약학박사 학위논문

Targeted classification, identification and metabolite profiling of triterpenoids in the genus *Gymnema* and *Gynostemma* by developing a building block strategy using UHPLC-QTOF/MS

UHPLC-QTOF/MS를 활용한 building block 전략 및 이를 활용한 *Gymnema* 및 *Gynostemma* 속 triterpenoids 대사체 프로파일링 연구

2019 년 8 월

서울대학교 대학원

약학과 생약학 전공

Pham Ha Thanh Tung

Abstract

Targeted classification, identification and metabolite profiling of triterpenoids in the genus *Gymnema* and *Gynostemma* by developing a building block strategy using UHPLC-QTOF/MS

The quality control of a medicinal plant requires a rapid, accurate, and comprehensive analysis of the chemical profile of the bioactive extract. Common HRMS-based compound annotations relied on the mass fragmentation analysis of detected peaks to identify the partial structure of the compounds before merging them together. Herein, we applied a universally accepted concept that most of the natural products of the same organic origin are generated by the same building blocks using several conservative biosynthetic processes. The detection of metabolites - both known and unknown - can be redefined by an inverse approach, in which, prebuilt natural products were predicted using key prefabricated LEGO-type building blocks and their biosynthetic construction rules. The conventional mass fragmentation pattern analysis and NMR experiments are for the confirmation of the existence of predicted compounds.

Furthermore, we proposed a new idea of multilayer metabolite profiling, which provides simple, distinct, and comprehensive chromatogram interface with not only original information about retention time of metabolites, but also considerable insight into structure formation as well as the chemical relationship among plant metabolites. In this study, the building block strategy was applied effectively to explore the triterpenoid composition of three medicinal plants

including *Gymnema sylvestre*, *Gymnema latifolium* and *Gynostemma longipes*. Overall, we expect this new approach can be used, with much practicality, for massive structural characterization and for exploring the biosynthetic relationships among various compounds in medicinal plants.

Part 1: Discrimination of different geographic varieties of *Gymnema sylvestre* and development of a building block strategy to classification, identification and metabolite profiling of its oleanane triterpenoids

The major class of bioactive metabolites in *Gymnema sylvestre*, a popular Ayurevedic medicinal plant for the treatment of diabetes mellitus, is oleanane triterpenoids. In this study, a targeted, biosynthesis-inspired approach using UHPLC-qTOF/MS was implemented to elucidate the whole chemical profile of the plant for the standardization of the Vietnamese *G. sylvestre* variety. The known compounds reported in the literature were first analyzed to identify the building blocks of the biosynthetic intermediates and the construction rules for synthesizing oleanane triterpenoids in the plant. These blocks were recombined to build up a theoretical virtual library of all reasonable compounds consistent with the deduced construction rules. Various techniques, including microfractionation, relative mass defect filtering, multiple key ion analysis, mass fragmentation analysis, and comparison with standard references, were applied to determine the presence of these predicted compounds. Conventional isolation and structure elucidation of 6 of the new compounds were carried out to identify new building blocks and validate the assignments. Consequently, 119 peaks were quickly assigned to oleanane triterpenoid, and among them, 77 peaks are predicted to be new compounds based on their molecular formulas and mass fragmentation patterns. All the identified

metabolites were then classified into different layers to analyze their logical relationships and construct a multilayered chemical profile of the oleanane triterpenoids.

Part 2: Oleanane triterpenoids from *Gymnema latifolium* and their PTP1B inhibitory activities

Natural products are promising sources of lead compounds that play significant roles in the discovery of new antidiabetic agents via the mechanism of PTP1B inhibition. In our ongoing research to find PTP1B inhibitors from natural products, hundreds of plant's extracts available in Korea Bioactive Natural Material Bank has been screened against this biological target. An extract of the *Gymnema latifolium* Wall. ex. Wight showed considerable PTP1B inhibitory activity. This result was in good agreement with the use of this plant in Vietnam as an antidiabetic herbal medicine, similar with its taxonomical relative *G*. *sylvestre*. It is also well-known that most species belonging to the same genus possess similar chemical composition and thus exhibit similar biological activities. Further chemical investigation of this plant led to the isolation of 14 new oleanane triterpenes Gymlatinosides GL1-GL14 and 6 known oleanolic acid analogs. The structures of the new compounds were elucidated using diverse spectroscopic methods. Among them, compounds Gymlatinosides GL2 and GL3 showed significant PTP1b inhibitory effect. The building block strategy was also applied successfully to predict the structures of 54 new compounds which may have similar skeleton and biosynthesis construction with those isolates.

Part 3: 12,23-Dione dammarane triterpenes from *Gynostemma longipes* and their muscle cell proliferation activities via activation of the AMPK pathway.

The aging population is growing rapidly around the world. Sarcopenia, characterized by decreased muscle mass, strength, and function, is a common feature of the elderly population. AMP-activated protein kinase (AMPK) is an essential sensor and regulator of glucose, lipid, and energy metabolism throughout the body. Previous studies have shown that AMPK pathway activation by regular exercise and appropriate dietary control have beneficial effects on skeletal muscle. In the process of searching for new AMPK activators from medicinal plants, we isolated and characterized eight new 12,23-dione dammarane triterpenoids (**1**–**3** and **5**–**9**), as well as one known gypentonoside A from *Gynostemma longipes*. Application of the building blocks and their construction rules also led to the assignment of 26 possible new compounds. All 9 isolated compounds were tested for their AMPK activation activities, seven compounds (**1** and **3**–**8**) were significantly activated AMPK phosphorylation in mouse C2C12 skeletal muscle cell lines. Since *G*. *longipes* contained a significant amount of active compound Longipenoside A1 (over 2.08% per dried raw plant), it suggested the potential of this plant to be developed as a functional food or botanical drug that enhances muscle proliferation by activating AMPK signaling pathways.

Keywords: *Gynostemma longipes*, *Gymnema sylvestre*, *Gymnema latifolium*, oleanane triterpenes, dammarane triterpenes, building block strategy, PTP1B, muscle cell proliferation.

Student Number: 2015-30877

TABLE OF CONTENTS

List of Abbreviations

$[\alpha]_D^{25}$	Specific rotation at 25 Celsius degrees
1D-NMR	One-dimensional nuclear magnetic resonance spectroscopy
2D-NMR	Two-dimensional nuclear magnetic resonance spectroscopy
Ac	Acetyl
Ahg	*β*-D-arabino-2-hexulopyranosyl(1→3)-*β*-D-glucuronopyranoside
AMPK	AMP-activated protein kinase
Amu	Atomic mass unit
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
BrdU	Bromodeoxyuridine
Bz	Benzoyl
^{13}C	Carbon nuclear magnetic resonance spectroscopy
Cald	Calculated for
CAS	Chemical abstract service (registry number)
CC	Column chromatography
CH_3CN	Acetonitrile
COSY	Correlated spectroscopy
d	Doublet
Da	Dalton (atomic mass unit)
dd	Double of doublet
DM	Diabetes mellitus
DMSO	Dimethyl sulfoxide

DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
EtOAc	Ethyl acetate
EtOH	Ethanol
FBS	Fetal bovine serum
GLa	*Gymnema latifolium*
GLo	*Gynostemma longipes*
GlcA	*β*-D-glucuronopyranosyl
Glc	*β*-D-glucopyranosyl
GPS	Global Positioning System
GS	*Gymnema sylvestre*
GS-C	Chinese *Gymnema sylvestre* variety
GS-I	Indian *Gymnema sylvestre* variety
GS-V	Vietnamese *Gymnema sylvestre* variety
^{1}H	Proton nuclear magnetic resonance spectroscopy
HMBC	Heteronuclear Multiple Bond Correlation
HNIP	Medicinal Herbarium of Hanoi University of Pharmacy
HPLC	High Performance Liquid Chromatography
HRESIMS	High-resolution electrospray ionization mass spectrometry
HSQC	Heteronuclear Single Quantum Coherence
Hx	Hexane
Hz	Hertz

IHD	Indices of hydrogen deficiency
IR	Infrared (spectrum)
ITS	Internal transcribed spacer (sequence)
KBNMB	Korea Bioactive Natural Material Bank
KIF	Key ion filtering
LC-MS	Liquid chromatography – mass spectroscopy
m	Multiplet
Mba	2-(*S*)-methylbutyroyl
MeOH	Methanol
MFA	Mass fragmentation analysis
mKIA	Multiple key ion analysis
MMP	Multilayer metabolite profile
mPIF	multiple product ion filtering
MTT	3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2*H*-tetrazolium bromide
m/*z*	Mass to charge ratio
NMR	Nuclear magnetic resonance
NOESY	Rotating frame nuclear Overhauser Effect Spectroscopy
NP	Normal-phase (silica gel)
Nic	Nicotinoyl
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PI	Propidium iodide
pNPP	Para-nitrophenylphosphate
PTP	Protein tyrosine phosphatases

PTP1b	Protein tyrosine phosphatase 1B
PUBMED	National Center for Biotechnology Information, U.S. National Library of Medicine
PVDF	Polyvinylidene fluoride
Ref	Reference standard compound
Rha	α-L-rhamnopyranosyl
RMD	Relative mass defect
RNA	Ribosomal ribonucleic acid
RP-C_{18}	C_{18}-reversed phase silica gel
Rt	Retention time
s	singlet
SNP	Single-nucleotide polymorphism
t	Triplet
T2DM	Type 2 diabetes mellitus
TIC	Total ion chromatogram
Tig	Tigloyl
TLC	Thin-layer chromatography
qTOF-MS	Quadrupole-time-of-flight mass spectrometry
t_R	Retention time
UPLC	Ultraperformance liquid chromatography
UV	Ultraviolet (spectrum, detector)
Xyl	β-D-xylopyranosyl

LIST OF FIGURES

LIST OF TABLES

Part 1. Discrimination of different geographic varieties of *Gymnema sylvestre* and development of a building block strategy to classification, identification and metabolite profiling of oleanane triterpenoids in its Vietnamese variety using UHPLC-qTOF/MS

1. Introduction

Gymnema sylvestre (Retz.) R.Br. ex Sm. (GS) is a well-known Ayurvedic medicinal plant that has a deep root in history and is being studied extensively for its effectiveness in the treatment of type 2 diabetes mellitus (T2DM) (Pothuraju et al., 2014). Preparations of this plant include a simple tea brew, tea bags, beverages and confectioneries (Tiwari et al., 2014).

It has also found significant applications in various food preparations for regulation of sugar homeostasis, control of obesity and blood cholesterol levels. A food supplement based on wheat (*Triticum aestivum),* blended with legumes, non-fat dry milk, vegetable oils, spices and hypoglycemic ingredients including GS, was formulated as a suitable supplement or meal alternative for non-insulin-dependent diabetes patients (Shobana et al., 2007).

Most research on GS to date was performed on material from India, and its main active components are a group of gymnemic acids with a β-glucuronic acid at C-3 and a hydroxyl substitution at C-23 on an oleane triterpene-type aglycon (Pothuraju et al., 2014). These gymnemic acids have been recognized for years for their role in suppressing selectively sweet taste sensations in humans (Warren and Pfaffmann, 1959) (Frank et al., 1992) (Gent et al., 1999). Recently, Kashima et al. suggested that the subjective sweet taste intensity in GS-administered volunteers was decreased

compared with a control group and revealed the role of a GS extract in delaying postprandial gastrointestinal blood flow and gastric emptying, which may affect subsequent glycemic metabolism (Kashima et al., 2017).

LC-MS analysis of the GS extract after acid hydrolysis from different geographic distributions (India, Vietnam and China) revealed the similarity in LC pattern between samples from Vietnam and China but showed a significant discrepancy with samples of Indian origin. Specifically, gymnemagenin, a 23-hydroxyl triterpene aglycon, was found in the Indian sample but not the Vietnamese and Chinese samples. This result is in good agreement with the proposed chemical variation of GS varieties from China, characterized by the lack of 23 hydroxyl functional groups, from their oleanane-type triterpene glycosides (Ye et al., 2001b).

The advent of liquid chromatography coupled mass spectrometry (LC/MS) with advantages of rapid analysis, high mass accuracy, and high resolution has provided powerful support for the rapid identification and discovery of secondary metabolites in complex natural extract (Wolfender et al., 2015). However, recognition of metabolites in automatically acquired mass spectrometric data confront plenty of obstacles: (1) large amounts of interference ions often mask the ions of analyzed compounds; (2) plants are always comprised by hundreds of different constituents that belong to different group of metabolites; and (3) the availability of reference standards is necessary for profiling the constituents in herbal medicine (Xing et al., 2015).

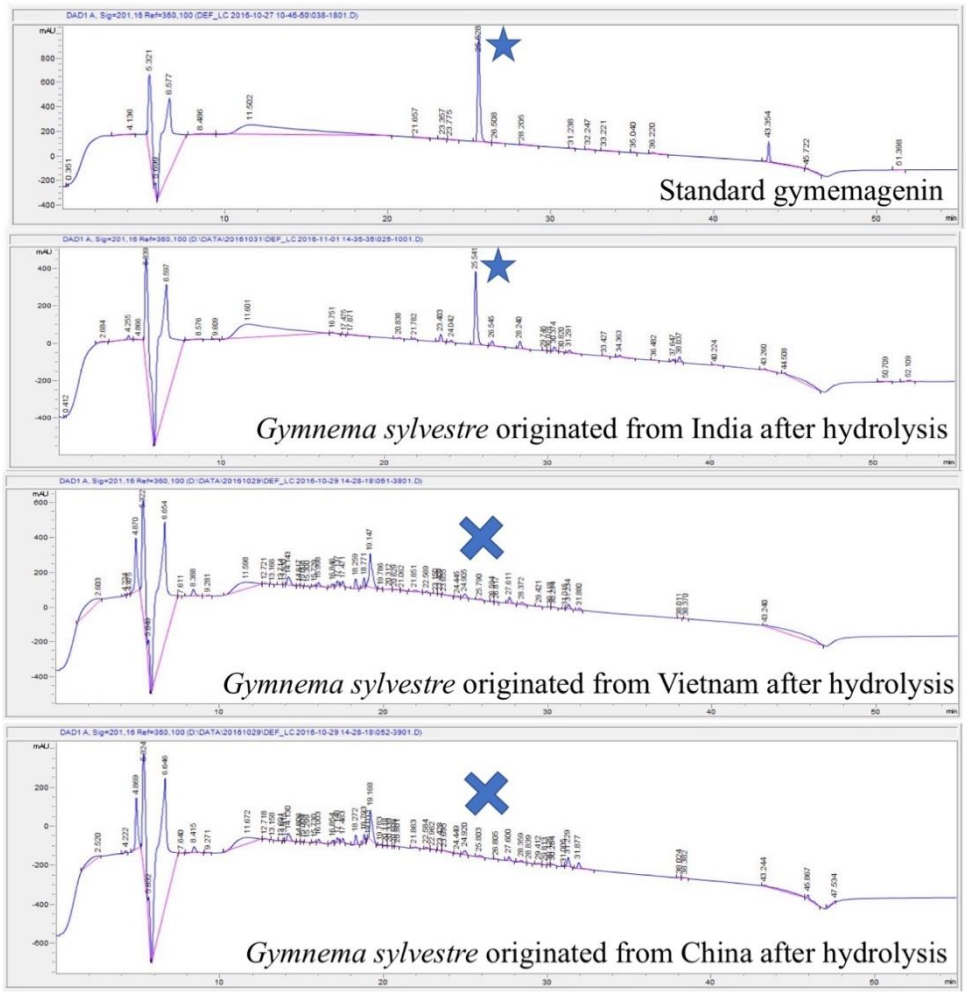


Figure 1: LC chromatogram of acid hydrolysis products of *G*. *sylvestre* originated from Vietnam, China, India and compare with standard gymnemagenin

In recent years, several approaches to postacquisition processing techniques for the detection and structural assignment of the chemical constitutes in medicinal plants have been developed. Traditional manual processing is generally slow, tedious, and likely to miss the trace metabolites (Xie et al., 2012). An improved method, relative mass defect (RMD) calculations, has enabled the rapid identification of an extensive range of different groups of metabolites in the complicated matrixes of plant extracts (Ekanayaka et al., 2015). Another approach

using key ion filtering (KIF) has been reported to be helpful in the identification of triterpenoid saponins that have a robust aglycone backbone and similar mass fragmentation patterns (Qi et al., 2012; Qiao et al., 2016; B. Yang et al., 2018). However, in many cases, the data from a single KIF of the total extract contained a substantial amount of noise, were nonselective, and were difficult to analyze. As a result, the multiple product ion filtering (mPIF) technique was developed and proved to be effective for identifying and classifying nontarget/target notoginsenosides in *P. notoginseng* extract (Xing et al., 2015). Recently, progress has been made in the deconvolution of complex extracts into their single constituents by the microfractionation of the extracts based on advanced high-performance liquid chromatography (HPLC) (Bohni et al., 2013). This technique is fast, simple, and effective for minimizing the loss of compounds of interest on the column and significantly amplifying the signals of low-content metabolites.

Notably, most natural products with the same organic origin are built from the same precursors, which could be considered the building blocks in the biosynthesis process. Additionally, knowing the structures of several of the key building blocks is sufficient to obtain an overview of the diversity of the natural products and give some insights into the metabolism (Franck, 1979). To explain how natural products are constructed, the structures can be broken down into their basic building blocks, and this fragmentation can be used to propose a reasonably construction pathway. This approach facilitates the identification of secondary metabolites and can be used to rationalize the elaborated structures and thus reveals logical relationships between apparently different chemical compounds (Dewick, 2009).

Any chemical composition variation can influence the pharmacological activity, safety and standardization of a medicinal plant, and thus, we have focused this study on the discrimination of two varieties of GS using different approaches, including morphological, anatomical and ITS1-5.8S-ITS2 DNA sequence comparison.

This research was conducted in 6 steps as shown in **Figure 2**. (1) A virtual mass library of predicted oleanane triterpenoids was established using their building blocks and the putative biosynthesis construction. (2) Microfractionation of the total extract was conducted using semipreparative HPLC. (3) Untargeted data of *G. sylvestre* extract and its microfractions were acquired by UPLC/qTOF-MS. (4) Integrated mass analysis techniques including RMD filtering were used for the rapid identification of the targeted oleanane triterpenoids and the elimination of fractions of nontargeted group of metabolites, and key ion extraction for the predicted compounds was used to confirm their presence in the extract and fractions; mKIA recognize aglycones, and MFA was used to confirm glycosidic and acyl linkages. (5) The compound determinations were validated by comparison with reference substances and the structure elucidation of selected compounds. (6) A multilayer metabolite profile of the *G. sylvestre* extract was constructed based on all the compound assignments to analyze the chemical relationships among the structures.

Purpose of research

- Identify the differences between Vietnamese and Indian *G. sylvestre* varieties.
- Establish a dereplication strategy to identify quickly the chemical composition of *G.sylvestre* Vietnamese variety

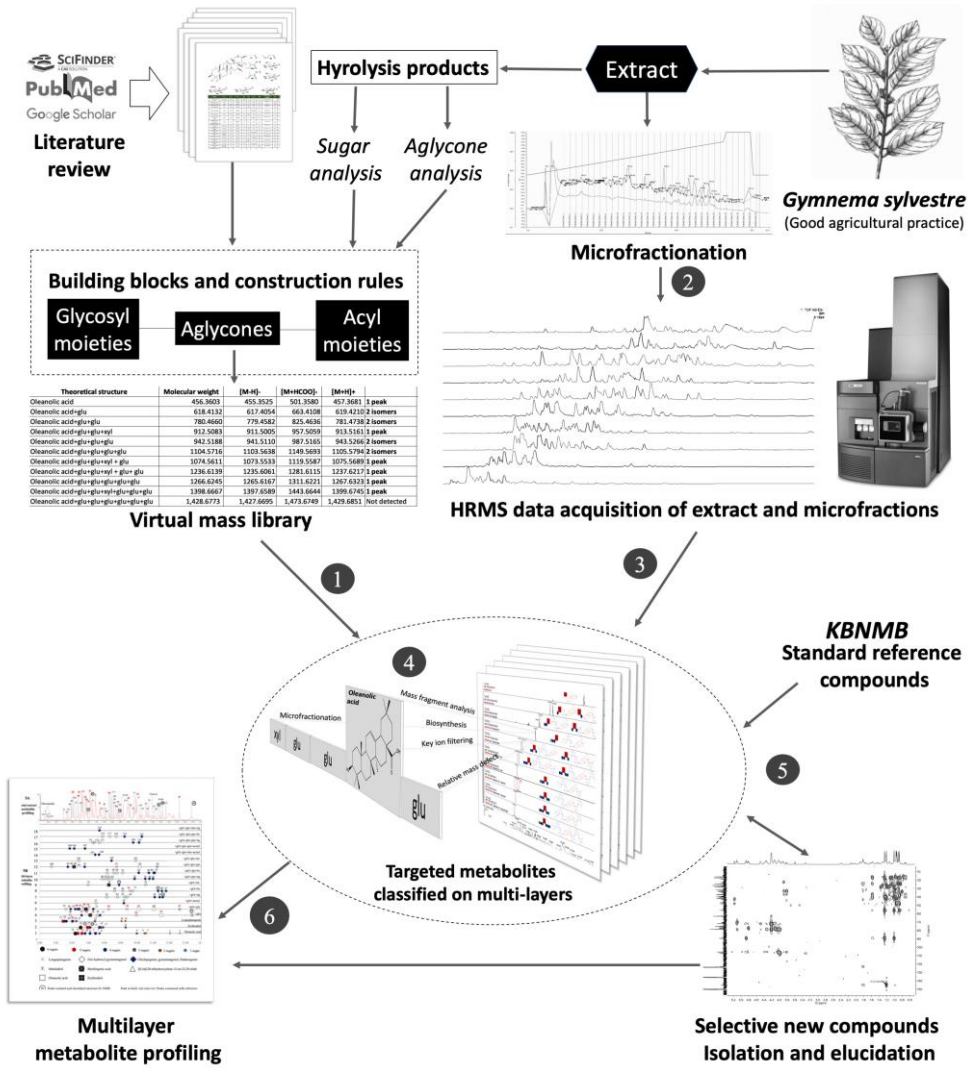


Figure 2: Research strategy of the study

2. Materials and methods

2.1. Plant materials

Samples of two varieties of *Gymnema sylvestre* originating from India (GS-I) and Vietnam (GS-V1 and G-V2) were cultivated in the same conditions at a WHO Good-Agriculture-Practice-certificated farm (GAP) in Nam Dinh and Thai Nguyen provinces, Vietnam. Samples were collected in September 2016. Voucher specimens of GS-I, GS-V1 and GS-V2 were deposited in the Medicinal Herbarium of Hanoi University of Pharmacy with the accession numbers HNIP-2016.05, HNIP-2016.06, and HNIP-2016.07, respectively. Samples of *Gymnema sylvestre* originating from Guangxi - China (GS-C) were collected in June 2017 and used in the hydrolysis experiment.

2.2. Morphology and anatomy analysis

An EZ4 Stereo Microscope (Leica, Germany) was used to analyze the characteristics including life form, stem, leaves, flowers, fruits and seeds. Photographs were obtained with a Canon SD4500IS or Canon EOS 60D + Canon 100 mm f2.8 IS Macro (Canon Inc., Japan). For anatomical analysis, cross sections of fresh mature stems and leaves were prepared using a rotatory microtome and double-stained with methylene blue and carmine red. A light microscope MBL200 (A.Krüss Optronic, Germany) connected to a Canon SD4500IS (Canon Inc., Japan) was used to visualize the results.

2.3. ITS1-5.8S-ITS2 sequence analysis

Total DNA was extracted from 200 mg of fresh plant leaves using a DNeasy Plant Mini Kit (QIAGEN, Germany) with some modifications. The internal

transcribed spacer sequence was amplified with forward primer ITS5 (5′-GGAAGTAAAAGTCGTAACAAGG-3′) and reverse primer ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) supplied by Bioneer (Bioneer Corporation, Korea) using a Mastercycler pro S (Eppendorf AG., Germany). PCR products were cleaned using a purification KIT from Thermo Fisher (USA), and sequencing was conducted by Macrogen Inc. (Seoul, Korea). DNA sequences were compared to published sequences available on GenBank (National Institutes of Health) using the Basic Local Alignment Search Tool (Blast) (Altschul et al., 1990).

Internal transcribed spacer (ITS1-5.8S-ITS2) sequences of two samples of *Gymnema sylvestre* collected in different regions of Vietnam (*G. sylvestre* V1 and V2) and one of Indian origin domesticated in Vietnam (GS-I), together with 18 ITS sequences of five species − *G. sylvestre*, *G. latifolium*, *G. inodorum*, *G. yunanense* and *Marsdenia tenacissima* (as an outgroup) − obtained from Blast analysis were aligned using Geneious alignment. Finally, Geneious DNA sequencing analysis software (version 8.1.8, Biomatters Ltd, New Zealand) was used to construct a neighbor-joining phylogenetic tree with resampling bootstrap values above 75% (expressed as percentages of 1,000 replications) (Kearse et al., 2012).

2.3. Establishment of a Virtual Mass Library based on Building Blocks

A comprehensive review of all triterpenoid compounds isolated from *G. sylvestre* was retrieved from publications indexed on Scifinder, Pubmed, and Google scholar. However, due to the differences of chemical composition by different geographic distributions (Pham et al., 2018; Ye et al., 2001b), only compounds isolated from *G. sylvestre* originated from China, Vietnam, and Taiwan were used for data processing. It is also noteworthy that species *G. alternifolium* collected from

Taiwain is a synonym of *G. sylvestre* based on its morphology (HaiLong, 1998) and chemical composition (K. Yoshikawa et al., 1998; 1999). This database was further interpreted to create a library of predicted compounds by combinations of building blocks following construction rules deduced from the known compounds. Based on the molecular formulae of predicted compounds, their molecular weights were calculated using Masslynx 4.1 (Waters MS Technologies, Manchester, UK).

