Accepted Manuscript

Title: A new liquid chromatography-high resolution Orbitrap mass spectrometry-based strategy to characterize Glucuronide Oleanane-type Triterpenoid Carboxylic Acid 3, 28-*O*-Bidesmosides (GOTCAB) saponins.A case study of *Gypsophila glomerata* Pall ex M. B. (Caryophyllaceae)

Authors: Reneta Gevrenova, Krum Bardarov, Sabine Bouguet-Bonnet, Yulian Voynikov, Vessela Balabanova, Dimitrina Zheleva-Dimitrova, Max Henry

PII: S0731-7085(18)31340-2
DOI: https://doi.org/10.1016/j.jpba.2018.07.041
Reference: PBA 12112

To appear in: *Journal of Pharmaceutical and Biomedical Analysis*

Received date: 5-6-2018
Revised date: 19-7-2018
Accepted date: 20-7-2018

Please cite this article as: Gevrenova R, Bardarov K, Bouguet-Bonnet S, Voynikov Y, Balabanova V, Zheleva-Dimitrova D, Henry M, A new liquid chromatography-high resolution Orbitrap mass spectrometry-based strategy to characterize Glucuronide Oleanane-type Triterpenoid Carboxylic Acid 3, 28-*O*-Bidesmosides (GOTCAB) saponins.A case study of *Gypsophila glomerata* Pall ex M. B. (Caryophyllaceae), *Journal of Pharmaceutical and Biomedical Analysis* (2018), https://doi.org/10.1016/j.jpba.2018.07.041

A new liquid chromatography-high resolution Orbitrap mass spectrometry-based strategy to characterize Glucuronide Oleanane-type Triterpenoid Carboxylic Acid 3, 28-*O*-Bidesmosides (GOTCAB) saponins. A case study of *Gypsophila glomerata* Pall ex M. B. (Caryophyllaceae).

Reneta Gevrenova[a*], Krum Bardarov[b], Sabine Bouguet-Bonnet[c], Yulian Voynikov[d], Vessela Balabanova[a], Dimitrina Zheleva-Dimitrova[a], Max Henry[e]

[a]*Department of Pharmacognosy, Faculty of Pharmacy, Medical University-Sofia, 2 Dunav str., 1000 Sofia, Bulgaria*
[b]*Institute of molecular biology and biotechnology, bul. Ruski 105, 4000 Plovdiv, Bulgaria*
[c]*Université de Lorraine, CNRS, CRM2, F-54000 Nancy, France*
[d]*Department of Chemistry, Faculty of Pharmacy, Medical University-Sofia, 2 Dunav str., 1000 Sofia, Bulgaria*
[e]*Université de Lorraine, CNRS, L2CM, F54000 Nancy, France*

**Corresponding author: Reneta Gevrenova*
Faculty of Pharmacy, Medical University-Sofia, 2 Dunav str., 1000 Sofia, Bulgaria; E-mail: rgevrenova@gmail.com

Highlights

- UHPLC-ESI/HRMS (Orbitrap) acquisition strategy for GOTCAB saponins was developed.
- 41 GOTCAB were elucidated in *Gypsophila glomerata* roots including 16 new saponins.
- MS/MS fragmentation pathways for three GOTCAB types were proposed.
- The structure of new saponin was established by 1D and 2D-NMR experiments.
- The saponin content was determined by means of prosaponin Gypsogenin-3-glucuronide.

Abstract

Glucuronide **O**leanane-type **T**riterpenoid **C**arboxylic **A**cid 3, 28-**B**idesmosides (GOTCAB) saponins are bioactive natural compounds spread in Caryophyllidae. The high complexity of GOTCAB occurring as closely related isobaric and positional isomers is a challenge in their separation and identification. A new liquid chromatography – high resolution Orbitrap mass spectrometry acquisition strategy would be important for the structural elucidation of GOTCAB in plant extracts. In this study, the fragmentation behaviors of GOTCAB from methanol-aqueous root extract of *Gypsophila glomerata* Pall ex M. B. (Caryophyllaceae) were investigated using ultra-high performance liquid chromatography (UHPLC) coupled with hybrid quadrupole-Orbitrap high resolution mass spectrometry (HRMS). A new saponin was isolated and its structure was established by 1D and 2D-NMR spectroscopic experiments as 3-*O*-β-D-galactopyranosyl-(1→2)-[α-L-arabinopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-α-L-arabinopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester. On the basis of the accurate mass measurements, fragmentation patterns in MS/MS analyses and comparison with previously

isolated authentic references, a total of 41 GOTCAB saponins were identified or tentatively elucidated in *G. glomerata* roots, including 14 pairs of isobars. Possible fragmentation pathways for three groups of GOTCAB are suggested. The group I appeared to be GOTCAB of gypsogenin with two carbohydrate chains: a branched trisaccharide at C-3 and tri- to hexa-saccharide attached to C-28 of the aglycone through a deoxyhexose residue. Saponins with monoacetylated (group II) or sulphated (group III) C-28 chain were evidenced, as well as quillaic and oleanolic acid GOTCAB. Sixteen GOTCAB were previously not described. The content of *Gypsophila* prosaponins, gypsogenin 3-*O*-glucuronide (7.4079 ± 0.0723mg/g dry weight, dw) and quillaic acid 3-*O*-glucuronide (4.4593 ± 0.1207 mg/g dw), was determined by solid phase extraction – high-performance liquid chromatography (SPE-HPLC). In this study is presented the first systematic investigation on the fragmentation patterns and diagnostic fingerprints of the fragment ions in the MS/MS spectra of the gypsogenin -, quillaic acid - and oleanolic acid – bidesmosides. A LC-HRMS Orbitrap acquisition strategy could give an insight in the GOTCAB containing taxa.

Keywords: *Gypsophila glomerata*; GOTCAB saponins; UHPLC-ESI/HRMS; Orbitrap, Gypsogenin 3-*O*-glucuronide

1. Introduction

Glucuronide **O**leanane-type **T**riterpenoid **C**arboxylic **A**cid 3, 28-*O*-**B**idesmosides (GOTCAB) are a group of the triterpenoid saponins [1]. They possess a glucuronic acid moiety at C-3 hydroxyl group of the aglycone (Fig. 1) and are among the highest glycosylated bidesmosides in advanced taxonomic groups Caryophyllidae, Rosidae and Asteridae. *Gypsophila* L. species (Caryophyllaceae) have been shown to accumulate GOTCAB saponins [2-5]. *Gypsophila* GOTCAB are known to possess various pharmaceutical properties such as cytostatic and cytotoxic effects on malignant tumor cells [6-9], synergistic enhancement of the immunotoxines activity [10], adjuvant properties for vaccines as immunostimulatory complexes [11].

Since GOTCAB saponins usually are a mixture of structurally related forms with very similar polarities, their separation remains a challenge [1]. In the past 20 years, a variety of GOTCAB have been isolated and characterized by chromatographic techniques [3-5, 7, 9]. After sample preparation step, preparative and semi-preparative liquid chromatographic methods have been performed under variable elution conditions. The isolation and characterization of GOTCAB is very tedious as low amount of purified saponins is obtained after purification procedures. Based on their complexity, the structural elucidation of GOTCAB is a very challenging task, requiring the use of several analytical techniques. Technological advances that combine high-performance liquid chromatography with mass spectrometers enable the rapid identification of bioactive saponins in plant extracts by distinguishing previously identified ones [12-14]. Although a large number of *Gypsophila* mono- and bidesmosides have been characterized, only the complete GOTCAB profiling of *Gypsophila trichotoma* has been established by LC-HRMS [15]. The use of LC-MS for saponin analysis is limited due to the lack of commercial standards and the availability of isolated pure saponins is essential for confirming the analytical results. Recently, eleven new GOTCAB saponins were isolated from *Gypsophila trichotoma* Wend. *var. trichotoma* roots native to Bulgaria [9] . The saponins have gypsogenin as a common aglycone substituted at C-3 with a branched trisaccharide and at C-28 with an ester-bonded oligosaccharide possessing acetyl, methoxycinnamoyl and (or) sulphate groups. The trisaccharide at C-3 consists of constant β-D-galactopyranose, β-D-glucuronopyranose and a third variable pentose unit that could be either β-D-xylopyranose (β-D-Xylp) or α-L-arabinopyranose (α-L-Arap). These two types of trisaccharide cause the appearance of isobaric pairs of saponins differing in a terminal sugar at C-3 carbohydrate chain. Additional pairs of

positional isomers resulted from the substitution with acetyl, sulphate and methoxycinnamoyl groups at the ester-bonded chain. The biological activity of the aforementioned newly isolated GOTCAB was first evaluated *in vitro* on human breast cancer cell line 25 MDA MB-231 and the synergistic cytotoxicity in combination with type I ribosome-inactivating protein (RIP-I) was used to derive a quantitative structure – activity relationship [16].

The medicinal and commercial importance attached to *Gypsophila* prompted us to investigate the saponins in *G. glomerata* Pall ex M.B. This perennial herb is native to eastern and south eastern regions in Bulgaria [17]. Our previous studies on intact roots and *in vitro* excised root cultures revealed *G. glomerata* as a promising source of saponins [18, 19].

Based on all abovementioned studies, we aimed at developing a new LC-ESI/HRMS data acquisition strategy for GOTCAB recognition. Thus, the chromatographic and ESI-MS/MS fragmentation behavior of previously isolated *G. trichotoma* GOTCAB references was studied and the structure-fragmentation relationships, as well as diagnostic fragments were established. As a part of our ongoing *Gypsophila* study [9, 15, 16, 19], in the present work we report the identification and tentative structural elucidation of 41 GOTCAB including 14 pairs of isobars and 17 new saponins in the methanol-aqueous extract from *G. glomerata* roots.

2. Materials and methods

2.1. Plant material

G. glomerata roots were collected in September 2010 at Ognyanovo village (Pazardjik region) (42°15' N – 24°42' E). The plants were identified by Dr. V. Balabanova (Faculty of Pharmacy, Medical University-Sofia, Bulgaria). Voucher specimen of *G. glomerata* SOM 171499 was deposed at Institute of Biodiversity and Ecosystems Research, Bulgarian Academy of Sciences, Sofia, Bulgaria.

2.2 GOTCAB references

Five GOTCAB saponins (**A**-**E**), previously isolated from *Gypsophila trichotoma* roots, were used as references. Saponins were characterized on the basis of extensive NMR analysis (^{1}H, ^{13}C NMR, COSY, HSQC, HMBC, TOCSY, HSQC-TOCSY, ROESY), completed by HR-ESI-MS and ESI-MS^{n} analyses [9]. The structures of the saponin references are given in Fig. 1. The purities of saponin references were over 90%, determined by HPLC/UV analyses at 210 nm and gradient program **II** described in 2.5.3. Each saponin reference (0.1 mg) was dissolved in 10 mL methanol and diluted to 1 μg/mL concentration. The solutions were stored at 4° C.

2.3. Chemicals

Acetonitrile (HPLC-gradient grade and hypergrade for LC-MS), formic acid (HPLC –grade), analytical-grade methanol, sulfuric and *ortho*-phosphoric acid, and sodium hydrogen carbonate were purchased from Merck (Darmstadt, Germany), while ethanol (95%) was provided by Sigma Aldrich (Deisenhofen, Germany).

2.4. Isolation and NMR analysis of a compound 1 from *G. glomerata* roots

*2.4.1. Extraction and isolation of compound **1***

Powdered air-dried roots of the plant (50 g) were extracted with 50% ethanol (1:10 w/v) by sonication for 30 min (each time) at room temperature to yield 21 g of the crude extract. An aliquot of the crude extract (1 g) was fractionated by a low pressure liquid chromatography over a C18 column (310 × 25 mm,40-63 μm) (Merck, Germany) using a binary gradient of methanol-water (10:90 – 100:0) to afford 20 fractions (each 25 ml). Fractions (10-13) were separated by repeated semi-preparative HPLC on Kinetex C18 (150 x 10 mm, 5 μm) (Phenomenex, USA) and gave 18 fractions (each 5ml). Semi-preparative HPLC was performed on a chromatographic system Varian equipped with a ternary pump Model 9012, a Rheodyne injector with 200 μL loop and an UV/VIS detector Model 9050 set at 210 nm. The mobile phase comprised acetonitrile (A) and water (B). The elution was performed employing a program from 15% A

to 84%A for 60 min at the flow rate 1.5 ml/min. Fractions (8-9) were further re-chromatographed by the aforementioned chromatographic conditions to give saponin **1** (t_R 28.2, 2.1 mg), 93% purity (HPLC).

