

Identification of Glycoside Compounds from Tobacco by High Performance Liquid Chromatography/Electrospray Ionization Linear Ion-Trap Tandem Mass Spectrometry Coupled with Electrospray Ionization Orbitrap Mass Spectrometry

Chunxiao Jia,^a Yonghua Zhu,^a Jinjie Zhang,^a Jing Yang,^b Chunping Xu^b and Duobin Mao^{*.b}

^aCollege of Material and Chemical Engineering and ^bCollege of Food and Biology Engineering, Zhengzhou University of Light Industry, 450002 Zhengzhou, China

In order to comprehensively screen and identify the glycoside compounds in tobacco, a simple, rapid and sensitive method of high performance liquid chromatography/electrospray ionization linear ion-trap tandem mass spectrometry (HPLC-ESI-LIT/MSⁿ) coupled with electrospray ionization orbitrap mass spectrometry (ESI-Orbitrap-MS) was developed for the first time. As a result, twenty-two glycoside compounds, including eleven alcoholic glycosides, eight phenolic glycosides, two ester glycosides and an indole glycoside, were reliably identified from tobacco with high mass accuracy (within 5 mDa). Among them, four compounds were confirmed as novel molecules and other four compounds, as far as we know, were not reported previously in tobacco. This study provided a useful tool to identify the new structures of glycoside compounds in natural products, especially when there were no reference compounds available.

Keywords: glycoside, tobacco, HPLC-ESI-LIT/MSⁿ, ESI-Orbitrap-MS

Introduction

Glycosides are a large class of low-molecular-weight secondary metabolites formed in the process of growth and development of plants. These components are active ingredient of many medicinal plants. For example, previous study has showed that stevioside isolated from the leaves of *Stevia rebaudiana* (Compositae) lowered blood pressure in spontaneously hypertensive rats when administered intravenously.¹ Amygdaline, the substance initially found in plants of the rosaceous family, has been widely used to treat asthma, aplastic anemia and tumors in oriental medicine.² Furthermore, it has been proved that glycosides readily release aglycones straightforwardly in the processing of plants maturation, gather and industrial processing because of enzymolysis.³ By enzymatic or acid hydrolysis, the glycosides in many vegetables, fruits and fragrant plants could decomposed into monoterpene alcohols and aromatic alcohols, which were important fruity aroma. Therefore, glycosides is known as "aroma precursor".⁴⁻⁷ Previous analyses have showed that many glycosidically-bound aroma volatiles, such as terpenols,

aromatic alcohols and fatty alcohols, are very common in tobacco.^{8,9} Thus, the effective method for the separation and structural identification of new glycosides in tobacco is of significant importance for the investigation of tobacco flavor and bioactivity of tobacco glycosides.

To the structural identification of most new glycosides in plants, traditional strategies that involved extraction, isolation, purification and identification by spectroscopic methods such as infrared (IR) and nuclear magnetic resonance (NMR) techniques in the previous literature were performed.¹⁰⁻¹³ Although it is the most efficient method for complete structural elucidation of the compounds, a relatively large amount of purified sample is required and the isolation procedure from a complicated mixture is laborious and costly. Currently, various analytical techniques used for separation and identification of glycosides in tobaccos include enzymatic hydrolysis or acid hydrolysis coupled with gas chromatography-mass spectrometry (GC-MS) method,^{9,14-17} and liquid chromatography-mass spectrometry (LC-MS) or liquid chromatography-tandem mass spectrometry (LC/MSⁿ) method.¹⁸⁻²¹ However, those methods above are indirect because they need comparison with standard references and literature data. Therefore it is desirable to develop a rapid and sensitive method for

*e-mail: duobinmao@163.com

identifying and characterizing the known or new structures. High resolution mass spectrometry (HRMS) and its hybrid or equipped with MSⁿ have proven to be very powerful tools for the structural characterization and identification of natural compounds in view of their complementary capacity of providing multistage fragmentations and accurate mass measurements for precise elemental compositions of molecular ions and subsequent product ions.²² This method is possible to rapidly screen many new glycosides in plants using a minor amount of material samples and to detect the trace compounds which might be ignored by NMR.

In this study, a method of high performance liquid chromatography electrospray ionization linear ion-trap tandem mass spectrometry (HPLC-ESI-LIT/MSⁿ or LC-ESI-LIT/MSⁿ) coupled with electrospray ionization orbitrap mass spectrometry (ESI-Orbitrap-MS) was used to separate and identify the glycosides in tobacco. The relationship between structure characteristics and MSⁿ fragmentation ions of five authentic glycosides was investigated and the fragmentation patterns of the five glycosides were proposed. On the basis of the patterns, a comprehensive screening of glycoside compounds in tobacco was performed through MSⁿ experiments, such as data-dependent and neutral losses scans. Then, the structures of the screened glycosides were elucidated by MSⁿ fragmentation ions and accurate mass measurements. As a result, twenty-two glycoside compounds were reliably identified with very good mass accuracy (within 5 mDa) from the tobacco. Among them, four compounds were confirmed as novel molecules and other four compounds were not reported previously in tobacco. It is the first time to rapidly screen new glycosides in tobacco, based on this technique with the fragmentation patterns of authentic glycosides.

Experimental

Chemicals, reagents and sample

The authentic glycosides, including benzyl β -*D*-glucopyranoside, 2-phenylethyl β -*D*-glucopyranoside, rutin, 3-oxo- α -ionyl β -*D*-glucopyranoside and kaempferyl-3-rutinoside were used as reference standards for identification purposes. Among them, benzyl β -*D*-glucopyranoside and 2-phenylethyl β -*D*-glucopyranoside were prepared by the modified Konigs-Knorr method.²³ The other three compounds were isolated from tobacco samples by semi-preparative HPLC in our laboratory. Briefly, the tobacco sample (500 g) was extracted with methanol by ultrasonic assistant extraction. The tobacco sample and the experiment method were the same as the description in Sample preparation sub-section, while the Amberlite AB-8 column

(4 × 60 cm) was employed and the volume of all reagents was enlarged 200 times. The resulting sample solution was filtered through a membrane filter (0.45 μ m, Millipore). Then two mL of the resulting filtrate was injected into the semi-preparative HPLC. The HPLC system equipped with a Waters 600 controller, a Waters Delta 600 pump, an ultraviolet detector (Waters 2489), and a SunFire™ prep C18 column (250 × 10 mm internal diameter (i.d.), 10 μ m, Waters Corp., USA). The mixed solvent CH₃CN-0.1% formic acid aqueous solution (70:30, v/v) was used as eluent at a flow rate of 4.0 mL min⁻¹. As a result, 45, 11 and 13 mg of rutin, 3-oxo- α -ionyl β -*D*-glucopyranoside and kaempferyl-3-rutinoside, respectively, were obtained from 500 g of a tobacco sample. The structures of the authentic glycosides (Figure 1) were unambiguously confirmed by spectroscopic methods (IR, MS, ¹H and ¹³C NMR). The purity of the standards was determined to be higher than 98% by normalization of the peak area by HPLC-ELSD (evaporative light scattering detector, ELSD). Stock solutions (0.4 mg mL⁻¹) of the five glycosides were prepared individually in methanol.

HPLC-grade acetonitrile and methanol were purchased from Merck (Darmstadt, Germany) and HPLC-grade formic acid was purchased from Amethyst Chemicals (Beijing, China). Deionized water (18 M Ω cm) was obtained from a Milli-Q water purification system (Millipore Corp., USA).

The flue-cured tobacco sample (*Nicotiana tabacum* L., Guangdong Nanxiong grade: CF-3-2) was kindly provided from R&D center of Hongta Tobacco (Group) Co., Ltd. (Yuxi, China). The sample was dried at 45 °C for 4 h and ground to pass through a 60-mesh sieve, and then stored in brown glass jars at 4 °C.