2.4. UPLC-ESI-MS^n Experiments

An extract of the Vietnamese *G. sylvestre* variety was prepared by extracting the plant materials with70% EtOH by Nam Duoc Pharmaceutical Jsc. (Nam Dinh province, Vietnam) in July 2017. In order to obtain good chromatographic peaks, this extract was then microfractionated into 20 subfractions (T1-T20) using preparative HPLC in time collection mode [Optima Pak C18, using a gradient elution of MeOH 55% - 90%, flow rate 2.0 mL/min, each fraction 2 min].

The total extract and subfractions from *G. sylvestre* were analyzed via a Waters Acquity UPLC system (Waters Co., Milford, MA, USA) coupled with a Waters Xevo G2 QTOF mass spectrometer (Waters MS Technologies, Manchester, UK) that was equipped with an electrospray interface (ESI). A mobile phase (solvent A) was H_2O containing 0.1% formic acid (v/v) and the organic phase (solvent B) was acetonitrile. A 20 min binary gradient elution was executed at a flow rate 0.2 mL/min for the separation: a linear gradient from 10% to 90% solvent B in 17 min; then followed by an isocratic 90% B in 3 min. The mass spectrometer was operated in the ESI (–) MS^1 and (+) MS^n mode. The parameters were as follows: wave length range: 200 to 500 nm; analyses of the samples 5 mg/mL (2.0 µL injected into the partial loop in the needle overfill mode) were performed in the negative ion modes

in the *m*/*z* 50 – 1200 range with acquisition times of 0.2 s in the centroid mode. The ESI conditions were set as follows: capillary voltage 2.0 kV, con voltage 20V, source temperature 120 °C, desolvation temperature 350 °C, and desolvation gas flow 600 L/h. The MS^n ramp energy was set from 25.0 to 65.0 eV.

RMD is calculated in ppm as (mass defect/measured monoisotopic mass) × 10^6. Key ion filterings applied for calculated mass of parent ions were applied in MS^1 of both negative and positive mode for the detection of predicted target compounds. Multiple key ions analysis was applied in positive HRMS (65 eV) to determine different aglycone types. Neutral fragment mass analysis was used to analyze and the positive HRMS data at 25 eV to identify glycosyl and acyl substitutions on the aglycones. The chemical formulae of detected compounds were finally confirmed using a formula predictor by setting the parameters as follows: C (20–100), H (40–200), O (20–100), and N (0 or 1) while other elements such as P, S, Cl, and Br were not considered.

2.5. Hydrolysis of Total Extract

Total extract (100 mg) was hydrolyzed with 2.0 N HCl (70% MeOH, 10 mL) at 90 °C for 1 h. The solution was neutralized by 10% NaOH, dried, and suspended in H_2O and partitioned with EtOAc. The EtOAc portion was dried, dissolved in MeOH, and filtered through a 0.2-μm Whatman hydrophilic membrane filter into an HPLC sample vials immediately before analysis by HRMS to identify different aglycones. The identification of each aglycone was confirmed by comparing retention time and HRMS with isolated references. The residual H_2O layer was concentrated and dissolved in pyridine (1.0 mL) and 5.0 mg of L-cysteine methyl ester hydrochloride was added. The mixture was kept for 1 h at 60 °C, and then 4.4

μL of phenylisothiocyanate (Sigma, St. Louis, MO, USA) was added. This solution was filtered through a 0.2-μm Whatman hydrophilic membrane filter into an HPLC sample vials immediately before LC-MS analysis. Analysis was performed using an Agilent 1200 HPLC system (Agilent Technologies, Palo Alto, CA, USA) with INNO C18 (4.6 × 250 mm inner diameter, 5 μm particle size; Young Jin Bio Chrom Co., Ltd) at a column temperature of 30 °C. Chromatographic condition was carried out using a mobile phase of 27% CH_3CN isocratic elution at a flow rate 0.6 mg/mL in 60 min. The sugar derivatives showed the retention time at 10.9, 14.9, 17.3, and 21.5 min, and were identical to the derivatives of authentic D-glucuronic acid, D-glucose, D-xylose, and L-rhamnopyranose, respectively.

2.6. Extraction and isolation schemes

2.6.1. General experimental procedures

Optical rotation was measured on a JASCO P-2000 polarimeter (JASCO International Co. Ltd., Tokyo, Japan). IR data were recorded on a Nicolet 6700 FT-IR spectrometer (Thermo Electron Corp., Waltham, MA, USA). The NMR data were analyzed using an AVANCE 500 MHz spectrometer (Bruker, Germany) or JNM-ECA 600 MHz spectrometer (Jeol, Japan) or AVANCE III 800 HD spectrometer coupled with a 5-mm CPTCI cryoprobe (Bruker, Germany). The HRESIMS values were analyzed via an Agilent Technologies 6130 Quadrupole LC/MS spectrometer equipped with an Agilent Technologies 1260 Infinity LC system (Agilent Technologies, Inc., Santa Clara, CA, USA) and INNO C18 column (4.6 × 150 mm, 5 μm particle size, 12 nm, J.K.Shah & Company, Korea). Silica gel (particle size: 63–200 μm) and RP-C_{18} (particle size: 40–63 μm), which were purchased from Merck (Darmstadt, Germany), and Sephadex LH-20 from Sigma-Aldrich (St. Louis,

MO, USA) were used for column chromatography (CC). Silica gel 60 F_{254} and RP-18 F_{254} TLC plates were from Merck (Darmstadt, Germany). A Gilson HPLC purification system, equipped with an Optima Pak C_{18} column (10 × 250 mm, 10 μm particle size; RS Tech, Seoul, Korea), was used with a flow rate of 2 ml/min and UV detection at 205 and 254 nm.

2.6.2. Isolation scheme

The crude extract obtained (1 kg) was suspended in water, absorbed on macroporous resin HP20, and washed out with water, 50% EtOH, 95% EtOH, and acetone in a sequential elution process. The 95% EtOH fraction (300 g) was subjected to silica gel column chromatography (15 × 45 cm; 63–200 μm particle size) using *n*-hexane/EtOAc (gradient from 10:1 to 0:1) then EtOAc/CH_3OH (gradient from 6:1 to 0:1) to give 8 fractions (A-G) based on the thin-layer chromatography profile. Fraction F, which was identified to be a triterpenoid-rich fraction by LC-MS analysis, was then chromatographed on reversed phase silica gel column eluting with CH_3OH/H_2O (v/v, from 2:3 to 1:0) to yield ten fractions (FI-FX).

Fraction FV (12 g) was subjected to a RP-18 column separation eluted by CH_3CN/H_2O (v/v, gradient 3:7 to 7:3) to yield five fractions (FV.N1 – FV.N5). Fraction FV.N1 (5.0 g) was re-chromatographed on silica gel column chromatography (5 × 20 cm; 40–63 μm particle size) eluting with CH_2Cl_2/CH_3OH(v/v, gradient from 6:1 to 0:1) to yield five sub-fractions FV.N1-A to FV.N1-E. Subfraction FV.N1-E was applied to a succession of Sephadex LH-20 (100% CH_3OH) and HPLC Optima Pak C_{18}, CH_3CN/H_2O (2:3), flow rate 2 mL/min to afford compound **1** (50.1 mg). Fraction F.V.N3 (1.2 g) was applied to a Sephadex

LH-20 (using 100% CH_3OH) and then purified by semi-preparative HPLC (Optima Pak C_{18}, 35% CH_3CN, flow rate 2 mL/min) to produce compounds **2** (21.1 mg).

Subfraction FIII (5.2 g) was re-chromatographed with silica gel column chromatography (5 × 20 cm; 40–63 μm particle size), eluting with CH_2Cl_2/CH_3OH (v/v, gradient from 10:1 to 0:1) to yield 5 sub-fractions FIII.N1 to FIII.N5. Fraction FIII.N3 (520 mg) was applied to a succession of Sephadex LH-20 (CH_3OH 100%) and HPLC (Optima Pak C_{18}, 50% CH_3CN, flow rate 2 mL/min) to afford compounds **3** (10.6 mg). Fraction FIII.N4 (1.5 g) was developed on a reversed phase silica gel column chromatography, eluting with CH_3CN/H_2O (v/v, from 1:4 to 1:0) to yield 6 subfractions (FIII.N4.R1-6). Subfraction F.III.N4.R3 was purified by HPLC (Optima Pak C_{18}, 55% CH_3CN, flow rate 1.5 mL/min) to afford compounds **4** (6.5 mg). Subfraction F.III.N4.R5 was purified by HPLC (Optima Pak C_{18}, 50% CH_3CN, flow rate 1.5 mL/min) to afford compounds **5** (7.8 mg). Fraction F.III.N4.R6 was applied to a succession of Sephadex LH-20 (CH_3OH 100%) and HPLC (Optima Pak C_{18}, 60% CH_3CN, flow rate 1.5 mL/min) to afford compounds **6** (5.0 mg).

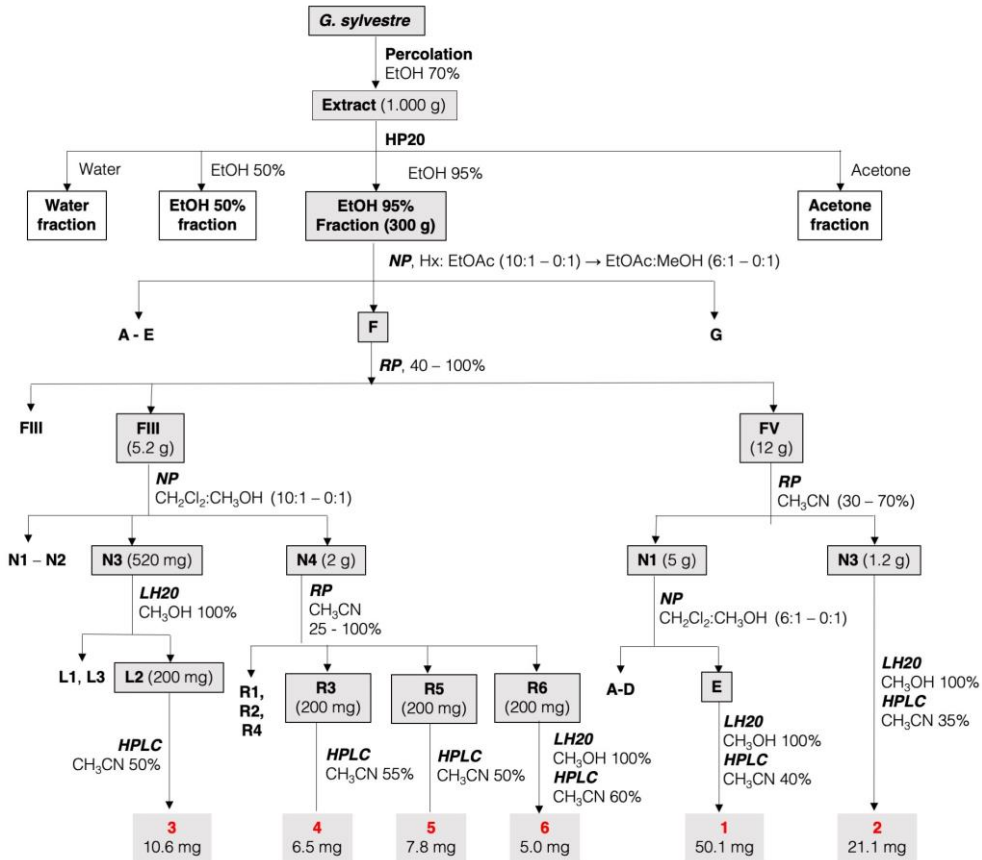


Figure 3: Isolation scheme of compounds **1 – 6** from *G. sylvestre*

2.6.3. Physical and chemical characteristics of isolated compounds

*Gymnemoside GS1 (**1**)*: amorphous powder; $[\alpha]_D^{25}$ –24.6 (c 0.2, MeOH); IR (KBr) ν_{max}: 3394, 1052, 1610 cm^{-1}. ^{13}C and ^{1}H NMR data, Tables 3 and 4; HRESIMS *m/z* 1089.5884 $[M - H]^-$ (calcd for $C_{54}H_{89}O_{22}$, 1089.5845).

*Gymnemoside GS2 (**2**)*: amorphous powder; $[\alpha]_D^{25}$ –26.3 (c 0.2, MeOH); IR (KBr) ν_{max}: 3397, 2954, 2296, 1647, 1348, 1017 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 3 and 4; HRESIMS *m/z* 1221.6298 $[M - H]^-$ (calcd for $C_{59}H_{97}O_{26}$ 1221.6268).

*Gymnemoside GS3 (**3**)*: Amorphous powder; $[\alpha]_D^{25}$ –14.0 (c 0.2, MeOH); IR (KBr) ν_{max}: 3389, 2949, 2298, 1663, 1053 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 3 and 4; HRESIMS 753.4235 $[M–H]^-$ (calcd for $C_{43}H_{61}O_{11}$ 753.4214).

*Gymnemoside GS4 (**4**)*: Amorphous powder; $[\alpha]_D^{25}$ –19.3 (c 0.2, MeOH); IR (KBr) ν_{max}: 3397, 2954, 2298, 1647, 1535, 1302, 1108, 1017 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 3 and 4; HRESIMS *m/z* 754.4158 $[M – H]^-$ (calcd for $C_{42}H_{60}NO_{11}$, 754.4166).

*Gymnemoside GS5 (**5**)*: Amorphous powder; $[\alpha]_D^{25}$ –22.0 (c 0.2, MeOH); IR (KBr) ν_{max}: 3394, 2980, 1704, 1014 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 3 and 4; HRESIMS *m/z* 731.4382 $[M – H]^-$ (calcd for $C_{41}H_{63}O_{11}$ 731.4370).

*Gymnemoside GS6 (**6**)*: Amorphous powder; $[\alpha]_D^{25}$ –11.0 (c 0.2, MeOH); IR (KBr) ν_{max}: 3391, 2971, 2843, 1703, 1515, 1050 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 3 and 4; HRESIMS *m/z* 617.4054 $[M – H]^-$ (calcd for $C_{36}H_{57}O_8$ 617.4053).

3. Results and discussion

3.1. Discrimination of different geographic varieties of G. sylvestre

3.1.1. Agronomical characteristics

Two varieties of *G.sylvestre*, which were growth up in the same environment conditions, exhibited several agronomical characteristic differences. The Vietnamese variety has flowers two times per year including Spring (February – April) and autumn (September – November) while the Indian variety only has flower in autumn (September – November). The number of seed per fruit of the Vietnamese variety (30 – 50) is also much more thatn that of the Indian variety (<10) (**Figure 4**).

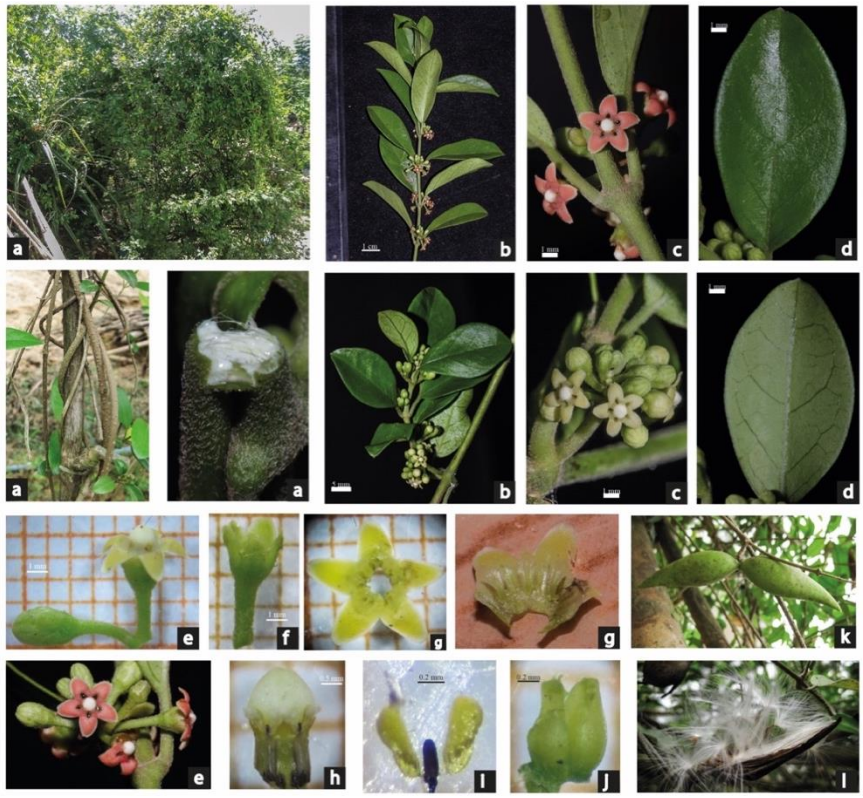
Vietnamese variety *Indian variety*

Figure 4: Agronomical characteristics of two varieties of *G. sylvestre*

3.1.2. Morphological and anatomical characteristics

Detailed descriptions of the macroscopic and microscopic characteristics of two samples, GS-V and GS-I, include many similar morphological traits and matched

with those of *Gymnema sylvestre* (Retz.) R.Br. ex Sm., described in Flora of China (Wu and Raven, 1995) (**Figure 5**).



A. GS-V

B. GS-I

Figure 5. Morphological characteristics of *G. sylvestre* originated from Vietnam (A. GS-V) and India (B. GS-I).

In addition to these similarities, some distinctive characteristics can be used to differentiate the two samples (**Figure 6**), including the following: (1) young branchlets that are glabrescent or pubescent in GS-V but densely pubescent in GS-I; (2) in both samples, leaf blades are diverse and can vary from obovate to ovate; however, they are more likely to be obovate and thick in GS-V yet ovate and thinner in GS-I; (3) adaxial and abaxial leaves are nearly glabrous and slightly pubescent at the mid-vein of GS-V, but they are found to be pubescent on both sides and densely pubescent at the midvein of GS-I; (4) GS-V has 4-5 pairs of lateral veins, while the venation system of GS-I consists of 3-4 pairs with three main veins converging at the base that are always more prominent; (5) follicle fruits are broadly lanceolate with an acuminated beak on top in GS-V; in contrast, they are smaller with narrowly lanceolate in outline, with no beak, in GS-I.

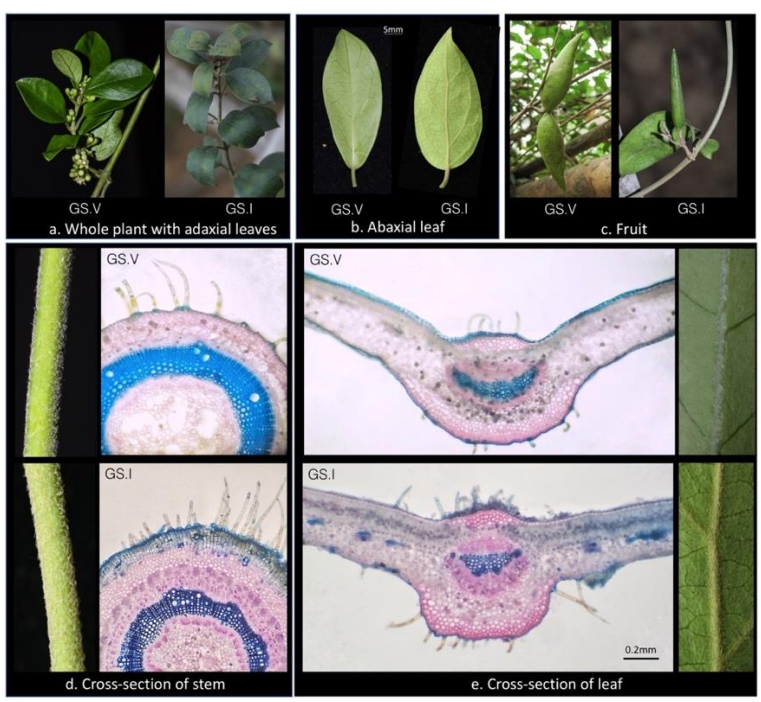
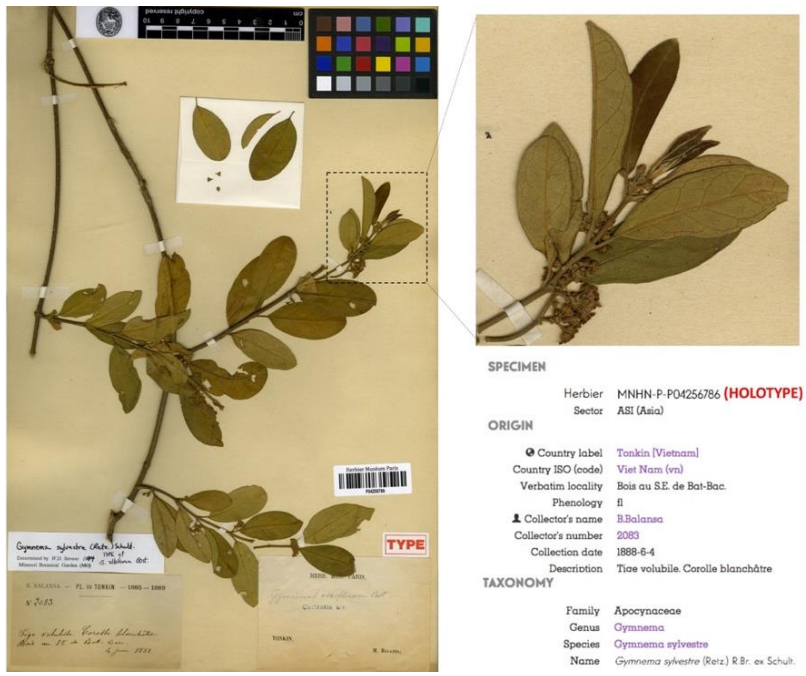


Figure 6. Selective morphological and anatomical characteristics to differentiate two varieties of *G. sylvestre*.

To confirm the scientific names of these two samples, we further compared their morphological characteristics with TYPE specimens of *Gymnema sylvestre* deposited at the Museum National d'Histoire Naturelle, Paris. Sample GS-V was identified as similar to HOLOTYPE specimen MNHN-P-P04256786 (**Figure 7-A**) collected in Tonkin, Vietnam, while sample GS-I was comparable to the "TYPE" specimen MNHN-P-P00645841 (**Figure 7-B**) from India. Another SYNTYPE specimen MNHN-P-P00442712 from Madagascar (Africa) also matched with GS-I. All specimens mentioned above are identified as *Gymnema sylvestre*. Although the two studied samples were determined to be the same species, their differences were apparent enough to be considered two varieties of the species *Gymnema sylvestre* (Retz.) R.Br. ex Sm.



A.

B.

Figure 7. A. HOLOTYPE specimen of *Gymnema sylvestre* (Retz.) R.Br. ex Schult (MNHN-P-P04256786) B. TYPE specimen of *Gymnema sylvestre* (Retz.) R.Br. ex Schult (MNHN-P-P00645841).

3.1.3. Internal transcribed spacer ITS1-5.8S-ITS2 sequence analysis

The ITS region encompasses the two noncoding regions, ITS1 and ITS2, which are separated by the highly conserved 5.8S rRNA gene (White et al., 1990). Multiple alignment analysis of 21 samples also illustrated the conservative property of the 5.8S region, with only 2 single nucleotide polymorphisms (SNPs). Variations between samples mainly occurred in the ITS1 and ITS2 regions (**Figure 8-A**) and thus promised significant separation of closely related species. Accordingly, the neighbor-joining phylogenetic tree showed clear divisions for all samples at the inter-species level, with pairwise genetic distances by identity that varied from 90.9% (between GS-I and *G. latifolium*) to 96.4% (between GS-V and *G. yunnanense*) (**Figure 8-B**). At the intra-species level, *G. sylvestre* samples were divided into two groups, which strongly referred to the native origins from Vietnam and India (**Figure 8-C**). The molecular differences between two groups of samples of *G. sylvestre* were consistent with morphological analysis and further supported the discrimination of two varieties.

A.

B.

