–96.
- W. Si, W. Yang, D. Guo, J. Wu, J. Zhang, S. Qiu, C. Yao, Y. Cui and W. Wu, *J. Pharm. Biomed. Anal.*, 2016, **117**, 510–521.
- R. S. Plumb, K. A. Johnson, P. Rainville, B. W. Smith, I. D. Wilson, J. M. Castro-Perez and J. K. Nicholson, *Rapid Commun. Mass Spectrom.*, 2006, **20**, 1989–1994.
- F. Fenaille, P. B. Saint-Hilaire, K. Rousseau and C. Junot, *J. Chromatogr. A*, 2017, **1526**, 1–12.
- M. Sundstrom, A. Pelander and I. Ojanpera, *J. Anal. Toxicol.*, 2017, **41**, 623–630.
- J. Boschmans, F. Lemiére and F. Sobott, *J. Chromatogr. A*, 2017, **1490**, 80–88.
- S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinou, R. H. Bateman, M. T. Bowers and J. H. Scrivens, *Int. J. Mass Spectrom.*, 2007, **261**, 1–12.
- G. Paglia and G. Astarita, *Nat. Protoc.*, 2017, **12**, 797–813.
- S. J. Valentine, M. D. Plasencia, X. Liu, M. Krishnan, S. Naylor, H. R. Udseth, R. D. Smith and D. E. Clemmer, *J. Proteome Res.*, 2006, **5**, 2977–2984.
- J. Far, C. Delvaux, C. Kune, G. Eppe and E. de Pauw, *Anal. Chem.*, 2014, **86**, 11246–11254.
- X. Qiao, R. Li, W. Song, W. Miao, J. Liu, H. Chen, D. Guo and M. Ye, *J. Chromatogr. A*, 2016, **1441**, 83–95.
- J. Zhang, Z. Wang, Y. Li, Y. Liu, W. Cai, C. Li, J. Lu and Y. Qiao, *Talanta*, 2016, **147**, 16–27.
- X. Cheng, J. Wan, P. Li and L. Qi, *J. Chromatogr. A*, 2011, **1218**, 5774–5786.
- K. P. Bateman, J. Castro-Perez, M. Wrona, J. P. Shockcor, K. Yu, R. Oballa and D. A. Nicoll-Griffith, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 1485–1496.
- J. Chen, Z. Shi, Y. Song, X. Guo, M. Zhao, P. Tu and Y. Jiang, *J. Chromatogr. A*, 2016, **1464**, 102–114.
- H. Ouyang, J. Li, B. Wu, X. Zhang, Y. Li, S. Yang, M. He and Y. Feng, *J. Chromatogr. A*, 2017, **1502**, 38–50.
- J. J. J. van der Hooft, J. Wandy, F. Young, S. Padmanabhan, K. Gerasimidis, K. E. V. Burgess, M. P. Barrett and S. Rogers, *Anal. Chem.*, 2017, **89**, 7569–7577.
- J. J. J. van der Hooft, J. Wandy, M. P. Barrett, K. E. V. Burgess and S. Rogers, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 13738–13743.
- J. Wandy, Y. Zhu, J. J. J. van der Hooft, R. Daly, M. P. Barrett and S. Rogers, *Bioinformatics*, 2018, **34**, 317–318.
- Y. Luo, J. Han, J. Xu and Y. Zhao, *Henan Traditional Chinese Medicine*, 2014, **34**, 628–629.
- N. Mi, T. Cheng, H. Li, P. Yang, X. Mu, X. Wang, X. Zu, X. Qi, X. Guo, J. Ye and W. Zhang, *J. Pharm. Biomed. Anal.*, 2018, **164**, 70–85.
- H. Y. Fan, F. H. Fu, M. Y. Yang, H. Xu, A. H. Zhang and K. Liu, *Thromb. Res.*, 2010, **126**, E17–E22.
- Y. Yao, A. H. Liu, W. Y. Wu, S. H. Guan, B. H. Jiang, M. Yang, K. S. Bi, X. Liu and D. A. Guo, *Phytochem. Lett.*, 2008, **1**, 135–138.
- Y. Yao, W. Y. Wu, A. H. Liu, S. S. Deng, K. S. Bi, X. Liu and D. A. Guo, *Am. J. Chin. Med.*, 2008, **36**, 313–328.
- Z. S. Huang, C. L. Zeng, L. J. Zhu, L. Jiang, N. Li and H. Hu, *J. Thromb. Haemostasis*, 2010, **8**, 1383–1393.
- Y. Chen, N. Zhang, J. Ma, Y. Zhu, M. Wang, X. Wang and P. Zhang, *J. Pharm. Biomed. Anal.*, 2016, **117**, 178–183.
- J. Cao, Z. Chen, Y. Zhu, Y. Li, C. Guo, K. Gao, L. Chen, X. Shi, X. Zhang, Z. Yang and A. Wen, *J. Ethnopharmacol.*, 2014, **155**, 1053–1060.
- J. Cao, J. Wei, Y. Hu, C. He, M. Chen, J. Wan and P. Li, *J. Chromatogr. A*, 2016, **1427**, 79–89.