Sample preparation

The tobacco sample (2.00 g) was extracted by ultrasonic assistant extraction with methanol (60 mL) for 40 min at 120 W. After filtration, the residue was washed with methanol (3 × 10 mL) and the resulting solutions were combined and concentrated by rotary evaporation to give a residue (0.53 g), and then dissolved in deionized water (10 mL). This solution was extracted with hexane-ether (1:1, v/v) (3 × 30 mL) to remove lipids, and the water phase was collected, concentrated under reduced pressure to give a residue (0.44 g), and then dissolved in methanol (5 mL). Afterwards, this solution was passed over an Amberlite AB-8 column (2 × 30 cm) and rinsed with 200 mL deionized water. Finally, the column was eluted with 300 mL methanol and the resulting solution was taken to dryness by rotary evaporation. The resulting residue (0.11 g) was dissolved in 5 mL methanol to give a sample

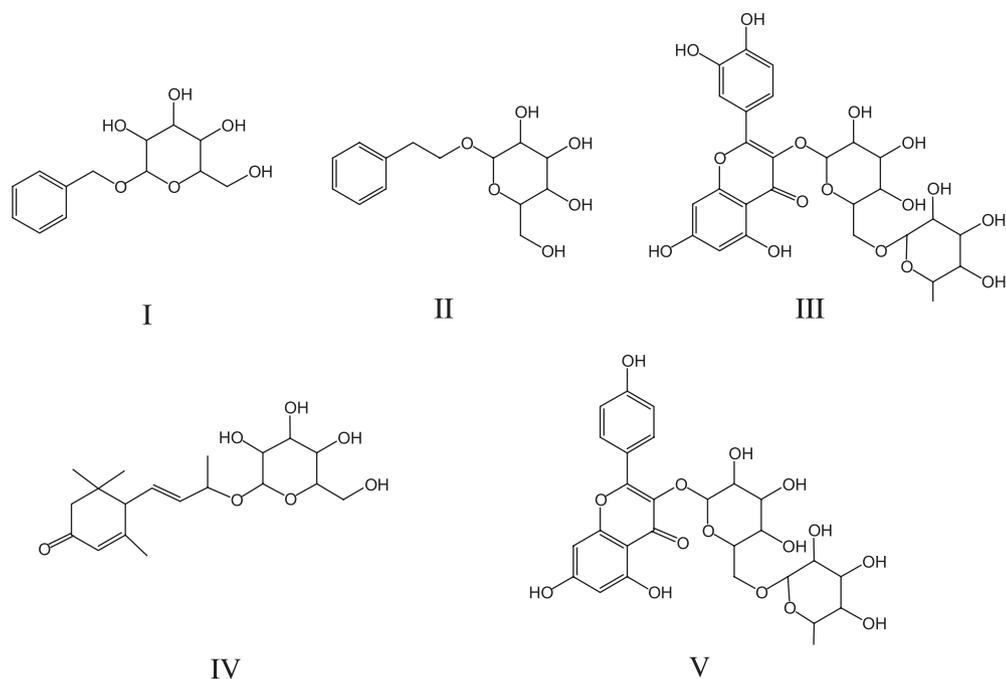


Figure 1. The chemical structures of the five authentic glycosides compounds: **I**: benzyl β-*D*-glucopyranoside; **II**: 2-phenylethyl β-*D*-glucopyranoside; **III**: rutin; **IV**: 3-oxo-α-ionyl β-*D*-glucopyranoside; **V**: kaempferyl-3-rutinoside.

solution of tobacco glycosides, and the sample solution was analyzed by LC-ESI-LIT/MSⁿ and ESI-Orbitrap-MS after filtering through a 0.45 μm organic membrane.

LC conditions

HPLC analysis was performed on a Thermo Fisher Scientific LTQ XL instrument coupled to a quaternary 1250 pump, an auto-sampler and a column department. HPLC separation was performed on a ACQUITY UPLC HSS T3 column (2.1 × 100 mm i.d., 1.8 μm; Waters, USA) with a flow rate of 300 μL min⁻¹ and a sample injection volume of 5 μL. The mobile phase was consisted of water (A) and acetonitrile (B). A gradient program was used as follows: 0-2 min, 90% A; 2-25 min, 90-10% A; 25-28 min, 10-90% A; 28-30 min, 90% A.

Mass spectrometric conditions

For LC-ESI-LIT/MSⁿ analysis, mass spectra were acquired using a LTQ ion trap mass spectrometer (Thermo Electron Corp., USA) equipped with an ESI source operated in positive ion and negative-ion mode. LC-MS settings were as follows: spray voltage 5.0 kV; sheath gas flow rate 30 (arbitrary units); auxiliary gas flow rate 10 (arbitrary units); temperature of the capillary 350 °C. High-purity nitrogen (N₂) was used as the sheath and auxiliary gas. Ultra-high-purity helium (He) was used as the collision gas. LC/MSⁿ analyses was carried out in data-dependent scan

mode with three scan events: event 1: full mass spectrum, range *m/z* 200-800 (unless otherwise specified); event 2: data-dependent MS² spectrum of most abundant ion from event 1; event 3: data-dependent MS³ spectrum of most abundant ion from event 2. Detection was achieved at an isolation width of 2.0 *m/z* units with collision-induced dissociation (CID) activation type, and fragmentation with normalized collisional dissociation energy of 25% for event 2 and 35% for event 3 unless otherwise specified.

High-resolution mass spectrometer (HRMS) experiment was performed with a Thermo Fisher Scientific Exactive-Orbitrap ion trap mass spectrometer equipped with an ESI source (ESI-Orbitrap-MS), operated in positive and negative-ion mode and using a syringe pump. The syringe speed was set at 8.00 μL s⁻¹. The ionization source-specific parameters were as follows: spray voltage 3.50 kV; sheath gas flow rate 30 (arbitrary units); auxiliary gas flow rate 10 (arbitrary units); capillary temperature 300 °C; capillary voltage 60 V. In the same way, high-purity nitrogen (N₂) was used as the sheath and auxiliary gas and ultra-high-purity helium (He) was used as the collision gas.

Results and Discussion

Investigation of authentic glycosides by LC-ESI-LIT/MSⁿ and ESI-Orbitrap-MS

The authentic glycosides **I-V** were analyzed by LC-ESI-LIT/MSⁿ and ESI-Orbitrap-MS in positive and

negative ion mode, respectively. The accurate molecular mass of the five compounds were acquired by HRMS analyses with high mass accuracy (less than 2 mDa) (data not shown). The results of MSⁿ experiments show that these five compounds can provide many characteristic fragment ions, both in positive-ion and negative-ion mode. Moreover, the compounds present similar fragmentations in MSⁿ experiments. As a representative example, the fragmentation patterns of compound **III** (rutin) were investigated for elucidation of the molecular structure.

The proposed fragmentation pathways (a) and mass spectra in positive ESI mode of rutin (b) are shown in Figure 2. In positive ESI mode, rutin generated a minor [M + H]⁺ ion at *m/z* 611 and a predominant [M + Na]⁺ ion at *m/z* 633. The ions at *m/z* 615 and 601 were formed by the losses of H₂O (18 Da) and CH₃OH (32 Da) from *m/z* 633. The ions at *m/z* 487 and 325 corresponded to the sequential losses of rhamnosyl residue (rha, 146 Da) and glucosyl (glu, 162 Da) residue from *m/z* 633. The ion at *m/z* 297 was formed by the loss of CO from *m/z* 325, then it yielded ions at *m/z* 269, 193 and 165 from the losses of CO, C₄H₅Na and CO, respectively. The ion at *m/z* 331 was the sodium adduct ion of disaccharide (glu-rha) residue ([308 + Na]⁺), then it produced the ion at *m/z* 185 ([162 + Na]⁺) and the ion at *m/z* 313 from the losses of a rhamnose residue and H₂O, respectively. In negative ESI mode (mass spectra not shown), rutin produced predominant [M - H]⁻ ion at 609. A fragment ion [M - H - 18]⁻ at *m/z* 591 was characterized by the loss of H₂O. An obvious fragment ion [M - H - 162 - 146]⁻ at *m/z* 301 was characterized from the losses of a neutral glucosyl residue and a rhamnosyl residue. In MS³, a fragment ion *m/z* 283 was characterized by the loss of H₂O from the ion at *m/z* 301. A fragment ion *m/z* 179 showed a glucose unit contained in the molecule.^{12,24-26}

Based on the analyses of the five authentic compounds, it was concluded that their fragmentation patterns were corresponded to their structural characteristics. They presented similar MSⁿ fragmentation behaviors such as the subsequent losses of H₂O (18 Da) and CO (28 Da) from the [M + H]⁺, [M + Na]⁺, [M + K]⁺ or [M - H]⁻ ions in ESI-MSⁿ conditions. The neutral loss of 162 Da revealed the presence of a hexose such as glucosyl residue in the compound, a 146 Da loss indicated the presence of a methyl pentose such as rhamnose residue, and a loss of 308 Da revealed the presence of a disaccharide residue (glu + rha), which was in agreement with the literature data.^{24,25,27} Furthermore, the [M - H]⁻, [M + H]⁺, [M + K]⁺ or [M + Na]⁺ ions readily eliminated the sugar moieties to produce the corresponding [aglycone - H]⁻, [aglycone + H]⁺, [aglycone + Na]⁺ or [aglycone + K]⁺ ions, and these ions were selected to trace the structural information of the

aglycone. Hence, according to these fragmentation patterns, the glycoside compounds could be screened out from tobacco comprehensively, and the structures of unknown tobacco glycosides could be proposed.