	FM178493.1 - Marsdenia tenacissima	G.sylvestre-India domesticated in Vietnam	G.sylvestre Vietnam V1	G.sylvestre Vietnam V2	KP126832 - Gymnema latifolium GL1 (GB)	KP126833 - Gymnema latifolium GL2 (GB)	KP126834 - Gymnema latifolium GL3 (GB)	KP126835 - Gymnema latifolium GL6 (GB)	KP126836 - Gymnema sylvestre GS1 (GB)	KP126837 - Gymnema sylvestre GS2 (GB)	KP126838 - Gymnema sylvestre GS3 (GB)	KP126839 - Gymnema sylvestre GS4 (GB)	KP126840 - Gymnema sylvestre GS5 (GB)	KP126841 - Gymnema sylvestre GS6 (GB)	KP126842 - Gymnema sylvestre GS7 (GB)	KP126843 - Gymnema inodorum GI1 (GB)	KP126844 - Gymnema yunnanense GY1 (GB)	KP126845 - Gymnema latifolium GL5 (GB)	KP163979 - Gymnema latifolium GL4 (GB)	KX277701 - Gymnema sylvestre India (GB)	KX277713 - Gymnema sylvestre India (GB)
FM178493.1 - Marsdenia tenacissima		73	62	63	58	55	55	55	61	63	62	62	62	62	62	59	68	55	55	66	65
domesticated in Vietnam	88.0		34	35	54	54	50	54	33	35	34	34	34	34	34	48	30	54	53	18	17
G.sylvestre Vietnam V1	89.8	94.2		1	40	39	35	39	1	1	0	0	0	0	0	40	21	39	40	29	28
G.sylvestre Vietnam V2	89.6	94.1	99.8		41	40	36	40	2	2	1	1	1	1	1	41	22	40	41	30	29
KP126832 - Gymnema latifolium GL1 (GB)	90.4	90.9	93.3	93.1		11	11	11	41	41	40	40	40	40	40	51	43	11	13	49	48
KP126833 - Gymnema latifolium GL2 (GB)	90.9	90.9	93.5	93.3	98.2		9	0	40	40	39	39	39	39	39	47	42	0	12	47	46
KP126834 - Gymnema latifolium GL3 (GB)	90.9	91.6	94.1	94.0	98.2	98.5		9	36	36	35	35	35	35	35	44	38	9	7	42	41
KP126835 - Gymnema latifolium GL6 (GB)	90.9	90.9	93.5	93.3	98.2	100.0	98.5		40	40	39	39	39	39	39	47	42	0	12	47	46
KP126836 - Gymnema sylvestre GS1 (GB)	90.0	94.4	99.8	99.7	93.1	93.3	94.0	93.3		2	1	1	1	1	1	40	22	40	41	28	27
KP126837 - Gymnema sylvestre GS2 (GB)	89.6	94.1	99.8	99.7	93.1	93.3	94.0	93.3	99.7		1	1	1	1	1	41	22	40	41	30	29
KP126838 - Gymnema sylvestre GS3 (GB)	89.8	94.2	100.0	99.8	93.3	93.5	94.1	93.5	99.8	99.8		0	0	0	0	40	21	39	40	29	28
KP126839 - Gymnema sylvestre GS4 (GB)	89.8	94.2	100.0	99.8	93.3	93.5	94.1	93.5	99.8	99.8	100.0		0	0	0	40	21	39	40	29	28
KP126840 - Gymnema sylvestre GS5 (GB)	89.8	94.2	100.0	99.8	93.3	93.5	94.1	93.5	99.8	99.8	100.0	100.0		0	0	40	21	39	40	29	28
KP126841 - Gymnema sylvestre GS6 (GB)	89.8	94.2	100.0	99.8	93.3	93.5	94.1	93.5	99.8	99.8	100.0	100.0	100.0		0	40	21	39	40	29	28
KP126842 - Gymnema sylvestre GS7 (GB)	89.8	94.2	100.0	99.8	93.3	93.5	94.1	93.5	99.8	99.8	100.0	100.0	100.0	100.0		40	21	39	40	29	28
KP126843 - Gymnema inodorum GI1 (GB)	90.3	91.9	93.3	93.1	91.5	92.2	92.7	92.2	93.3	93.1	93.3	93.3	93.3	93.3	93.3		39	47	48	41	40
KP126844 - Gymnema yunnanense GY1 (GB)	88.8	94.9	96.4	96.3	92.8	92.9	93.6	92.9	96.3	96.3	96.4	96.4	96.4	96.4	96.4	93.4		42	41	27	24
KP126845 - Gymnema latifolium GL5 (GB)	90.9	90.9	93.5	93.3	98.2	100.0	98.5	100.0	93.3	93.3	93.5	93.5	93.5	93.5	93.5	92.2	92.9		12	47	46
KP163979 - Gymnema latifolium GL4 (GB)	90.9	91.1	93.3	93.1	97.8	98.0	98.8	98.0	93.1	93.1	93.3	93.3	93.3	93.3	93.3	92.0	93.1	98.0		48	47
KX277701 - Gymnema sylvestre India (GB)	89.1	96.9	95.1	94.9	91.8	92.1	93.0	92.1	95.2	94.9	95.1	95.1	95.1	95.1	95.1	93.1	95.4	92.1	92.0		3
KX277713 - Gymnema sylvestre India (GB)	89.3	97.1	95.2	95.1	92.0	92.3	93.1	92.3	95.4	95.1	95.2	95.2	95.2	95.2	95.2	93.3	95.9	92.3	92.1	99.5	

Percentage of bases/residues which are identical
Number of bases/residues which are not identical

C.

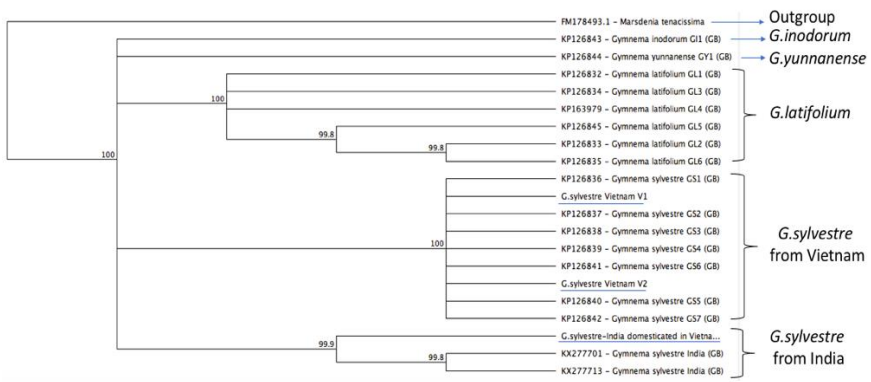


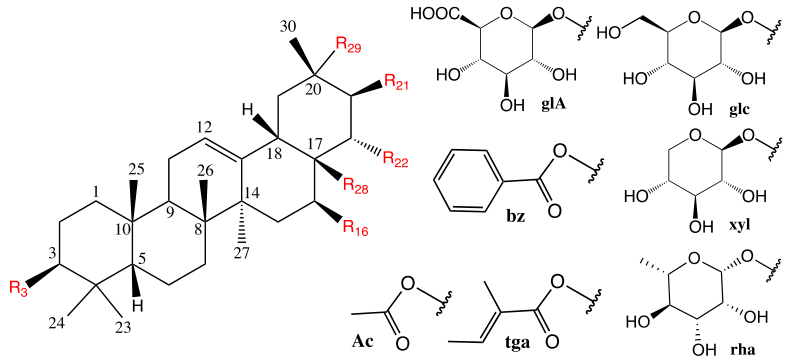
Figure 8. Neighbor-joining phylogenetic tree based on internal transcribed spacer (ITS) sequences of nuclear ribosome DNA of samples in the genus *Gymnema*.

Bootstrap values expressed as percentages of 1,000 replications (>75%) are shown above branches. Underlined samples were directly sequenced in this study, and other samples sequences were obtained from GenBank via Blast analysis. *Marsdenia tenacissima* was used as an outgroup sample for this study.

3.2. Development of a building block strategy to classification, identification and metabolite profiling of oleanane triterpenoids in the Vietnamese G. sylvestre variety using UHPLC-qTOF/MS

3.2.1. Building Blocks and Construction Rules of Triterpenoids isolated from G. Sylvestre.

A comprehensive review of publications up to December 2018 on all the oleanane-type triterpenoids isolated from *G. sylvestre* indexed on SciFinder, PubMed, and Google Scholar (12 studies) resulted in a total of 55 compounds (**Figure 9**).



No.	R_3	R_{16}	R_{21}	R_{22}	R_{28}	R_{29}	CAS number
1	-O-*β*-glcA	-OH	-H	-OH	-CH_2-O-α-rha	-CH_3	2253651-94-2
2	-OH	-H	-H	-H	-COOH	-CH_3	508-02-1
3	-OH	-OH	-H	-H	-CH_2OH	-CH_3	465-94-1
4	-OH	-OH	-OH	-H	-CH_2OH	-CH_3	474-15-7
5	-OH	-OH	-H	-H	-CH_2OH	-CH_2OH	122746-52-
6	-OH	-OH	-OH	-H	-CH_2OH	-CH_3	53187-93-2
7	-OH	-OH	-H	-OH	-CH_2OH	-COOH	862377-55-7
8	-OH	-H	-H	-H	-CO-O-*β*-glu	-CH_3	14162-53-9
9	-O-*β*-glcA	-OH	-H	-H	-CH_2OH	-CH_3	287390-11-8
10	-O-*β*-glcA	-OH	-OH	-H	-CH_2OH	-CH_3	2229693-57-4
11	-O-*β*-glcA	-OH	-H	-H	-CH_2OH	-CH_2OH	2229693-58-5
12	-O-*β*-glc^6-O-Me*	-OH	-OH	-H	-CH_2OH	-CH_3	1096581-47-3
13	-O-*β*-glcA	-OH	-H	-H	-CH_2OH	-COOH	873799-50-9
14	-O-*β*-glcA	-OH	-O-Bz	-H	-CH_2OH	-CH_3	287389-94-0
15	-O-*β*-glcA	-OH	-H	-OH	-CH_2O-bz	-CH_2OH	2229693-59-6
16	-O-*β*-glc^6-O-*β*-glc	-H	-H	-H	-COOH	-CH_3	240140-86-7
17	-O-*β*-glc	-H	-H	-H	-COO-*β*-glc	-CH_3	78454-20-3
18	-O-*β*-glcA^3-O-*β*-glc	-OH	-H	-H	-CH_3	-CH_3	2229693-55-2

No.	R_3	R_{16}	R_{21}	R_{22}	R_{28}	R_{29}	CAS number
19	-O-β-glcA	-OH	-H	-H	-CH_2-O-β-glc	-CH_3	256510-00-6
20	-O-β-glcA^3-O-β-glc	-OH	-OH	-H	-CH_2OH	-CH_3	330595-34-1
21	-O-β-glcA^3-O-β-glc	-OH	-H	-OH	-CH_2OH	-CH_3	212775-47-8
22	-O-β-glcA	-OH	-H	-OH	-CH_2-O-β-glc	-CH_3	212775-48-9
23	-[O-β-glcA^3-O-β-glc].K	-OH	-H	-H	-CH_2OH	-CH_2OH	330595-36-3
24	-O-β-glcA^3-O-β-glc	-OH	-OH	-H	-CH_2OH	-CH_3	2229693-54-1
25	-O-β-glcA	-OH	-H	-H	-CH_2OH	-CH_2-O-β-glc	2229693-56-3
26	-O-β-glcA^3-O-β-glc	-OH	-H	-H	-CH_2OH	-COOH	2229693-52-9
27	-O-β-glcA^3-O-β-gal	-OH	-H	-H	-CH_2OH	-COOH	2229693-53-0
28	-O-β-glcA	-OH	-H	-O-Ac	-CH_2-O-α-rha	-CH_3	212775-40-1
29	-O-β-glcA	-O-Ac	-H	-OH	-CH_2-O-α-rha	-CH_3	212775-43-4
30	-O-β-glcA	-OH	-H	-O-tga	-CH_2-O-β-rha	-CH_3	256509-97-4
31	-O-β-glcA^3-O-β-glc	-OH	-H	-O-tga	-CH_2OH	-CH_3	256509-77-0
32	-O-β-glcA^3-O-β-glc	-OH	-H	-OH	-CH_2-O-tga	-CH_3	256509-82-7
33	-O-β-glcA^3-O-β-glc	-O-tga	-H	-OH	-CH_2OH	-CH_3	256509-83-8
34	-O-β-glc^6-O-β-glc^6-O-β-xyl	-H	-H	-H	-COOH	-CH_3	287389-96-2
35	-O-β-glc^6-O-β-glc^6-O-β-xyl	-H	-H	-H	-COOH	-CH_3	287389-96-2
36	-O-β-glc^6-O-β-glc^6-O-β-xyl	-OH	-H	-H	-CH_2OH	-CH_3	164177-55-3
37	-O-β-glc^6-O-β-glc^6-O-β-xyl	-OH	-H	-OH	-CH_3	-CH_3	1422031-87-5
38	-O-β-glcA^3-O-β-glc	-OH	-O-Bz	-H	-CH_2OH	-CH_3	330595-32-9
39	-O-β-man^3-O-β-glc^6	-OH	-H	-OH	-CH_2-O-β-glc	-CH_3	1422031-89-7
40	-O-β-glc^6-O-β-glc^6-O-β-xyl	-OH	-H	-OH	-CH_2OH	-CH_3	163456-79-9
41	-O-β-glc^6-O-β-glc	-H	-H	-H	-CO-O-β-glc	-CH_3	287389-95-1
42	-O-β-glc^6-O-β-glc	-OH	-H	-H	-CH_2-O-β-glc	-CH_3	1096581-44-0
43	-O-β-glcA^3-O-β-glc	-OH	-H	-OH	-CH_2-O-α-rha	-CH_3	212775-51-4
44	-O-β-glcA^3-O-β-glc	-OH	-H	-H	-CH_2-O-β-glc	-CH_3	212775-50-3
45	-O-β-glcA^3-O-β-glc	-OH	-H	-OH	-CH_2-O-β-glc	-CH_3	212775-45-6
46	-O-β-glcA^3-O-β-glc	-OH	-H	-O-Ac	-CH_2-O-α-rha	-CH_3	212775-19-4
47	-O-β-glcA^3-O-β-glc	-O-Ac	-H	-OH	-CH_2-O-α-rha	-CH_3	212775-23-0
48	-O-β-glcA^3-O-β-glc	-O-Ac	-H	-OH	-CH_2-O-α-rha	-CH_3	330595-38-5
49	-O-β-glcA^3-O-β-glc	-OH	-H	-O-tga	-CH_2-O-β-xyl	-CH_3	256509-93-0
50	-O-β-glcA^3-O-β-glc	-OH	-H	-O-tga	-CH_2-O-α-rha	-CH_3	212775-27-4
51	-O-β-glcA^3-O-β-glc	-OH	-H	-O-tga	-CH_2-O-β-glc	-CH_3	256509-85-0
52	-O-β-glc^6-O-β-glc^6-O-β-xyl	-H	-H	-H	-CO-O-β-glc	-CH_3	287389-97-3
53	-O-β-glc^6-O-β-glc^6-O-β-xyl	-H	-H	-H	-CO-O-β-glc	-CH_3	287389-96-2
54	-O-β-glc^6-O-β-glc^6-O-β-xyl	-OH	-H	-H	-CH_2-O-β-glc	-CH_3	256510-01-7
55	-O-β-glc^6-O-β-glc	-H	-H	-H	-CO-O-β-glc^6-O-β-glc	-CH_3	287389-98-4

Figure 9. Library of 55 compounds isolated from *G. sylvestre* reported in the literature

Based on the literature data, we anticipated that the triterpenoid saponins of the Vietnamese variety of *G*. *sylvestre* were constructed from three components: a main aglycone, 1 or 2 glycosidic linkages at C-3 and/or C-28, and one acyl moiety attached to a hydroxy group of the main aglycone. The triterpenoid aglycones reported from *G*. *sylvestre* were longispinogenin (L), chichipegenin (C), sitakisogenin (S), gymnemagenol (G), 22*α*-hydroxygymnemagenol (HG), myrtilogenic acid (My), maniladiol (M), oleanolic acid (O), and 3β,16β,28-trihydroxyolean-12-en-22,29-olide (La). The glycosidic side chains included glucuronic acid (glcA), glucose (glc), rhamnose (rha), and xylose (xyl). The acyl fragments could be acetyl, tigloyl (tig), and benzoic acid (bz). Two new building blocks, erythrodiol (E) and nicotinoyl (nic), were identified from three new compounds isolated in this study (**Figure 10**)

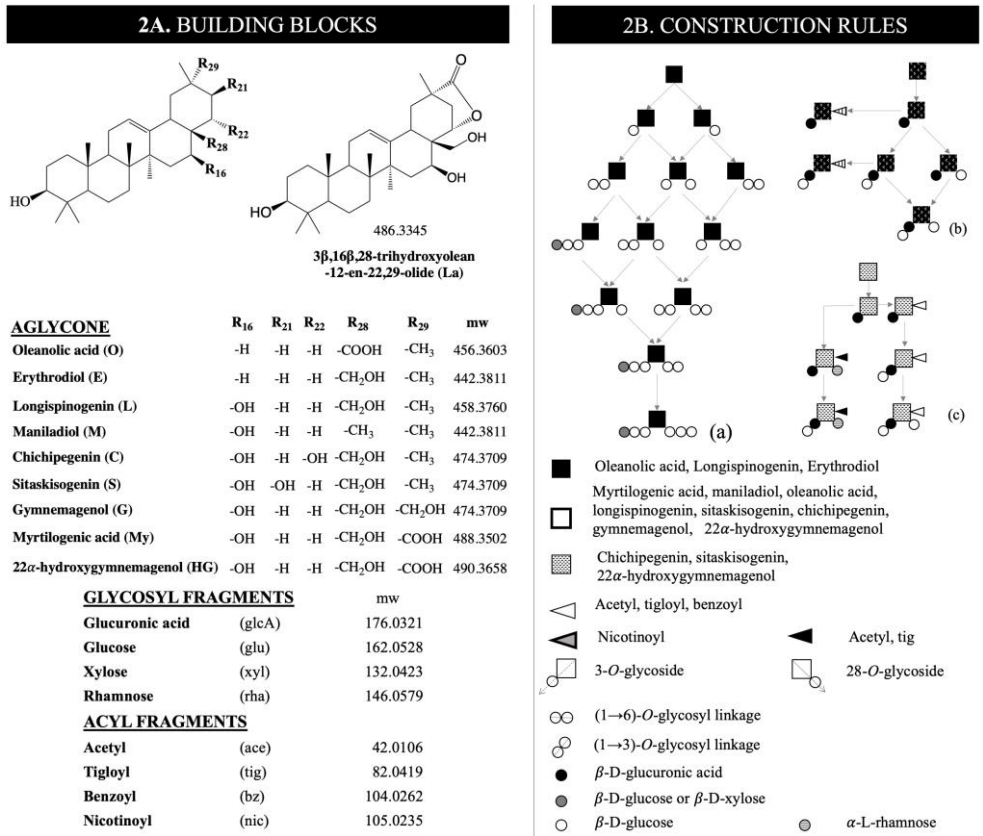


Figure 10. Building blocks (A) and construction rules (B) of oleanane triterpenoids of *G*. *sylvestre*.

The presence of aglycone and glycosidic building blocks was confirmed by comparing the retention times of the hydrolyzed extract of *G*. *sylvestre* with reference aglycones and the trimethylsilyl-L-cysteine derivatives of the water-soluble fraction of the hydrolysate with authentic sugar derivatives (**Figure 11**).

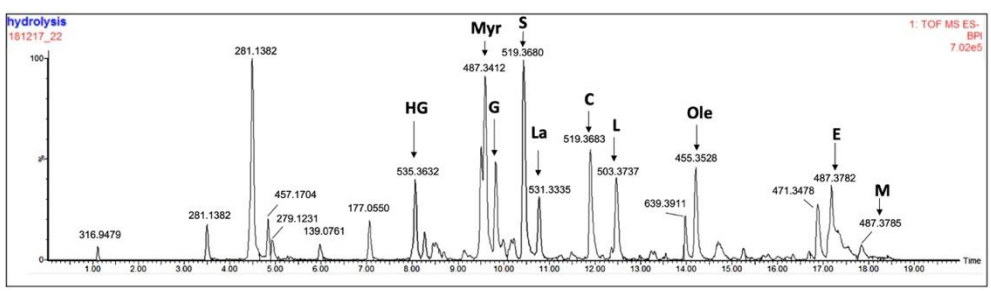


Figure 11. Detection of aglycones of oleanane triterpenes in the hydrolysate

The construction rules of oleanane triterpenoids comprising these building blocks were deduced from the structures of the known compounds (**Figure 10**). A library of predicted structures and their theoretical deprotonated masses ($[M - H]^-$) was created using these rules (**Table 1**).

Table 1. Virtual deprotonated mass $[M - H]^-$ "library" of compounds constructed by building block rules and their detection in extract or microfractions.

	O	E	La	My	L	CGS	HG	M
3-*O*-glcA								
+glcA	631.3846	617.4054	661.3588	663.3745	633.4003	649.3952	665.3901	617.4054
+glcA+acetyl	673.3952	659.4159	703.3694	705.3850	675.4109	691.4058	707.4007	659.4159
+glcA+tig	713.4265	699.4472	743.4007	745.4163	715.4422	731.4371	747.4320	699.4472
+glcA+bz	735.4108	721.4316	765.3850	767.4007	737.4265	753.4214	769.4163	721.4316
+glcA+nic	736.4081	722.4289	766.3823	768.3980	738.4238	754.4187	770.4136	722.4289
+glcA+rha	777.4425	763.4633	807.4167	809.4324	779.4582	795.4531	811.4480	763.4633
+glcA+acetyl+rha	819.4531	805.4738	849.4273	851.4429	821.4688	837.4637	853.4586	805.4738
+glcA+glc	793.4375	779.4582	823.4116	825.4273	795.4531	811.4480	827.4429	779.4582
+glcA+glc+acetyl	835.4480	821.4688	865.4222	867.4379	837.4637	853.4586	869.4535	821.4688
+glcA+glc+tig	875.4793	861.5001	905.4535	907.4692	877.4950	893.4899	909.4848	861.5001
+glcA+glc+benz	897.4637	883.4844	927.4379	929.4535	899.4793	915.4742	931.4691	883.4844
+glcA+glc+nic	898.4610	884.4817	928.4351	930.4508	900.4766	916.4715	932.4664	884.4817
+glcA+glc+glc	955.4903	941.5110	985.4644	987.4801	957.5059	973.5008	989.4957	941.5110
+glcA+glc+glc+tig	1037.5322	1023.5529	1067.5063	1069.5220	1039.5478	1055.5427	1071.5376	1023.5529
+glcA+glc+glc+bz	1059.5165	1045.5372	1089.4906	1091.5063	1061.5321	1077.5270	1093.5219	1045.5372
+glcA+glc+glc+acetyl	997.5009	983.5216	1027.4750	1029.4907	999.5165	1015.5114	1031.5063	983.5216
+glcA+glc+rha	939.4954	925.5161	969.4695	971.4852	941.5110	957.5059	973.5008	925.5161
+glcA+acetyl +rha	981.5059	967.5267	1011.4801	1013.4958	983.5216	999.5165	1015.5114	967.5267
+glcA+glc+tig+rha	1021.5372	1007.5580	1051.5114	1053.5271	1023.5529	1039.5478	1055.5427	1007.5580
3-*O*-glc								
+glc	617.4054	603.4261	647.3795	649.3952	619.4210	635.4159	651.4108	603.4261
+glc+glc	779.4582	765.4789	809.4324	811.4480	781.4738	797.4687	813.4636	765.4789
+glc+glc+glc	941.5110	927.5318	971.4852	973.5008	943.5267	959.5216	975.5165	927.5318
+glc+glc+glc+glc	1103.5638	1089.5846	1133.5380	1135.5537	1105.5795	1121.5744	1137.5693	1089.5846
+glc+glc+glc+glc+glc	1265.6165	1251.6373	1295.5907	1297.6064	1267.6322	1283.6271	1299.6220	1251.6373
+glc+glc+glc+glc+glc+glc	1427.6693	1413.6901	1457.6435	1459.6592	1429.6850	1445.6799	1461.6748	1413.6901
+glc+glc+xyl	911.5005	897.5212	941.4746	943.4903	913.5161	929.5110	945.5059	897.5212
+glc+glc+xyl+glc	1073.5533	1059.5740	1103.5275	1105.5431	1075.5689	1091.5638	1107.5587	1059.5740
+glc+glc+xyl+glc+glc	1235.6061	1221.6268	1265.5803	1267.5959	1237.6218	1253.6167	1269.6116	1221.6268
+glc+glc+xyl+glc+ glc+glc	1397.6589	1383.6796	1427.6331	1429.6487	1399.6746	1415.6695	1431.6644	1383.6796

2B-a 2B-b 2B-c Construction rules scheme (Figure 2B)

Abbreviation:

O: Oleanolic acid; E: Erythrodiol; M: Maniladiol; My: Myrtilogenic acid;

La: 3β,16β,28-trihydroxyolean-12-en-29,22-olide; L: Longispinogenin;

CGS: Chichipegenin/gymnemagenol/Sitaskisogenin; HG: 22-hydroxylgymnemagenol

glcA: glucuronic acid; glc: glucose; rha: rhamnose; xyl: xylose;

acetyl: acetyl; tig: tigloyl; bz: benzoyl; nic: nicotinoyl;

3.2.2. Relative mass filtering

RMD filtering is a useful method for the rapid classification of various groups of metabolites by interpreting their chromatography-mass spectrometry data and calculating their molecular and fragment ions as well as their retention times (Ekanayaka et al., 2015). The theoretical RMD range of oleanane-type triterpenes of *G. sylvestre* was 537.4 ± 62.2 based on the calculated molecular weights of the triterpenoids in the predicted virtual library. A simple RMD filtering was applied to the predominant parent ions of all the detected compounds in negative MS^1 mode to facilitate the exclusion of nontargeted constituents such as flavonoids and fatty acids in the extract. Thus, the retention times of the targeted oleanane triterpenoids were identified, and they ranged from 5–14 min using the chromatographic conditions of this study. The RMD values also provided remarkable information related to the structures of the triterpenoid glycosides. Conjugation with polar glycoside moieties (e.g., glucosylation, $C_6H_{10}O_5$) lowered the hydrogen content, decreasing the RMD value as the number of attached sugars increased. On the other hand, the attachment of an acyl group may maintain a high hydrogen ratio and keep the RMD higher than that of the same core terpenoid structure with a glycosyl substituent (**Figure 12**).