2.4.2. NMR experiments

NMR experiments were performed on a spectrometer operating at 14.1Tesla (Bruker Avance III, frequencies of 600MHz and 150 MHz for ^{1}H and ^{13}C respectively), using a Bruker 5mm TBI probe. Pulse widths were 11.6 and 14.7 μs for ^{1}H and ^{13}C, respectively. The sample was dissolved in a volume of 500 μL of a mixture of 1/3 D_2O and 2/3 methanol-d_4 (used as chemical shift reference: CH_3 signals at 3.31 ppm for ^{1}H and 49.15 ppm for ^{13}C), and all experiments were performed at 300K. Experiments were processed using the Bruker software package. The ^{1}H, ^{13}C decoupled from proton, ^{13}C JMOD, ^{1}H-^{1}H DQF-COSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{13}C HSQC-TOCSY NMR experiments were run under standard conditions.

Acquisition parameters: ^{1}H spectral width 6000 Hz, ^{13}C spectral width 33200 Hz, repetition time 2 s. *Double Quantum Filtered COSY experiment*: 2K complex points x 256 increments; 16 scans per increment. *HSQC experiment*: 2K x 256 data set, 128 scans per increment. *HMBC experiment*: a 50 ms waiting period was used for the evolution of long-range coupling, and a value of 3.4 ms for the low-pass *J* filter. 2K x 512 data set, 320 scans per increment. *NOESY experiment* (recorded on a spectrometer Bruker Avance III at 9.4Tesla (400MHz 1H resonance frequency): 2K * 256 data set, 48 scans per increment, 300 ms of mixing time. *HSQC-TOCSY experiment*: homonuclear Hartman-Hahn mixing was made with a DIPSI2 scheme (100ms). 1K x 256 data set, 128 scans per increment.

2.5. Analysis of saponins in *Gypsophila glomerata* roots

2.5.1. Plant extract preparation and saponin mixture purification

Air-dried powdered roots of the plant (1 g) were extracted with 50 mL 50% ethanol (× 3) by sonication for 5 min at room temperature. The extract, after filtering, was concentrated in vacuo at 35°C and lyophilized to yield 410.7 mg of crude extract. The latter was dissolved in 3 mL 10% ethanol. Solid phase extraction (SPE) procedure was accomplished on Varian Vac Elut 10 vacuum manifold using cartridges Bond Elut C18, 1 g, 6 mL (Varian, CA, USA). After loading sample on previously conditioned cartridge (10 mL ethanol and 10 mL water), and washing step with 10 mL water, the saponin mixture was eluted from the cartridge with 1 mL ethanol (× 3). The final eluates were evaporated under gentle nitrogen stream and each residue was dissolved in 1 mL methanol. The solutions were filtered through a 0.45 μm syringe filter disc (Polypure II, Alltech, Lokeren, Belgium) and subjected to UHPLC-HRMS.

2.5.2. Acid hydrolysis of the Gypsophila glomerata extract

The lyophilized root extract samples (50 mg) were dissolved in 500 μl water by sonication for 15 min (× 3). A mild acid hydrolysis for obtaining prosaponins was performed according to Gevrenova et al., 2014 [20]. Briefly, 200 μl 4N H_2SO_4 were added to each test-tube (final concentration 0.57M) and the samples were heated to 100°C for 1.5 hour. The hydrolysis mixtures were neutralized by 10% sodium hydrogen carbonate (pH 7.0-7.2). Prosaponins were purified by SPE on Bond Elut C18, 1g, 6 mL cartridges (Varian, MA, USA). After loading samples, and washing step with water, the prosaponins were eluted from the cartridges with 1 mL EtOH (× 2). Prosaponins quantification was performed by HPLC-UV with gradient program **I**.

2.5.3. Analytical high-performance liquid chromatography (HPLC)

Analytical HPLC was performed on a Shimadzu chromatographic system (Kyoto, Japan) equipped with a Symmetry Shield RP18, 5 μm, 250 × 4.0 mm ID (Waters, Milford, MA, USA) at 210nm. All data were acquired and processed with Shimadzu CLASS-VP (version 4.3). The binary mobile phase consisted of solvent A (sA): 25% acetonitrile with 0.1% *ortho*-phosphoric acid (v/v) and solvent B (sB): 84% acetonitrile with 0.1% *ortho*-phosphoric acid (v/v). The following gradient programme **I** was used to analyse the prosaponin content: from 80% sA/20%

sB to 100% sB for 30 min, an isocratic step (100% sB) for 15 min, and then return to initial condition in 1 min. The gradient program **II**, used for the reference purity, began at 100% sA followed by a linear gradient to 100% sB for 15 min, an isocratic step (100% sB) for 5 min and then changed to the initial conditions in 5 min. In both gradient programs the flow rate was 1 mL/min. The oven temperature was set at 40°C.

2.5.4. Quantitative determination of gypsogenin 3-O-glucuronide and quillaic acid 3-O-glucuronide. Validation of the HPLC method.

Gypsogenin 3-*O*-glucuronide (96% purity, HPLC) was obtained from commercial Saponin Pure White (Merck, Germany). The quantification of gypsogenin 3-*O*-glucuronide was carried out using the external standard method. Triplicate HPLC analyses were performed for each concentration (2, 1, 0.5, 0.25 and 0.125 mg/mL) at 210 nm. Quillaic acid 3-*O*-glucuronide concentration was estimated using the gypsogenin 3-*O*-glucuronide calibration curve. Because of the similar molecular structure, the responses of quillaic acid 3-*O*-glucuronide was related to gypsogenin 3-*O*-glucuronide, assuming the responses at 210 nm to be equal. Thus, the amounts of quillaic acid 3-*O*-glucuronide are expressed as gypsogenin 3-*O*-glucuronide equivalent. The complete assay was performed in triplicate and standard deviation was calculated. The repeatability was established by injecting the gypsogenin 3-*O*-glucuronide solution (0.25 mg/mL) six times over one day, while the reproducibility was determined over 10 days by three injections per day of the aforementioned solution. The limit of detection (LOD) and limit of quantification (LOQ) were calculated as follows: LOD =3.3 σ /S and LOQ =10 σ /S, where σ is the standard deviation of the Y-value distribution around the regression line and S is the slope of the calibration curve. The linearity was evaluated from the calibration curve using a least-squares linear regression analysis method. Regarding the recovery study, the SPE-HPLC procedure was done with a control *G. glomerata* hydrolysis mixture spiked with the gypsogenin 3-*O*-glucuronide solution (0.25 mg/mL). The percentage recovery was calculated by subtracting the values measured for the control matrix from those samples that had been spiked with the standard, divided by the amount added of standard and multiplied by 100. The mean recovery and the standard deviation were intended.

2.6. Ultra high-performance liquid chromatography – electrospray ionization/high resolution mass spectrometry (UHPLC-ESI/HRMS)

The LC-MS analysis was carried-out on a Q-Exactive heated electrospray ionization (H-ESI)-HRMS (Thermo Fisher Scientific, Waltham, MA, USA) equipped with an Accela quaternary UHPLC pump and Accela autosampler. Data were processed using Xcalibur 3.0 ® (Thermo Scientific Co, USA) instrument control/data handling software.

2.6.1. UHPLC separation

LC separation was performed on a Poroshell C18 150 x 3 mm 2.7 µm column (Agilent, Santa Clara, CA, USA). The mobile phase consisted of: (A) acetonitrile/water (0.08% formic acid) 5:95 (v/v) and (B) acetonitrile (0.08% formic acid)/water 97:3 (v/v). RP-UHPLC was achieved by a gradient starting at 100% A for 0.5 min followed by a linear gradient for 18.5 min to 100% B, isocratic elution for 10 min and the mobile phase was returned to the initial conditions in 1 min, and the column was equilibrated for 5 min. The flow rate was 250 µL/min. The solvents were filtered through 0.2 µm filterers (Millipore, Watford, Ireland) prior the analyses.

2.6.2. ESI-MS and MS/MS detection

Heated electrospray ionization (H-ESI). 250^{o}C HESI vaporizer temperature, spray voltage at 3kV, ion transfer tube temperature at 300^{o}C, sheath gas pressure 35psi, auxiliary gas flow 10 (arbitrary units) were adjusted for the interphase.

2.6.3. Mass spectrometry conditions

The LC-MS method was operating in Full scan-ddMS^2/Top 5 scan mode with stepped collision energy. The full MS scan event (without collision energy) was followed by the data dependent scans with fragmentation energy applied to the top 5 most intensive ion peaks in the scan

spectrum. It was performed in negative ion monitoring mode with the following settings: 35000 FWHM resolution in Full MS from 200 to 2000 *m/z*, 80 msec maximal trap filling time, 17500 FWHM resolution for fragment spectra scans with 4 *m/z* quadrupole isolation window of precursor ions, loop count 5, MSX count 1. Higher energy collision-induced dissociation (hcd) was set at 20, 25, 30 and 35 eV. The hcd 25 eV with stepped collision energy 40% was selected to assist in the structure elucidation. Thus, the mass spectrometer performed a two step fragmentation on the precursor ions: 25-(25×40%) and 25+(25×40%). All fragments created in the two steps were sent to the Orbitrap analyzer for one scan detection.
The calculation of the exact masses and mass measurement errors, prediction of the molecular formulas and simulation of isotopic profiles were carried out with Xcalibur 3.0 software (ThermoScientific).

3. Results and discussion

3.1. A newly discovered saponin 1 from *G.glomerata* roots

The ethanol-aqueous extract of *G. glomerata* roots was purified by the chromatographic techniques low pressure chromatography and semi-preparative high-performance liquid chromatography to yield a compound **1**. It was obtained as a white amorphous powder. The HRMS in negative ion mode gave deprotonated molecule $[M-H]^-$ at *m/z* 1495.6603 (calcd 1495.6598 for $C_{69}H_{107}O_{35}$, 0.309 ppm) indicating the molecular formula $C_{69}H_{108}O_{35}$. 1D and 2D NMR techniques were performed in order to elucidate the chemical structure of **1**. The proton and carbon resonances of the aglycone moiety were assigned from the analysis of ^{1}H-^{1}H COSY, HSQC and HMBC spectra.
The ^{1}H NMR spectrum showed signals ascribed to six tertiary methyl groups at δ_H 1.17, 0.98, 0.76, 1.16, 0.90 and 0.92 ppm. An olefinic proton was observed at δ_H 5.26 (H-12) showing HSQC correlation with δ_C 123.3. The ^{1}H-NMR data indicated one aldehyde proton at δ_H 9.44 and one methine signal at δ_H 3.92. Complete assignment of the aglycone moiety is given in Table 1. Thus, the aglycone was unambiguously identified as gypsogenin by comparison with literature data [9]. The chemical shift of the carbonyl C-28 at δ_C 178.2 and the deshielded carbon at δ_C 86.0 of the gypsogenin indicated bidesmosidic saponin.
The presence of seven sugar moieties in **1** was evidenced by the anomeric protons at δ_H 4.42 ppm, 4.78, 4.79, 5.33, 5.37, 4.50 and 4.63 displaying correlations with anomeric carbon signals at δ_C 104.2, 104.6, 103.4, 95.0, 101.0, 105.6 and 104.9, respectively. Complete assignment of each sugar resonances was achieved by considering the ^{1}H-^{1}H COSY and HSQC-TOCSY spectra. Evaluation of spin-spin couplings and chemical shifts [9] allowed the identification of one β-fucose (Fuc, Fuc-H1 at 5.33 ppm), two α-arabinose (Ara1, Ara-H1 at 4.79 ppm; Ara2, Ara2-H1 at 4.50 ppm), one β-xylose (Xyl, Xyl-H1 at 4.63 ppm), one α-rhamnose (Rha, Rha-H1 at 5.37 ppm), one β-galactose (Gal, Gal-H1 at 4.78 ppm) and one β-glucuronic acid (GlcA, GlcA-H1 at 4.42 ppm) (assignments in Table 2). The common D configuration for Gal, GlcA, Xyl and Fuc and the L configuration for Rha and Ara were assumed according to those most often encountered among the plant glycosides. Glycosidation shifts were observed for Fuc-C2 (δ_C 74.4 ppm), Rha-C3 (δ_C 81.6 ppm), Rha-C4 (δ_C 79.1 ppm), GlcA-C2 (δ_C 78.5 ppm) and GlcA-C3 (δ_C 85.8 ppm). The cross peak in the HMBC experiment between GlcA-H1 (δ_H 4.42 ppm) and C-3 of the aglycon (δ_C 86.0 ppm) showed that the glucuronic acid was linked to the aglycon at C-3 position. Chemical shifts of Fuc-H1 (δ_H 5.33 ppm) and Fuc-C1 (δ_C 95.0 ppm) suggested that this sugar was involved in an ester linkage with the C-28 carboxylic group of the aglycone, this is confirmed by the HMBC cross peak between Fuc-H_1 (δ_H 5.33 ppm) and C-28 of the aglycone (δ_C 178.2 ppm) (Fig. 2). The position and the sequence of the sugar residues were also defined by the HMBC experiment which showed key correlation peaks between the proton signal at δ_H 4.78 ppm (Gal-H1) and the carbon resonance at δ_C 78.5 ppm (GlcA-C2), between the proton signal at δ_H 5.04 ppm (Ara1-H1) and the carbon resonance at δ_C 85.8 ppm