Identification of glycosides in the sample solution of flue-cured tobacco

The sample solution (see Sample preparation sub-section) of flue-cured tobacco was analyzed by LC-ESI-LIT/MSⁿ and ESI-Orbitrap-MS in positive ion mode and negative ion mode. The LC-ESI-LIT/MSⁿ analyses were carried out in data-dependent scan mode and neutral loss scan mode (glu, rha, and glu-rha, 162, 146 and 308 Da, respectively) in order to screen out the glycosides compounds (including unknown glycosides) comprehensively from tobaccos. The total ion chromatogram (TIC) and the extracted ion chromatogram (EIC) of sample solution in positive ion mode were acquired by LC-ESI-LIT/MSⁿ, which were shown in Figure 3. According to the fragmentation patterns of the authentic compounds, the total of 22 glycoside compounds were preliminarily screened out and designed as compound **1** to **22** according to the retention times sequence (see Figure 3). The MSⁿ fragmentation ions and the accurate *quasi*-molecular ion mass of the 22 compounds acquired by LC-ESI-LIT/MSⁿ and ESI-Orbitrap-MS analyses are presented in Table 1.

Among the 22 compounds detected in this study, compounds **2**, **3**, **8**, **9** and **10** were unambiguously identified as benzyl β-*D*-glucopyranoside, 2-phenylethyl β-*D*-glucopyranoside, rutin, 3-oxo-α-ionyl β-*D*-glucopyranoside and kaempferyl-3-rutinoside, respectively, by comparing the retention time (*t_R*) and fragmentation patterns with those of the authentic glycosides **I-V**.

Compound **1** eluted in HPLC at 0.88 min, presented a [M + H]⁺ ion at *m/z* 343, and a significant fragment in MS² at *m/z* 281 was ascribable to the loss of a glucosyl residue [M + H - 162]⁺. The MS² fragmentation showed peak at *m/z* 325 due to the loss of H₂O from *m/z* 343. The minor fragment ion ([M + H - 44]⁺) at *m/z* 299 was characterized by the loss of C₂H₄O. The ion at *m/z* 268 was formed from the ion at *m/z* 325 by McLafferty rearrangement. Based on the analyses above and combined with literature data,²⁸ compound **1** could be reasonably assigned as coniferin.

Compound **4** eluted in the chromatography at 4.10 min, exhibited a predominant [M + K]⁺ ion at *m/z* 311. The product ions at *m/z* 293 and 275 were formed by the successive losses of H₂O. The ions at *m/z* 279 and 251 were formed by the successive losses of CH₃OH (32 Da) and CO (28 Da) from *m/z* 311, respectively. The ions at *m/z* 163 and

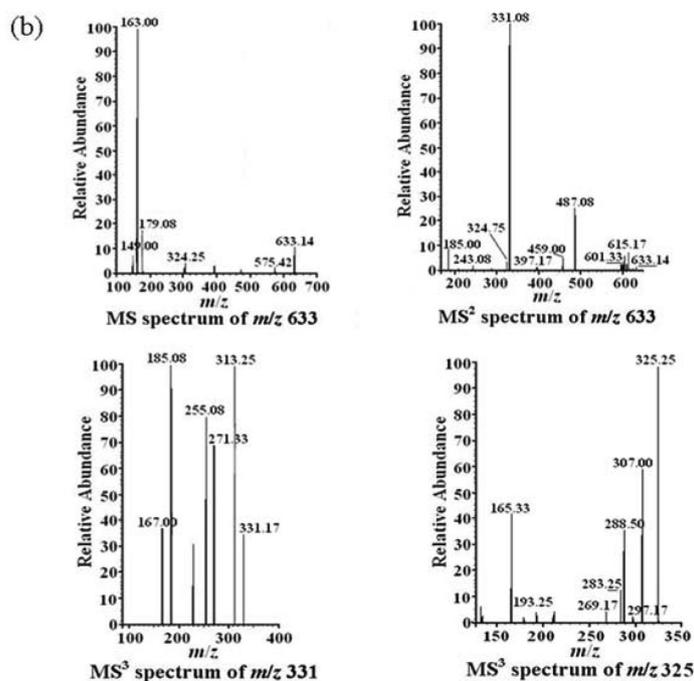
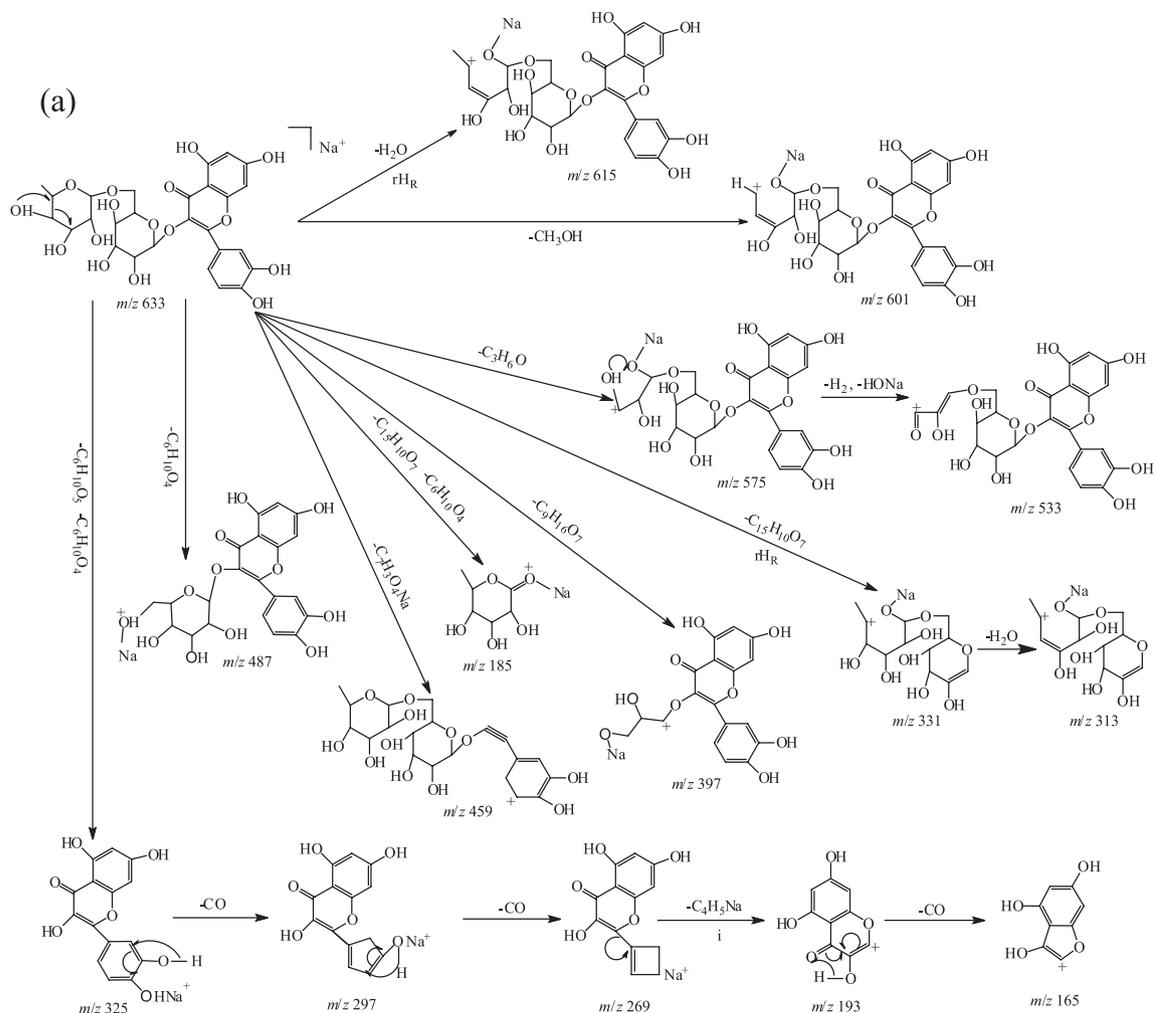


Figure 2. The fragmentation pathway of rutin (a) and its ESI-MSⁿ spectrum in positive ESI mode (b).