- S. T. Yang, X. Wu, W. Rui, J. Guo and Y. F. Feng, *Acta Chromatogr.*, 2015, **27**, 711–728.
- X. Chen, Z. Lou, H. Zhang, G. Tan, Z. Liu, W. Li, Z. Zhu and Y. Chai, *Rapid Commun. Mass Spectrom.*, 2011, **25**, 1661–1674.
- C. Li, H. Song, S. Liu, Q. Wang, G. Dai, X. Ding and W. Ju, *J. Sep. Sci.*, 2016, **39**, 1099–1109.
- L. Qi, J. Cao, P. Li, Q. Yu, X. Wen, Y. Wang, C. Li, K. Bao, X. Ge and X. Cheng, *J. Chromatogr. A*, 2008, **1203**, 27–35.



- 37 C. Chu, H. Cai, M. Ren, E. H. Liu, B. Li, L. Qi and P. Li, *J. Sep. Sci.*, 2010, **33**, 570–581.
- 38 M. Liebeke, N. Strittmatter, S. Fearn, A. J. Morgan, P. Kille, J. Fuchser, D. Wallis, V. Palchykov, J. Robertson, E. Lahive, D. J. Spurgeon, D. McPhail, Z. Takats and J. G. Bundy, *Nat. Commun.*, 2015, **6**, 7869.
- 39 H. Dong, J. Ren, J. Wang, L. Ding, J. Zhao, S. Liu and H. Gao, *J. Evidence-Based Complementary Altern. Med.*, 2016, 1–11.
- 40 Z. Zhu, H. Zhang, L. Zhao, X. Dong, X. Li, Y. Chai and G. Zhang, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 1855–1865.
- 41 Y. Zhou, G. Xu, F. F. K. Choi, L. Ding, Q. Bin Han, J. Z. Song, C. F. Qiao, Q. Zhao and H. Xu, *J. Chromatogr. A*, 2009, **1216**, 4847–4858.
- 42 Z. Ma, M. Zhang and Z. Song, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 2857–2866.
- 43 M. Yang, A. H. Liu, S. H. Guan, J. H. Sun, M. Xu and D. Guo, *Rapid Commun. Mass Spectrom.*, 2006, **20**, 1266–1280.
- 44 P. O. Donkor, Y. Chen, L. Ding and F. Qiu, *J. Ethnopharmacol.*, 2016, **194**, 530–548.
- 45 X. Ran, L. Ma, C. Peng, H. Zhang and L. Qin, *Pharm. Biol.*, 2011, **49**, 1180–1189.
- 46 X. Lei, G. Li, L. Cheng, X. Wang and F. Meng, *J. Pharm. Biomed. Anal.*, 2018, **154**, 123–137.
- 47 X. Z. Zhang, H. B. Xiao, Q. Xu, X. L. Li, L. N. Wang and X. M. Liang, *J. Chromatogr. Sci.*, 2003, **41**, 428–433.
- 48 L. Qi, X. Wen, J. Cao, C. Li, P. Li, L. Yi, Y. Wang, X. Cheng and X. Ge, *Rapid Commun. Mass Spectrom.*, 2008, **22**, 2493–2509.
- 49 J. Zhang, X. Xu, W. Xu, J. Huang, D. Zhu and X. Qiu, *J. Chromatogr. Sci.*, 2015, **53**, 945–952.
- 50 X. Zhang, H. Xiao, X. Xue, Y. Sun and X. Liang, *J. Sep. Sci.*, 2007, **30**, 2059–2069.
- 51 T. Yamagaki and A. Sato, *Anal. Sci.*, 2009, **25**, 985–988.
- 52 W. Liu, J. Fan, Y. Li, X. Qiu, X. Liu, R. Su and Z. Zhao, *Chin. J. Pharm. Anal.*, 2014, **34**, 1417–1421.
- 53 Y. Zhang, W. Dong, J. Huo and W. Wang, *Chin. Tradit. Herb. Drugs*, 2017, **48**, 252–262.
- 54 N. Khaselev and R. C. Murphy, *J. Am. Soc. Mass Spectrom.*, 2000, **11**, 283–291.
- 55 N. Noda, R. Tanaka, M. Nishi, S. Inoue and K. Miyahara, *Chem. Pharm. Bull.*, 1993, **41**, 1366–1368.
- 56 R. C. Murphy, J. Fiedler and J. Hevko, *Chem. Rev.*, 2001, **101**, 479–526.
- 57 A.-H. Liu, H. Guo, M. Ye, Y. Lin, H. Sun, M. Xu and D. Guo, *J. Chromatogr. A*, 2007, **1161**, 170–182.
- 58 M. N. Clifford, J. Kirkpatrick, N. Kuhnert, H. Roozendaal and P. R. Salgado, *Food Chem.*, 2008, **106**, 379–385.
- 59 C. Xie, K. Yu, D. Zhong, T. Yuan, F. Ye, J. A. Jarrell, A. Millar and X. Chen, *J. Agric. Food Chem.*, 2011, **59**, 11078–11087.
- 60 M. N. Clifford, S. Knight and N. Kuhnert, *J. Agric. Food Chem.*, 2005, **53**, 3821–3832.