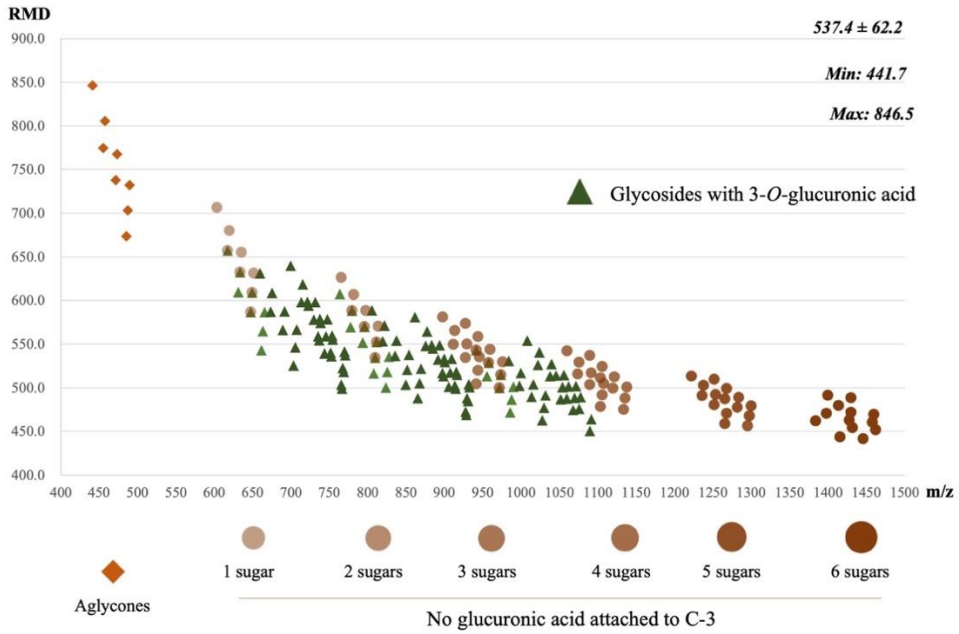


Figure 12. Relative mass defect (RMD) values of predicted triterpenoids of *G. sylvestre*

3.2.3. Identification of Targeted Metabolites by UPLC-MS^n.

The compound detection and structural assignment of the targeted triterpenoids were implemented by a combination of three techniques, KIF, mKIA, and MFA. It has been widely reported that saponins are not degraded by a single TOF in negative mass detection (Qi et al., 2012). Therefore, the masses of the predicted targeted compounds were extracted effectively and selectively in the negative MS^1 mode, where the dominant precursor ions $[M - H]^-$ or $[M + HCOO]^-$ could be obtained with high specificity and sensitivity (**Figure 13**). In addition, the rigid polycyclic structures of oleanane triterpenoids in a specific plant usually vary over a limited number of aglycones, and each of these can be characterized by a set of key MS^n ions produced by dehydration during mass fragmentation. As a result, the mKIA technique utilizing various sets of key positive MS^n ions can be used to effectively

discriminate different aglycones in G. sylvestre. Furthermore, MFA of *G. sylvestre* in TOF-positive fragmentation mode showed the typical neutral losses of acyl and sugar moieties. **Figure 13** illustrates an example of the integration of these techniques in peak picking and structure elaboration using oleanolic acid glycoside (peak 21). Consequently, 119 peaks of oleanane triterpenoids were quickly identified in the same manner, and among them, 77 peaks are predicted to be new compounds based on their molecular formulas and mass fragmentation patterns. To validate the compound assignment, three levels of assignment fidelity were proposed in this study (**Table 2**). Level (A): the compound was unambiguously determined by comparison with reference standard compounds or isolated and elucidated from its NMR data. Level (B): the compound has been previously reported from *G. sylvestre* and was consistent with the building block mass data, a rational retention time relationship, mKIA, and MFA. Level (C): the compound was constructed from the building blocks, and its structure was supported by mKIA and MFA analysis, but no obvious evidence to differentiate structural isomer was available. Herein, the compound detection and structure determination of all the targeted triterpenoids will be discussed. As mentioned above, the oleanane triterpenoids formed in *G. sylvestre* could be broadly classified into two major subgroups according to the existence of a glucuronic acid linkage at C-3 in the structure.

3.2.3.1. Oleanane Triterpenes without Glucuronic Acid.

These compounds were constructed following Scheme 2B-a (**Figure 10**). In brief, 1-2 glycosyl chains are formulated starting with a glucose attached to the aglycone at C-3 and/or C-28. These glycoside chains were then elongated by the addition of 1-2 glucose units in a (1→6)-glycosyl pattern, and the third sugar of the

C-3 glycoside chain could be either glucose or xylose. The three subgroups of glycosides following this construction rule were oleanolic acid glycosides, erythrodiol glycosides and longispinogenin glycosides. Fourteen peaks of oleanolic acid (O) and its analogs of this type were filtered from the data of the *G. sylvestre* extract or microfractions with mass deviations of less than 5 ppm (**Figure 11**). Of the compounds constructed following rule **Figure 10B-a**, 8 have been reported in the literature (**Table 2**).

Oleanolic acid O#456 (peak ***1***) was confirmed by comparison with standard oleanolic acid. As illustrated in **Figure 11**, the positive mass fragmentation data of oleanolic acid showed a set of key ions 457, 439, 421, 411, and 393, which can be used to characterize this compound and its glycosides. Two oleanolic monoglucoside peaks, ***2*** and ***3***, possessing the molecular formula $C_{42}H_{68}O_{13}$, were detected by ion extraction (663.4110 $[M + HCOO]^-$). Both spectra showed key diagnostic ions of oleanolic acid as well as the loss of an *O*-glycosidic hexosyl moiety, resulting in their tentative determinations as oleanolic acid 3-*O*-β-D-glucopyranoside and oleanolic acid 28-*O*-β-D-glucopyranoside, respectively. The presence of oleanolic acid 28-*O*-β-D-glucopyranoside (***2***) in *G. sylvestre* has been reported in the literature (Ye et al., 2000a), while oleanolic acid 3-*O*-β-D-glucopyranoside (***3***) was confirmed by comparison with a reference standard.

The targeted extraction of compounds similar of oleanolic acid diglycoside ($C_{42}H_{68}O_{13}$) revealed two peaks, ***4*** and ***5***. mKIA analysis showed apparent key ions and mass losses corresponding to 2 glucose units. Notably, two compounds, 3-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl oleanolic acid and 3-*O*-β-D-

glucopyranosyl oleanolic acid 28-*O*-β-D-glucopyranosyl ester, were isolated from the Chinese variety of *G*. *sylvestre* (X. Liu et al., n.d.; Ye et al., 2000a). Since these were the only two detectable peaks with this molecular weight , it was reasonable to assign them to the two aforementioned compounds.

Peak ***6*** (t_R = 10.004 min), which showed losses of a xylose (132 Da) and two other glucose fragments (2 × 162), was determined as oleanolic acid 3-*O*-β-D-xylopyranosyl-(1→6)-β-D-glucopyranosyl(1→6)-β-D-glucopyranoside by referring to the literature (M.-Q. Zhang et al., 2012) and comparing with a standard reference. Peak ***6*** was the only high-abundance isomer of mass extraction 911.5025 $[M - H]^-$, indicating the substantial accumulation of glc^6-glc^6-xyl as a major glycosyl chain at C-3 via biosynthesis. This observation was confirmed by the detection of peaks ***9*** and ***12***, which were identified as 3-*O*-β-D-xylopyranosyl (1→6)-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl oleanolic acid 28-*O*-β-D-glucopyranosyl ester (K. Yoshikawa et al., 1999) and 3-*O*-β-D-xylopyranosyl (1→6)-β-D-glucopyranosyl(1→6)-β-D-glucopyranosyl oleanolic acid 28-*O*-β-D-glucopyranosyl(1→6)-β-D-glucopyranosyl ester, respectively. Their structrures were determined by mKIA and MFA and unambiguously confirmed by comparison with reference compounds. Using the same approach, a single peak (***14***, $C_{65}H_{106}O_{32}$) was found by filtering by predominant ion 1397.6590 $[M - H]^-$. The similar fragmentation pattern to that of ***12***, together with the presence of an additional glucose moiety, supported the structural assignment of ***14*** as 3-*O*-β-D-xylopyranosyl-(1→6)-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl oleanolic

acid 28-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl ester.

Peaks ***10*** (t_R = 6.801 min) and ***11*** (t_R = 7.108 min) can be regarded as two isomers. They exhibited deprotonated ions $[M - H]^-$ at 1103.5664 and 1103.5653 and showed similar sets of key aglycone ions of oleanolic acid. The main difference between these two compounds is the sugar cleavage pattern. Peak ***10*** was identified as 3-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl oleanolic acid 28-*O*-β-D-glucopyranosyl(1→6)-β-D-glucopyranosyl ester by referring to the literature (Ye et al., 2000a) and comparing with reference compounds. Peak ***11*** exhibited a strong signal at 943.5230, suggesting an oleanolic acid core with a sugar side chain comprising 3 glucose units at C-3 after the loss of a glucose linkage at the carboxylic C-28 position. Consequently, this compound was tentatively identified as 3-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl(1→6)-β-D-glucopyranosyl oleanolic acid 28-*O*-β-D-glucopyranosyl ester, proposing it a new compound.

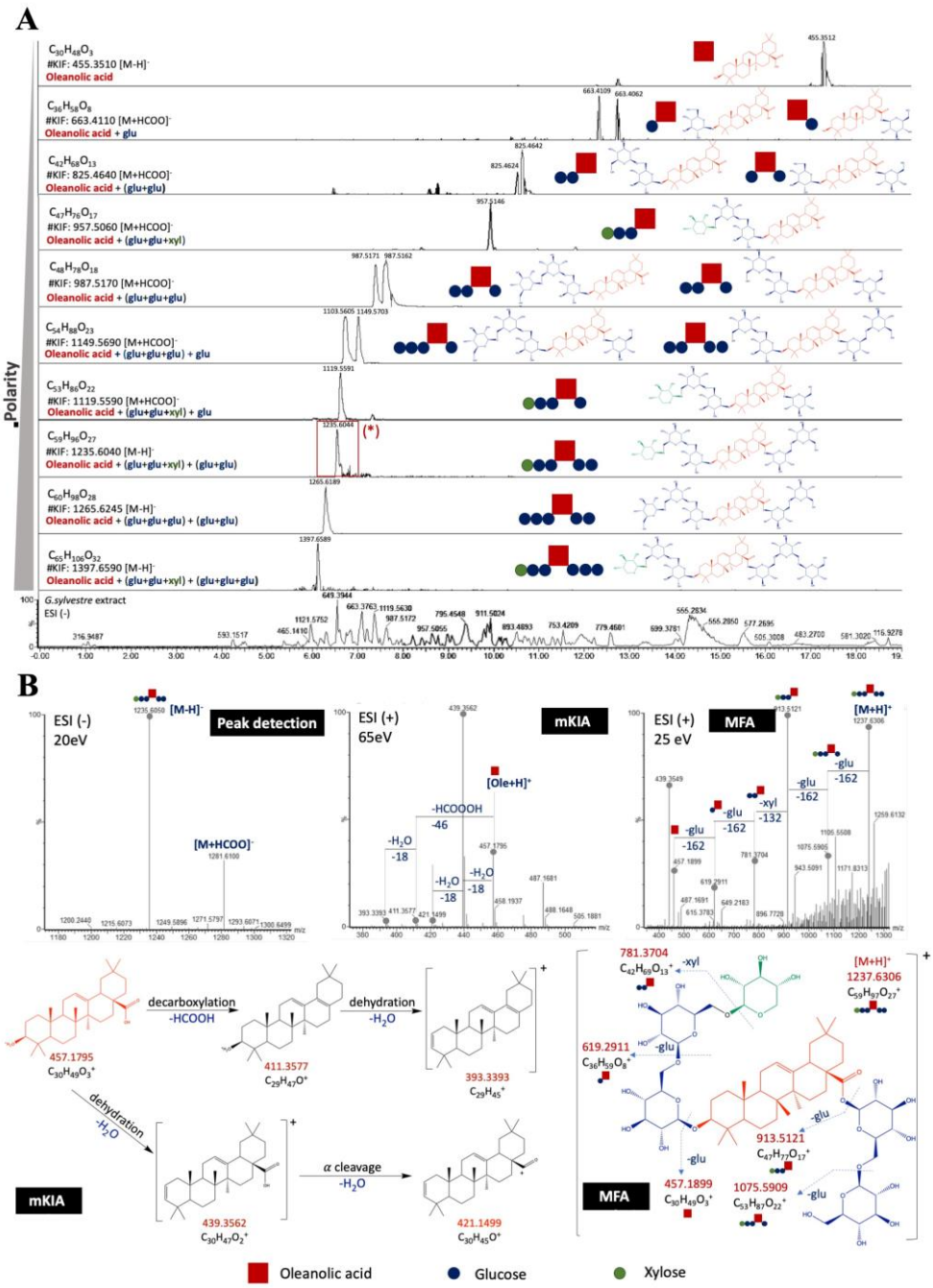


Figure 13. Application of building block strategy for the identification of oleanolic acid triterpenoids from *G. sylvestre*.

Table 2. Compound assignments for detected metabolites by using building block library.

Peak	R_t (#)	*m/z* (-)	Cald. *m/z (-)*	Diff (ppm)	RMD	Formula	Exact mass	Block-based compound assignment	Assignment fidelity (*)	CAS Number (&)
1	17.194	455.3520	455.3525	-1.1	773.0	$C_{30}H_{48}O_3$	456.3598	O	A, Ref	508-02-1
2	12.063	617.4038	617.4053	-2.4	654.0	$C_{36}H_{58}O_8$	618.4116	O+28-(glc)	B, (Ye et al., 2000a)	3391-80-8
3	12.660	617.4016	617.4053	-3.7	650.5	$C_{36}H_{58}O_8$	618.4094	O+3-(glc)	A, Ref	14162-53-9
4	10.485	779.4576	779.4582	-0.8	587.1	$C_{42}H_{68}O_{13}$	780.4654	O+3-(glc^6+glc)	B, (Ye et al., 2000a)	240140-86-7
5	10.564	779.4592	779.4582	1.3	589.1	$C_{42}H_{68}O_{13}$	780.4670	O+3-(glc)+28-(glc)	B, (X. Liu et al., n.d.)	78454-20-3
6	10.004	911.5024	911.5004	2.2	551.2	$C_{47}H_{76}O_{17}$	912.5102	O+3-(glc^6+glc^6+xyl)	A, Ref, (M.-Q. Zhang et al., 2012)	287389-96-2
7	7.485	941.5104	941.5110	-0.6	542.1	$C_{48}H_{78}O_{18}$	942.5182	O+3-(glc^6+glc^6+glc)	C	No report
8	7.695	941.5114	941.5110	0.4	543.2	$C_{48}H_{78}O_{18}$	942.5192	O+3-(glc^6+glc)+28-(glc)	A, Ref, (Ye et al., 2000b)	287389-95-1
9	7.441	1073.5521	1073.5532	-1.0	514.3	$C_{53}H_{86}O_{22}$	1074.5599	O+3-(glc^6+glc^6+xyl)+28-(glc)	A, Ref, (Ye et al., 2000a)	287389-97-3
10	6.801	1103.5605	1103.5638	-3.3	507.9	$C_{54}H_{88}O_{23}$	1104.5683	O+3-(glc^6+glc)+28-(glc^6+glc)	A, Ref, (Ye et al., 2000a)	287389-98-4
11	7.108	1103.5610	1103.5638	-2.5	508.4	$C_{54}H_{88}O_{23}$	1104.5688	O+3-(glc^6+glc^6+glc)+28-(glc)	C	No report
12	6.617	1235.6089	1235.6061	2.3	492.8	$C_{59}H_{96}O_{27}$	1236.6167	O+3-(glc^6+glc^6+xyl)+28(glc^6+glc)	A, Ref, [31]	Not available
13	6.362	1265.6189	1265.6166	1.8	489.0	$C_{60}H_{98}O_{28}$	1266.6267	O+3-(glc^6+glc^6+glc)+28-(glc^6+glc)	C	No report
14	6.195	1397.6608	1397.6589	1.4	472.8	$C_{65}H_{106}O_{32}$	1398.6686	O+3-(glc^6+glc^6+xyl)+28-(glc^6+glc^6+glc)	C	No report
15	9.318	897.5163	897.5153	1.1	575.3	$C_{47}H_{78}O_{16}$	898.5241	E+3-(glc^6+glc^6+xyl)	C	No report
16	7.801	927.5278	927.5317	-4.2	569.0	$C_{48}H_{80}O_{17}$	928.5356	E+3-(glc^6+glc^6)+28-(glc)	C	No report
17	7.915	927.5345	927.5317	3.0	576.3	$C_{48}H_{80}O_{17}$	928.5423	E+3-(glc^6+glc^6+glc)	C	No report
18	7.695	1059.5743	1059.574	0.3	542.0	$C_{53}H_{88}O_{21}$	1060.5821	E+3-(glc^6+glc^6+xyl)+28-(glc)	C	No report
19	6.994	1089.585	1089.5845	0.5	536.9	$C_{54}H_{90}O_{22}$	1090.5928	E+3-(glc^6+glc^6+glc)+28-(glc)	C	No report
20	7.117	1089.5853	1089.5845	0.7	537.2	$C_{54}H_{90}O_{22}$	1090.5931	E+3-(glc^6+glc)+28-(glc^6+glc)	A, NMR	No report
21	6.950	1221.6289	1221.6268	1.7	514.8	$C_{59}H_{97}O_{26}$	1,222.6367	E+3-(glc^6+glc^6+xyl)+28-(glc^6+glc)	A, NMR	No report
22	6.529	1383.6791	1383.6796	-0.4	490.8	$C_{65}H_{108}O_{31}$	1384.6869	E+3-(glc^6+glc^6+xyl)+28-(glc^6+glc^6+glc)	C	No report
23	9.073	781.4726	781.4738	-1.5	604.8	$C_{42}H_{70}O_{13}$	782.4804	L+3-(glc)+28-(glc)	C	No report
24	9.301	781.4731	781.4738	-0.9	605.4	$C_{42}H_{70}O_{13}$	782.4809	L+3-(glc^6+glc)	C	No report
25	7.415	913.5130	913.5161	-3.4	561.6	$C_{47}H_{78}O_{17}$	914.5208	L+3-(glc^6+glc^6+xyl)	B, (K. Yoshikawa et al., 1999)	164177-55-3
26	6.046	943.5244	943.5266	-2.3	555.8	$C_{48}H_{80}O_{18}$	944.5322	L+3-(glc^6+glc)+28-(glc)	B, (Zhu et al., 2008)	1096581-44-0
27	6.257	943.5276	943.5266	1.1	559.2	$C_{48}H_{80}O_{18}$	944.5354	L+3-(glc^6+glc^6+glc)	C	No report
28	6.029	1075.5725	1075.5689	3.3	532.3	$C_{53}H_{88}O_{22}$	1076.5803	L+3-(glc^6+glc^6+xyl)+28-(glc)	A, Ref	256510-01-7
29	5.529	1105.5813	1105.5795	1.6	525.8	$C_{54}H_{90}O_{23}$	1106.5891	L+3-(glc^6+glc)+28-(glc^6+glc)	C	No report
30	5.643	1105.5793	1105.5795	-0.2	524.0	$C_{54}H_{90}O_{23}$	1106.5871	L+3-(glc^6+glc^6+glc)+28-(glc)	C	No report
31	5.441	1237.6240	1237.6217	1.9	504.2	$C_{59}H_{98}O_{27}$	1238.6318	L+3-(glc^6+glc^6+xyl)+28-(glc^6+glc)	C	No report
32	13.423	617.4055	617.4053	0.3	656.8	$C_{36}H_{58}O_8$	618.4133	M+3-(glcA)	A, NMR	No report
34	9.976	633.4016	633.4003	2.1	634.0	$C_{36}H_{58}O_9$	634.4094	L+3-(glcA)	A, Ref, (Ye et al., 2000a)	287390-11-8
35	6.590	649.3959	649.3952	1.1	609.6	$C_{36}H_{58}O_{10}$	650.4037	G+3-(glcA)	A, Ref, (Pham et al., 2018)	2229693-58-5

Peak	R_t (#)	*m/z* (-)	Cald. *m/z (-)*	Diff (ppm)	RMD	Formula	Exact mass	Block-based compound assignment	Assignment fidelity (*)	CAS Number (&)
36	7.090	649.3968	649.3952	2.5	611.0	$C_{36}H_{58}O_{10}$	650.4046	S+3-(glcA)	A, Ref, (Pham et al., 2018)	2229693-57-4
37	7.730	649.3964	649.3952	1.8	610.4	$C_{36}H_{58}O_{10}$	650.4042	C+3-(glcA)	A, Ref, (Pham et al., 2018)	No report
38	7.564	661.3594	661.3588	0.9	543.4	$C_{36}H_{54}O_{11}$	662.3672	La+3-(glcA)	C	No report
39	7.126	663.3752	663.3744	0.2	565.6	$C_{36}H_{55}O_{11}$	664.3830	My+3-(glcA)	A, Ref, (Pham et al., 2018)	873799-50-9
40	5.880	665.3923	665.3901	3.3	589.6	$C_{36}H_{57}O_{11}$	666.4001	HG+3-(glcA)	C	No report
41	12.643	779.4579	779.4582	-0.4	587.5	$C_{42}H_{68}O_{13}$	780.4657	M+3-(glcA+glc)	A, Ref, (Pham et al., 2018)	2229693-55-2
42	10.112	793.4370	793.4374	-0.5	552.8	$C_{42}H_{66}O_{14}$	794.4464	O+3-(glcA)+28-(glc)	C	25406-56-8
43	11.072	793.4386	793.4374	1.5	552.8	$C_{42}H_{66}O_{14}$	794.4464	O+3-(glcA^3+glc)	A, Ref, (Alabdul Magid et al., 2006)	No report
44	8.266	795.4545	795.4531	1.8	571.4	$C_{42}H_{68}O_{14}$	796.4623	L+3-(glcA)+28-(glc)	B, (K. Yoshikawa et al., 1999)	330595-34-1
45	9.407	795.4545	795.4531	1.8	571.4	$C_{42}H_{68}O_{14}$	796.4623	L+3-(glcA^3+glc)	A, Ref, (Ye et al., 2000a)	256510-00-6
46	5.915	811.4473	811.4480	-0.9	551.2	$C_{36}H_{58}O_{10}$	812.4551	G+3-glcA+29-glc	A, Ref, (Pham et al., 2018)	2229693-56-3
47	6.195	811.4482	811.4480	0.2	552.3	$C_{36}H_{58}O_{10}$	812.4560	S+3-(glcA)+(glc)	C	No report
48	6.362	811.4489	811.4480	1.1	553.2	$C_{36}H_{58}O_{10}$	812.4567	C+3-glcA+28-glc	A, (K. Yoshikawa et al., 1998)	212775-48-9
49	6.451	811.4460	811.4480	-2.5	549.6	$C_{36}H_{58}O_{10}$	812.4538	G+3-(glcA^3+glc)	A, Ref, (Ye et al., 2001a)	330595-36-3
50	6.827	811.4477	811.4480	-0.4	551.7	$C_{36}H_{58}O_{10}$	812.4555	S+3-(glcA^3+glc)	A, (Pham et al., 2018)	2229693-54-1
51	7.319	811.4496	811.4480	2.0	554.1	$C_{36}H_{58}O_{10}$	812.4574	C+3-(glcA^3+glc)	B, (K. Yoshikawa et al., 1998)	212775-47-8
52	6.801	823.4119	823.4116	0.4	500.2	$C_{42}H_{64}O_{16}$	824.4197	La+3-(glcA)+28-(glc)	C	No report
53	6.932	823.4111	825.4234	-0.6	499.3	$C_{42}H_{64}O_{16}$	824.4189	La+3-(glcA^3+glc)	C	No report
54	6.862	825.4308	825.4272	4.2	521.9	$C_{42}H_{66}O_{16}$	826.4386	My+3-(glcA^3+glc)	A, Ref, (Pham et al., 2018)	2229693-52-9
55	5.617	827.4431	827.4429	0.2	535.5	$C_{42}H_{68}O_{16}$	828.4509	HG+3-(glcA^3+glc)	C	No report
56	9.204	691.4057	691.4057	-1.4	586.8	$C_{38}H_{60}O_{11}$	692.4135	C+3-(glcA)+acetyl	C	No report
57	7.476	707.4021	707.4007	2	568.4	$C_{38}H_{60}O_{12}$	708.4099	HG+3-(glcA)+acetyl	C	No report
58	11.107	731.438	731.4370	1.4	598.8	$C_{41}H_{64}O_{11}$	732.4458	CS+3-(glcA)+tig	C	No report
59	11.423	731.4382	731.4370	1.6	599.1	$C_{41}H_{64}O_{11}$	732.4460	CS+3-(glcA)+tig	C	No report
60	11.607	731.4378	731.4370	1.1	598.5	$C_{41}H_{64}O_{11}$	732.4456	S+3-(glcA)+21-tig	A, NMR	No report
61	8.161	747.4316	747.4320	-0.5	577.4	$C_{41}H_{64}O_{12}$	748.4394	HG+3-(glcA)+tig	C	No report
62	8.810	747.4351	747.4320	4.1	582.1	$C_{41}H_{64}O_{12}$	748.4429	HG+3-(glcA)+tig	C	No report
63	9.406	747.432	747.4320	0	578.0	$C_{41}H_{64}O_{12}$	748.4398	HG+3-(glcA)+tig	C	No report
64	9.266	747.433	747.4320	1.3	579.3	$C_{41}H_{64}O_{12}$	748.4408	HG+3-(glcA)+tig	C	No report
65	11.327	753.4228	753.4214	1.9	561.2	$C_{43}H_{62}O_{11}$	754.4306	C+3-(glcA)+28-bz	A, NMR	No report
66	11.555	753.4223	753.4214	1.2	560.5	$C_{43}H_{62}O_{11}$	754.4301	CS+3-(glcA)+bz	C	No report
67	11.809	753.4224	753.4214	1.3	560.6	$C_{43}H_{62}O_{11}$	754.4302	CS+3-(glcA)+bz	C	No report
68	8.634	769.4189	769.4163	3.4	544.4	$C_{43}H_{62}O_{12}$	770.4267	HG+3-(glcA)+bz	C	No report
69	9.134	769.4188	769.4163	3.2	544.3	$C_{43}H_{62}O_{12}$	770.4266	HG+3-(glcA)+bz	C	No report
70	9.599	769.4158	769.4163	3.2	540.4	$C_{43}H_{62}O_{12}$	770.4236	HG+3-(glcA)+28-bz	A, Ref, (Pham et al., 2018)	2229693-59-6
71	9.739	769.4183	769.4163	2.6	543.7	$C_{43}H_{62}O_{12}$	770.4261	HG+3-(glcA)+bz	C	No report
72	10.555	738.4244	738.4217	3.7	574.7	$C_{42}H_{61}NO_{10}$	739.4322	L+3-(glcA)+nic	C	No report
73	8.906	754.4214	754.4166	6.4	558.6	$C_{42}H_{61}NO_{11}$	755.4292	C+3-(glcA)+28-nic	A, NMR	No report
74	8.274	768.3981	768.3959	2.9	518.1	$C_{42}H_{59}NO_{12}$	769.4059	My+3-(glcA)+Nic	C	No report