(GlcA-C3), between the proton signal at δ_H 5.37 ppm (Rha-H1) and the carbon resonance at δ_C 74.4 ppm (Fuc-C2), between the proton signal at δ_H 4.50 ppm (Ara2-H1) and the carbon resonance at δ_C 81.6 ppm (Rha-C3), between the proton signal at δ_H 4.63 ppm (Xyl-H1) and the carbon resonance at δ_C 79.1 ppm (Rha-C4). Furthermore, HMBC correlations were confirmed by NOESY cross peaks at δ_H/δ_H 4.78/3.69 (Gal-H1/GlcA-H2), δ_H/δ_H 4.79/3.72 (Ara1-H1/GlcA-H3), δ_H/δ_H 5.37/3.82 (Rha-H1/Fuc-H2), δ_H/δ_H 4.50/3.84 (Ara2-H1/Rha-H3) and δ_H/δ_H 4.63/3.65 (Xyl-H1/Rha-H4). Thus, the saponin was identified as 3-*O*-β-D-galactopyranosyl-(1→2)-[α-L-arabinopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-α-L-arabinopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester. **1** was previously not described in the literature.

3.2. Fragmentation patterns of GOTCAB references A-E by LC-HRMS

In this study, five GOTCAB saponins (**A-E**), previously isolated from *G. trichotoma* Wend. served as saponin references (Fig.1) [9]. LC-HRMS analyses of the available references allowed to deduce the fragmentation patterns of GOTCAB. In these saponins, the sugar chains are attached to the aglycone *via* the hydroxyl groups at the C-3 and C-28 positions. Since a carboxyl group is presented in the aglycone gypsogenin, the LC-HRMS analyses of the GOTCAB references were performed in negative ion mode [14, 15].

The fragmentation patterns of **A-E,** as well as the accurate and exact masses of the deprotonated molecules $[M-H]^-$ and fragment ions are presented in Table S1. In full scan mode, the accurate mass values were within 5 ppm of the exact masses. MS/MS spectra of the reference **A** $[M-H]^-$ at different higher-energy collisional dissociation (HCD) values are compared in Fig. 3. At HCD 20 and 25, GOTCAB typically produced abundant $[M-H]^-$ and fragment ions indicating loss of branched oligosaccharide chains, as well as sequential loss of sugar residues (Fig. 3A, B, Table S1). Thus, 25 HCD (Fig. 3B) could be taken appropriate fragmentation energy giving adequate abundance of the deprotonated molecules and fragment ions. Two different fragmentation pathways occurred. One pathway gave [(M-H)-ester-bonded chain]$^-$ at MS/MS spectra (**A**-**D**); the base peak ions were derived from [gypsogenin-H]$^-$ at *m/z* 469.332 (Table S1). The other pathway afforded a base peak [ester-bonded chain-H]$^-$ (**E**). The MS and MS/MS data were consistent with previous publications [9, 15] but numerous fragments have not been reported previously. Fragment ions with abundance < 10% of the base peak are also reported as they are needed for comparison [21].

Fragmentation pathways of **A** and **C/D** are proposed in Fig. 4 and 5, respectively. The commonly used nomenclature for triterpenoid saponins was used to signify the fragment ions [22]. As the C-28 bond cleavage first occurred under MS/MS conditions, the ester-bonded chain at C-28 was defined as the α chain, whereas the oligosaccharide at C-3 was defined as the β chain. Cross-ring cleavages (X) were designated by superscript numbers indicating the two bonds cleaved. Thus, **A-E** were deglycosylated to form two fragment ions at *m/z* 939.460 [(M-H)-α chain]$^-$ ($Y_{0\alpha}$) and 469.332 [gypsogenin-H]$^-$ ($Y_{0\alpha}/(Y_{0\beta})$ (Table S1, Fig. 3-5). However, the abundance of the [(M-H)-α chain]$^-$ is higher because it is generated by a C-28 ester bond cleavage, which indicated the location of this oligosaccharide (Fig. 3A, B, C). The fragment ion at *m/z* 469.332 corresponded to the loss of α and β chains and, at the same time, indicated the deprotonated aglycone (Fig. 4). The fragmentation pathway of the [gypsogenin-H]$^-$ ($Y_{0\alpha}/(Y_{0\beta})$ involved separately losses of water at *m/z* 451.322 $[C_{30}H_{43}O_3]^-$ and H_2CO at *m/z* 439.322 $[C_{29}H_{43}O_3]^-$. Moreover, abundant ion at *m/z* 423.327 $[C_{29}H_{43}O_2]^-$ (11-12%) resulted from the HCO_2H elimination (Table S1, Fig. 4). In addition, the gypsogenin yielded low abundant ion at *m/z* 405.316 $[C_{29}H_{41}O]^-$ (up to 3%) by concomitant loss of H_2O and HCO_2H (64 Da). As the carboxyl group at C-28 in GOTCAB was the probable ionization site, it was suggested that its transformation resulted in the formation of an olefinic bond (Fig. 4).

The fragmentation of $Y_{0\alpha}$ at *m/z* 939.460 yielded the fragment ions at *m/z* 789.407 [(M-H-α chain)-Pent-H_2O]$^-$ ($Y_{0\alpha}/Z_{1\beta'}$) and 759.396 [(M-H)-α chain-Hex-H_2O]$^-$ ($Y_{0\alpha}/Z_{1\beta}$) by losing arabinose

and galactose moieties, respectively, indicating both *O*-glycosylation and branched β chain (Figure 4 and 5). All examined references lost CO_2 together with water molecule(s), resulting in abundant fragment ions at *m/z* 583.364 [(M-H)-α chain-Hex-Pent-H_20-CO_2]$^-$ (15.9%, **A**) and 565.3524 [(M-H)-α chain-Hex-Pent-$2H_20$-CO_2]$^-$ (29.9%, **A**) (Table S1). Formic acid (46 Da) and CO_2 (44 Da) have been reported as two possible elimination pathways of a carboxylic function of triterpenic acids in ESI-MS/MS [23]. Thus, fragmentation pathway of $Y_{0\alpha}$ yielded decarboxylated ions at *m/z* 537.359 [(M-H)-α chain-Hex-Pent-H_20-HCO_2H-CO_2]$^-$ and 519.348 [(M-H)-α chain-Hex-Pent-$2H_20$-HCO_2H-CO_2]$^-$. Hexuronic acid was deduced from the fragment ions at *m/z* 551.338 [(M-H)-α chain-Hex-Pent-$2H_20$-58]$^-$ (20.9%, **A**) and 511.343 [(M-H-α chain)-Hex-Pent-134]$^-$ (3.4%, **A**) resulting from the internal cleavage of the GlcA - $^{4,5}A_{0\beta}$ and $^{0,2}A_{0\beta}$, respectively (Fig. 5). Nevertheless, $Y_{0\alpha} \rightarrow Y_{0\alpha}/Z_{1\beta}$ fragmentation (*m/z* 759) together with $Y_{0\alpha}/^{0,2}X_{0\beta}/Y_{1\beta}/Y_{1\beta'}$ (*m/z* 511) contributed to the assignment of the glycosylation on GlcA. An interglycosidic linkage (1→2) of the galactose was deduced from the more abundant ion at *m/z* 759 (9.2%, **A**; 21.7%, **D**), whereas the arabinose moiety was linked at C-3 of the GlcA witnessed by the low abundant ion at *m/z* 807.417 [(M-H)-α chain-Pent]$^-$, (3.9%, **A**; 15.8%, **D**), supported by 745 and 727 (Table S1).

Both β chains linked at C-3 of the gypsogenin, with terminal α-L-Arap (references **C**) and β-D-Xylp (reference **D**) [9], are identical in mass and fragmentation pattern. Nevertheless, the pentose moiety at C-3 trisaccharide modified the chromatographic behavior, thus favoring lower retention time for the saponin **D**.

Concerning the C-28 α chain, the loss of 586.211 Da was in agreement with a tetrasaccharide (reference **A**). The loss of 628.222 and 670.232 Da could be attributed to the monoacetylated (reference **B**) and diacetylated tetrasaccharide (references **C** and **D)**, respectively (Table S1). In MS/MS spectra **C** and **D** showed mass differences at 42.011 (acetyl group) and 204.063 (glucose + acetyl group) indicating that the associated compounds possess terminal acetylated glucose moiety (Table S1, Fig. 5). It should be note that aforementioned fragment ions were not seen in the fragmentation pattern at high HCD (35). **B**-**D** yielded the fragment ion at *m/z* 981.470 indicating the loss of a trisaccharide (Rha, Glc, Ara) together with ($^{0,2}X_{0\alpha}$+Ac) (**B**) or $Y_{1\alpha}/^{0,2}X_{0\alpha}$ (**C** and **D**) (Table S1, Fig. 5).

Regarding **E**, three mass intervals of 79.958 (sulphate group), 162.053 (glucose) and 242.011 (glucose + sulphate group) were observed in MS/MS spectra (Table S1). The base peak corresponded to the [α chain-H]$^-$ (monoacetylated and sulphated tetrasaccharide at *m/z* 707) which further undergoes cross-ring cleavages of the terminal acetylated fucose: $^{3,5}A_{0\alpha}$ (-100), $^{0,3}A_{0\alpha}$ (-116) and $^{1,3}A_{0\alpha}$ (-128) resulting in the ions at *m/z* 607.120, 591.127 and 579.124, respectively (Table S1). The fragment ions at *m/z* 503.109 [(α chain-H)-Pent-$^{3,4}A_{0\alpha}$]$^-$ and 527.163 [(α chain-H)-SO_3-$^{3,5}A_{0\alpha}$]$^-$suggested fucose C-4 acetylation, supported by the abundant fragment ion at *m/z* 519.103 [(α chain-H)-AcdHex]$^-$ (31.2%). These observations lead to the conclusion that the acetyl residue is easier to remove in comparison with the sulphate group. The xylose moiety was linked at C-4 of the rhamnose unit witnessed by the ion at *m/z* 329.019 [707-Pent-AcdHex-$^{3,5}A_{1\alpha}$]$^-$. Glucose attachment at C-3 was confirmed by the ion at *m/z* 301.023 [707-Pent-AcdHex-$^{1,3}X_{1\alpha}$]$^-$supported by 371.067 [707-AcdHex-$Z_{2\alpha'}$]$^-$. The ion at *m/z* 241.002 [(α chain-H)-Pent-dHex-AcdHex]$^-$ (17%) is consistent with the sulphated glucose [$GlcSO_3$]$^-$.