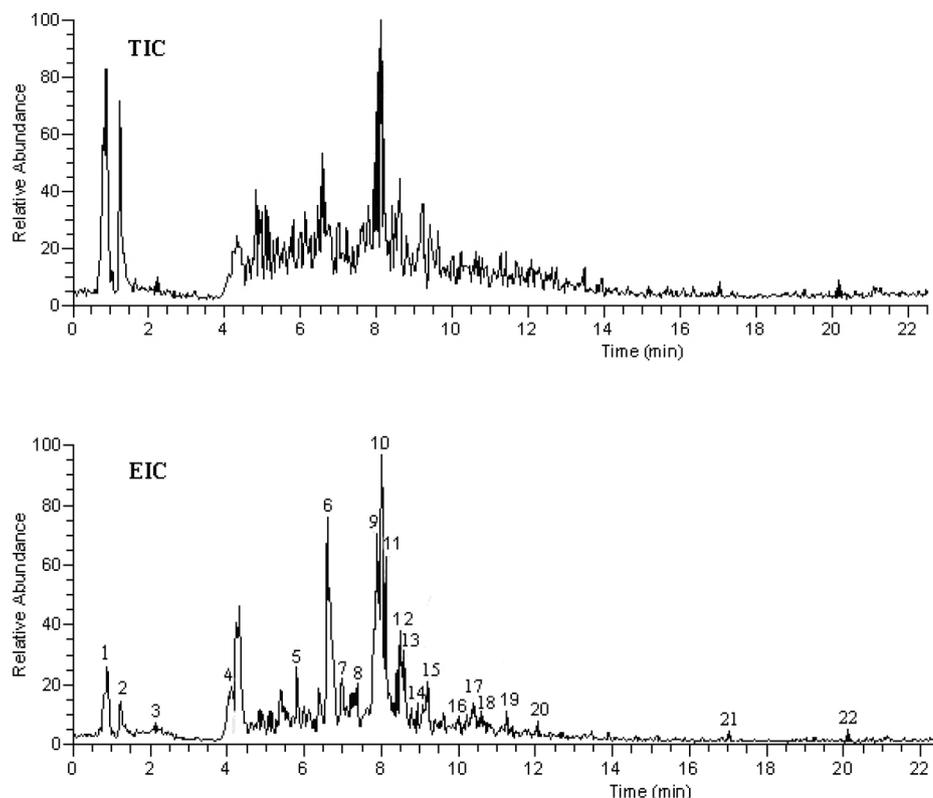


Figure 3. The LC-MS total ion chromatogram (TIC) and elective ion chromatogram (EIC) of sample solution from flue-cured tobacco in positive-ion mode.

133 proved the presence of a glucose unit. Based on the above MSⁿ data, compound **4** could be assigned as arbutin, which has been identified in flue-cured tobacco.^{20,29}

Compound **5** eluted in the chromatography at 5.81 min, exhibited a predominant [M + Na]⁺ ion at *m/z* 471. Successive losses of two H₂O molecules and a glucosyl residue (162 Da) produced another three minor ions at *m/z* 453, 435 and 309, respectively. An unobvious fragment ion ([M + Na – 294]⁺) at *m/z* 177 was characterized by the loss of the primeverose residue (294 Da). By comparing these results with the literature data,³⁰ compound **5** could be identified as linalyl β-primeveroside.

Compound **6** yielded a [M + Na]⁺ ion at *m/z* 302 (*t_r* = 6.59 min), which was greater than [M + H]⁺ ion at *m/z* 280. The product ions at *m/z* 284 and 266 were formed by the successive losses of H₂O from ion at *m/z* 302. The ion at *m/z* 139 was formed by the loss of a glucosyl residue from *m/z* 302, and the ion at *m/z* 163 showed the existence of a glucose unit. On the basis of these information and the literature data,³¹ compound **6** was identified as 3-indoxyl-β-*D*-glucopyranoside.

Compound **7** (*t_r* = 7.01 min) showed a major ion peak at *m/z* 409 [M + Na]⁺ and a prominent fragment at *m/z* 247 ascribable to the loss of a glucosyl residue ([M + Na – 162]⁺). The prominent ions at *m/z* 391 and 373 were formed by the successive losses of H₂O from

m/z 409. The ion at *m/z* 352 was formed from the ion at *m/z* 409 by rearrangement reaction and the losses of H₂O, OH and ·ONa (57 Da). The signals at *m/z* 179, 163 and 134 proved the presence of a glucosyl group. Based on the MSⁿ data, blumenyl A β-*D*-glucopyranoside had been analyzed and its structure was precisely accorded with our deduced structure of compound **7**.²⁸ Therefore, compound **7** could be unequivocally identified as vomifoliol β-*D*-glucopyranoside.

The molecular formula of compound **11** (*t_r* = 8.10 min) was calculated as C₂₅H₄₀O₇ with an error of –0.58 mDa on the basis of an accurate *quasi*-molecular ion mass at *m/z* 453.2841 for [M + H]⁺. This ion produced another two minor ions at *m/z* 435 and 417, respectively, by the successive losses of H₂O. The signals at *m/z* 179 and 163 proved the presence of a glucosyl group. The prominent aglycone ion at *m/z* 291, corresponding to the neutral loss of a glucosyl residue (162 Da), could be observed in its MS² spectrum. An obvious fragment ion [M + H – 162 – 18]⁺ at *m/z* 273 was characterized by the loss of H₂O from the aglycone ion. In the same way, an obvious fragment ion [M + H – 162 – 18 – 18]⁺ at *m/z* 255 was characterized by the losses of two H₂O molecules from the aglycone ion. The ion at *m/z* 193 was formed by the loss of C₆H₁₀O (98 Da) from the aglycone ion by a hydrogen rearrangement and *i*-cracking reaction, then it yielded *m/z* 175 by the loss

Table 1. Characterization of 22 glycoside compounds in the sample solution from flue-cured tobacco by LC-ESI-LIT/MS^a and ESI-Orbitrap-MS

Compound	MS protonated molecule (<i>m/z</i>)	MS ² fragments (<i>m/z</i>)	MS ³ and MS ⁴ fragments (<i>m/z</i>)	Molecular formula	Calculated <i>m/z</i>	Error / mDa	Identification
1	343.1374 ^a [M + H] ⁺	325, 260, 181, 163	268, 242, 216, 177, 163, 106	C ₁₆ H ₂₂ O ₈	343.1387	-1.34	coniferin
2	293.11 [M + Na] ⁺ 271.1179 ^a [M + H] ⁺	275, 247, 185, 151	257, 247, 229, 167, 127	C ₁₃ H ₁₈ O ₆	271.1176	0.29	benzyl β- <i>D</i> -glucopyranoside
3	307.1160 ^a [M + Na] ⁺	289, 279, 247, 217, 185, 179, 163	271, 259, 243, 206, 163, 132	C ₁₄ H ₂₀ O ₆	307.1152	0.79	2-phenylethyl β- <i>D</i> -glucopyranoside
4	311.0529 ^a [M + K] ⁺	293, 279, 251, 192, 163, 133	275, 265, 227, 161	C ₁₂ H ₁₆ O ₇	311.0528	0.14	arbutin
5	471.2194 ^a [M + Na] ⁺	453, 414, 309, 177, 163	435, 396, 370, 163	C ₂₁ H ₃₆ O ₁₀	471.2201	-0.67	linalyl β-primeveroside
6	302.1014 ^a [M + Na] ⁺	284, 270, 245, 163, 139	266, 256, 162	C ₁₄ H ₁₇ O ₃ N	302.0999	1.51	3-indoxyl-β- <i>D</i> -glucopyranoside
7	409.1842 ^a [M + Na] ⁺	394, 391, 352, 247, 179, 163, 134	373, 247, 209, 217, 175, 163, 134	C ₁₉ H ₃₀ O ₈	409.1833	0.91	vomifoliol β- <i>D</i> -glucopyranoside
8	633.1404 ^a [M + Na] ⁺	615, 487, 459, 368, 331, 185	325, 313, 307, 285, 269, 255, 193, 165	C ₂₇ H ₃₀ O ₁₆	633.1426	-2.18	rutin
9	393.20 [M + Na] ⁺ 371.2057 ^a [M + H] ⁺	375, 303, 273, 231, 203, 185, 163	357, 331, 317, 285, 273, 203	C ₁₉ H ₃₀ O ₇	371.2064	-0.73	3-oxo-α-ionyl β- <i>D</i> -glucopyranoside
10	617.08 [M + Na] ⁺ 595.1661 ^a [M + H] ⁺	593, 562, 512, 471, 308, 331, 253, 179	313, 243, 187, 185, 132	C ₂₇ H ₃₀ O ₁₅	595.1657	0.35	kaempferol-3- <i>O</i> -rutinoside
11 ^b	453.2841 ^a [M + H] ⁺	435, 421, 417, 291, 257, 273, 179, 163	291, 273, 255, 193, 175	C ₂₅ H ₄₀ O ₇	453.2847	-0.58	5-androstene-3,17-diol 17- <i>O</i> -β- <i>D</i> -glucopyranoside
12	373.2217 ^a [M + H] ⁺	355, 316, 290, 257, 179, 163	337, 272, 211, 192, 179, 163, 132	C ₁₉ H ₃₂ O ₇	373.2221	-0.38	7,8-dihydro-3-oxo-α-ionol β- <i>D</i> -glucopyranoside
13 ^c	535.2152 ^a [M + H] ⁺	517, 504, 478, 452, 355, 179, 163	460, 418, 348, 284, 193, 179	C ₂₇ H ₃₄ O ₁₁	535.2174	-2.19	phillyrin
14 ^c	319.1165 ^a [M + Na] ⁺	301, 287, 262, 229, 179, 163, 157	273, 244, 218, 179, 163, 149	C ₁₅ H ₂₀ O ₆	319.1152	1.29	<i>trans</i> -4-(1-propenyl)-phenyl β- <i>D</i> -glucopyranoside
15 ^b	489.2259 ^a [M + K] ⁺	471, 458, 432, 327, 319, 203, 179, 163	453, 401, 327, 309, 281, 179, 163	C ₂₅ H ₃₈ O ₇	489.2249	0.99	17-hydroxy-17-methyl-4-estrene-3-one 17- <i>O</i> -β- <i>D</i> -glucopyranoside
16	389.2174 ^a [M + H] ⁺	371, 357, 332, 274, 227, 179, 163, 133	353, 314, 311, 288, 209, 163, 132	C ₁₉ H ₃₂ O ₈	389.2170	0.41	3-hydroxy-5,6-epoxy-β-ionyl β- <i>D</i> -glucopyranoside
17	711.23 [2M + Na] ⁺ 359.1741 ^a [M + H] ⁺	693, 679, 651, 549, 341	675, 597, 531, 211	C ₁₇ H ₂₆ O ₈	359.1700	4.06	lolidide-β- <i>D</i> -glucopyranoside
18	437.1080 ^a [M + H] ⁺	419, 405, 363, 359, 275, 257, 179, 163	401, 383, 275, 257, 231, 215, 167, 123	C ₂₀ H ₂₀ O ₁₁	437.1078	0.16	bellidifolin-8- <i>O</i> -β- <i>D</i> -glucopyranoside
19 ^c	625.2077 ^a [M + H] ⁺	607, 568, 479, 471, 463, 325, 179, 163	589, 463, 179, 163	C ₂₉ H ₃₆ O ₁₅	625.2127	-5.00	verbascoside
20 ^c	517.14 [M + K] ⁺ 479.1573 ^a [M + H] ⁺	499, 471, 460, 343, 179, 163	453, 428, 414, 179, 163	C ₂₃ H ₂₆ O ₁₁	479.1548	2.51	3,4-dihydroxy-phenethylalcohol-3- <i>O</i> -caffeoyl-β- <i>D</i> -glucopyranoside
21 ^b	391.1733 ^a [M + Na] ⁺	373, 359, 321, 229, 163	355, 345, 186, 147, 133	C ₁₉ H ₂₈ O ₇	391.1727	0.58	9-hydroxy-megastigma-3,5,8-trien-7-one 9- <i>O</i> -β- <i>D</i> -glucopyranoside
22 ^b	623.2865 ^a [M + K] ⁺	605, 591, 579, 540, 477, 347, 419, 217	587, 569, 315, 297, 201, 163	C ₃₀ H ₄₈ O ₁₁	623.2828	3.68	17-hydroxy-estran-3-one rutinoside