Peak	R_t (#)	*m/z* (-)	Cald. *m/z (-)*	Diff (ppm)	RMD	Formula	Exact mass	Block-based compound assignment	Assignment fidelity (*)	CAS Number (&)
75	6.994	770.4136	770.4116	2.6	536.9	$C_{42}H_{61}NO_{12}$	771.4214	HG+3-(glcA)+nic	C	No report
76	9.862	893.4913	893.4899	1.6	549.9	$C_{47}H_{74}O_{16}$	894.4991	C+3-($glcA^3$+glc)+22-tig	B, (K. Yoshikawa et al., 1999)	256509-77-0
77	10.528	893.4917	893.4899	2	550.3	$C_{47}H_{74}O_{16}$	894.4995	C+3-($glcA^3$+glc)+28-tig	A, Ref, (K. Yoshikawa et al., 1999)	256509-82-7
78	10.783	893.4913	893.4899	1.6	549.9	$C_{47}H_{74}O_{16}$	894.4991	C+3-($glcA^3$+glc)+16-tig	B, (K. Yoshikawa et al., 1999)	256509-83-8
79	11.003	893.4940	893.4899	1.8	552.9	$C_{47}H_{74}O_{16}$	894.5018	CS+3-($glcA^3$+glc)+tig	C	No report
80	7.862	909.4854	909.4848	0.7	533.7	$C_{47}H_{74}O_{17}$	910.4932	HG+3-($glcA^3$+glc)+tig	C	No report
81	8.248	909.4849	909.4848	0.1	533.2	$C_{47}H_{74}O_{17}$	910.4927	HG+3-($glcA^3$+glc)+tig	C	No report
82	8.459	909.4865	909.4848	1.9	534.9	$C_{47}H_{74}O_{17}$	910.4943	HG+3-($glcA^3$+glc)+tig	C	No report
83	8.073	909.4852	909.4848	0.4	533.5	$C_{47}H_{74}O_{17}$	910.4930	HG+3-($glcA^3$+glc)+tig	C	No report
84	10.125	915.4756	915.4742	1.5	519.5	$C_{49}H_{72}O_{16}$	916.4834	CS+3-($glcA^3$+glc)+bz	C	No report
85	10.748	915.4756	915.4742	1.5	519.5	$C_{49}H_{72}O_{16}$	916.4834	CS+3-($glcA^3$+glc)+bz	C	No report
86	11.048	915.4737	915.4742	-0.5	517.4	$C_{49}H_{72}O_{16}$	916.4815	CS+3-($glcA^3$+glc)+bz	C	No report
87	11.107	915.4760	915.4742	2	519.9	$C_{49}H_{72}O_{16}$	916.4838	CS+3-($glcA^3$+glc)+21-bz	B, (Ye et al., 2001a)	330595-32-9
88	8.169	931.4698	931.4691	0.8	504.4	$C_{49}H_{71}O_{17}$	932.4776	HG+3-($glcA^3$+glc)+bz	C	No report
89	8.581	931.4702	931.4691	1.2	504.8	$C_{49}H_{71}O_{17}$	932.4780	HG+3-($glcA^3$+glc)+bz	C	No report
90	8.774	931.4708	931.4691	1.8	505.4	$C_{49}H_{71}O_{17}$	932.4786	HG+3-($glcA^3$+glc)+bz	C	No report
91	9.125	931.4724	931.4691	3.5	507.2	$C_{49}H_{71}O_{17}$	932.4802	HG+3-($glcA^3$+glc)+bz	C	No report
92	11.265	941.5110	941.5110	0	542.7	$C_{48}H_{77}O_{18}$	942.5188	M+3-($glcA^3$+glc)+28-(glc)	C	No report
93	10.195	955.4911	955.4903	0.8	514.0	$C_{48}H_{76}O_{19}$	956.4989	O+3-($glcA^3$+glc)+ 28-(glc)	C, (Alabdul Magid et al., 2006)	26020-29-1
94	8.755	957.5073	957.5059	1.5	529.8	$C_{48}H_{77}O_{19}$	958.5151	L+3-($glcA^3$+glc)+28-(glc)	A, Ref, (K. Yoshikawa et al., 1998)	212775-50-3
95	5.740	973.5030	973.5008	2.3	516.7	$C_{48}H_{78}O_{20}$	974.5108	C+3-($glcA^3$+glc)+28-(glc)	B, (K. Yoshikawa et al., 1998)	212775-45-6
96	6.134	973.5017	973.5008	0.9	515.4	$C_{48}H_{78}O_{20}$	974.5095	G+3-($glcA^3$+glc)+28-(glc)	C	No report
97	6.783	973.5005	973.5008	-0.3	514.1	$C_{48}H_{77}O_{20}$	974.5083	S+3-($glcA^3$+glc)+28-(glc)	C	No report
98	6.134	985.4604	985.4586	1.8	467.2	$C_{48}H_{73}O_{21}$	986.4682	La+3-($glcA^3$+glc)+28-(glc)	C	No report
99	6.117	987.4801	987.4801	0	486.2	$C_{48}H_{76}O_{21}$	988.4879	My+3-($glcA^3$+glc)+28-(glc)	C	No report
100	4.722	989.4953	989.4957	-0.4	500.6	$C_{48}H_{77}O_{21}$	990.5031	HG+3-($glcA^3$+glc)+28-(glc)	C	No report
101	9.932	900.4773	900.4745	3.1	530.1	$C_{48}H_{71}NO_{16}$	901.4851	L+3-($glcA^3$+glc)+nic	C	No report
102	8.441	916.4713	916.4695	2	514.3	$C_{48}H_{71}NO_{16}$	917.4791	CS+3-($glcA^3$+glc)+nic	C	No report
103	6.415	932.4634	932.4644	-1.1	497.0	$C_{48}H_{70}NO_{17}$	933.4712	HG+3-($glcA^3$+glc)+nic	C	No report
104	7.432	837.4810	837.4789	2.5	574.3	$C_{44}H_{70}O_{15}$	838.4888	C+3-(glcA)+28-(rha)+22-acetyl	B, (K. Yoshikawa et al., 1998)	212775-43-4
105	7.670	837.4770	837.4789	-2.3	569.6	$C_{44}H_{70}O_{15}$	838.4848	C+3-(glcA)+28-(rha)+ 16-acetyl	B, (K. Yoshikawa et al., 1998)	212775-43-4
106	6.134	999.5154	999.5165	-1.1	515.6	$C_{50}H_{80}O_{20}$	1000.5232	C+3-($glcA^3$+glc)+28-(rha)+22-acetyl	B, (K. Yoshikawa et al., 1998)	212775-19-4
107	6.801	999.5170	999.5165	0.5	517.2	$C_{50}H_{80}O_{20}$	1000.5248	C3-($glcA^3$+glc)+28-(rha)+16-acetyl	B, (K. Yoshikawa et al., 1998)	212775-23-0
108	5.862	1015.5121	1015.5114	0.7	504.3	$C_{50}H_{80}O_{21}$	1016.5199	C+3-($glcA^3$+glc)+28-(glc)+acetyl	C	No report

Peak	R_t (#)	*m/z* (-)	Cald. *m/z (-)*	Diff (ppm)	RMD	Formula	Exact mass	Block-based compound assignment	Assignment fidelity (*)	CAS Number (&)
109	9.783	1055.5441	1055.5427	1.3	515.5	$C_{53}H_{83}O_{21}$	1056.5519	CS+3-(glcA^3+glc)+28-(glc)+tig	C	No report
110	10.143	1055.5457	1055.5427	2.8	517.0	$C_{53}H_{83}O_{21}$	1056.5535	CS+3-(glcA^3+glc)+28-(glc)+22-tig	B, (K. Yoshikawa et al., 1998)	256509-85-0
111	7.283	1071.5386	1071.5376	0.9	502.6	$C_{53}H_{83}O_{22}$	1072.5464	HG+3-(glcA^3+glc)+28-(glc)+tig	C	No report
112	7.380	1071.5381	1071.5376	0.5	502.2	$C_{53}H_{83}O_{22}$	1072.5459	HG+3-(glcA^3+glc)+28-(glc)+tig	C	No report
113	7.678	1071.5393	1071.5376	1.6	503.3	$C_{53}H_{83}O_{22}$	1072.5471	HG+3-(glcA^3+glc)+28-(glc)+tig	C	No report
114	10.310	1077.5294	1077.5270	2.2	491.3	$C_{55}H_{82}O_{21}$	1078.5372	CS+3-(glcA^3+glc)+28-(glc)+bz	C	No report
115	9.248	1077.5298	1077.5270	2.6	491.7	$C_{55}H_{82}O_{21}$	1078.5376	CS+3-(glcA^3+glc)+28-(glc)+bz	C	No report
116	8.722	1093.5260	1093.5219	3.7	481.0	$C_{55}H_{82}O_{22}$	1094.5338	HG+3-(glcA^3+glc)+28-(glc)+bz	C	No report
117	8.327	1093.5243	1093.5219	2.2	479.5	$C_{55}H_{82}O_{22}$	1094.5321	HG+3-(glcA^3+glc)+28-(glc)+bz	C	No report
118	7.888	1093.5231	1093.5219	1.1	481.0	$C_{55}H_{82}O_{22}$	1094.5338	HG+3-(glcA^3+glc)+28-(glc)+bz	C	No report
119	7.695	1039.5466	1039.5478	-1.2	525.8	$C_{53}H_{84}O_{20}$	1040.5544	C+3-(glcA^3+glc)+28-(rha)+22-tig	B, (K. Yoshikawa et al., 1998)	212775-27-4

(#) R_t : Retention time.

(&) CAS number: Chemical Abstract Service registration number. The compounds with "No report" were presumed to be new compounds after looking up its possible structures on Scifinder by chemical formula and no result was found to have similar building block components and mass fragmentation patterns. Structure of peak ***12*** was reported but has not been indexed with a CAS number.

(*) Rate of assignment fidelity:

(A) Unambiguous identification by reference compounds (Ref) from KBNMB or structure isolation and elucidation of new compounds (NMR).

(B) Evidence of building block mass data, mKIA, MFA, and retention time analysis; Compound should be reported isolated from *G. sylvestre*. The assignment can be interchanged between unidentified isomers of the same molecular weight and mass behavior characteristics.

(C) Evidence of building block mass data, mKIA, MFA and retention time analysis.

3.2.3.2. Erythrodiol Glycosides

Peaks ***20*** (t_R = 7.117 min) and ***21*** (t_R = 6.950 min) displayed mass fragmentations of glucoside chains similar to those of the oleanolic acid glycosides, but they each possessed a set of key ions at 425 and 407. Because the structures of the compounds reported in the literature could not explain these key ions, direct isolation by conventional column chromatography and structure elucidation by NMR spectroscopy were performed for these two peaks. Consequently, peaks ***20*** and ***21*** were identified as two new compounds, 3-*O*-β-D-glucopyranosyl-(1→6)-*O*-β-D-glucopyranosyl erythrodiol 28-*O*-β-D-glucopyranosyl-(1→6)-*O*-β-D-glucopyranoside and 3-*O*-β-D-xylopyranosyl-(1→6)-*O*-β-D-glucopyranosyl-(1→6)-*O*-β-D-glucopyranosyl erythrodiol 28-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside, respectively (**Figure 14**). These findings highlight the existence of erythrodiol (E) as an aglycone of *G. sylvestre*, and its glycosides followed a construction rule similar to **Figure 10B-a** with oleanolic acid. Similar target ion filtrations of predicted compounds led to the detection of 6 other peaks (***15-19*** and ***22***) from erythrodiol glycosides, and their structures were tentatively assigned. All these assumed erythrodiol glycosides were suggested to be new compounds by searching for their molecular formulas and structures in SciFinder.

3.2.3.3. Longispinogenin (L) Glycosides

Longispinogenin has been reported as the main aglycone of the Chinese and Vietnamese varieties of *G. sylvestre* (Pham et al., 2018),(Ye et al., 2001b) and it can be practically recognized by based on its characteristic skeleton ions at *m/z* 441, 423, and 405. Nine peaks of longispinogenin derivatives (***23***–***31***) displayed the same type

of glycoside construction as was seen with oleanolic acid and erythrodiol (**Figure 10-a**). Peak ***28*** (t_R = 6.029 min) was determined to be alternoside XIX (K. Yoshikawa et al., 1999) by comparison with a reference compound. Peak ***31*** (t_R = 5.441 min) was expected to be a new compound with one additional glucose compared to ***28***. Peaks ***25*** and ***26*** were tentatively assigned as sitakisoside X (Y. Liu et al., 2014) and gymnemoside W1 (Zhu et al., 2008), respectively.

Three new compounds, longispinogenin 3-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside (***27***), 3-*O*-β-D-glucopyranosyl (1→6)-β-D-glucopyranosyl longispinogenin 28-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside (***29***) and 3-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranosyl(1→6)-β-D-glucopyranosyl longispinogenin 28-*O*-β-D-glucopyranoside (***30***), were deduced similarly. Considering oleanolic acid and erythrodiol (**Figure 21-B**), the carboxylic group at C-28 of oleanolic acid (layer 1) caused a weak increase in the polarities of all compounds compared to their erythrodiol counterparts (layer 2), which had the same glycoside pattern except for the modification of the hydroxy group at C-28. In addition, longispinogenin glycosides (layer 3), which have one addition hydroxy group on the aglycone at C-16 relative to erythrodiol (layer 2), showed remarkably higher polarities.

3.2.3.4. Oleanane Triterpenes with Glucuronic Acid

These compounds characterized by a glucuronic acid attached to C-3, and they were diversified by two factors: glycosylation and acylation. Glycosylation frequently occurs at C-3' and C-28 of glucuronic acid (**Figure 10B-b**), while acylation is the attachment of an acyl unit to a hydroxy group on rings D and E of

the oleanane skeleton (**Figure 10B-c**). The key characteristic of the nonacylated triterpene glycosides is the presence of a glucuronic acid at C-3 of the main aglycone. The building block model for [aglycone + glcA] predicted 7 possible deprotonated mass values (**Figure 11**), and their key ion extraction led to the detection of peaks ***32***-***41*** (**Table 2**). Among them, peaks ***33***-***36*** and ***39*** were unambiguously identified by comparison with reference compounds, and three peaks ***37***, ***38***, and ***40*** were tentatively assigned based on their mass fragmentation patterns. The ion at *m*/*z* 617.4057 of peak ***32*** can be the 3-*O*-β-D-glucuronopyranosyl analogue of either erythrodiol or maniladiol because they share the same molecular formula ($C_{36}H_{58}O_8$). NMR spectroscopy was used to clarify the structure of peak ***32***, and it was revealed to be maniladiol 3-*O*-β-D-glucuronopyranoside.

Considering that all these detected glycosides possess similar glucuronic acid linkages at C-3, the difference in their polarities would be related to the structure of their aglycones. As seen in **Figure 21** (layer 4), the retention times of these peaks were in the following order of their aglycones, and the order matched the number of hydroxy or carboxylic groups in their structure: 22α-hydroxygymnemagenol (5 × OH) > gymnemagenol (4 × OH) > sitakisogenin (4 × OH) > myrtilogenic acid (3 × OH + COOH) > 3β,16β,28-trihydroxyolean-12-en-22,29-olide (3 × OH + C=O) > chichipegenin (4 × OH) > longispinogenin (3 × OH) > oleanolic acid (OH + COOH) > maniladiol (2 × OH). Similarly, the 9 possible deprotonated mass values predicted for [aglycone + glcA + glc] were extracted (**Figure 11**), resulting in the detection of 15 peaks (***41***-***55***) (**Figure 21-B**). The structure of compounds ***41***, ***43***-***45***, ***47***, ***48***, ***50***, and ***54*** (confirmed by comparison with reference standards) and compounds ***46*** and ***49*** (tentatively assigned by comparing with literature data) suggested that an

additional glucose can be attached to C-3' of the glucuronic acid moiety to extend the C-3 linkage or connected to the aglycone skeleton at C-28 (or C-29 in gymnemagenol).

An investigation of 3 known glycoside peaks (***34***, ***44***, and ***45***) of the same aglycone longispinogenin showed a shift in the retention time of compound ***44***, which has a second glucose unit attached to C-28 (-1.7 min), to a time much earlier than that of compound ***45***, in which the second sugar is linked to C-3' of glucuronic acid (-0.6 min). A low variation in the retention times (0.4-0.8 min) was also observed in aglycones in which the second sugar was connected to C-3' of the glucuronic acid, and these included maniladiol (peaks ***32*** and ***41***, -0.8 min), myrtilogenic acid (peaks ***39*** and ***54***, -0.4 min), and sitakisogenin (peaks ***36*** and ***48***, -0.7 min) (***Figure 21***). On the other hand, the large shift in the retention time (-1.7 min) of peak ***42***, which is known to have the structure [O + 3-(glcA) + 28-(glc)] relative to that of peak ***33*** [O + 3-(glcA)], further suggested that unknown peak ***43*** was [O + 3-(glcA^3-glc)] based on the difference between its retention time with that of peak ***33*** (0.7 min). For longer sugar chains, three compounds have been reported to be constructed from three units following the construction rule [aglycone + 3-(glcA^3-glc) + 28-(glc)] **(K. Yoshikawa et al., 1998)**. Building block analysis and peak extraction resulted in the detection of 8 peaks (**Figure 21-B**, layer 12). The difference between the retention times of the reference compound ***94*** [L + 3-(glcA^3-glc) + 28-(glc)] and ***45*** [L + 3-(glcA^3-glc)] suggested a polar shift of approximately 1.3 min due to the new glucose linkage at C-28. As a result, a range of retention time shifts of 1-1.4 min was observed between peaks ***92*** (M), ***93*** (O), ***98*** (My), and ***100*** (HG) (**Figure** 21**-B**, layer 12) and their corresponding analogues [aglycone + 3-

(glcA^3-glc)]. Peaks ***41*** (M), ***42*** (O), ***54*** (My), and ***45*** (HG) (**Figure 21-B,** layer 5) were assigned as aglycone + 3-(glcA^3-glc) + 28-(glc).

3.2.3.5. Triterpenes Glycosides with Tigloyl and Benzoyl Substituents

The acylation of the glycosides of 22α-hydroxygymnemagenol, chichipegenin, and sitakisogenin by tigloyl and benzoyl moieties following **Figure 10B-c** has been reported. The acylation of the triterpenoid glycoside core structure significantly reduces the polarity of the compound. As illustrated in **Figure 21-B**, three peaks (***58-60***) were assigned to CS + 3-(glcA) + tigloyl derivatives and 4 peaks (***61-64***) were assigned to HG + 3-glcA + tigloyl derivatives (layer 7). The other isomers with the same molecular weight were presumed to be aglycone variants or be acylated at different positions. Interestingly, a corresponding number of benzoyl derivatives (peaks ***65-71***, **Figure 21-B**, layer 8) were observed, and they showed slight reductions in polarity (0.2 min) compared with their tigloyl derivatives (**Figure 21-B**, layer 7). Since compound ***65*** was elucidated as 28-benzoyl-chichipegenin 3-*O*-β-D-glucopyranoside, it is likely that peak ***58*** was 28-tigloyl-chichipegenin 3-*O*-β-D-glucopyranoside. A similar analysis of peak ***67*** suggested that it was 21-*O*-tigloyl-sitaskisogenin 3-*O*-β-D-glucuronopyranoside based on the assignment of ***60*** as 21-*O*-tigloyl-sitaskisogenin 3-*O*-β-D-glucuronopyranoside, which was determined from NMR evidence. Peak ***70*** was confirmed as 28-benzoyl-22α-hydroxygymnemagenol 3-*O*-β-D-glucopyranoside by comparison to reference compounds, so its tigloyl analogue (peak ***63***) could be assigned as 28-tigloyl-22α-hydroxygymnemagenol 3-*O*-β-D-glucopyranoside (**Figure 21-B**, layers 10 and 11).

3.2.3.6. Triterpenes Glycosides with Nicotinoyl Substituents.

The determination that peak ***73*** had an odd exact mass suggested that the compound possessed a nitrogen moiety attached to the aglycone. Because no similar compound has been reported from *G. sylvestre*, the compound was isolated and elucidated to be 28-nicotinoyl chichipegenin 3-*O*-β-D-glucuronopyranoside. This finding suggested that the nicotinoyl moiety is another biosynthetic building block for *G. sylvestre* compounds. Building block peak extraction was then applied and revealed 7 nicotinoyl derivatives of longispinogenin, chichipegenin, myrtilogenic acid and 22α-hydroxygymnemagenol (peaks ***72-75*** and ***101-103***) (**Figure 21-B**, layers 9 and 13). The construction rule and shift in retention times suggested that their structures can be tentatively assigned as [aglycone + 3-(glcA) + 28-(nic)] and [aglycone + 3(glcA^3-glc) + 28-(nic)] derivatives, respectively (**Table 2**).

3.2.3.6. Triterpenes Glycosides with Acetyl Substituents

Four acylated compounds were reported from *G. sylvestre*, with the main aglycone being chichipegenin (K. Yoshikawa et al., 1998). The deprotonated ion mass extraction also suggested the presence of these compounds in the fractionated matrixes (peaks ***104-107***). Building block prediction suggested that acetylation is not common in this plant, and the derivatives of chichipegenin and 22α-hydroxygymnemagenol were observed in low quantities. The following building block structures could be found on the basis of the determined construction rules: peaks ***56*** and ***57*** [aglycone + 3-(glcA) + acetyl], peaks ***104*** and ***105*** [aglycone + 3-(glcA^3-glc) + 28-(rha) + acetyl], and peaks ***106-108*** [aglycone + 3-(glcA^3-glc) + 28-(glc) + acetyl]. Substitution on a rhamnose moiety occurred only with an acetyl or tigloyl moiety was available since the extraction of the predicted glycosides containing nonacylated rhamnose moieties revealed no results (**Table 1**).

3.3. Structure elucidation of selective isolation of new compounds predicted by virtual block library

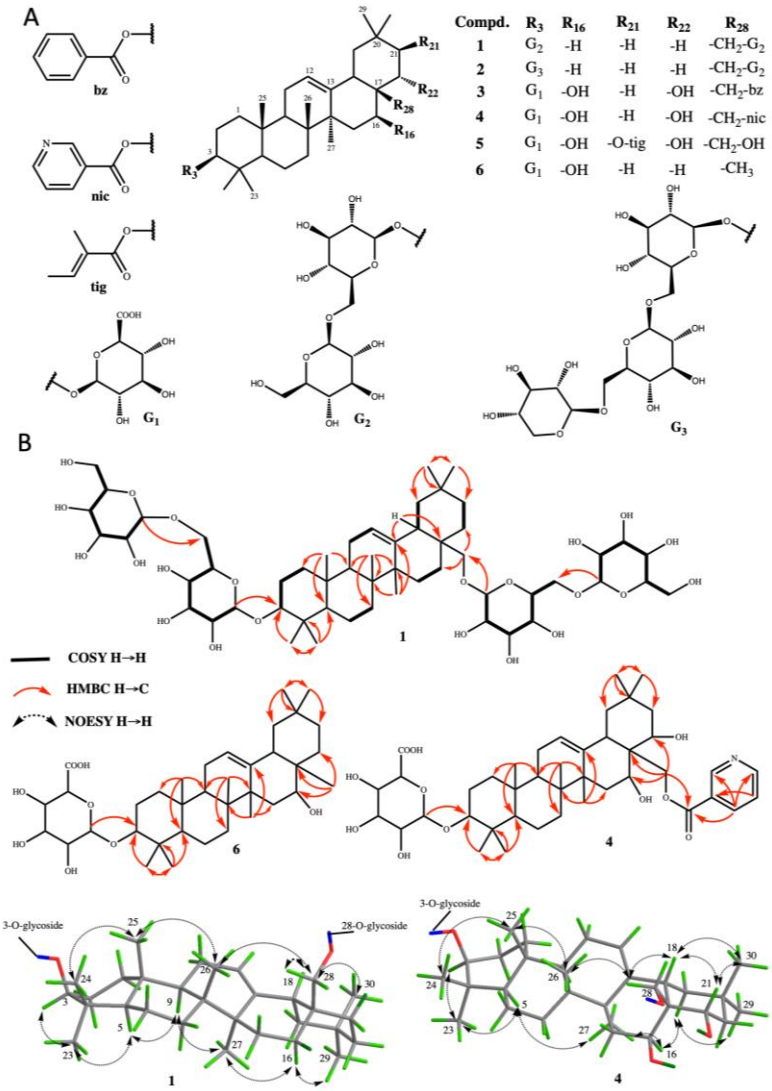


Figure 14. Structures of new isolated compounds **1–6** (A) and key 2D-NMR correlations (B) of new isolated compounds **1**, **4**, and **6**.