3.3. Profiling of GOTCAB saponins in *Gypsophila glomerata* roots (GGR) by UHPLC-ESI/HRMS

Based on the common approach, an UHPLC-ESI/HRMS method involving binary solvent system (acetonitrile and 0.1% aqueous formic acid) and gradient elution was applied for the GOTCAB profiling of GGR [20]. It has been reported that the use of a mobile phase containing low concentration of formic acid (0.03-0.1%) increased triterpenoid mono- and bidesmoside ESI response [13-15]. By analogy with the glycyrrhizic acid, the glucuronic acid in the GOTCAB saponins promotes the formation of deprotonated molecules [M-H]$^-$ [24]. In negative

ESI mode, most oleanane type triterpenoid saponins afforded abundant $[M-H]^-$; the adduct process is not reproducible [25]. It has been reported that the $[M-H]^-$ are the major ions generated in the negative ion mode of UHPLC-ESI-QTOF-MS analyses [23, 26]. In our study, the UHPLC mobile phase was modified with 0.1% of formic acid and negative ionization mode for the ESI-MS/MS detection of *Gypsophila* saponins was used due to the sufficient ion response and the detection of $[M-H]^-$ signals without adduct formation. Herein, we demonstrate our approach using the reference saponins **A**-**E** and the newly isolated saponin **1**. A survey full scan MS of GGR is followed by data dependent (dd) MS^2 scans where the stepped collision energy (-40%/+40%) was applied to the selected "n" number most intense precursor ions for MS/MS (n=5).

The total ion chromatogram (TIC) of GGR and extracted ion chromatograms of some saponins are shown in Fig. 6. Both full scan (MS) and fragment spectra (MS/MS) allowed the differentiation and characterization of GOTCAB. The GGR saponin identification and tentative elucidation were based on HRMS and MS/MS data, comparison with fragmentation fingerprints observed for the GOTCAB references and literature data [8, 9, 15, 20, 27].

As triterpenoid saponins have different ester-bonded oligosaccharides and showed different fragmentation patterns, they were classified into three groups: I - GOTCAB saponins with a C-28 oligosaccharide; II - GOTCAB saponins with a C-28 oligosaccharide substituted with an acetyl group; III - GOTCAB saponins with a C-28 oligosaccharide substituted with a sulphate group.

Type I - GOTCAB saponins with a C-28 oligosaccharide (tri-, tetra-, penta- and hexasaccharide)

In UHPLC-HRMS analysis, the abundant deprotonated molecule $[M-H]^-$ at *m/z* 1495.658 (**1**) and 1495.659 (**2**) (calc. for $C_{69}H_{107}O_{35}$, ppm -1.0 and -0.1, respectively) were first used for saponins recognition (Fig. 6D, Fig. 7B). **1** gave $[aglycone-H]^-$ as a MS/MS base peak at *m/z* 469.332, supported by the abundant ions at *m/z* 939.460 $[(M-H)-\alpha\ chain]^-$and *m/z* 565.353 $[(M-H)-\alpha\ chain-Pent-Hex-2H_2O-CO_2]^-$, as previously observed for saponin references **A**-**D** (Table S2). The loss of α chain (556 Da) indicated 30 mass units lower than that of reference **A** (586 Da). This suggested the presence of a pentose unit instead of a hexose in **A**. Thus, **1** and **2** possess a C-28 ester-bonded chain consisting of 2 desoxyhexoses and 2 pentoses. Fragment ions at *m/z* 759.396, 745.417 and 511.341 (**1**) were diagnostic for hexuronic acid substituted with both hexose at C-2 and pentose at C-3 (Table S2, Fig. 7B). Both molecules share the same aglycone, gypsogenin, witnessed by the ions at *m/z* 451.321, 439.321 and 423.327 (**1**) by losing H_2O, CH_2O and HCO_2H, respectively, together with 405.315 $[(aglycone-H)-HCO_2H-H_2O]^-$ (Fig. 7E). The isotopic peak profile of $[M-H]^-$ at *m/z* 1495.658 (1495.658, 100%; 1496.662, 85.6%; 1497.665, 37.1%, 1498.672, 2.1%) corroborated the predicted molecular formula of $C_{69}H_{107}O_{35}$, as the relative intensities of isotopic peaks matched closely with the simulated profile (1495.659, 100%; 1496.663, 74.6%; 1497.667, 27.4%; 1498.667, 5.4%). The structure of a newly discovered saponin **1** was established on the basis of extensive NMR analysis (Table 1 and 2, Fig. 2). By analogy with the behavior of the references **C/D**, it is reasonable to conclude that **1** and **2** are isomers (Fig. 6D). They could be differentiated by the pentose unit at C-3 trisaccharide (Arabinose in **1** and Xylose in **2**). This conclusion is consistent with the results from the previous studies, where the variable pentose in C-3 trisaccharides causes the appearance of characteristic pairs of isobars [9, 15]. Thus, **2** matched compound 6 reported in *Gypsophila oldhamiana* Miq. or 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-α-L-arabinopyranosyl-(1→3)-β-D-xylopyranosyl-(1→4)-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester [28].

MS data of compounds **3** and **4** showed the $[M-H]^-$ at *m/z* 1363.617 corresponding to 132 Da less than that of **1** and **2** (Table S2). Peak **3** generated prominent fragment ions at *m/z* 807.418 $[(M-H)-\alpha\ chain]^-$, 627.353 $[(M-H)-\alpha\ chain-Hex-H_20]^-$ and 469.331 $[aglycone-H]^-$ suggesting

that **3** possesses a C-3 disaccharide consisting of hexose and hexuronic acid. The fragment at *m/z* 511.342 allowed situating the hexose at C-2 of the hexuronic acid, supported by the ion at *m/z* 583.363. Related GOTCAB saponin is not described in the literature.

Compound **4** gave the fragment ion at *m/z* 939.459 $[(M-H)-2dHex-Pent]^-$ indicating the lack of one pentose moiety in the C-28 ester-bonded chain in comparison with those in **1/2.** Related GOTCAB was previously identified in *Gypsophila pacifica* Kom. roots as compound 3, or 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-β-D-xylopyranosyl-(1→4)-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester [29].

Two isobaric saponins **5** and **6** shared the same $[M-H]^-$ at *m/z* 1511.655 consistent with molecular formula $C_{69}H_{107}O_{35}$ (Table S2, Fig. 6B). The precursor ion of **5** generated a base peak at *m/z* 955.455 $[(M-H)-2dHex-2Pent]^-$ indicating the same α chain as in **1** (Fig. 7A). Fragmentation pathway of **5** is proposed in Fig. 8. Fragment ions at *m/z* 805.400 $[(M-H)$-α chain-Pent-$H_20]^-$, 775.392 $[(M-H)$-α chain-Hex-$H_20]^-$and 625.336 $[(M-H)$-α chain-Pent-Hex-$2H_20]^-$ suggested a branched C-3 chain supported by the higher relative abundance of the ion at *m/z* 775 (Fig. 7A). Another ions at *m/z* 581.347 $[(M-H)$-α chain-Pent-Hex-$2H_20$-$CO_2]^-$ and 563.339 $[(M-H)$-α chain-Pent-Hex-$3H_20$-$CO_2]^-$ derived from the elimination of water molecules together with CO_2 (Fig. 8). Because this behavior was consistent with the fragmentation features of $Y_{0\alpha}$ (references **A**-**D**), it is sufficiently robust to identify C-3 oligosaccharide as 3-*O*-hexosyl-(1→2)-[pentosyl-(1→3)]-hexuronoside.

Compound **5** afforded a fragment ion at *m/*z 485.327 $[aglycone-H]^-$ characteristic for quillaic/gypsogenic acid, having the same molecular weight. Especially, the fragmentation of the $[aglycone-H]^-$ in (-) ESI-MS/MS displayed features including both forms of decarboxylation (CO_2 and HCO_2H). Thus, the ion at *m/z* 485.327 $[C_{30}H_{45}O_5]^-$ (calc.485.327) yielded fragment ion deriving from the loss of H_2O at *m/z* 467.318 $[C_{30}H_{43}O_4]^-$, while the ions at *m/z* 439.322 $[C_{29}H_{43}O_3]^-$ and 421.310 $[C_{29}H_{41}O_2]^-$were generated after the separate losses of HCO_2H and (HCO_2H+H_2O) (Fig. 7). Furthermore, the concomitant elimination of $2H_2O$ and CO_2 led to the prominent ion at *m/z* 405.316 $[C_{29}H_{41}O]^-$ (24.2%). Similar MS/MS behaviour was previously observed with the related triterpenic acids [23]. This is consistent with the presence of two tertiary hydroxyl groups at C-3 and C-16 (Fig. 8), and the structure of quillaic acid, regularly identified in *Gypsophila* species in GOTCAB forms [30-33]. Quillaic acid was distinguished by its unique MS/MS fragmentation pattern including diagnostic ions at *m/z* 421.311 and 405.316 (Table S2). In the mass fingerprint of **6** appeared three significant fragments at *m/z* 939.460 $[(M-H)-dHex-2Pent-Hex]^-$, 469.332 $[aglycone-H]^-$ and 451.322 $[(aglycone-H)-H_2O]^-$ which can be attributed to the α chain (572 Da) and the aglycone gypsogenin (Table S2). A comparison of the MS/MS data of saponins **6** with those of **1** suggested that there was a hexose moiety instead of one deoxyhexose in the chain.

Concerning **7**, $[M-H]^-$ at *m/z* 1481.679 generated major fragment at *m/z* 925.480 $[(M-H)-2dHex-2Pent]^-$ indicating the same α chain as in **1** (Table S2, Fig. 7C). The fragmentation pattern of $[aglycone-H]^-$ at *m/z* 455.353 $[C_{30}H_{47}O_3]^-$ (calc.455.353) and the comparison with the literature data were consistent with the structure of oleanolic acid [34]. The fragmentation of the [aglycone-H] ion yielded the characteristic fragment ion deriving from the concomitant loss of H_2CO and H_2O (48 Da) at *m/z* 407.330 $[C_{29}H_{43}O]^-$ (calc.407.332) (Fig. 7F). Thus, compounds **7** and **8** were ascribed to oleanolic acid GOTCABs.

The assignments of the MS data of **9** and **10** were in good agreement with those of saponin reference **A** (Table S2, Fig. 6C). By comparing the retention times, compound **10** was ascribed to saponin **A** or 3-*O*-β-D-galactopyranosyl-(1→2)-[α-L-arabinopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-(β-D-glucopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester. The MS data and chromatographic behavior of **9** allowed us to deduce an isobar of **10**. Their structural difference

could be attributed to the pentose unit at C-3 trisaccharide – xylose (**9**) and arabinose (**10**). Thus, **9** could correspond to 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-(β-D-glucopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester, previously isolated from the roots of *G. paniculata* L.[30] and *G. oldhamiana* [8, 28].

Peak **11** showed prominent fragment ions at *m/z* 955.455 $[(M-H)-2dHex-Pent-Hex]^-$ and 485.327 $[aglycone-H]^-$, suggesting the same ester chain as in the reference **A** (Table S2, Fig. 6A). **11** could be related to compound G1 reported in *Gypsophila paniculata* as 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl quillaic acid 28-*O*-(β-D-glucopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester [30].

Mass fragmentation of **12** disclosed also two ions at *m/z* 807.418 $[(M-H)-2dHex-Pent-Hex]^-$ and 469.332 $[gypsogenin-H]^-$. MS/MS spectrum indicated the presence of hexose (1→2) hexuronic acid (β chain) seen previously in **3** (Table S2). By analogy with the chromatographic behavior of **1**/**2** and **9**/**10**, compounds **12**/**13** were tentatively assigned to two isomers, previously undescribed in the literature.

Concerning compounds **15/16** ($[M-H]^-$ at *m/z* 1539.685), there was mass difference of 14 Da between their α chain and that in reference **A** (Table S2, Fig. 6F). Thus, the α chain was deduced as 3dHex-Hex. The fragmentation pattern of **17**/**18** could be related to those of compound 1 reported in *G. pacifica* or 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-α-L-arabinopyranosyl-(1→4)-α-L-arabinopyranosyl-(1→3)-β-D-xylopyranosyl-(1→4)-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester [29]. We tentatively assigned **19** – **22** to four isomers related to the previously isolated *Gypsophila* saponin 1 from *Gypsophila paniculata* or 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-β-D-glucopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→3)-β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester [35].

Compounds **27/28** could be associated with saponin 2 from the *G. arrostii* or 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-β-D-xylopyranosyl-(1→4)-[β-D-glucopyranosyl-(1→3)]-α-L-rhamnopyranosyl-(1→2)-[β-D-glucopyranosyl-(1→4)]-β-D-fucopyranosyl ester [4]. **29**/**30** could be related to GOTCAB isobars saponin 1 from *G. arrostii*, identified as 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl quillaic acid 28-*O*-β-D-xylopyranosyl-(1→4)-[β-D-glucopyranosyl-(1→3)]-α-L-rhamnopyranosyl-(1→2)-[β-D-glucopyranosyl-(1→4)]-β-D-fucopyranosyl ester [4] and saponin 5 from *G. pacifica* or 3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl quillaic acid 28-*O*-β-D-glucopyranosyl-(1→3)-[β-D-glucopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-[α-L-arabinopyranosyl-(1→4)]-β-D-fucopyranosyl ester [32].