^aAcquired by HRMS; ^bnovel compounds; ^ccompounds reported for the first time in this plant; MS: mass spectrometry.

of H₂O. Based on the analyses above and combined with mass spectrometry analysis software of Mass Frontier 7.0, compound **11** was assigned as 5-androstene-3,17-diol 17-*O*-β-*D*-glucopyranoside, which was reported here for the first time. The proposed fragmentation pathways for this compound are shown in Figure 4.

Compound **12** ($t_R = 8.54$ min) showed a major ion peak at m/z 373 [M + H]⁺ and a prominent aglycone ion at m/z 211 [M + H - 162]⁺, ascribable to the loss of a glucosyl residue. The produced ions at m/z 355 and 337 were formed by the successive losses of two H₂O molecules from m/z 373. The ion at m/z 316 was formed from m/z 373 by the loss of C₄H₉ (57 Da). The signals at m/z 179, 163 and 132 proved the presence of a glucosyl group. According to the fragmentation pathways obtained above and the literature data,³² compound **12** could be unequivocally identified as 7,8-dihydro-3-oxo-α-ionyl-β-*D*-glucopyranoside.

The molecular formula of compound **13** ($t_R = 8.61$ min) was calculated as C₂₇H₃₄O₁₁ on the basis of an accurate quasi-molecular ion mass at m/z 535.2152 for [M + H]⁺, and the ion at m/z 478 was formed from it by the loss of C₃H₅O (57 Da). In addition, the ions at m/z 504 and 517 were characterized by the losses of OCH₃ and H₂O from m/z 535, respectively. The signals at m/z 179 and 163 proved the presence of a glucosyl group. An obvious fragment ion [M + H - 162 - 18]⁺ at m/z 355 was characterized by the loss of neutral glucosyl residue and H₂O. According to the information obtained above and the literature data,³³

compound **13** could be identified as phillyrin, which was first found in tobacco.

Compound **14** eluted in the chromatography at 8.98 min, exhibited a predominant [M + Na]⁺ ion at m/z 319, which produced ion at m/z 301 from the loss of H₂O. The ion at m/z 157 was characterized by the loss of a neutral glucosyl residue ([M + Na - 162]⁺), that is the aglycone ion [*trans*-4-(1-propenyl)-phenol + Na]⁺. The obvious ions at m/z 179 and 163 proved the presence of a glucosyl group. According to the fragmentation pathways obtained above and the literature data,^{34,35} the structure of compound **14** was identified as *trans*-4-(1-propenyl)-phenyl β-*D*-glucopyranoside, which was first time to be found in tobacco.

The accurate mass value of the [M + K]⁺ ion corresponding to compound **15** ($t_R = 9.21$ min) was 489.2259 consistent with an elemental composition of C₂₅H₃₈O₇K⁺ with an error of 0.99 mDa. The [M + K]⁺ ion produced prominent ions at m/z 471, 453 and 435, which were formed by the successive losses of three H₂O molecules. The obvious signals at m/z 179 and 163 proved the presence of a glucosyl group. The ion at m/z 432 was formed from the ion at m/z 489 by rearrangement reaction and the losses of H₂O and K atom. The aglycone ion at m/z 327 corresponded to the neutral loss of a glucosyl residue (162 Da) from m/z 489, which suggests that 17-hydroxy-17-methyl-4-estrene-3-one (C₁₉H₂₈O₂, 288 Da) was the aglycone compound. Then the ion at m/z 327

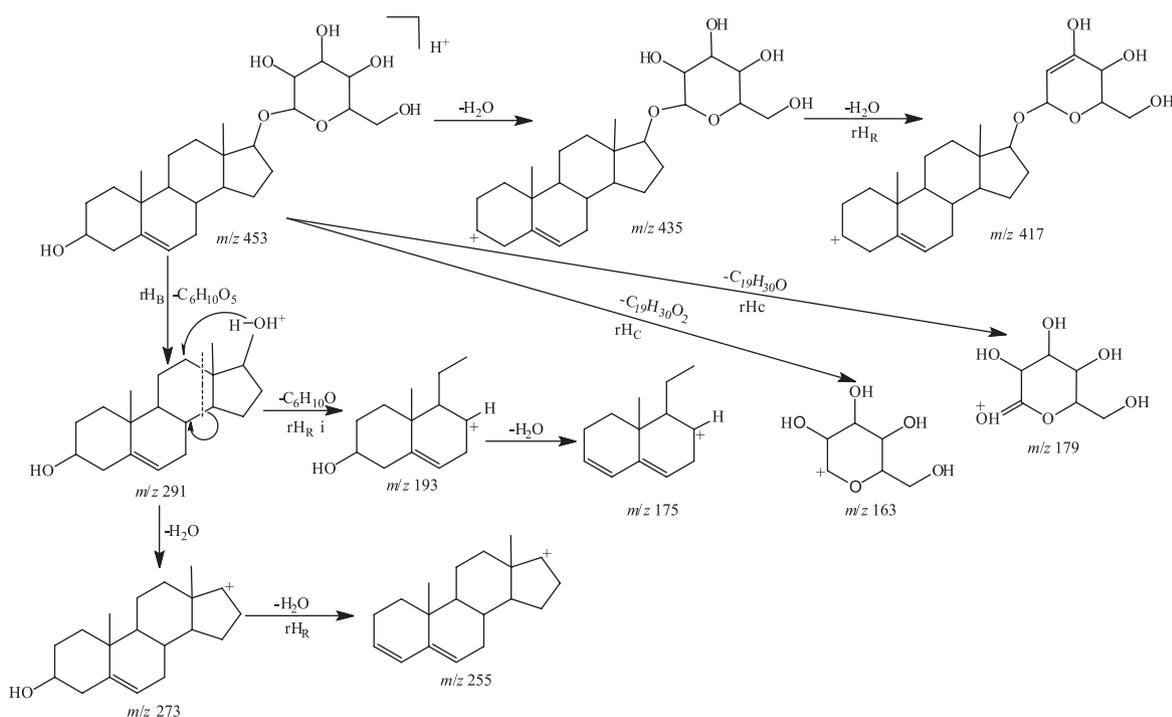


Figure 4. Proposed fragmentation pathways for compound **11**.

yielded two product ions at m/z 309 and 281 from the successive losses of H_2O and C_2H_4 (28 Da). Based on the analyses above, the literature data,³⁶ and mass spectrometry analysis software of Mass Frontier 7.0, compound **15** was assigned as 17-hydroxy-17-methyl-4-estrene-3-one 17-*O*- β -*D*-glucopyranoside, which was reported for the first time. The proposed fragmentation pathways for this compound are shown in Figure 5.