Table 3. ^{13}C NMR spectroscopic data for compounds Gymnemosides **GS1**–**6** (in Pyridine -d_5).

position	1 δ_C, type, 150 Mhz	2 δ_C, type, 150 Mhz	3[#] δ_C, type, 125 Mhz	4[##] δ_C, type, 200 Mhz	5 δ_C, type, 150 Mhz	6 δ_C, type, 125 Mhz
1	39.1, CH_2	39.1, CH_2	39.1, CH_2	39.1, CH_2	39.1, CH_2	39.1, CH_2
2	27.0, CH_2	27.0, CH_2	26.9, CH_2	26.8, CH_2	27.0, CH_2	27.0, CH_2
3	89.4, CH	89.4, CH	89.3, CH	89.3, CH	89.3, CH	89.4, CH
4	39.8, C	39.8, C	39.9, C	39.8, C	39.9, C	39.9, C
5	56.0, CH	56.0, CH	55.9, CH	55.9, CH	56.0, CH	56.0, CH
6	18.8, CH_2	18.8, CH_2	18.7, CH_2	18.7, CH_2	18.8, CH_2	18.8, CH_2
7	33.0, CH_2	33.0, CH_2	33.2, CH_2	33.1, CH_2	33.3, CH_2	33.4, CH_2
8	40.4, C	40.4, C	40.6, C	40.6, C	40.5, C	40.6, C
9	48.3, CH	48.2, CH	47.4, CH	47.4, CH	47.4, CH	47.4, CH
10	37.1, C	37.1, C	37.0, C	37.0, C	37.1, C	37.1, C
11	24.0, CH_2	24.1, CH_2	24.2, CH_2	24.2, CH_2	24.2, CH_2	24.2, CH_2
12	123.1, CH	123.1, CH	124.2, CH	124.4, CH	123.5, CH	122.8, CH
13	144.9, C	144.8, C	142.3, C	142.3, C	143.1, C	144.9, C
14	42.1, C	42.1, C	42.9, C	42.9, C	44.1, C	44.3, C
15	26.5, CH_2	26.5, CH_2	36.5, CH_2	36.4, CH_2	36.3, CH_2	36.8, CH_2
16	22.4, CH_2	22.3, CH_2	66.6, CH	66.6, CH	67.2, CH	64.8, CH
17	37.4, C	37.3, C	44.8, C	44.7, C	43.5, C	38.4, C
18	43.4, CH	43.4, CH	44.2, CH	44.2, CH	44.1, CH	49.9, CH
19	47.0, CH_2	47.0, CH_2	46.4, CH_2	46.3, CH_2	47.4, CH_2	47.5, CH_2
20	31.4, C	31.4, C	32.4, C	32.4, C	37.1, C	31.5, C
21	34.7, CH_2	34.7, CH_2	44.4, CH_2	44.3, CH_2	76.7, CH	35.1, CH_2
22	33.0, CH_2	33.0, CH_2	69.9, CH	69.9, CH	30.3, CH_2	31.6, CH_2
23	28.5, CH_3	28.5, CH_3	28.5, CH_3	28.5, CH_3	28.6, CH_3	28.5, CH_3
24	17.3, CH_3	17.3, CH_3	17.3, CH_3	17.3, CH_3	17.3, CH_3	17.3, CH_3
25	16.1, CH_3	16.0, CH_3	16.0, CH_3	15.9, CH_3	16.1, CH_3	16.0, CH_3
26	17.4, CH_3	17.3, CH_3	17.5, CH_3	17.4, CH_3	17.3, CH_3	17.4, CH_3
27	26.5, CH_3	26.5, CH_3	28.0, CH_3	28.0, CH_3	27.4, CH_3	27.8, CH_3
28	77.3, CH_2	77.3, CH_2	63.6, CH_2	64.1, CH_2	67.2, CH_2	22.8, CH_3
29	33.7, CH_3	33.7, CH_3	33.7, CH_3	33.6, CH_3	29.6, CH_3	33.9, CH_3
30	24.3, CH_3	24.3, CH_3	25.2, CH_3	25.2, CH_3	19.1, CH_3	24.5, CH_3
GlcA-1'			107.6, CH	107.4, CH	107.4, CH	107.6, CH
GlcA-2'			75.8, CH	75.8, CH	75.8, CH	75.9, CH
GlcA-3'			78.5, CH	78.5, CH	78.6, CH	78.6, CH
GlcA-4'			73.9, CH	73.9, CH	73.6, CH	73.9, CH
GlcA-5'			78.5, CH	78.4, CH	78.5, CH	78.0, CH
GlcA-6'			172.6, COOH	173.9, COOH	173.2, COOH	173.4, COOH
Glc-SA1-1'	107.2, CH	107.2, CH				
Glc-SA1-2'	75.9, CH	75.9, CH				
Glc-SA1-3'	78.8, CH	78.7, CH				
Glc-SA1-4'	72.0, CH	71.9, CH				
Glc-SA1-5'	77.5, CH	77.5, CH				
Glc-SA1-6'	70.8, CH_2	70.6, CH_2				
Glc-SA2-1'	105.7, CH	105.7, CH				
Glc-SA2-2'	75.5, CH	75.4, CH				
Glc-SA2-3'	78.7, CH	78.8, CH				
Glc-SA2-4'	71.9, CH	71.9, CH				
Glc-SA2-5'	77.3, CH	77.3, CH				
Glc-SA2-6'	63.0, CH_2	70.0, CH_2				
Xyl-SA3-1'		106.3, CH				
Xyl-SA3-2'		75.2, CH				
Xyl-SA3-3'		78.4, CH				
Xyl-SA3-4'		71.4, CH				
Xyl-SA3-5'		67.3, CH_2				
Glc-SB1-1'	105.8, CH	105.6, CH				
Glc-SB1-2'	75.5, CH	75.4, CH				
Glc-SB1-3'	78.7, CH	78.7, CH				
Glc-SB1-4'	72.0, CH	71.9, CH				
Glc-SB1-5'	77.3, CH	77.3, CH				
Glc-SB1-6'	70.3, CH_2	70.3, CH_2				
Glc-SB2-1'	105.7, CH	105.2, CH				
Glc-SB2-2'	75.3, CH	75.3, CH				

	1	**2**	**3**[#]	**4**[##]	**5**	**6**
position	δ_C, type, 150 Mhz	δ_C, type, 150 Mhz	δ_C, type, 125 Mhz	δ_C, type, 200 Mhz	δ_C, type, 150 Mhz	δ_C, type, 125 Mhz
Glc-SB2-3'	78.7, CH	78.7, CH				
Glc-SB2-4'	71.9, CH	71.8, CH				
Glc-SB2-5'	77.3, CH	77.3, CH				
Glc-SB2-6'	63.0, CH_2	63.0, CH_2				

[#] 28-*O*-benzoyl substitution: Bz-1 (δ_C167.1, C), Bz-2 (δ_C131.5, C), Bz-3,7 (δ_C130.0, CH), Bz-4,6 (δ_C129.3, CH), Bz-5 (δ_C133.7, CH).

[##] 28-*O*-nicotinoyl substitution: COO- (δ_C167.1, C); Nic-2 (δ_C151.1, CH); Nic-3 (δ_C127.1, C); Nic-4 (δ_C137.3, CH); Nic-5 (δ_C123.1, CH); Nic-6 (δ_C154.2, CH).

***** 21 *O*-tigloyl moiety: -COO- (δ_C167.9, C); Tig-2 (δ_C 130.0, CH); Tig-3 (δ_C137.1, C); Tig-4 (δ_C14.6, CH); Tig-5 (δ_C12.7, CH).

Table 4. ^{1}H NMR Spectroscopic Data for Gymnemosides **GS1**–**6** (in Pyridine -d_5).

position	**1** δ_H (*J* in Hz), 600 Mhz	**2** δ_H (*J* in Hz), 600 Mhz	**3**[#] δ_H (*J* in Hz), 500 Mhz	**4**[##] δ_H (*J* in Hz), 800 Mhz	**5** δ_H (*J* in Hz), 600 Mhz	**6** δ_H (*J* in Hz), 500 Mhz
1	1.55, overlap 1.05, overlap	1.55, overlap 1.05, overlap	1.39, overlap 0.84, overlap	1.32, overlap 0.79, overlap	1.38, overlap 0.87, overlap	1.41, overlap 0.89, ovelap
2	2.39, m 1.84, overlap	2.39, m 1.84, overlap	2.22, overlap 1.85, overlap	2.22, overlap 1.78, overlap	2.19, overlap 1.84, overlap	2.08, overlap 1.86, overlap
3	3.31, dd (11.5, 4.3)	3.31, dd (11.5, 4.3)	3.36, dd (11.5, 4.0)	3.34, dd (11.5, 4.0)	3.41, dd (11.5, 4.0)	3.36, dd (4.0, 11.5)
5	0.72, d (11.7)	0.72, d (11.7)	0.74, overlap	0.77, overlap	0.80, d (12.0)	0.77, d (11.7)
6	1.47, overlap 1.31, overlap	1.48, overlap 1.32, overlap	1.50, overlap 1.30, overlap	1.48, overlap 1.26, overlap	1.51, overlap 1.34, overlap	1.49, overlap 1.34, overlap
7	1.41, overlap 1.24, overlap	1.43, overlap 1.25, overlap	1.57, overlap 1.33, overlap	1.56, overlap 1.30, overlap	1.51, overlap 1.32, overlap	1.54, overlap 1.34, overlap
9	1.58, overlap	1.59, overlap	1.57, overlap	1.56, overlap	1.58, overlap	1.57, overlap
11	2.05, overlap 1.85, overlap	2.04, overlap 1.84, overlap	1.81, overlap 1.58, overlap	1.81, overlap 1.56, overlap	1.84, overlap 1.58, overlap	1.86, overlap 1.57, overlap
12	5.17, br s	5.17, br s	5.33, br s	5.36, br s	5.33, br s	5.30, br s
15	1.80, overlap 0.90, overlap	1.80, overlap 0.90, overlap	2.15, t (13.0) 1.71, dd (13.0, 4.0)	2.12, overlap 1.71, overlap	2.19, t (13.0) 1.71, dd (13.0, 4.0)	2.06, overlap 1.62, overlap
16	1.91, overlap 1.63, overlap	1.91, overlap 1.62, overlap	5.16, dd (11.3. 4.5)	5.16, dd (11.3. 4.5)	4.71, dd (11.3. 4.5)	4.57, overlap
18	2.13, dd (12.0, 3.0)	2.13, dd (12.0, 3.0)	2.94, dd (14.0, 4.0)	2.92, dd (14.0, 4.0)	2.73, dd (14.0, 4.0)	2.30, dd (13.8, 4.0)
19	1.83, overlap 1.14, overlap	1.84, overlap 1.14, overlap	2.04, overlap 1.23, overlap	2.03, overlap 1.23, overlap	2.09, t (14.0) 1.42, overlap	1.90, overlap 1.15, overlap
21	1.55, overlap 1.13, overlap	1.55, overlap 1.13, overlap	2.08, overlap 1.95, overlap	2.06, overlap 1.95, overlap	5.51, dd (13.0, 4.0)	2.08, overlap 1.62, overlap
22	1.81, overlap 1.76, overlap	1.82, overlap 1.76, overlap	4.84, dd (12.0, 3.0)	4.80, dd (12.0, 3.0)	3.09, dd (13.0, 4.0) 2.12, t (13.0)	2.45, td (14.0, 2.6) 1.14, overlap
23	1.01, s	1.01, s	0.97, s	1.09, s	1.02, s	1.00, s
24	1.26, s	1.26, s	1.30, s	1.30, s	1.34, s	1.30, s
25	0.88, s	0.87, s	0.80, s	0.80, s	0.85, s	0.86, s
26	0.96, s	0.95, s	1.10, s	1.11, s	0.98, s	1.02, s
27	1.21, s	1.21, s	1.43, s	1.43, s	1.35, s	1.38, s
28	3.88, d (10.5) 3.81, d (10.5)	3.88, d (10.5) 3.81, d (10.5)	5.41, d (10.5) 4.93, d (10.5)	5.43, d (10.5) 4.80, d (10.5)	4.35, d (10.5) 3.83, d (10.5)	1.14, s
29	0.91, s	0.91, s	1.15, s	0.97, s	1.05, s	0.97, s
30	0.94, s	0.94, s	1.03, s	1.02, s	1.23, s	0.92, s
GlcA-1'			4.97, d (7.7)	4.91, d (7.7)	5.01, d (7.7)	5.01, d (7.7)
GlcA-2'			4.12, overlap	4.08, overlap	4.13, overlap	4.13, overlap
GlcA-3'			4.32, t (7.7)	4.29, t (7.7)	4.33, t (7.7)	4.33, t (7.7)
GlcA-4'			4.58, t (7.7)	4.49, t (7.7)	4.58, overlap	4.58, overlap
GlcA-5'			4.66, d (7.7)	4.55, d (7.7)	4.66, d (7.7)	4.66, d (7.7)
Glc-SA1-1'	4.89, d (7.5)	4.87, d (7.5)				
Glc-SA1-2'	3.97, overlap	3.97, overlap				
Glc-SA1-3'	4.10, overlap	4.10, overlap				
Glc-SA1-4'	4.16, overlap	4.16, overlap				
Glc-SA1-5'	3.99, overlap	3.99, overlap				
Glc-SA1-6'	4.29, overlap 4.89, overlap	4.29, overlap 4.89, overlap				
Glc-SA2-1'	5.14, d (7.5)	5.10, *d* (7.5)				
Glc-SA2-2'	3.96, overlap	3.96, overlap				
Glc-SA2-3'	4.16, overlap	4.16, overlap				
Glc-SA2-4'	4.09, overlap	4.09, overlap				
Glc-SA2-5'	3.99, overlap	3.99, overlap				
Glc-SA2-6'	4.75, d (7.5)	4.77, d (7.5)				

	1	**2**	**3**[#]	**4**[##]	**5**	**6**
position	δ_H (*J* in Hz), 600 Mhz	δ_H (*J* in Hz), 600 Mhz	δ_H (*J* in Hz), 500 Mhz	δ_H (*J* in Hz), 800 Mhz	δ_H (*J* in Hz), 600 Mhz	δ_H (*J* in Hz), 500 Mhz
	4.30, overlap	4.30, overlap				
Xyl-SA3-1'		4.94, d (7.5)				
Xyl-SA3-2'		3.97, overlap				
Xyl-SA3-3'		4.10, overlap				
Xyl-SA3-4'		4.16, overlap				
Xyl-SA3-5'		4.29, overlap 3.62, t (7.5)				
Glc-SB1-1'	4.83, d (8.0)	4.76, d (7.5)				
Glc-SB1-2'	4.12, overlap	4.12, overlap				
Glc-SB1-3'	4.17, overlap	4.17, overlap				
Glc-SB1-4'	4.29, overlap	4.29, overlap				
Glc-SB1-5'	4.09, overlap	4.09, overlap				
Glc-SB1-6'	4.33, overlap	4.33, overlap				
Glc-SB2-1'	5.14, d (7.5)	5.04, d (7.5)				
Glc-SB2-2'	3.96, overlap	3.96, overlap				
Glc-SB2-3'	4.09, overlap	4.09, overlap				
Glc-SB2-4'	4.16, overlap	4.16, overlap				
Glc-SB2-5'	4.09, overlap	4.09, overlap				
Glc-SB2-6'	4.45, dd (11.0, 2.0) 4.31, overlap	4.45, dd (11.0, 2.0) 4.31, overlap				

[#] 28-*O*-Benzoyl substitution: Bz-3,7 [each 1H, δ_H 8.25, (d, *J*= 7.5 Hz)], Bz-4,6 [each 1H, δ_H 7.43, (t, *J*= 7.5 Hz)], Bz-5 [1H, δ_H 7.52, (t, *J*= 7.5 Hz)].

[##] 28-*O*-nicotinoyl substitution: Nic-2 (1H, δ_H 9.50, s); Nic-4 [1H, δ_H 8.36, (d, *J*= 7.5 Hz)]; Nic-5 [1H, δ_H 7.35 (dd, *J*= 7.5, 1.8 Hz)]; Nic-6 [1H, δ_H 8.85 (d, *J*= 1.8 Hz)].

***** 21 *O*-tigloyl moiety: Tig-3 [1H, δ_H 7.00, (q, *J*= 7.0 Hz)]; Tig-4 [1H, δ_H 1.66, (d, *J*= 7.0 Hz); Tig-5 (1H, δ_H 1.89, s).

*3.3.1. Gymnemoside GS1 - Compound **1***

Compound **1** (peak 20) was obtained as an amorphous powder with $[\alpha]_D^{25}$ 24.6 (*c* 0.1, MeOH). The IR absorption bands at 3394, 1052, and 1610 cm^{-1} indicated the presence of a hydroxy group and an olefin. Its molecular formula ($C_{54}H_{90}O_{22}$), which corresponds to 10 indices of hydrogen deficiency (IDHs), was deduced based on its HRESIMS ion peak at *m/z* 1089.5884 $[M - H]^-$ (calcd for $C_{54}H_{89}O_{22}$, 1089.5845). The HPLC-MS fragment ions obtained in positive mode suggested the presence of 4 glucoses (4 ×162 Da) and an aglycone with a molecular weight of 422 with 2 hydroxy groups, which was indicated by the three key ions at 443, 425 and 407. The ^1H, ^{13}C, and HSQC NMR spectra showed seven methyl proton singlets at δ_H 1.26, 1.21, 1.01, 0.96, 0.94, 0.91, and δ_H 0.88 (each 3H, s), and these signals corresponded to the carbon signals at δ_C 17.3, 26.5, 28.5, 17.4, 24.3, 33.7, and 16.1, respectively. Furthermore, one broad singlet at δ_H 5.17 (1H, br s) for an olefinic proton and two olefinic carbon signals at δ_C 144.9 and 123.1 were observed. These NMR resonances, together with the HMBC and NOESY correlations (**Figure 15**), suggested that **1** was an erythrodiol aglycone due to the olefinic linkage at C_{12}-C_{13} and the presence of the seven methyl groups of an oleanane triterpenoid backbone. A β-oriented hydroxy group at C_3 (δ_C 89.4) was determined from the HMBC cross peaks from H-23 (δ_H 1.26, s) and H-24 (δ_H 1.01, s) to C_3 (δ_C 89.4) and the NOESY correlation between H-3α (δ_H 3.31, dd, *J*=11.5, 4.3 Hz) and H-5α (δ_H 0.72, d, *J*=11.7 Hz). The assignment of the aglycone was confirmed by comparison with the NMR data of erythrodiol glycosides reported in the literature (Ito et al., 2004). The presence of four sugar residues was clearly indicated by the four anomeric resonances at [δ_C 107.2, δ_H 4.89 (1H, d, *J* = 7.5 Hz)], [δ_C 105.8, δ 4.83(1H, d, *J* = 7.5 Hz)], [δ_C 105.7, δ_H 5.12 (1H, d, *J* = 7.5 Hz)] and [105.7, δ_H 5.14 (1H, d, *J* = 7.5 Hz)].

The acid hydrolysis of **1** yielded a sugar that was identified as D-glucose by comparison with an authentic sugar standard. The coupling constants (all $J = 7.5$ Hz) of the anomeric protons suggested that these glucose moieties were in the β isomer form. Importantly, the shifts of the glycosylated carbons were observed at C_3 (δ_C 89.4) and C_{28} (δ_C 77.3), which supported the substitution of sugar chains at C-3 (GlcSA) and C-28 (GlcSB). These connections were confirmed by the HMBC cross peaks between H-1_{GlcSA1} [δ_H 4.84 (1H, d, $J = 7.5$ Hz)] and C-3 (δ_C 89.4) and between H-1_{GlcSB1} [δ_H 4.83 (1H, d, $J = 7.5$ Hz)] and C-28 (δ_C 77.3). Additional investigation of the saccharide portion by ^{13}C NMR spectroscopy (**Table 3**) revealed that only two of the four CH_2OH groups of the glucose residues were remarkably downfield shifted (δ_C 70.8, and 70.3), suggesting 1→6 linkages between the glucose units. Furthermore, the HMBC spectrum of 1 showed clear correlations between H-1_{GlcSA2} [δ_H 5.12 (1H, d, $J = 7.5$ Hz)] and C-6_{GlcSA1} (δ_C 70.8) and between H-1_{GlcSB2} [δ_H 5.14 (1H, *d*, $J = 7.5$ Hz)] and C-6_{GlcSB1} (δ_C 70.3). Taken together, compound **1** was elucidated to be 3-*O*-β-D-glucopyranosyl-(1→6)-*O*-β-D-glucopyranosyl erythrodiol 28-*O*-β-D-glucopyranosyl-(1→6)-*O*-β-D-glucopyranoside. This is the first reported erythrodiol analogue in *G. sylvestre* and thus suggests erythrodiol is a building block in this system and its glycosides would be constructed in a manner similar to oleanane glycosides (**Figure 10-B**).

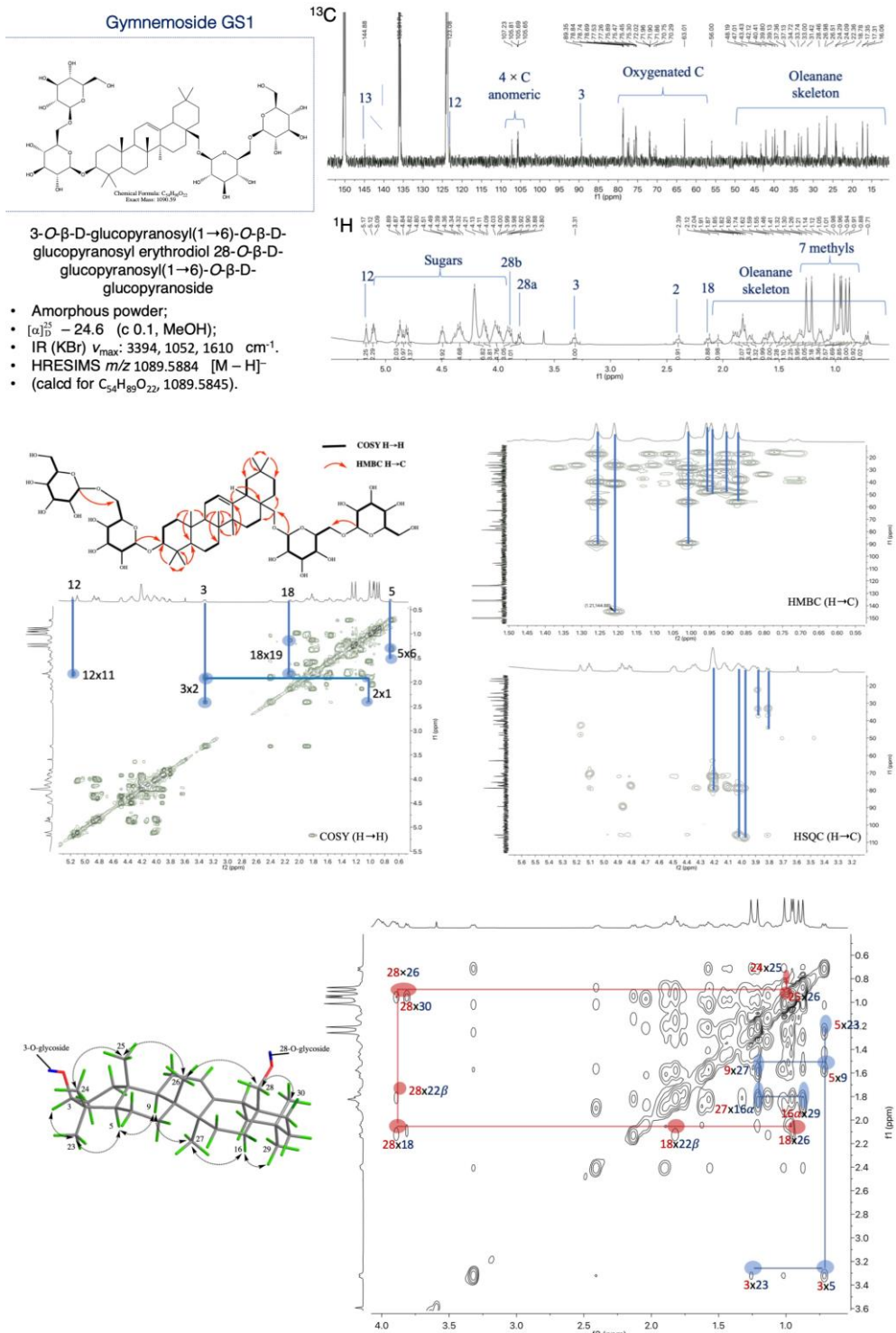


Figure 15. Gymnemoside GS1

*3.3.1. Gymnemoside GS2 - Compound **2***

Compound **2** (peak **21**) was obtained as an amorphous powder with $[\alpha]_D^{25}$ 26.3 (c 0.2, MeOH). Its molecular formula was determined to be $C_{59}H_{98}O_{26}$ based on its HRESIMS peak at *m/z* 1221.6298 $[M - H]^-$ (calcd for $C_{59}H_{97}O_{26}$, 1221.6268). Evidence from KIA, MFA, and the construction rule in **Figure 10B-a** suggested that the structure of **2** may be erythrodiol + 3-(glc^6-glc^6-xyl) + 28-(glc^6-glc). The ^{1}H and ^{13}C NMR spectroscopic data of **2** (Tables 3 and 4) were similar to those of **1**, except for the addition of one xylose unit based on the signals of an anomeric proton Xyl-SA3-1' (δ_H 4.92, d, J = 7.5 Hz) and 6 carbons at Xyl-SA3-1' (δ_C 106.3), Xyl-SA3-2' (δ_C 75.2), Xyl-SA3-3' (δ_C 78.4), Xyl-SA3-4' (δ_C 71.4), and Xyl-SA3-5' (δ_C 67.3). Furthermore, a downfield shift of Glc-SA2-6' (δ_C 70.0, +7.0 ppm) was observed relative to **1**, which supported the elongation of the sugar chain by continuing the 1→6 linkage pattern through the attachment of a xylopyranosyl unit to C-6 of Glc-SA2. Further HMBC analysis indicated a connection between Xyl-SA3-1' (δ_H 4.92, d, J = 7.5 Hz) and C-Glc-SA2-6' (δ_C 70.0). Accordingly, compound **2** was elucidated to be 3-*O*-β-D-xylopyranosyl-(1→6)-*O*-β-D-glucopyranosyl(1→6)-*O*-β-D-glucopyranosyl erythrodiol 28-*O*-β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside.