The aforementioned fragment ions were favored to differentiate saponins **31** (GOTCAB of gypsogenin) and **32** (GOTCAB of quillaic acid), in which the α chain is hexasaccharide composed of 2dHex-3Pent-Hex. **31** yielded the fragment ion at *m/z* 1229.5203 indicating a loss of (2Pent-Hex-134), prior of the α chain. This implied that three sugar moieties are linked to the secondary dHex, evidenced by intramolecular breakage $^{0,1}A_{0\alpha}$ (-134). Compounds **15**/**16**, **31** and **32** were not described in the literature and the type of supplementary sugar residues must be determined.

Type II - GOTCAB saponins with C-28 oligosaccharide substituted with an acetyl group

Generally, this type (**33-37**) produced a MS/MS base peak derived from the $[(M-H)-\alpha\ chain]^-$ at *m/z* 939 (gypsogenin GOTCAB) and 955 (quillaic acid GOTCAB). The compound **33** gave $[M-H]^-$ at *m/z* 1567.678 consistent with molecular formula $C_{72}H_{111}O_{37}$ (calc. 1567.681, -2.2

ppm) (Table S2). The MS data assignments of **33** were in good agreement with those of saponin reference **B,** identified as 3-*O*-β-D-galactopyranosyl-(1→2)-[α-L-arabinopyranosyl-(1→3)]-β-D-glucuronopyranosyl gypsogenin 28-*O*-(β-D-glucopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-4-*O*-acetyl-β-D-fucopyranosyl ester [9]. In the (-) ESI-MS spectrum of **34**, the $[M-H]^-$ was observed at *m/z* 1583.675 indicating 16 mass units higher than that of reference **B**. The fragment ions at *m/z* 1113.544 [(M-H)-Pent-Hex-HexA]$^-$ and 955.459 [(M-H)-2dHex-Pent-Hex-Ac]$^-$ showed the presence of the same α and β chains as in **33,** while the ion formation at *m/z* 485.327 [aglycone-H]$^-$ was favored for quillaic acid. Thus, **34** could be associated with either compound 2 reported from *G. altissima* L. roots (3-*O*-β-D-galactopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucuronopyranosyl quillaic acid 28-*O*-(6-O-acetyl)-β-D-glucopyranosyl-(1→3)-[β-D-xylopyranosyl-(1→4)]-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranosyl ester) or its positional isomer [33].

In the same way, the saccharide chains and the aglycone type of compounds **35** and **36** were obtained using MS/MS of $[M-H]^-$ at *m/z* 1699.722 and 1715.718, respectively. This allowed us to deduce an α chain composed of 2dHex-2Pent-Hex-Ac, linked at C-28 of the gypsogenin (**35**) and quillaic acid (**36**) (Table S2). Regarding **35**, the MS/MS spectrum gave prominent fragment ions at *m/z* 1657.709 [(M-H)-Ac]$^-$, 1495.6663[(M-H)-AcHex]$^-$ and 1187.592 [(M-H)-2Pent-AcHex-CO_2]$^-$. These ions were accompanied by the fragment at *m/z* 981.466 [(M-H)-2Pent-dHex-AcHex-104]$^-$, indicating a branched C-28 pentasaccharide with both terminal pentose and acetylated hexose, as previously observed for references **C**/**D**. Cross ring cleavage ^{0,2}dHex (-104) is diagnostic for C-2 substituted primary dHex. Similarly, MS/MS spectrum of **37 (**[M-H]$^-$ at *m/z* 1847.7595) indicated the presence of an supplementary pentose unit in α chain in comparison with **36** (Table S2). In addition, **37** is derivative of **32** with acetylated α chain. Compounds **35**-**37** are newly described and the location of the supplementary sugar units should be determined.

Type III - GOTCAB saponins with a C-28 oligosaccharide substituted with sulphate group

This type (**38**-**41**) was characterized by a base peak corresponding to the α chain (Table S2). All four compounds examined lost the terminal sugar units, yielding [(α chain-H)-dHex]$^-$, [(α chain-H)-Pent/Hex]$^-$ or [(α chain-H)-dHex-Pent/Hex]$^-$ (Table S2). Furthermore, typical ions of deoxyhexose cross-ring cleavage were observed: $^{1,3}X_{0\alpha}$ ([(α chain-H)-86]$^-$) and $^{3,5}X_{0\alpha}$ ([(α chain-H)-58], suggesting either C-2 or C-3 substitution of the terminal dHex (Fig. 9). Consistent with the reference **E** and previously published data [15], C-2 substitution rather than C-3 would be expected. By analogy with the fragment at *m/z* 329 (reference **E**), the occurrence of the ion at 297 [(α chain-H)-dHex-Pent/Hex-60]$^-$ ($^{1,3}X_{1\alpha}/C_{1\alpha}$) contributed to the assignment of the pentose/hexose unit at C-3 of the secondary deoxyhexose. The sulphate group was linked to a pentose unit witnessed by the ion at *m/z* 210.992 ($B_{2\alpha'}$) and 229.002 ($C_{2\alpha'}$) (Fig. 9). Thus, the ester chain in **38**-**40** could be assessed as *O*-sulphate-pentosyl-(1→4)-[pentosyl/hexosyl-(1→3)]-deoxyhexosyl-(1→2)-deoxyhexoside.

Regarding **38**, it was a sulphated derivative of **1**. The parent ions at *m/z* 1605.628 (**39)** and *m/z* 1621.626 (**40)** (Fig. 6G, E) showed fragment ions at *m/z* 1135.499 and 1151.494 [(M-H)-Pent-Hex-HexA]$^-$, as well as at *m/z* 939.456 and 955.453 [(M-H)-2dHex-Pent-Hex-SO_3]$^-$, respectively, indicating consecutive losses of α and β chains (Table S2). In addition, the elimination of 79.958 Da derived from the loss of sulphate group yielding fragment at *m/z* 1525.679 (**39**) and 1541.668 (**40**). **39** and **40** corresponded probably to the sulphated derivatives of compounds **9**/**10** and **11**, respectively, and were new compounds.

3.4. Quantification of GOTCAB saponins in *Gypsophila glomerata* roots (GGR) by means of their prosaponin content

Quantitative analysis of saponins in GGR was performed by means of two representative prosaponins of *Gypsophila* species, gypsogenin 3-*O*-glucuronide and quillaic acid 3-*O*-

glucuronide. After a mild acid hydrolysis of the GGR, prosaponins were purified by solid phase extraction (SPE) and identified by UHPLC-ESI/HRMS. Deprotonated molecules $[M–H]^-$ at *m/z* 645.365 (calc. for $C_{36}H_{53}O_{10}$, 0.9 ppm) and 661.359 (calc. for $C_{36}H_{53}O_{11}$, 0.2 ppm) were first used for prosaponins recognition [20]. Glycosidic bond at C-3 of the prosaponins showed facile cleavage in negative ion mode of ESI-MS/MS to form fragment ions at *m/z* 485.327 [quillaic acid-H]$^-$ and 469.331 [gypsogenin-H]$^-$, respectively. These signals corresponded to the loss of glucuronic acid residue [(M-H)-176.124]$^-$. A RP-HPLC method involving gradient elution and UV detection for the quantification of these two prosaponins was applied following pretreatment of the samples by SPE [20]. Peaks of gypsogenin 3-*O*-glucuronide and quillaic acid 3-*O*-glucuronide were assigned in the HPLC chromatograms by comparing individual peak retention times with these of the authentic references, as well as by spiking techniques. In HPLC analyses, the retention time of gypsogenin 3-*O*-glucuronide was 19.44 ± 0.05 min. Calibration curve was constructed from peaks areas versus analyte concentrations. The regression equation of gypsogenin 3-*O*-glucuronide standard curve was $y = 3355.5\ x – 306.16$, $r^2 = 0.9994$ (Analytik-Software STL, Leer, Germany). The uncertainties in the slope and intercept were estimated to be 0.7% and 8.2%, respectively. The F-statistic was 19029 and the regression was justified. The LOD and LOQ for gypsogenin 3-*O*-glucuronide were 2.7 µg/mL and 8.2 µg/mL. The RSD of the repeatability and reproducibility was calculated to be 1.02% and 2.60%, respectively. The recovery of gypsogenin 3-*O*-glucuronide was 90±2%. In general, the proposed method was sensitive to detect and quantify the prosaponins in the samples. The content of gypsogenin 3-*O*-glucuronide was 7.4079 ±0.0723 mg/g dry weight (dw), while quillaic acid 3-*O*-glucuronide was substantially lower – 4.4593 ± 0.1207 mg/g dw.

3.5. A strategy to screen and characterize GOTCAB saponins

In the present study, chromatographic peaks were identified or tentatively elucidated by detailed analyses of their MS and MS/MS data, and by comparison with literature data. Peaks **10** and **33** in GGR were attributed to saponin references **A** and **B**, respectively, by comparing the retention time and MS/MS with those of reference GOTCAB saponins.

GOTCAB fragmentation patterns displayed that a majority of saponins share the same aglycone, gypsogenin. Derivatives of the aglycones quillaic and oleanolic acid were also observed in the GOTCAB profiling. Saponins are 3,28-*O*-bidesmosides containing two saccharide chains: α chain ester-bonded at C-28 of the aglycone and β chain linked by a glycosidic bond at C-3.

We report the parameters of Full scan-ddMS^2 scan mode including hcd set to 25, and two step fragmentation on the precursor ion with stepped collision energy 40%, which favored the formation of characteristic fragment ions (Table S1). This allowed do deduced the monosaccharide and aglycone type, as well as oligosaccharide linkages in each GOTCAB. Deprotonated molecules $[M-H]^-$ were first used for saponin recognition. The accurate $[M-H]^-$ ion masses differed from the exact masses by not more than 5 ppm in full scan mode (Table S2).

Further, our strategy for GOTCAB recognition suggested fragmentation rules for three major groups GOTCAB and possible fragmentation pathways were proposed (Fig. 5, 7-9). Saponins with different C-28 ester-bonded chains showed different mass spectrometry fragmentation patterns in negative ion mode. Generally, in the MS/MS spectra of type I, the precursor ions $[M-H]^-$ yield abundant fragment ions by losing ester-bonded chain, while the base peaks correspond to the aglycone. Concerning the relative abundance of both deprotonated molecules and fragment ions, for some GGR GOTCAB (**5**, **7**, **11**), the MS/MS experiments gave spectra differ slightly from the fragmentation patterns of the references (Table S2). Fragment ions attributed to the loss of one acetyl group, acetylated hexose, and pentose together with the base peak at *m/z* 939.460 [(M-H)-α chain]$^-$ (GOTCAB of gypsogenin) and 955.454 [(M-H)-α chain]$^-$ (GOTCAB of quillaic acid) could be considered as diagnostic ions for C-28 branched

oligosaccharide, ending with both pentose and acetylated hexose units (**35**-**37**). Besides the general fragmentations, the base peak [α chain-H]$^-$ can be used to differentiate the sulphated GOTCAB, in which the production of above mentioned base peak seems to be favored for this type saponin derivatives. The corresponding ions $^{3,5}A_{0\alpha}$ (-58), and $^{1,3}A_{0\alpha}$ (-86) from the internal cleavages were indicative of deoxyhexose units (Fig. 9). Particularly, the fragment ions at *m/z* 210.9907 ($C_5H_9O_8S$) and 229.0013 ($C_5H_7O_7S$) witnessed sulphated pentose moiety in this group.

Within all groups (except for **3** and **12**), the β chain consists of pentose, hexose and hexuronic acid discerned by the relevant ions differing according as the aglycones were different (Fig. 10). The low abundant fragments $Y_{0\alpha}/Y_{1\beta'}$, $Y_{0\alpha}/Z_{1\beta'}/CO_2$ and $Y_{0\alpha}/Z_{1\beta'}/H_2O/CO_2$ allowed situating the pentose moiety at C-3 of the hexuronic acid, while the C-2 linked hexose was discerned by the more abundant ion $Y_{0\alpha}/Z_{1\beta}$. The aglycones were easily located in the spectra as the known gypsogenin, quillaic and oleanolic acid witnessed by the fragment ions at *m/z* 469.332, 485.327 and 455.353, respectively. The ion formation at *m/z* 423.327 [gypsogenin-HCO_2H-H]$^-$ was favored for gypsogenin, while quillaic and oleanolic acid yielded characteristic fragment ions at *m/z* 405.316 and 407.332, respectively, arising from the loss of ($2H_2O + CO_2$) (quillaic acid) and H_2CO (oleanolic acid) (Fig. 10).