Compound **16** occurred at a retention time of 9.89 min and exhibited a $[\text{M} + \text{H}]^+$ ion at m/z 389. The produced prominent ions at m/z 371, 353 and 335 were formed by the successive losses of three H_2O molecules. In MS^2 , a ion at m/z 311 was characterized by the losses of H_2O and CH_3COOH ($[\text{M} + \text{H} - 18 - 60]^+$). The ion at m/z 332 was formed from m/z 389 by the loss of C_4H_9 (57 Da). The prominent aglycone ion at m/z 227, corresponding to the neutral loss of a glucosyl residue (162 Da) from m/z 389, born a obvious fragment ion $[\text{M} + \text{H} - 162 - 18]^+$ at m/z 209 by the loss of H_2O molecule. This suggests that 3-hydroxy-5,6-epoxy- β -ionol ($\text{C}_{13}\text{H}_{22}\text{O}_3$, 226 Da) was the

aglycone compound. The signals at m/z 179, 163 and 133 indicated the presence of a glucosyl group. By comparing these results with the literature data,²⁷ compound **16** was identified as 3-hydroxy-5,6-epoxy- β -ionyl β -*D*-glucopyranoside.

Compound **17** exhibited a $[\text{M} + \text{H}]^+$ ion at m/z 359 and a $[2\text{M} + \text{Na}]^+$ ion at m/z 711 ($t_{\text{R}} = 10.42$ min). The product ion at m/z 693 was formed by the loss of H_2O from m/z 711. The ions at m/z 679 and 651 were formed by the successive losses of CH_3OH (32 Da) and CO (28 Da) from m/z 711. The ion at m/z 549 and 341 were the fragment ions $[2\text{M} + \text{Na} - 163]^+$ and $[\text{M} + \text{H} - \text{H}_2\text{O}]^+$, respectively. The ions at m/z 675 and 531 were formed by the losses of H_2O from the ion at m/z 693 and 549, respectively. Based on the analyses above and the literature data,¹⁹ compound **17** was identified as lolidide- β -*D*-glucopyranoside.

Compound **18** ($t_{\text{R}} = 10.60$ min) generated a molecular ion peak with great intensity at m/z 437.1080 for $[\text{M} + \text{H}]^+$ in positive ion mode, which corresponded to the molecular formula of $\text{C}_{20}\text{H}_{20}\text{O}_{11}$ with an error of 0.16 mDa. The

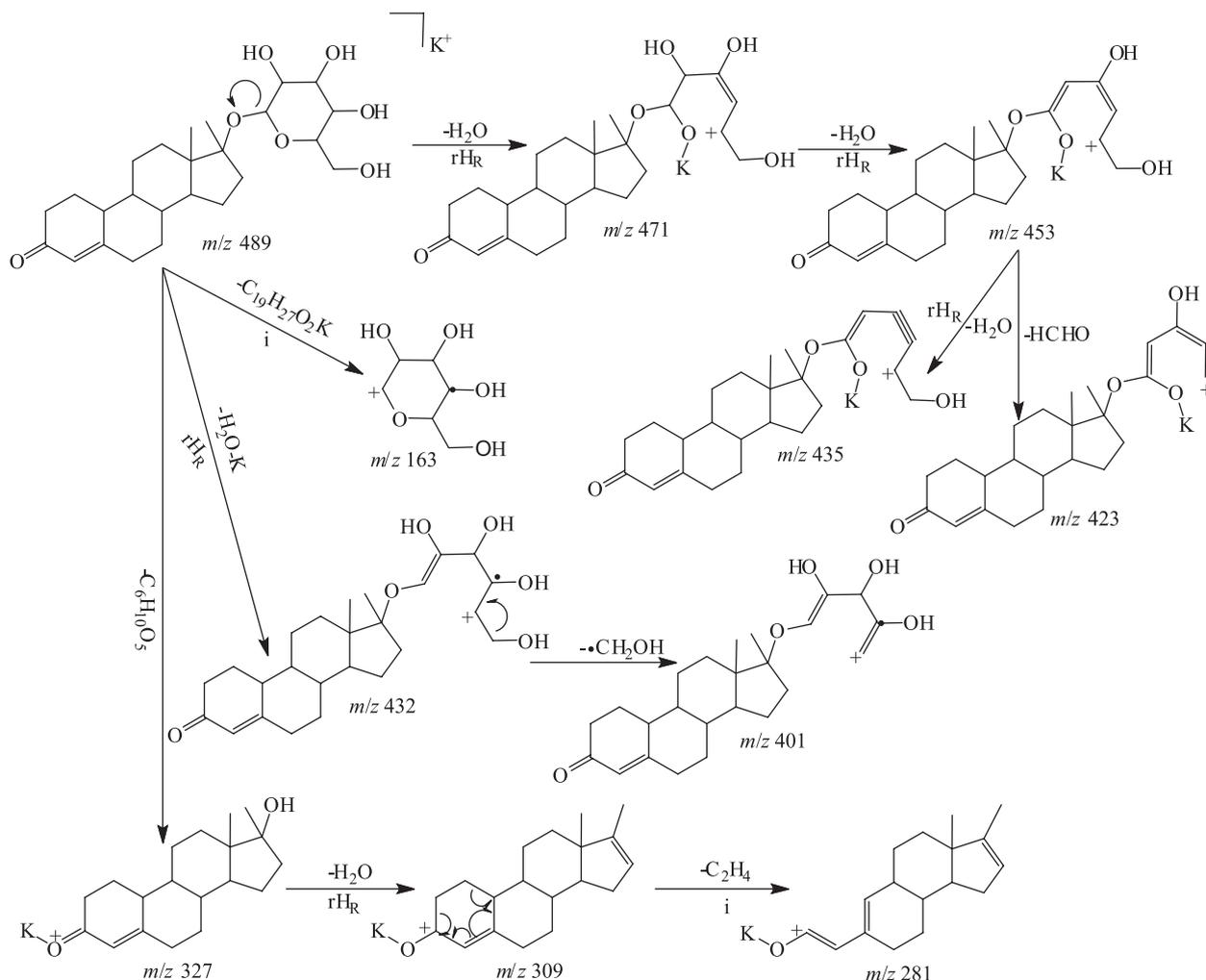


Figure 5. Proposed fragmentation pathways for compound **15**.

$[M + H]^+$ ion produced three major ions at m/z 419, 401 and 383 by the successive losses of H_2O . The signals at m/z 179 and 163 proved the presence of a glucosyl group and m/z 134 was formed by the loss of CHO (29 Da) from m/z 163. The prominent aglycone ion at m/z 275 corresponded to the neutral loss of a glucosyl residue (162 Da). Then the aglycone ion yielded the product ion at m/z 257 from the loss of H_2O ($[M + H - 162 - 18]^+$). These data suggested the presence of a bellidifolin molecule ($C_{14}H_{10}O_6$, 274 Da). Compound **18** was therefore characterized as bellidifolin-8- β -D-glucopyranoside, which was also identified in *Swertia punctata* and *Scutellaria baicalensis*.

Compound **19** ($t_R = 11.26$ min) exhibited a $[M + H]^+$ ion at m/z 625 and the product ion at m/z 607 was formed by the loss of H_2O . An obvious fragment ion at m/z 479 was characterized by the loss of a rhamnosyl residue ($[M + H - 146]^+$), and the ion at m/z 471 was formed by the loss of $C_8H_{10}O_3$ (154 Da) ($[M + H - 154]^+$). Another ion at m/z 325 was formed by the loss of rhamnosyl residue (146 Da) from m/z 471, which produced the highest relative abundance. Then the ion at m/z 325 generated ion at m/z 163 ($[caffeoyl]^+$) by the loss of glucosyl residue (162 Da). By comparing these results with the literature data,³⁷ the structure of compound **19** was identified as verbascoside, which was first found in tobacco.