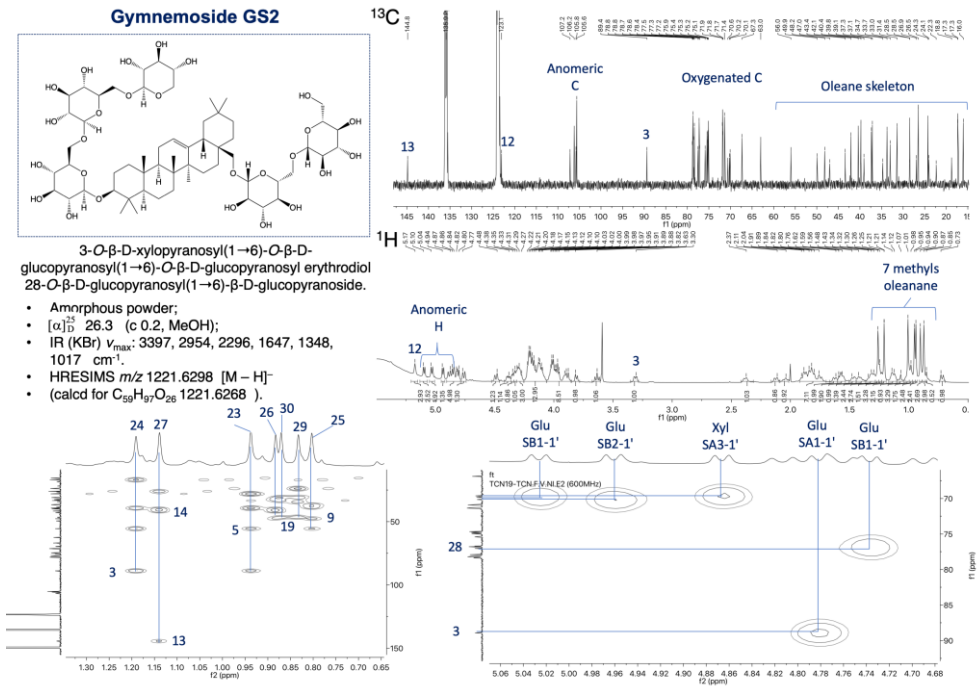


Figure 16. Gymnemoside GS2

*3.3.3. Gymnemoside GS3 - Compound **3***

Compound **3** (peak 65) was obtained as an amorphous powder with $[\alpha]_D^{25}$ 14.0 (c 0.2, MeOH). It possessed the molecular formula $C_{41}H_{62}O_{11}$, as determined from its negative ion HRESIMS peak $[M - H]^-$ at *m/z* 753.4235 (calcd for $C_{43}H_{61}O_{11}$, 753.4214). Its molecular formula and fragmentation were consistent with the structure of CS + (3-glcA) + bz formulated from the building block library. The ^{1}H NMR spectrum showed seven methyl singlets at δ_H 1.43, δ_H 1.30, δ_H 1.15, δ_H 1.10, δ_H 1.03, δ_H 0.97, and δ_H 0.80 (each 3H, s). The ^{13}C NMR spectrum showed signals for 42 carbons, including two carboxylic groups at δ_C 172.6 and 167.1, one anomeric carbon, 8 aromatic and olefinic carbons, 7 oxygenated carbons, and 24 methine or methylene carbons. The signals of a double bond at $C_{12\text{-}13}$ δ_C 123.3 and 143.4 and resonances of 4 oxygenated carbons were observed at δ_C 63.6, 66.6, 69.9, and 89.3,

which are similar to the characteristic peaks of chichipegenin.(K. Yoshikawa et al., 1999; 1994a) The shift in the resonance of C-3 due to glycosylation (δ_C 89.4, +10.9 ppm) relative to the same carbon in chichipegenin (K. Yoshikawa et al., 1994a) and the HMBC cross peak from GlcA-H1′ δ_H 4.97 (d, J = 7.7 Hz) to C-3 (δ_C 89.4) confirmed the structure and linkage of the sugar portion to C-3. The presence of a benzoyl group was evident by comparing the aromatic signals [bz-1 (δ_C167.1, C), bz-2 (δ_C131.5, C), bz-3,7 (δ_C130.0, CH), bz-4,6 (δ_C129.3, CH), and bz-5 (δ_C133.7, CH)] to those in the literature (Pham et al., 2018; Ye et al., 2001a) and the location of this group at C-28 on the aglycone was deduced from the HMBC correlation between H-28 (δ_H 5.41, d, J = 10.5 Hz; δ_H 4.93, d, J = 10.5 Hz) and bz-1 (δ_C 167.1). Based on these data, compound **3** was elucidated as 28-*O*-benzoyl-chichipegenin 3-*O*-β-D-glucopyranoside.

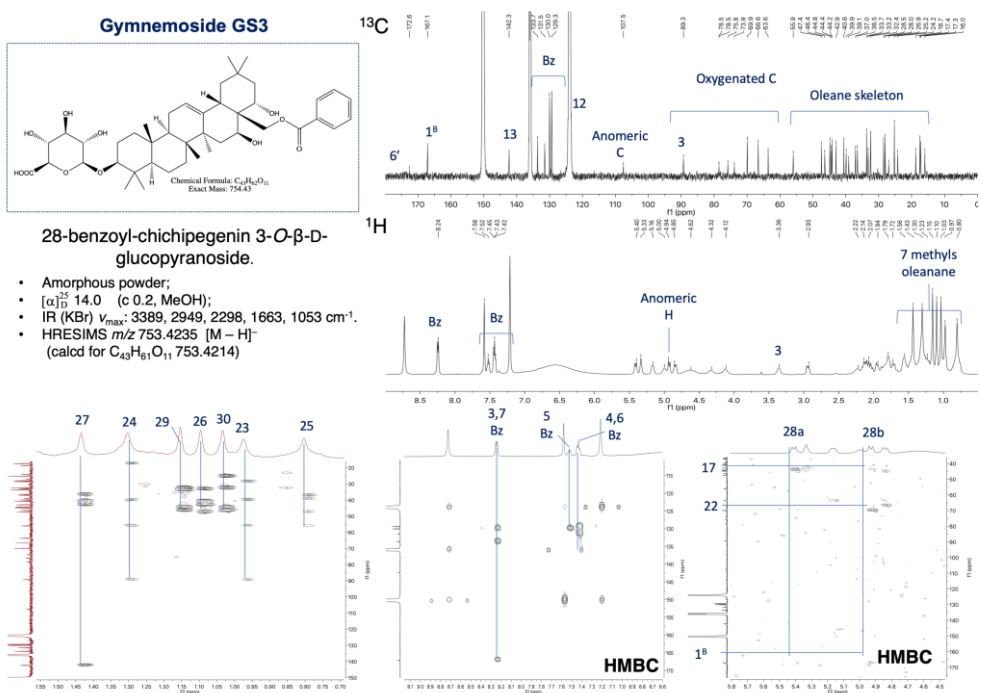


Figure 17. Gymnemoside GS3

*3.3.4. Gymnemoside GS4 - Compound **4***

Compound **4** (peak 73) was obtained as an amorphous powder with $[\alpha]_D^{25}$ 19.3 (*c* 0.1, MeOH), and its molecular formula of $C_{42}H_{61}NO_{11}$ was determined based on its negative HRESIMS ion peak *m/z* 754.4158 $[M - H]^-$ (calcd for $C_{42}H_{60}NO_{11}$, 754.4166). HPLC-MS experiments in positive mode suggested the existence of a glucuronic acid moiety (corresponding to a loss of 176 Da), a nitrogen-containing acyl moiety (107 Da) and 4 hydroxy groups on the triterpene skeleton (corresponding to 4 × 18 Da). Comparison of the ^1H and ^{13}C NMR data of **4** with those of **3** revealed that the two compounds were structurally similar in their aglycone and glycosyl portions, and the benzoyl group was replaced by a 3-nicotinoyloxy group, as indicated by the NMR data (nic-1 [δ_C 167.1], nic-2 (δ_C 151.1), nic-3 (δ_C 124.3), nic-4 (δ_C 137.3), nic-5 (δ_C 123.1) and nic-6 (δ_C 154.2)). The 3-nicotinoyloxy linkage at C-28 was established by the HMBC cross peaks from the carbonol protons H-28a (δ_H 5.40, d, J = 10.5 Hz) and H-28b (δ_H 4.90, d, J = 10.5 Hz) to nic-1 (δ_C 165.7). Therefore, compound **4** was identified as 28-*O*-(3-nicotinoyloxy)-chichipegenin-3-*O*-β-D-glucuropyranoside. This is the first time a compound bearing a nicotinoyl moiety has been found in *G. sylvestre*, and it can be incorporated as a building block of oleanane triterpenoids in this plant.

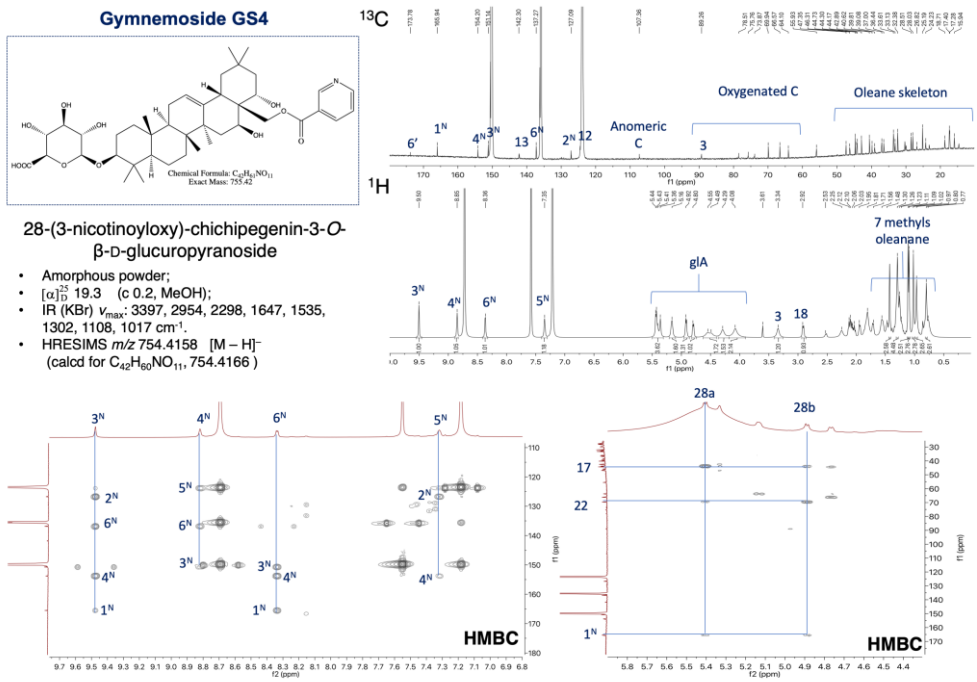


Figure 18. Gymnemoside GS4

*3.3.5. Gymnemoside GS5 – Compound **5***

Compound **5** (peak 60) was obtained as an amorphous powder with $[\alpha]_D^{25}$ 22.0 (c 0.2, MeOH). Its molecular formula ($C_{41}H_{64}O_{11}$) was determined from its quasimolecular ion peak at *m/z* 731.4382 $[M - H]^-$ (calcd for $C_{41}H_{63}O_{11}$, 731.4370) in HRESIMS. The ^1H, ^{13}C, and HSQC NMR spectroscopic data of **5** (**Tables 3 and 4**) were similar to those of gymnemoside ND6 with an additional tigloyl unit, which was indicated by the resonances of tig-1 (δ_C 167.9), tig-2 (δ_C 130.0), tig-3 (δ_C 137.1), tig-4 (δ_C 14.6), and tig-5 (δ_C 12.7) as well as the downfield shift of C-21 (δ_C 76.7) compared with the corresponding carbon in gymnemoside ND6 and the HMBC cross peak from H-21 (δ_H 5.51) to tig-1 (δ_C 167.9). Consequently, compound **5** was elucidated as 21-*O*-tigloyl-sitaskisogenin 3-*O*-β-D-glucuronopyranoside.

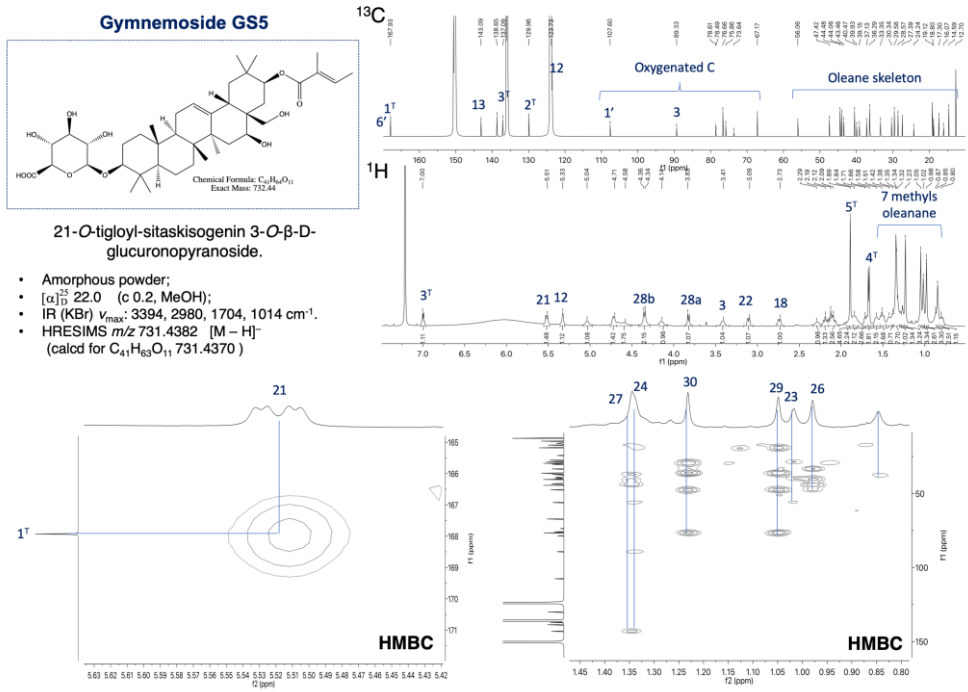


Figure 19. Gymnemoside GS5

*3.3.6. Gymnemoside GS6 - Compound **6***

Compound **6** (peak 32) was obtained as an amorphous powder with $[\alpha]_D^{25}$ 11.0 (c 0.2, MeOH), and it possessed a molecular formula of $C_{36}H_{58}O_8$ based on its HRESIMS peak at *m/z* 617.4054 $[M - H]^-$ (calcd for $C_{36}H_{57}O_8$, 617.4053). The ^1H and ^{13}C NMR data (**Table 3**) together with the HSQC and HMBC data suggested that 6 was structurally similar to gymnemoside ND4 (Pham et al., 2018) but lacked a glucose in the glycoside chain. Thus, compound **6** was elucidated as maniladiol-3-*O*-β-D-glucuronopyranoside. The elucidation of **6** and the detection of gymnemoside ND4 in the extract suggested that these compounds and maniladiol glycosides were constructed by **Figure 10B-b** (unlike erythrodiol glycosides which were constructed following **Figure 10B-a**, although both possess the same molecular weight).

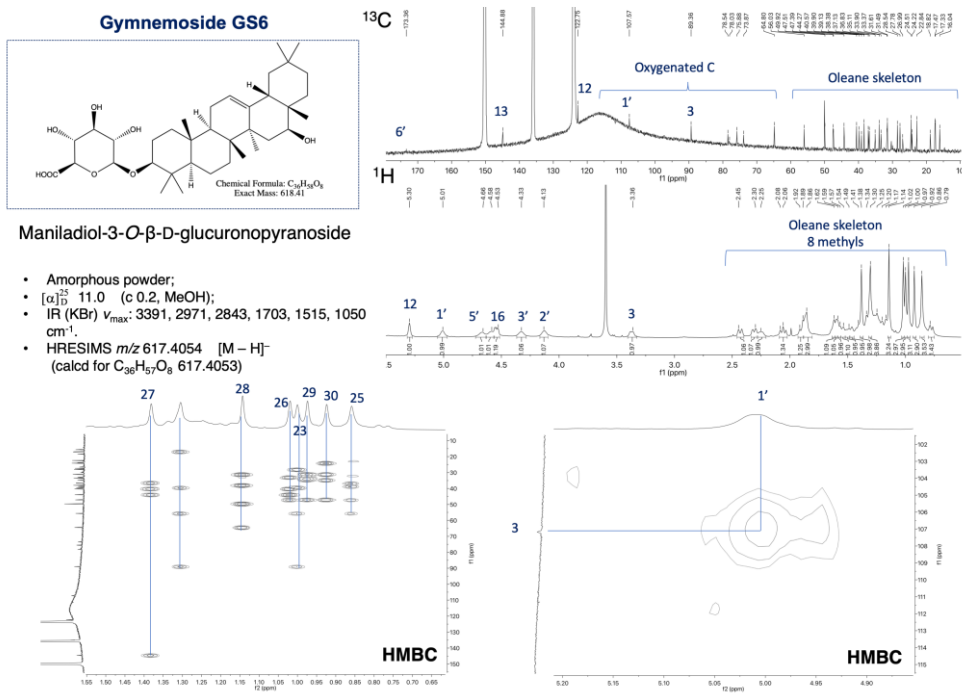


Figure 20. Gymnemoside GS6

3.4. Metabolite Profiling of Oleanane Triterpenoids

The shortcomings of conventional metabolite profiling by manual investigation of chromatographic spectra have been clearly recognized as unsystematic, untargeted, and time-consuming (Wolfender et al., 2015; Xie et al., 2012; Xing et al., 2015). Our new approach to metabolite assignment in this study provided a clear improvement on the current method. The remarkable advantage of the simple microfractionation technique using in this study was the ability to magnify signals of minor or trace compounds, led to the detection of 119 peaks while only 60 peaks could be found in a single UPLC-MS scan of the total extract.

A new concept of multilayer metabolite profiling – MMP was developed as an addition to the typical total ion chromatography (TIC) chemical profile (**Figure 21**). The advantage of MMP is that it not only displayed the retention time of

compounds in a straightforward chromatogram interface but also provided the information of the structural formation of metabolites. Another strength of MMP is its ability to underline considerable insight into the chemical relationship of oleanane saponins by categorization of all detected peaks into homogeneous structural sub-groups (layers). The relationship between detected peaks can be analyzed on each layer or comparing peaks of different layers.

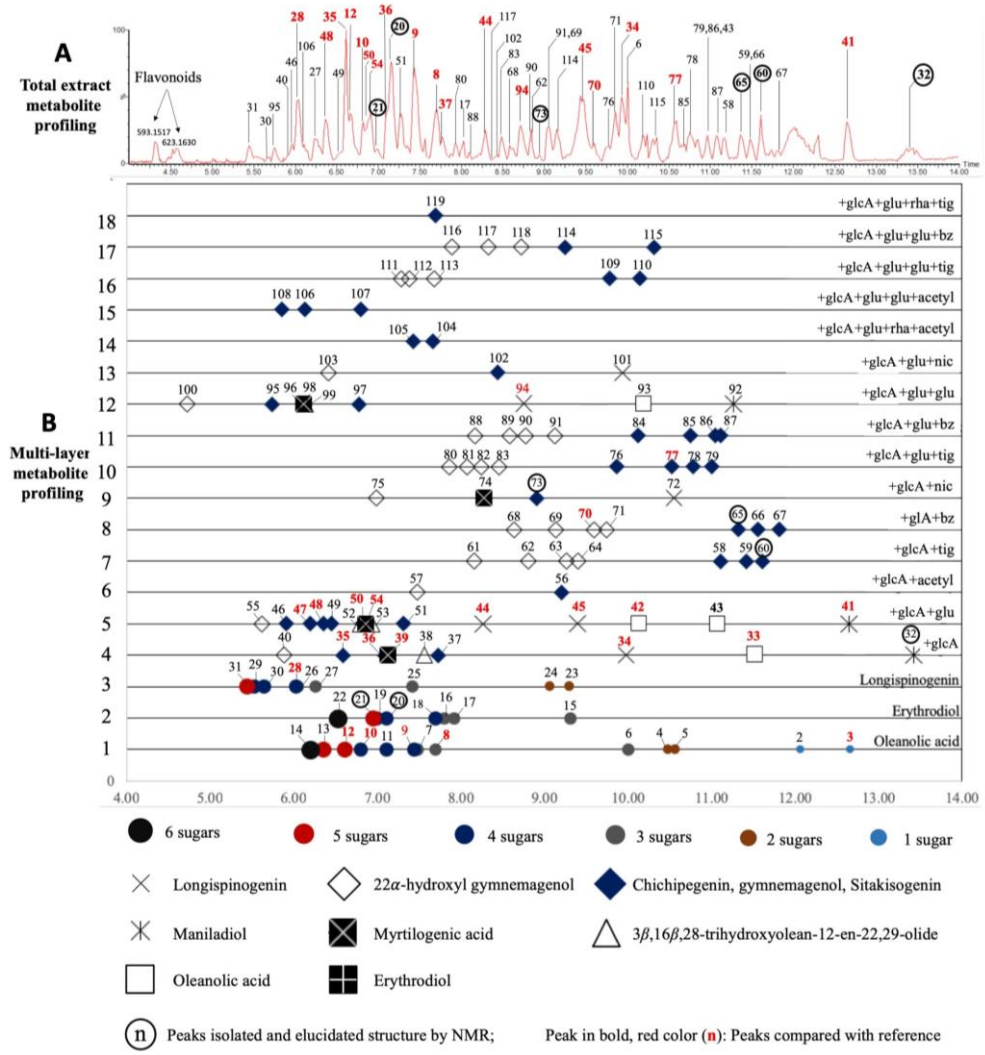


Figure 21. Metabolite profiling of oleanane triterpnenoids from *G. sylvestre*.

4. Conclusions

The quality control of a medicinal plant requires a rapid and accurate analysis of the comprehensive chemical profile of the active extract (Mohn et al., 2009). In this study, we described a new, reliable, and generally applicable approach to analyze the 70% EtOH extract of *G. sylvestre* by ultra-performance liquid chromatography (UPLC) with quadrupole-time-of-flight mass spectroscopy (UHPLC/qTOF-MS). This strategy is realized by integrating building blocks analysis with rapid HPLC microfractionation of extract, and multiple key ion analysis (mKIA), and mass fragmentation analysis (MFA) to allowing a systematic structure assignment and dereplication of the triterpenoids saponins of *G. sylvestre*. Spectra and mass data, which are obtained from the 24 standard reference compounds isolated from Vietnamese *G. sylvestre* variety available in Korea Bioactive Natural Material Bank (KBNMB), provided unambiguous evidence for the metabolite profiling. Targeted isolation for selective compounds was also implemented to confirm the structural annotations. Finally, all identified compounds were categorized into multiple structural layers to examine the biosynthesis relationship of different scaffolds within the chemical diversity of crude extracts.

Part 2. Oleanane triterpenoids from *Gymnema latifolium* and their PTP1B inhibitory activities

1. Introduction

Type 2 diabetes mellitus (T2DM) has emerged as a global health issue which correlates closely to the widespread occurrence of obesity. The International Diabetes Federation estimated that there were 382 million patients of T2DM aged 20–70 years worldwide in 2013, and the number was anticipated to surge up to 592 million by 2035 (DeFronzo et al., 2015). T2DM is defined as a complex metabolic disorder characterized by high blood glucose levels caused by insulin deficiency, insulin resistance or both (DeFronzo et al., 2015). Poor glycaemic control may lead to a variety of diabetes complications, including cardiovascular disease, diabetic nephropathy, diabetic neuropathy, diabetic retinopathy, stroke, and lower limb amputations (Hossain et al., 2007; Marcovecchio et al., 2019; J. L. Yang et al., 2017).