Thus, the strategy for GOTCAB recognition was based on the preferential fragmentation leading to relevant ions corresponding to each aglycone (Fig. 10). In addition, the relative abundance of the ions of $Y_{0\alpha}$ fragmentation allowed the differentiation of the glycosylation position on hexuronic acid. It was possible to discriminate between the acetylation on dHex and terminal Hex of the α chain. The occurrence of the ions at *m/z* 210.992 and 241.002 characterized a sulphated pentose and hexose, repectively.

In this study, adequate resolution of many GOTCAB was achieved using an UHPLC analysis. It appears that certain *G. glomerata* GOTCAB occur as isobaric pairs in the profile. In ten pairs, saponins with matching mass fingerprints differ in retention times (Fig. 6). This is consistent with the results from our previous study on *G. trichotoma* GOTCAB, where both isobaric pairs and positional isomers of GOTCAB saponins were reported [3, 9; 15]. However, a larger variety of ester-bonded chains (tri-to hexasaccharide) at C-28 of gypsogenin, quillaic and oleanolic acid was characterized in *G. glomerata* in comparison with *G. trichotoma*. The data acquired on acetylated saponins revealed three newly described saponins with terminal acetylated hexose. Despite that sulphated GOTCABs were found in both species, mass spectral fragmentation patterns suggested that these peaks detected in *G. glomerata* originate from both gypsogenin and quillaic acid. The sulphate group was linked at the pentose moiety in *G. glomerata* GOTCABs, while the saponins in *G. trichotoma* possess a sulphate on a hexose unit.

The main finding of this study was that 41 GOTCAB including 14 pairs of isobars were identified or tentatively identified in *G. glomerata* roots by UHPLC-ESI/HRMS. In the majority of saponin pairs the structural difference could be associated with the pentose unit in C-3 trisaccharide, either xylose or arabinose as was seen in *G. trichotoma* [15]. Based on the exact masses and fragmentation behaviors, gypsogenin- and quillaic acid-GOTCAB could be deduced to share the same oligosaccharides as follows: **1/2** and **5**; **9/10** and **11**, **19-22** and **23/24**, **27/28** and **29/30**, **31** and **32**, **33** and **34**, **35** and **36, 39** and **40** (Table S2). In addition, similar C-28 tetrasaccharide was evidenced in **1/2** (gypsogenin GOTCAB), **5** (quillaic acid derivative) and **7/8** (oleanolic acid derivatives). To the best of our knowledge, sixteen GOTCAB core structures have not been previously reported in *Gypsophila* genus. A new gypsogenin-GOTCAB **1** was isolated from GGR and its structure was established by NMR experiments completed with HRMS and MS/MS spectra.

Isobaric pairs **3**/**4**, **9**/**10**, **17**/**18** and **29**/**30** could be related to GOTCAB, previously reported in *G. pacifica* [29]. In addition, related saponins of **1**/**2** and **9**/**10** were found in *G. oldhamiana*

[28] and **27**/**28** – in *G. arrostii* [4]. **11** and **19**-**22** could be associated with saponins, identified in *G. paniculata* [30, 31, 35].
Recently, new GOTCAB saponins with C-28 acetylated ester chain were isolated from the roots of *G. oldhamiana* [28], *G. repens* [36], *G. altissima* [33], *G. perfoliata* [27] along with the GOTCAB references **B**, **C** and **D** from *G. trichotoma* [9]. Saponins **33** was previously identified in *G. trichotoma* (reference **B**), while related saponin of **34** was found in *G. altissima*. Indeed, 3-*O*-*β*-sulphated gypsogenin, oleanolic and quillaic acid monodesmosides have been reported from the roots of *G. bermejoi* G. Lopez, *G. trichotoma* and *G. pacifica*, together with a 23-*O*-sulphated hederagenin glycoside from *G. bermejoi* [37-39]. In addition, new sulphated triterpenoids were isolated from *G. pacifica* [39]. Although 11 sulphated bidesmosides of gypsogenin were already identified or tentatively elucidated in the *G. trichotoma* roots [15], this is the first report of GOTCAB with sulphated pentose moiety in the ester-bonded chain from the genus *Gypsophila*.

4. Conclusion

This study presents a strategy based on reverse-phase LC-HRMS to analyze GOTCAB saponins in *Gypsophila* species. On the basis of exact masses, fragmentation patterns and comparison with saponin references, a total of 41 GOTCAB were identified or tentatively elucidated in the GGR extract. They were classified into three types of saponins, depending on the substitution pattern of C-28 ester chain: unsubstituted three- to hexasaccharide, acetylated or sulphated oligosaccharide. Type I consisted of GOTCAB with unsubstituted ester chain (**1**-**32**) including a GOTCAB **1** isolated from GGR for the first time. One of them (**10**) was identified as reference **A**, while **19**-**22**, **27**/**28** and **31** corresponded to its derivatives. Acetylated GOTCAB from type II (**33**-**37**) included reference **B** and its derivative **35** along with **37** (acetylated derivative of **32)**. Type III GOTCAB with sulphated pentose unit in the α chain was identified for the first time in the genus *Gypsophila*. They are sulphated derivatives of compounds **1**/**2**, **9**/**10**, **11**, **27**/**28**, and to the best of our knowledge, they have not been described previously in the literature. However, it should be noted that mass spectrometric method has some limitations in differentiating the linkage position of sugar moieties. In this study is described the first systematic investigation on the fragmentation patterns and diagnostic fingerprints of the fragment ions in the MS/MS spectra of GOTCAB in *G. glomerata* roots. For the first time, large number of triterpenoid saponins in *G. glomerata* roots was characterized.

Acknowledgments. This study was financially supported by Grant 41/2016 by Council of Medical Science at Medical University – Sofia (Bulgaria) and Grant H2020 Widespread-Teaming Phase 2 project PlantaSYST No739582.

Disclosure statement
No potential conflict of interest was reported by the authors.

REFERENCES

[1] M. Henry, Saponins and phylogeny: Example of the “Gypsogenin group” saponins, Phytochem. Rev. 4 (2005) 89-94.

[2] S. Bottger, M. F. Melzig, Triterpenoid saponins of the Caryophyllaceae and Illecebraceae family, Phytochem. Lett. 4 (2010) 59-68.

[3] R. Gevrenova, L. Voutquenne-Nazabadioko, D. Harakat, E. Prost, M. Henry, Complete ^{1}H and ^{13}C NMR assignments of saponins from roots of *Gypsophila trichotoma* Wend., Magn. Res. Chem. 44 (2006) 686-691.

[4] D. Pertuit, S. Avunduk, A.-C. Mitaine - Offer, T. Miyamoto, Ch. Tanaka, T.Paululat, S. Delemasure, P. Dutartre, M.-A. Lacaille-Dubois, Triterpenoid saponins from the roots of two *Gypsophila* species, Phytochemistry102 (2014) 182-188.

[5] I. Arslan, A. Celik, M. F. Melzig, Nebulosides A–B, novel triterpene saponins from under-ground parts of *Gypsophila arrostii* Guss. var. *Nebulosa*, Bioorg. Med. Chem. 21(2013) 1279-1283.

[6] C. Bachran, S. Bachran, M. Sutherland, D. Bachran, H. Fuchs, Saponins in tumor therapy, Mini. Rev. Med. Chem. 8 (2008) 575-584.

[7] I. Arslan, A. Celik, J. H. Chol, A cytotoxic triterpenoid saponin from under-ground parts of *Gypsophila pilulifera* Boiss.& Heldr, Fitoterapia. 83 (2012) 699-703.

[8] Zhang W, J.-G. Luo, C. Zhang, L.-Y. Kong, Different apoptotic effects of triterpenoid saponin-rich *Gypsophila oldhamiana* root extact on hamun hepatoma SMMC-7721 and normal hamun hepatic L02 cells, Biol. Pharm. Bull. 36 (2013) 1080-1087.

[9] L. Voutquenne-Nazabadioko, R. Gevrenova, N. Borie, D. Harakat, Ch. Sayagh, A. Weng, M. Thakur, M. Zaharieva, M. Henry, Triterpenoid saponins from the roots of *Gypsophila trichotoma* Wender, Phytochemistry. 90 (2013) 114-127.

[10] H. Fuchs, N. Niesler, A. Trautner, S. Sama, G. Jerz,H. Panjideh, A.Weng, Glycosylated triterpenoids as endosomal escape enhancers in targeted tumor therapies, Biomedicines. 5 (2017) 1-25.

[11] C. D. Skene, P. Sutton, Saponin-adjuvanted particulate vaccines for clinical use, Methods (San Diego, Calif.) 40 (2006) 53-59.

[12] L.-P. Kang, Y. Zhao, X. Pang, H.-S. Yu, C.-Q. Xiong, J. Zhang, Y. Gao, K. Yu, C. Liu, B.-P. Ma, Characterization and identification of steroidal saponins from the seeds of *Trigonella foenum-graecum* by ultra high-performance liquid chromatography and hybrid time-of-flight mass spectrometry, J. Pharm. Biomed. Anal. 74 (2013) 257-267.

[13] G. Montoya, G. J. Arango, J. R. Ramırez-Pineda, Rapid differentiation of isobaric and positional isomers of structurally related glycosides from *Phytolacca bogotensis*. Rapid Commun. Mass Spectrom. 23 (2009) 3361–3371.

[14] M. Henry, M. Kowalczyk, M. Maldini, S. Piacente, A. Stochmal, W. Oleszek.. Saponin inventory from *Argania spinosa* kernel cakes by liquid chromatography and mass spectrometry, Phytochem. Anal. 24 (2013) 616-622.

[15] R. Gevrenova, V. Bardarov, K. Bardarov, L. Voutquenne-Nazabadioko, M. Henry, Selective profiling of saponins from *Gypsophila trichotoma* Wend. by HILIC separation and HRMS detection. Phytochem. Anal. 29 (2018) 250-274.

[16] R. Gevrenova, A. Weng, L. Voutquenne-Nazabadioko, M. Thakur, I. Doytchinova, QSAR study on saponins as cytotoxicity enhancers, Lett. Drug Des. Discov. 12 (2015) 166-171.

[17] St. Vulev, Gypsophila L., in: D. Jordanov (Ed.), Fl. Reipubl. Popularis Bulgaricae. Vol. 6, Aedibus Acad. Sci. Bulgaricae, Serdicae, Sofia, 1966, pp. 177-181.

[18] R. Gevrenova, T. Stancheva, Y. Voynikov, D. Laurain – Mattar, M. Henry, Root in vitro cultures of six *Gypsophila* species and their saponins contents, Enzyme Microb. Technol. **47** (2010), 97-104.

[19] R. Gevrenova, K. Bardarov, Y. Voynikov, D. Zheleva-Dimitrova, V. Balabanova, M. Henry, Preliminary study on sulphated GOTCAB saponins from two Bulgarian *Gypsophila* L. (Caryophyllaceae) species. C. R. Acad. Bulg. Sci. 70 (2017) 1539-1548.

[20] R. Gevrenova, O. Joubert, Tsv. Mandova, M. Zaiou, Y. Chapleur, M. Henry, Cytotoxic effects of four Caryophyllaceae species extracts on macrophages cell lines. Pharm. Biol. 52 (2014) 919-925.

[21] C. Li, C. Huang, T. Lu, L. Wu, S. Deng, R. Yang, J. Li, Tandem mass spectrometric fragmentation behavior of lignans, flavonoids and triterpenoids in *Streblus asper*. Rapid Commun. Mass Spectrom. 28 (2014) 2363–2370.

[22] B. Domon, C. Costello, A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS Spectra of Glycoconjugates, Glycoconjugate J 5 (1988) 397-409.

[23] L. P. Sandjo, M. V. P. dos Santos Nascimento, L. A. L. da Silva, A. C. M. Munhoz, L. A. E. Pollo, M. W. Biavatti, B. T. Ngadjui, T. Opatz, T. S. Fröde, ESI-MS^2 and Anti-inflammatory Studies of Cyclopropanic Triterpenes. UPLC-ESI-MS and MS^2 Search of Related Metabolites from *Donella ubanguiensis*, Phytochem. Anal. 28 (2017) 27-41.