Compound **20** ($t_R = 12.17$ min) gave a $[M + K]^+$ ion at m/z 517 and a $[M + H]^+$ ion at m/z 479. The $[M + K]^+$ ion gave rise to several fragment ions at m/z 499 ($[M + K - H_2O]^+$), 471 ($[M + K - C_2H_6O]^+$), 460 ($[M + K - C_3H_5O]^+$) and 453 ($[M + K - 2H_2O - CO]^+$). The ions at m/z 179 and 163 corresponded to the fragments of a glucosyl group. By comparing these results with the literature data,³⁸ this compound was identified as 3,4-dihydroxyl phenylethyl-3- O -caffeoyl- β -D-glucopyranoside, which was first found in tobacco.

Compound **21** eluted in the chromatography at 17.02 min, exhibited a predominant $[M + Na]^+$ ion at m/z 391. The product ions at m/z 373 and 359 were formed by the losses of H_2O and CH_3OH (32 Da), respectively. The aglycone ion at m/z 229 corresponded to the neutral loss of a glucosyl residue (162 Da) from m/z 391. Then the aglycone ion yielded two product ions at m/z 186 and 147 by the successive losses of $\cdot COCH_3$ (43 Da) and $\cdot ONa$ (39 Da). The fragmentation patterns of the aglycone ion suggested that 9-hydroxy-megastigma-3,5,8-trien-7-one ($C_{13}H_{18}O_2$, 206 Da) was the aglycone compound. Thus, this compound could be assigned as 9-hydroxy-megastigma-3,5,8-trien-7-one 9- O - β -D-glucopyranoside. The proposed fragmentation pathways for compound **21** are shown in Figure 6. As far as we know, this compound was reported for the first time.

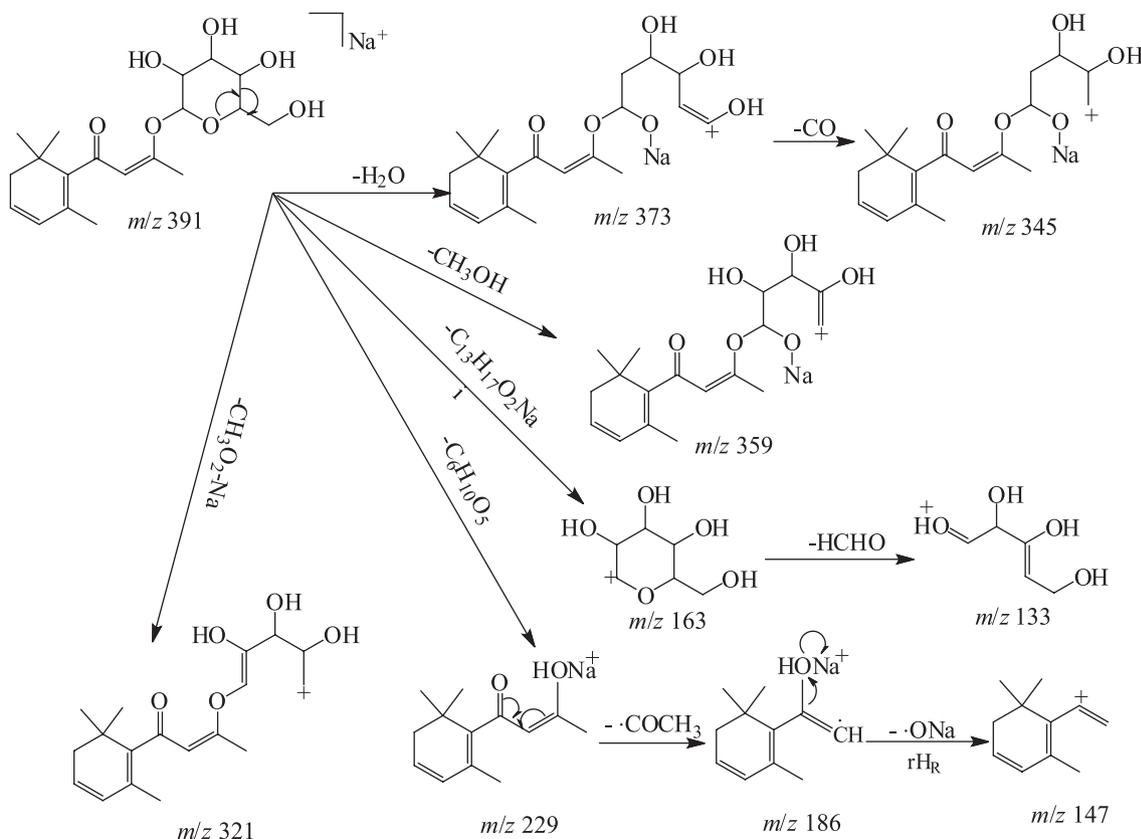


Figure 6. Proposed fragmentation pathways for compound **21**.

Compound **22** ($t_r = 20.20$ min) produced a molecular ion peak at m/z 623.2865 for $[M + K]^+$ corresponding to the molecular formula of $C_{30}H_{48}O_{11}$ with an error of 3.68 mDa. The produced prominent ions at m/z 605, 587 and 569 were formed by the successive losses of H_2O . The fragment ions at m/z 477 ($[M + K - 146]^+$) and m/z 315 ($[M + K - 146 - 162]^+$) corresponded to the sequential losses of a rhamnosyl residue (rha, 146 Da) and a glucosyl residue (glu, 162 Da) from m/z 623, which suggested that the ion at m/z 315 was the aglycone ion. Then the aglycone ion yielded two product ions at m/z 297 and 161 by the successive losses of H_2O and $C_{10}H_{16}$ (136 Da). The fragmentation patterns of the aglycone ion suggested that 17-hydroxy-estran-3-one ($C_{18}H_{28}O_2$, 276 Da) might be the aglycone compound. In addition, the ion at m/z 347 produced by the loss of aglycone compound from the $[M + K]^+$ (m/z 623) might be the potassium adduct ion of the sugar moiety $[glu + rha + K]^+$. Then the ion at m/z 347 yielded a product ion at m/z 201 by the loss of a rhamnosyl residue. By comparing with the fragmentation patterns of the reference of rutin, the sugar moiety is a rutinose. Thus, this compound was considered to be 17-hydroxy-5-estran-3-one-17-rutinoside. The proposed fragmentation pathways for compound **22** are shown in Figure 7. As far as we know, this compound was reported for the first time.

Conclusions

A simple, reliable and sensitive method by means of LC-ESI-LIT/MSⁿ coupled with ESI-Orbitrap-MS has been used for the comprehensive separation and identification of glycoside compounds in tobacco for the first time. In this study, five known authentic glycosides have been used to investigate the relationship between structure characteristics and MSⁿ fragmentation ions in detail. The fragmentation patterns were proposed and used for the identification of unknown glycosidic structures in a complex mixture, based on the MSⁿ data of the above-mentioned known compounds. On the other hand, HRMS instruments (ESI-Orbitrap-MS) offer a capability of unequivocal identification of compounds by means of accurate mass measurements. As a result, twenty-two glycoside compounds were reliably identified with high mass accuracy (within 5 mDa) from tobacco. Among them, four compounds were confirmed as novel molecules and other four compounds are not previously reported in tobacco samples. In addition, the results of this study clearly demonstrate the potential of mass spectrometry for the structural identification of glycosides and open perspectives for similar studies on other natural compounds.

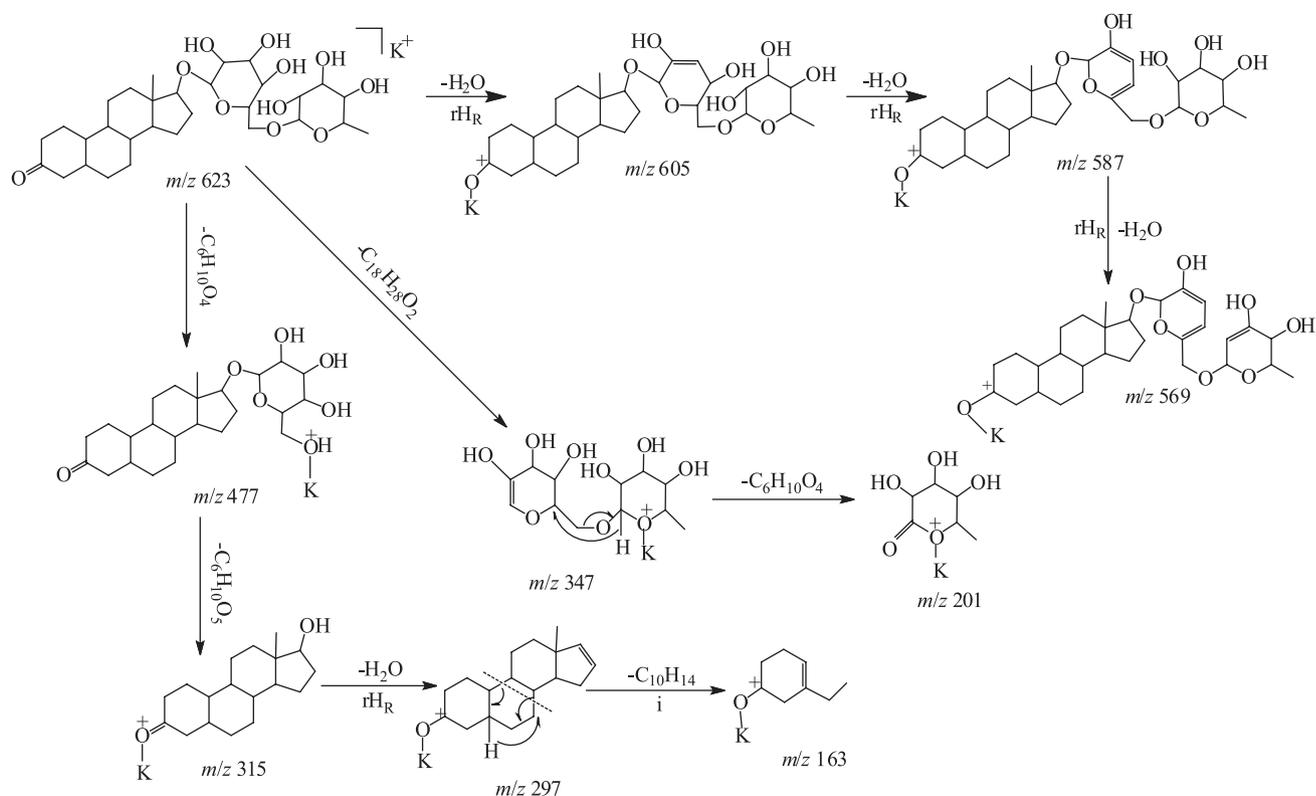


Figure 7. Proposed fragmentation pathways for compound **22**.

Acknowledgments

The research work was supported by the National Science Foundation (No. 21176229/B060804) of P. R. China.

References

1. Lee, C. N.; Wong, K. L.; Liu, J. C.; Chen, Y. J.; Cheng, J. T.; Chan, P.; *Planta Med.* **2001**, *67*, 796.
2. Ge, B. Y.; Chen, H. X.; Han, F. M.; Chen, Y.; *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2007**, *857*, 281.
3. Kilic, A.; Kollmannsberger, H.; Nitz, S.; *J. Agric. Food Chem.* **2005**, *53*, 2231.
4. Ma, S. J.; Naoharu, W.; Akihito, Y.; Kanzo, S.; *Phytochemicals* **2001**, *56*, 819.
5. Osorio, C.; Duque, C.; Fujimoto, Y.; *J. Agric. Food Chem.* **1999**, *47*, 1641.
6. Wu, P.; Kuo, M. C.; Ho, C. T.; *J. Agric. Food Chem.* **1990**, *38*, 1553.
7. Loscos, N.; Hernandez-Orte, P.; Cacho, J.; Ferreira, V.; *J. Agric. Food Chem.* **2007**, *55*, 6674.
8. Loughrin, J. H.; Hamilton-Kemp, T. R.; Burton, H. R.; Andersen, R. A.; Hildebrand, D. F.; *Phytochemicals* **1992**, *31*, 1537.
9. Cai, J. B.; Liu, B. Z.; Ling, P.; Su, Q. D.; *J. Chromatogr. A* **2002**, *947*, 267.
10. Shinozaki, Y.; Tobita, T.; Mizutani, M.; Matsuzaki, T.; *Biosci., Biotechnol., Biochem.* **1996**, *60*, 903.
11. Pieri, V.; Belancic, A.; Morales, S.; Stuppner, H.; *J. Agric. Food Chem.* **2011**, *59*, 4378.
12. Chen, J. X.; Leng, H. Q.; Duan, Y. X.; Zhao, W.; Yang, G. Y.; Guo, Y. D.; Chen, Y. K.; Hu, Q. F.; *Phytochem. Lett.* **2013**, *6*, 144.
13. Snook, M. E.; Johnson, A. W.; Severson, R. F.; Teng, Q.; White, R. A.; Sisson, J. A.; Jackson, D. M.; *J. Agric. Food Chem.* **1997**, *45*, 2299.
14. Green, C. R.; Colby, D. A.; Cooper, P. J.; Heckman, R. A.; Lyerly, L. A.; Thoma, F. A.; *Recent Adv. Tob. Sci.* **1980**, *6*, 123.
15. Loughrin, J. H.; Hamilton-Kemp, T. R.; Andersen, R. A.; Hildebrand, D. F.; *J. Agric. Food Chem.* **1990**, *38*, 455.
16. Liu, B. Z.; Xu, Y. T.; Sun, Z. J.; Zhu, X. L.; Chen, J. L.; Kan, Y.; *Acta Tab. Sin.* **1998**, *4*, 1.
17. Tian, Z. F.; Qu, X. Z.; Fang, D.; Chen, C.; Wang, K. B.; *Tob. Sci. Technol.* **2005**, *7*, 262.
18. Wu, X. H.; Zhu, R. Z.; Ren, Z. Y.; Wang, K.; Mou, D. R.; Wei, W. Z.; Miao, M. M.; *Chin. J. Chromatogr.* **2009**, *27*, 820.
19. Li, W. W.; Dai, Y. H.; Wu, M. J.; Chen, X. Q.; *J. Instrum. Anal.* **2009**, *28*, 1428.
20. Wu, X. H.; Zhu, R. Z.; Wang, K.; Ren, Z. Y.; Mou, D. R.; Wei, W. Z.; Miao, M. M.; *J. Chin. Mass Spectrom. Soc.* **2010**, *31*, 88.
21. Torras-Claveria, L.; Jáureguib, O.; Codina, C.; Tiburcio, A. F.; Bastida, J.; Viladomat, F.; *Plant Sci.* **2012**, *182*, 71.
22. Qi, L. W.; Gu, X. J.; Li, P.; Liang, Y.; Hao, H. P.; Wang, G. J.; *Rapid Commun. Mass Spectrom.* **2009**, *23*, 2151.
23. Ackermann, I. E.; Banthorpe, D. V.; Fordharn, W. D.; Kinder, J. P.; Poots, I.; *Liebigs Ann. Chem.* **1989**, *1*, 79.
24. Li, L.; Yang, J.; Dai, Y.; Long, J.; Li, J. H.; Zheng, J.; *Acta Tab. Sin.* **2008**, *14*, 13.
25. Long, J. J.; Yang, T.; *Chin. Med. J.* **2002**, *9*, 39.
26. Yang, K.; Yang, J.; Wang, H. J.; Bian, B. L.; *Chin. J. Exp. Tradit. Med. Form.* **2011**, *17*, 127.
27. Wang, Y. L.; Yang, Y. Q.; He, C. H.; Wang, E. W.; Welbeck, S. W.; Bligh, A.; Wang, Z. T.; *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1767.
28. Ito, K.; Tanabe, Y.; Kato, S.; Yamamoto, T.; Saito, A.; Mori, M.; *Biosci., Biotechnol., Biochem.* **2000**, *64*, 584.
29. Schieber, A.; Keller, P.; Carle, R.; *J. Chromatogr. A* **2001**, *910*, 265.
30. Wu, X. H.; Zhu, R. Z.; Ni, C. M.; Wang, Y.; Miao, M. M.; *Yunnan Chem. Technol.* **2009**, *36*, 62.
31. Huang, L.; Xu, Y. B.; Tian, Z. F.; Li, L.; Zhu, D. L.; Zhu, X. L.; *Tob. Sci. Technol.* **2012**, *1*, 34.
32. Wang, Y.; Liu, Z. H.; Liu, C. B.; Chen, Y. K.; Jiang, L. H.; Wang, Y. M.; *J. Instrum. Anal.* **2012**, *31*, 22.
33. Zuo, Y.; Li, H.; *China Pharm.* **2009**, *12*, 921.
34. Liu, B. M.; Lu, W. J.; Ya, Q. K.; Chen, J. Y.; *Guangxi Sci.* **2005**, *12*, 214.
35. Liu, B. M.; Lu, W. J.; Lin, X.; Chen, J. Y.; Ya, Q. K.; *J. Instrum. Anal.* **2006**, *25*, 98.
36. Cong, P. Z.; Su, K. M.; *Analytical Chemistry Handbook*; Chemical Industry Press: Beijing, 2000.
37. Wan, Y.; Xie, Y. M.; Zeng, X. X.; *Food Sci.* **2007**, *6*, 77.
38. Chen, W. J.; Li, D. P.; Chen, Y. Y.; Huang, Y. L.; Wen, Y. X.; *Guihaia* **2010**, *30*, 269.

Submitted: April 14, 2016

Published online: July 18, 2016