[24] Z. Yan, G. Lin, Y. Ye, Y. Wang, R. Yan, Triterpenoid saponins profiling by adducts-targeted neutral loss triggered enhanced resolution and product ion scanning using triple quadrupole linear ion trap mass spectrometry. Anal. Chim. Acta 819 (2014) 56–64.

[25] B. Khakimov, L.H.Tseng, M. Godejohann, S. Bak, S.B. Engelsen, Screening for triterpenoid saponins in plants using hyphenated analytical platforms. Molecules 21 (2016) 1614-1632.

[26] S. Ji, Q. Wang, X. Qiao, H.-c. Guo, Y.-f. Yang, T. Bo, C. Xiang, D.-a. Guo, M. Ye, New triterpene saponins from the roots of *Glycyrrhiza yunnanensis* and their rapid screening by LC/MS/MS. J. Pharm. Biomed. Anal. 90 (2014) 15-26.

[27] Q. Chen, J.G. Luo, L. Y. Kong, New triterpenoid saponins from the roots of *Gypsophila perfoliata* Linn., Carbohydr. Res. 346 (2011) 2206-2212.

[28] J.-G. Luo, L. Ma, L.-Y. Kong, New triterpenoid saponins with strong [alpha]- glucosidase inhibitory activity from the roots of *Gypsophila oldhamiana*., Bioorg. Med. Chem. 16 (2008) 2912–2920.

[29] W. Nie, J.-G. Luo, X.-B. Wang, L.-Y. Kong, An insight into enrichment and separation of oleanane-type triterpenoid saponins by various chromatographic materials, Sep. Purif. Technol. 65 (2009) 243–247.

[30] D. Frechet, B. Christ, B. Monegier du Sorbier, H. Fisher, M. Vuilhorgne, Four triterpenoid saponins from dried roots of Gypsophila species, Phytochemistry. 30 (1991) 927-931.

[31] S. Yao, L. Ma, J.-G. Luo, J.-S. Wang, L.-Y. Kong, New Triterpenoid Saponins from the Roots of *Gypsophila paniculata* L., Helv. Chim. Acta. 93 (2010) 361-374.

[32] W. Nie, J.-G. Luo, L.-Y. Kong, New triterpenoid saponins from the roots of *Gypsophila pacifica* Kom., Carbohydr. Res. 345 (2010) 68–73.

[33] Q. Chen, J.G. Luo, L. Y. Kong, Triterpenoid Saponins from *Gypsophila altissima* L., Chem. Pharm. Bull. 58 (2010) 412—414.

[34] G. Verardo, A. Gorassini, D. Riccic, D. Fraternalec, High Triterpenic Acids Production in Callus Cultures from Fruit Pulp of Two Apple Varieties, Phytochem Anal. 28 (2017) 5-15.

[35] A. Weng, K. Jenett-Siems, P. Schmieder, D. Bachran, C. Bachran, C. Gorick, M. Thakur, A convenient method for saponin isolation in tumour therapy, J. Chromatogr. B 878 (2010) 713-718.

[36] M. Elbandy, T. Miyamoto, M. A. Lacaille-Dubois, New triterpenoidal saponins from *Gypsophila repens*, Helv. Chim. Acta 90 (2007) 260-270.

[37] B. Acebes, A. M. Dıaz-Lanza, M. Bernabe. A saponin from the roots of *Gypsophila bermejoi*. *Phytochemistry* **1998**, *49*, 2077–2079.

[38] I. Krasteva, K. Jenett-Siems, M. Kaloga, S. Nikolov. 3-*O*-Sulfo-triterpenoid saponins from *Gypsophila trichotoma* Wend. *Z. Naturforsch. B* **2009**, *6*4, 319–322.

[39] J.-G. Luo, W. Nie, L.-Y. Kong. Three new sulphated triterpenoids from the roots of *Gypsophila pacifica*. *J. Asian Nat. Prod. Res.* **2011**, *13*, 529-533.

Table 1. ^{1}H and ^{13}C NMR data of the aglycone of **1** at 600MHz and 150 MHz, respectively.

Atom	δ(H) ppm	δ(C) ppm	Atom	δ(H) ppm	δ(C) ppm
CH_2(1)	1.69/1.09	38.9	CH_2(16)	2.03/1.59	23.7
CH_2(2)	2.03/1.80	25.4	C(17)	-	47.4
H-C(3)	3.92	86.0	H-C(18)	2.80	42.5
C(4)	-	56.3	CH_2(19)	1.70/1.13	46.9
H-C(5)	1.33	48.6	C(20)	-	31.1
CH_2(6)	1.49/0.96	21.3	CH_2(21)	1.36/1.21	34.5
CH_2(7)	1.47/1.33	33.1	CH_2(22)	1.75/1.58	32.7
C(8)	-	40.6	H-C(23)	9.44	211.8
H-C(9)	1.67	48.5	Me(24)	1.17	10.7
C(1)0	-	36.5	Me(25)	0.98	16.0
CH_2(11)	1.91	24.2	Me(26)	0.76	17.6
H-C(12)	5.26	123.3	Me(27)	1.16	26.1
C(13)	-	144.2	C(28)	-	178.2
C(14)	-	42.6	Me(29)	0.90	33.3
CH_2(15)	1.56/1.22	28.9	Me(30)	0.92	24.0

Table 2. Table 1. ^{1}H and ^{13}C NMR data of the sugar moieties of **1** at 600MHz and 150 MHz, respectively.

Trisaccharide in C-3 of gypsgenin			Tetrasaccaharide in C-28 of gypsogenin		
Atom	δ(H)	δ(C)	Atom	δ(H)	δ(C)
β-D-GlcA			*α*-D- Fuc		
H-C(1)	4.42	104.2	H-C(1)	5.33	95.0
H-C(2)	3.69	**78.5**	H-C(2)	3.82	**74.4**
H-C(3)	3.72	**85.8**	H-C(3)	3.63	73.2
H-C(4)	3.60	70.7	H-C(4)	3.75	75.9
H-C(5)	3.57	76.5	H-C(5)	3.72	72.5
C(6)OOH	-	169.9	Me(6)	1.22	16.3
β-D- Gal			*α*-L-Rha		
H-C(1)	4.78	104.6	H-C(1)	5.37	101.0
H-C(2)	3.32	74.7	H-C(2)	4.12	71.1
H-C(3)	3.39	77.6	H-C(3)	3.84	**81.6**
H-C(4)	3.87	69.9	H-C(4)	3.65	**79.1**
H-C(5)	3.50	74.8	H-C(5)	3.85	69.0
CH_2(6)	3.81/3.70	62.0	Me(6)	1.25	18.3
β-D-Ara1			*β*-D-Ara2		
H-C(1)	4.79	104.91	H-C(1)	4.50	105.6
H-C(2)	3.48	87.2	H-C(2)	3.33	77.6
H-C(3)	3.51	73.1	H-C(3)	3.56	72.6
H-C(4)	3.71	73.3	H-C(4)	3.55	74.2
CH_2(6)	3.83/3.64	66.9	CH_2(6)	3.92/3.60	67.3
			β-D-Xyl		
			H-C(1)	4.63	104.9
			H-C(2)	3.05	75.5
			H-C(3)	3.29	78.3
			H-C(4)	3.50	70.8
			CH_2(6)	3.84/3.20	66.5

Figure captions

Fig. 1. Structures of saponin references A-E.

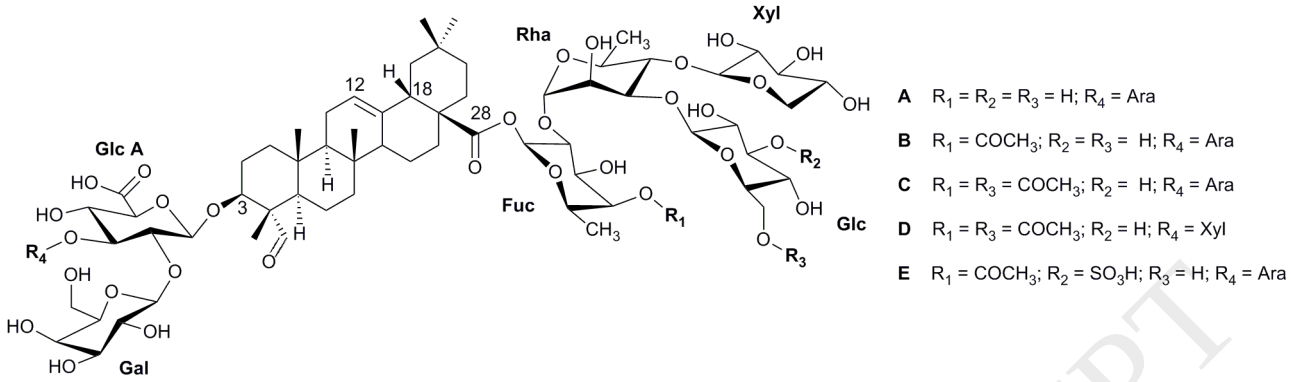

Fig. 2. HMBC correlations of **1**.

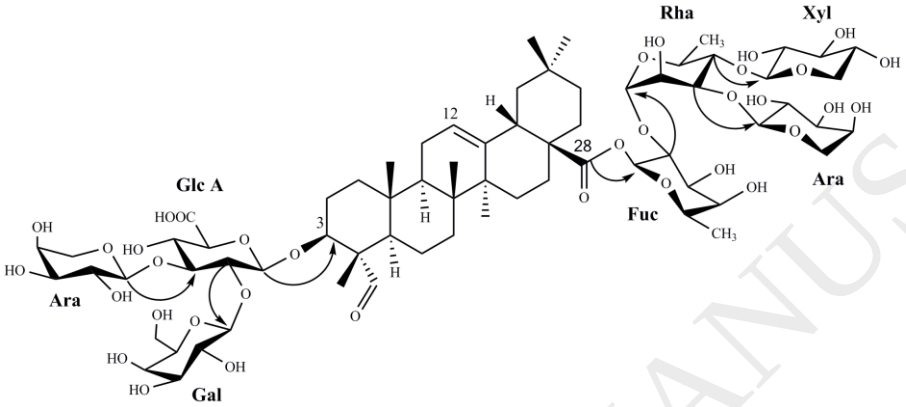

Fig. 3. UHPLC-ESI/MS/MS spectra of saponin reference **A**, $[M-H]^-$ at *m/z* 1525.6698 at hcd 20 (A), 25 (B), 30 (C) and 35 (D).

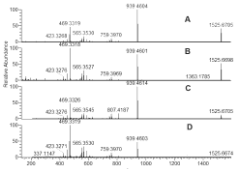

Fig. 4. Proposed fragmentation pathway of saponin reference **A** (see Fig. 1 and Table 1) (-ESI).

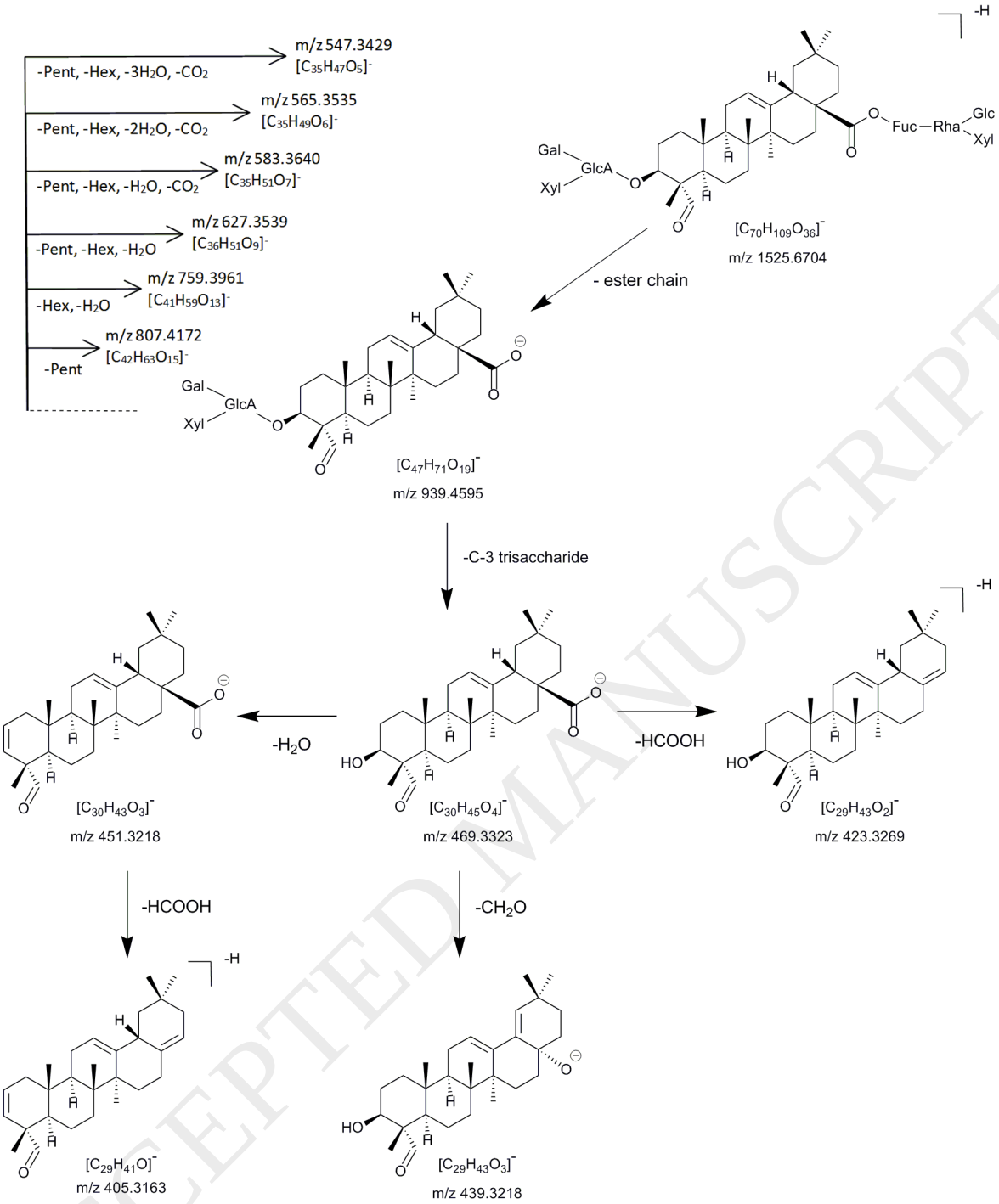

Fig. 5. Fragmentation pattern of saponin references **C/D** (see Fig. 1 and Table 1).

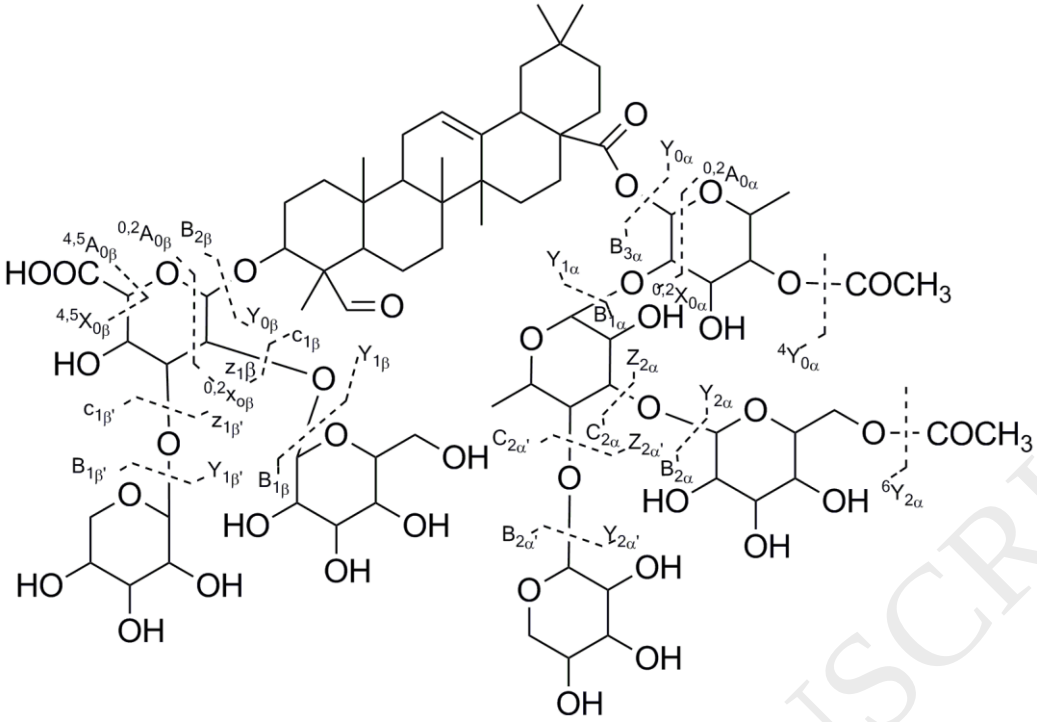

Fig. 6. Total ion chromatogram (TIC) of GGR saponins isolated using SPE and analyzed by UHPLC-ESI/HRMS in negative ion mode. Extracted ion chromatograms (mass tolerance of 5 ppm) of $[M-H]^-$ at *m/z* 1541.6653 (1541.6615-1541.6769) (A); *m/z* 1511.6548 (1511.6480-1511.6632) (B); *m/z* 1525.6704 (1525.6617-1525.6769) (C); *m/z* 1495.6598 (1495.6512-1495.6662) (D); *m/z* 1621.6221 (1621.6129-1624.6341) (E); *m/z* 1539.6861 (1539.6772-1539.6926) (F); *m/z* 1605.6272 (1605.6181-1605.6341) (G).

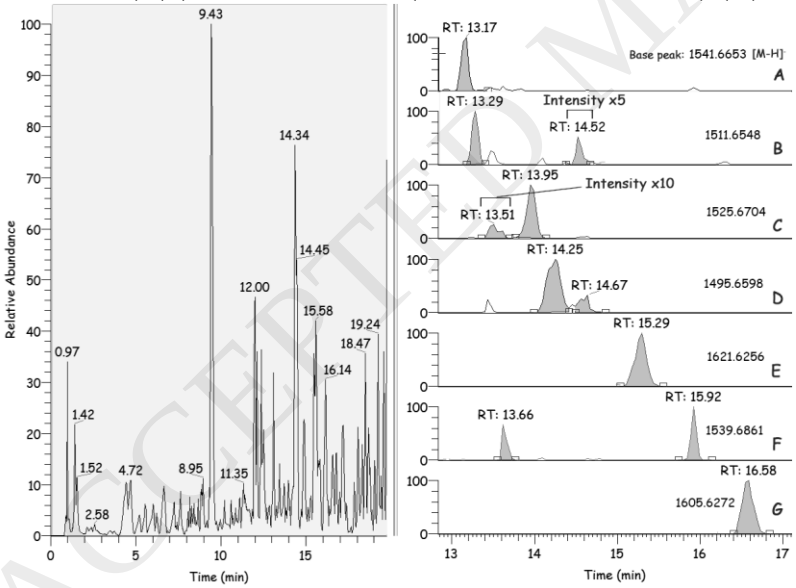

Fig. 7. UHPLC-ESI/MS/MS spectra of compound **5** (A); **1** (B) and **7** (C). The same spectra between *m/z* 380 and 490 and proposed fragmentation patterns for quillaic acid (D); gypsogenin (E); olenolic acid (F).

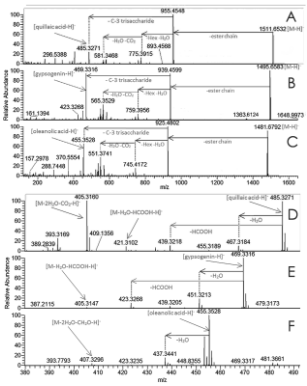

Fig. 8. Proposed fragmentation pathway of saponin **5** (see Table 2) (ESI-)

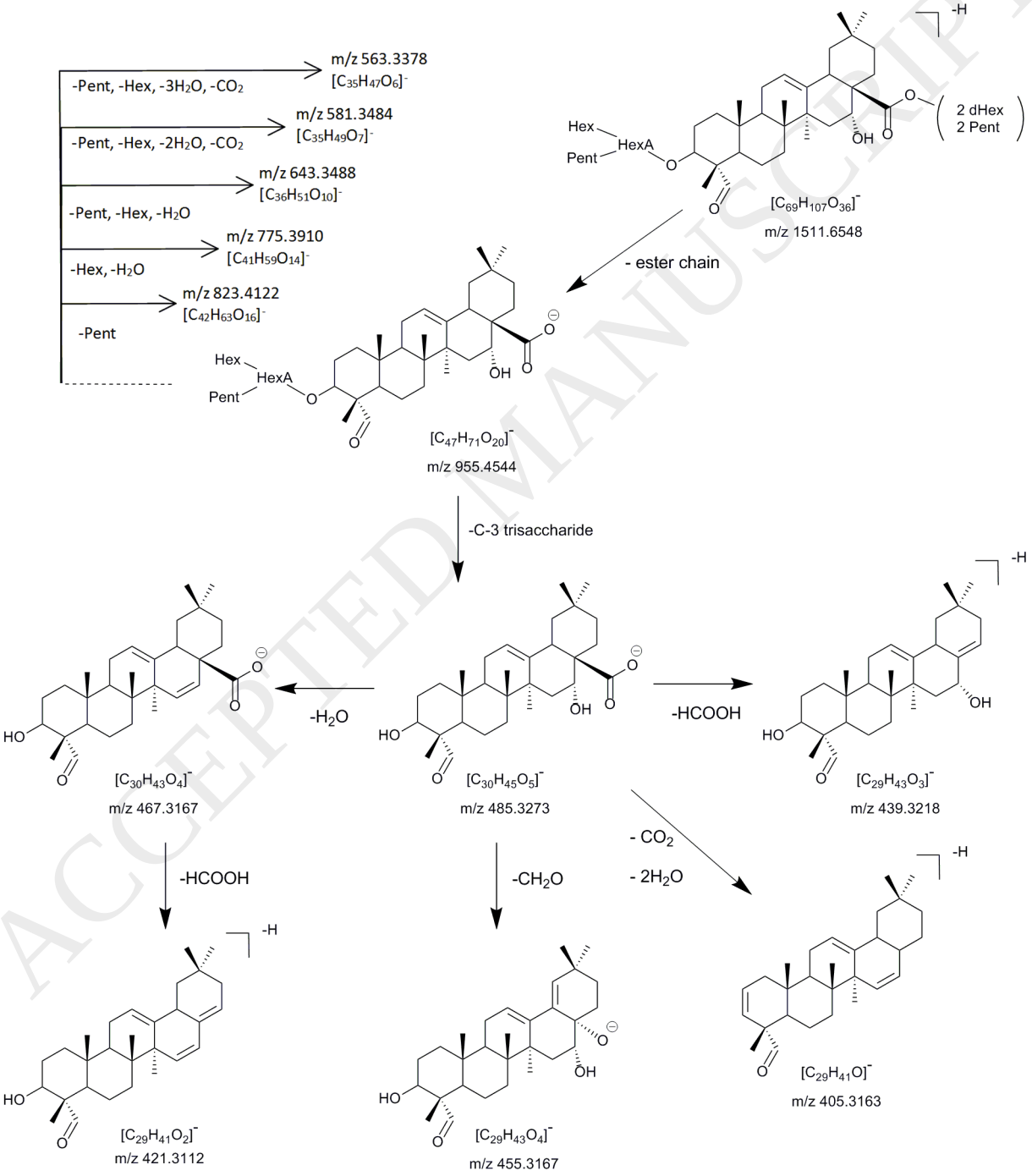

Fig. 9. Fragmentation of *O*-sulphate-pentosyl-(1→4)-[hexosyl-(1→3)]-deoxyhexosyl-(1→2)-deoxyhexoside (saponin **39**).

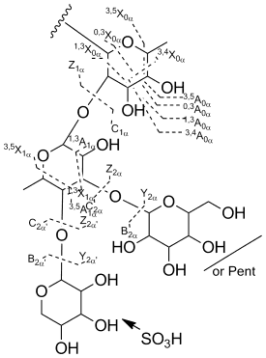

Fig. 10. Fragmentation key to screen and characterize GOTCAB saponins from *Gypsophila* species