Protein tyrosine phosphatases (PTPs) are considered as drug targets to treat type 2 diabetes and obesity by their role in regulating the tyrosine phosphorylation of target proteins (Fischer et al., 1991; He et al., 2014). PTP1B, a member of the PTP superfamily, has been proven to be a promising therapeutic target in the treatment of T2DM by regulating insulin and leptin signaling pathways (Zabolotny et al., 2002). PTP1B gene deletion also protected mice against the development of obesity, diabetes, and cardiovascular dysfunction (Cho, 2013; Coquerel et al., 2014).

Natural products are promising sources of lead compounds that play a significant role in the discovery of new antidiabetic agents via the mechanism of PTP1B inhibition. Approximately 300 new or known natural products from various resources were reported as potential PTP1B inhibitors, and many candidates

exhibited promising *in vitro*, *in vivo* activities and selectivity profiles (Jiang et al., 2012). Among them, naturally occurring triterpenes isolated from *Astibilbe koreana*, *Symplococos paniculata*, *Gynostemma pentaphyllum* displayed remarkable PTP1B inhibition activities (Hung et al., 2009; Na et al., 2006a; 2006b). These inhibitors, therefore, could be considered for further research and development as clinical drug candidates for treating diabetes, obesity and related metabolic syndromes (Jiang et al., 2012).

In our ongoing research to find PTP1B inhibitors from natural products, hundreds of plant extracts available in Korea Bioactive Natural Material Bank have been screened against this biological target. An extract of the *Gymnema latifolium* Wall. ex Wight showed considerable PTP1B inhibitory activity (>70% inhibition at 30 μg/mL). The result was in agreement with the use of this plant in Vietnam as an antidiabetic herbal medicine. Some studies have recommended that 3β-hydroxy oleanane derivatives were potential PTP1B inhibitor (Qian et al., 2010; Y.-N. Zhang et al., 2008). Various varieties of *Gymnema sylvestre*, a species of the same genus *Gymnema*, have been studied intensively on their extracts and their oleanane triterpenoid composition for the hypoglycaemic effects (Leach, 2007; Pham et al., 2018; Tiwari et al., 2017). It is also well-known that most species belonging to the same genus possess similar chemical composition and thus exhibit similar biological activities (Jürgens and Dötterl, 2004). As expected, in this study, we isolated 14 new and 6 known oleanane analogs from *G. latifolium* and evaluated their PTP1B inhibitory activities.

Purpose of research

- Authenticate the scientific name of the medicinal plant
- Identify the the chemical composition of *Gymnema latifolium*.
- Evaluate the PTP1B inhibitory effect of isolates from *G*. *latifolium*

2. Materials and methods

2.1. Plant materials

Plant materials were cultivated in an herb farm in the Thai Nguyen province of Vietnam (GPS 21°52'47.5"N 105°44'35.7"E) and collected in August 2017. A voucher specimen was deposited in the Medicinal Herbarium of Hanoi University of Pharmacy with the accession number HNIP/18068.

2.2. Extraction and isolation schemes

2.2.1. General experimental procedures

Optical rotations were measured on a JASCO P-2000 polarimeter using a 1-cm cell (JASCO International Co. Ltd., Tokyo, Japan). IR data were recorded on a Nicolet 6700 FT-IR spectrometer (Thermo Electron Corp., Waltham, MA, USA). NMR data were analyzed using an AVANCE 500 MHz spectrometer (Bruker, Billerica, MA USA) or a JNM-ECA 600 MHz spectrometer (JEOL Ltd., Tokyo, Japan). HRESIMS were analyzed using an Agilent Technologies 6130 Quadrupole LC/MS spectrometer equipped with an Agilent Technologies 1260 Infinity LC system (Agilent Technologies, Inc., Santa Clara, CA, USA) and an INNO C_{18} column (4.6 × 150 mm, 5 μm particle size, 12 nm, J.K. Shah & Company, Korea). Silica gel 60 F_{254} and RP-18 TLC plates and deuterated pyridine for NMR analysis were purchased from Merck (Darmstadt, Germany). Sephadex LH-20 from Sigma-Aldrich (St. Louis, MO, USA) was used for column chromatography (CC). A Gilson HPLC semi-preparative purification system, equipped with an Optima Pak C_{18} column (10 × 250 mm, 10 μm particle size; RS Tech, Seoul, Korea), was used at a flow rate of 2 mL/min and UV detection at 205 or 254 nm. Extra-pure grade solvents

for extraction, fractionation, and isolation were purchased from Dae Jung Pure Chemical Engineering Co. Ltd. (Siheung, Korea).

2.2.2. Isolation scheme

The aerial parts of *G. latifolium* (5.0 kg) were powdered and extracted with 70% EtOH (3 times × 10 L, for 3 h each) with ultrasonication, and the extract was concentrated in vacuo. The crude extract obtained (900 g) was suspended in water, absorbed onto Sephabeads SP70 resin and washed with water, 50% EtOH, 100% EtOH, and acetone, in a sequential elution process. The 100% EtOH fraction (150 g) GL.100 was subjected to silica gel column chromatography (15 × 45 cm; 63–200 μm particle size) using *n*-hexane/EtOAc (gradient from 10:1 to 0:1) and then EtOAc/MeOH (gradient from 10:1 to 0:1) to give 6 fractions (N1-N6) based on the thin-layer chromatography profile. Fraction N3 was chromatographed over C_{18}-reversed phase silica gel (RP-C_{18}) column chromatography, eluted with CH_3OH/H_2O (2:3 to 4:1) to obtain 12 fractions N3.M1-M12.

Fraction N3.M7 (800 mg) was applied to a Sephadex LH-20, eluted with CH_3OH/H_2O (7:10) to yield 7 subfractions N3.M7.L1−L7. Subfraction N3.M7.L3 (258 mg) was subjected to semi-preparative reversed-phase HPLC on an Optima Pak C_{18} column, CH_3OH/H_2O (4:6 to 7:10), flow rate 2 mL/min, to yield compound **15** (24.7 mg) and **17** (12.6 mg).

Fraction N3.M8 (850 mg) was applied to a sequential separation by Sephadex LH-20 (MeOH 70%) to yield 3 subfractions N3.M8.L1−L3. Subfraction N3.M8.L3 (520 mg) over C_{18}-reversed phase silica gel (RP-C_{18}) column chromatography, eluted with CH_3CN/H_2O (3:7 to 1:1) to obtain 5 fractions N3.M8.L3.A1−A5. Compounds **20** (14.6 mg) were obtained by semi-preparative

HPLC using a CH_3CN/H_2O (3:7) solvent system from fraction N3.M8.L3.A1 (90 mg). Compounds **18** (22.5 mg) and **19** (15.5 mg) were purified from fraction N3.M8.L3.A2 (150 mg) by semi-preparative HPLC using CH_3CN/H_2O (4:6), flow rate 2 mL/min. Subfraction N3.M8.L3.A3 (200 mg) was applied to a HPLC (Optima Pak C_{18}, CH_3CN/H_2O (gradient 2:8 to 4:6, flow rate 2 mL/min) to produce compounds **12** (16.5 mg) and **13** (9.5 mg). Subfraction N3.M8.L3.A4 (80 mg) was subjected to a HPLC (Optima Pak C_{18}, CH_3CN/H_2O (isocratic 4:6, flow rate 2 mL/min) to produce compounds **14** (8.2 mg) and **16** (10.2 mg).

Fraction N3.M9 (320 mg) was applied to a Sephadex LH-20, eluted with CH_3OH/H_2O (7:10) to yield 5 subfractions N3.M9.L1-L5. Compounds **5** (19.1 mg) were obtained by semi-preparative HPLC using a CH_3CN/H_2O (4:6) solvent system from fraction N3.M9.L3 (200 mg).

Fraction N3.M10 (500 mg) was applied to a sequential separation by Sephadex LH-20 (MeOH 70%) and HPLC (Optima Pak C_{18}, CH_3CN/H_2O (gradient 3:7 to 5:5, flow rate 2 mL/min) to produce compounds **6** (22.5 mg) and **7** (15.5 mg). Compound **8** (5.0 mg) were purified from fraction N3.M10.L2 (150 mg) by semi-preparative HPLC using CH_3CN/H_2O (9:11), flow rate 2 mL/min.

Fraction N3.M12 (500 mg) was applied to a Sephadex LH-20, eluted with CH_3OH/H_2O (7:10) to yield 7 subfractions N3.M12.L1-L7. Subfraction N3.M12.L2 (180mg) was subjected to semi-preparative reversed-phase HPLC on an Optima Pak C_{18} column, CH_3CN/H_2O (6:4), flow rate 2 mL/min, to yield compound **1** (21.4 mg) and **2** (15.5 mg). Subfraction N3.M12.L3 (120 mg) was purified by reversed-phase HPLC on an Optima Pak C_{18} column, using sequential separation by CH_3OH/H_2O

(3:1), flow rate 2 mL/min, to yield compound **3** (12.5 mg) and **4** (13.2 mg). Subfraction N3.M12.L4 (80 mg) was separated by another semi-preparative HPLC using CH_3CN/H_2O (11:9) to yield compound **9** (18.4 mg).

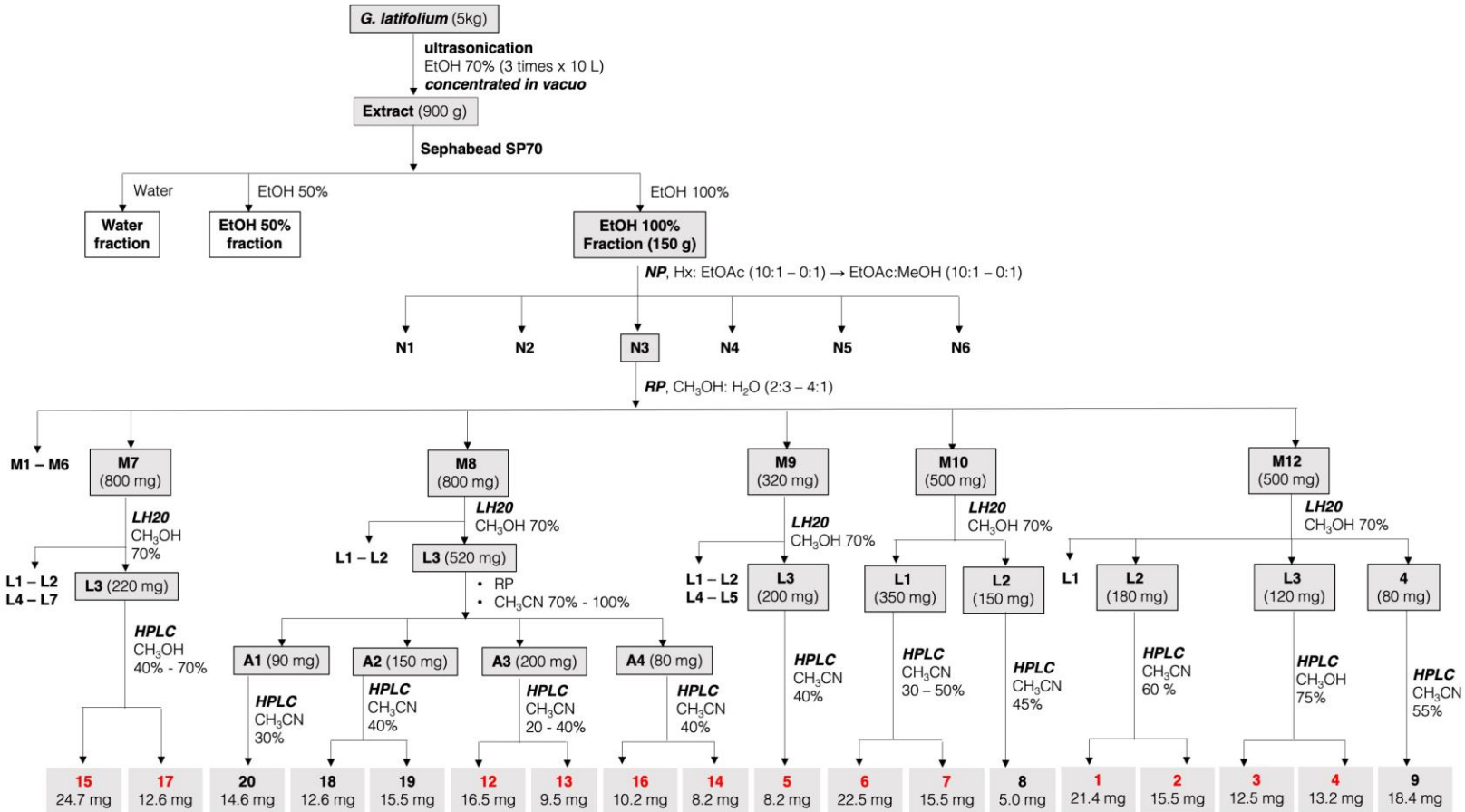


Figure 22. Isolation scheme of compounds **1** – **20** from *G. latifolium*

2.2.3. Physical and chemical characteristics of isolated compounds

Gymlatinoside GL1 (**1**): White amorphous powder, $[\alpha]_D^{25}$ +15.2 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3389, 2948, 2898, 1711, 1641, 1266, 1008 cm^{-1}; ^1H and ^{13}C NMR data, Tables 5 and 6; HRESIMS *m/z* 967.4896 $[M - H]^-$ (calcd for $C_{49}H_{75}O_{19}$, 967.4903).

Gymlatinoside GL2 (2): White amorphous powder, $[\alpha]_D^{25}$ +9.6 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3389, 2948, 1711, 1646, 1441, 1389, 1266, 1140 cm^{-1}; ^1H and ^{13}C NMR data, Tables 5 and 6; HRESIMS *m/z* 965.4752 $[M - H]^-$ (calcd for $C_{49}H_{73}O_{19}$, 965.4746).

Gymlatinoside GL3 (**3**): White amorphous powder, $[\alpha]_D^{25}$ +10.1 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3384, 2893, 2307, 1716, 1275, 1095 cm^{-1}; ^1H and ^{13}C NMR data, Tables 5 and 6; HRESIMS *m/z* 987.4581 $[M - H]^-$ (calcd for $C_{51}H_{71}O_{19}$, 987.4590).

Gymlatinoside GL4 (**4**): White amorphous powder, $[\alpha]_D^{25}$ +22.1 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3429, 2948, 1726, 1080, 1040 cm^{-1}; ^1H and ^{13}C NMR data, Tables 5 and 6; HRESIMS *m/z* 1007.4826 $[M - H]^-$ (calcd for $C_{51}H_{75}O_{20}$, 1007.4852).

Gymlatinoside GL5 (**5**): White amorphous powder, $[\alpha]_D^{25}$ +13.4 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3399, 2943, 1075, 1035 cm^{-1}; ^1H and ^{13}C NMR data, Tables 5 and 6; HRESIMS *m/z* 781.4746 $[M - H]^-$ (calcd for $C_{42}H_{69}O_{13}$, 781.4738).

Gymlatinoside GL6 (**6**): White amorphous powder, $[\alpha]_D^{25}$ +5.4 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3454, 2953, 1736, 1250, 1030 cm^{-1}; ^1H and ^{13}C NMR data, Tables 5 and 6; HRESIMS *m*/*z* 791.4225 $[M - H]^-$ (calcd for $C_{42}H_{63}O_{14}$, 791.4218).

Gymlatinoside GL7 (**7**): White amorphous powder, $[\alpha]_D^{25}$ +12.4 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3386, 2957, 2891, 1743 cm^{-1}; ^1H and ^{13}C NMR data, Tables 5 and 6; HRESIMS *m*/*z* 847.4508 $[M - H]^-$ (calcd for $C_{45}H_{67}O_{15}$, 847.4480).

Gymlatinoside GL8 (**12**): White amorphous powder, $[\alpha]_D^{25}$ +35.4 (*c* 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 325 nm (3.2); IR (KBr) ν_{max} 3364, 2957, 2891, 1743 cm^{-1}; ^1H and ^{13}C NMR data, Tables 7 and 8; HRESIMS *m*/*z* 1121.5549 $[M - H]^-$ (calcd for $C_{57}H_{85}O_{22}$, 1121.5532).

Gymlatinoside GL9 (**13**): White amorphous powder, $[\alpha]_D^{25}$ +24.3 (*c* 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 325 nm (3.5); IR (KBr) ν_{max} 3361, 2957, 2881, 1733 cm^{-1}; ^1H and ^{13}C NMR data, Tables 7 and 8; HRESIMS *m*/*z* 1151.5658 $[M - H]^-$ (calcd for $C_{58}H_{87}O_{23}$, 1151.5638).

Gymlatinoside GL10 (**14**): White amorphous powder, $[\alpha]_D^{25}$ +10.1 (*c* 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 325 nm (3.1); IR (KBr) ν_{max} 3365, 2955, 2884, 1733, 1232 cm^{-1}; ^1H and ^{13}C NMR data, Tables 7 and 8; HRESIMS *m*/*z* 827.4585 $[M - H]^-$ (calcd for $C_{46}H_{67}O_{13}$, 827.4582).

Gymlatinoside GL11 (**15**): White amorphous powder, $[\alpha]_D^{25}$ +26.3 (*c* 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 325 nm (3.4); IR (KBr) ν_{max} 3429, 2948, 1726,

1080, 1040 cm^{-1}; ^{1}H and ^{13}C NMR data, Tables 7 and 8; HRESIMS *m*/*z* 973.5167 $[M - H]^-$ (calcd for $C_{52}H_{77}O_{17}$, 973.5161).

Gymlatinoside GL12 (**16**): White amorphous powder, $[\alpha]_D^{25}$ +15.2 (*c* 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 325 nm (3.5); IR (KBr) ν_{max} 3455, 2948, 1737, 1252, 1030 cm^{-1}; ^{1}H and ^{13}C NMR data, Tables 7 and 8; HRESIMS *m*/*z* 1149.5629 $[M - H]^-$ (calcd for $C_{52}H_{77}O_{17}$, 973.5161).

Gymlatinoside GL13 (**17**): White amorphous powder, $[\alpha]_D^{25}$ +25.9 (*c* 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 325 nm (3.5); IR (KBr) ν_{max} 3366, 2957, 2884, 1733, 1232, 1030 cm^{-1}; ^{1}H and ^{13}C NMR data, Tables 7 and 8; HRESIMS *m*/*z* 1267.6147 $[M - H]^-$ (calcd for $C_{63}H_{95}O_{26}$, 1267.6147).

Gymlatinoside GL14 (**18**): White amorphous powder, $[\alpha]_D^{25}$ +12.5 (*c* 0.2, MeOH); IR (KBr) ν_{max} 3454, 2953, 1250, 1030 cm^{-1}; ^{1}H and ^{13}C NMR data, Tables 7 and 8**;** HRESIMS *m*/*z* 1091.5645 $[M - H]^-$ (calcd for $C_{53}H_{87}O_{23}$, 1091.5638).

2.3. Acid Hydrolysis

The total extract (100 mg) was hydrolyzed with 2.0 N HCl (70% MeOH, 10 mL) at 90 °C for 1 h. The solution was neutralized with 10% NaOH, dried, suspended in H_2O and partitioned with EtOAc. The residual H_2O layer was concentrated and dissolved in pyridine (1.0 mL), and 5.0 mg of L-cysteine methyl ester hydrochloride was added. The mixture was kept for 1 h at 60 °C, and then 4.4 μL of phenylisothiocyanate (Sigma, St. Louis, MO, USA) was added. This solution was filtered through a 0.2-μm Whatman hydrophilic membrane filter into an HPLC sample vial and immediately analyzed by LC-MS. The analysis was performed using an Agilent 1200 HPLC system (Agilent Technologies, Palo Alto, CA, USA) with

INNO C18 (4.6 × 250 mm inner diameter, 5 μm particle size; Young Jin Bio Chrom Co., Ltd), and the column temperature was 30 °C. The chromatographic separations were carried out using a mobile phase of 27% CH_3CN with isocratic elution at a flow rate of 0.6 mg/mL over 60 min. The sugar derivatives showed retention times of 10.9, 14.9 min, which were identical to the derivatives prepared with authentic D-glucuronic acid, D-glucose, respectively. Similar procedures also applied to identified the sugar portions of compound **1**, **5**, **6**, **7**, **12** .

2.4. PTP1B Assay

PTP1B (human, recombinant) was purchased from BIOMOL International LP (Plymouth Meeting, PA). The enzyme activity was measured using p-nitrophenyl phosphate (pNPP), as described previously. To each of 96 wells in a microtiter plate (final volume: 100 μL) was added 2 mM pNPP and PTP1B (0.05-0.1 μg) in a buffer containing 50 mM citrate (pH 6.0), 0.1 M NaCl, 1 mM EDTA, and 1 mM dithiothreitol (DTT), with or without test compounds. Following incubation at 37 °C for 30 min, the reaction was terminated with 10 M NaOH. The amount of produced p-nitrophenol was estimated by measuring the absorbance at 405 nm. The nonenzymatic hydrolysis of 2 mM pNPP was corrected by measuring the increase in absorbance at 405 nm obtained in the absence of PTP1B enzyme.

For the enzyme kinetic determination, Lineweaver-Burk was used to determine the inhibition type of PTP1B. On the basis of Lineweaver-Burk double reciprocal plots, the PTP1B inhibition mode was obtained at various concentrations of p-NPP substrate (1, 2, 4, and 8 mM) without or with different test compound concentrations (10, 20, 30, and 40 μM). The 50% inhibition concentration (IC50) and the inhibition

type were evaluated using the Sigma Plot Statistical Analysis software (11.0 software, SPCC Inc., Chicago, IL, USA). Data were represented as a mean of triplicate assays ± SD.

3. Results and discussion

3.1. Authentication of Gymnema latifolium Wall ex. Wight

Lianas are up to 6 m high. Stems corky lenticellate, old stems basally wing-like corky. The whole plant contains bright, yellow latex. Young branchlets yellowish green, densely pubescent, stalks densely hairy. Leaves opposite, broadly ovate, 8–13 cm × 5–8 cm, apex acute, base truncate to rounded, margin entire, orange-yellow pubescent abaxially, 5–7 lateral veins per side; petiole 1.5–4 cm long, densely pubescent. Inflorescences paired at node, multi-flowered, umbrella-shape cymes, pubescent. Peduncle, 3–8 mm long. Flowers yellow, 3 mm × 3 mm. Sepals ovate, puberulous outside. Corolla yellowish, campanulate, outside glabrous, densely pubescent inside. Gynostegium concealed in corolla, cylindrical with slightly swollen base, laterally with ten oval nectary patches; anther appendages shorter than stigma head. Pollinia oblong, erect, top enlargement; ovary densely pubescent, stigma conical divided into two. Follicles are in pair or solitary, lanceolate cylindrical, beaked, 7–10cm long, 0.3–0.6 cm in diameter, apex acuminate, base dilated, densely pubescent. Seeds many, oblong-lanceolate, winged with a thin edge; coma 3–4 cm long, white (**Figure 23-A**).

The morphology is closely matched the description of *Gymnema latifolium* by Flora of China (Wu and Raven, 1995) with a slight difference on the fruit shape (7-10 cm x 0,3-0.6 cm) compared with 4.5-5.5 × 1.5-2 cm. Also, the special characteristic of the wing-like corky can be found in the description of *Gymnema khadalense* by Subhash (Deokule et al., 2013) and *Gymnema kollimalayanum* by Ramachandran (Ramachandran and Viswanathan, 2009). However, two species

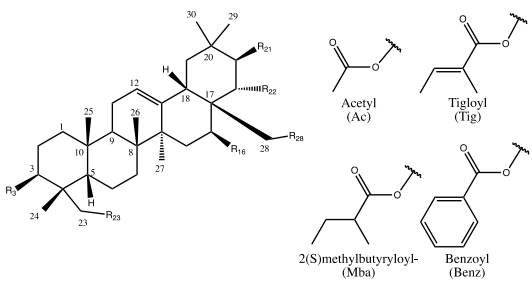
Gymnema khandalense and *Gymnema kollimalayanum* were considered as synonyms of *Gymnema latifolium* by Meve and Alejandro (Meve and Alejandro, 2011). In this study, by comparing the morphological characteristics with the specimens of syntype K000872841 (**Figure 23-B**) and lectotype specimen K000872839 (**Figure 23-C**) of *Gymnema latifolium* stored in the herbarium at the Royal Botanic Gardens Kew and the description by Meve and Alejandro, we identified the sample in Vietnam as *Gymnema latifolium* Wall. ex Wight. A DNA sequence of the ITS1-5.8S-ITS2 internal transcribed spacer of the sample was also deposited at GeneBank (National Institutes of Health) with the accession number KP163979 as a genetic reference for other studies in the future.

A. B. C.

Figure 23. A. Morphological characteristics of the studied *Gymnema latifolium* Wall. ex. Wight; B. Lectotype specimen of *Gymnema latifolium* Wall. ex. Wight (K000872839); C. Lectotype specimen of *Gymnema latifolium* Wall. ex. Wight (K000872839)

3.2. Oleanane triterpenoids isolated from G.latifolium

A 70% EtOH extract of *G. latifolium* was subjected to various chromatographic columns and then to purification by preparative high-performance liquid chromatography (HPLC) to afford 14 new oleanane triterpenoid glycosides, Gymlatinosides GL1-GL14 (1-7, 12-18) and 6 known compounds Gymnemic acid IX (8) (Liu et al., 1992), Gymnemic acid IV (**9**) (K. Yoshikawa et al., 1989), Gymnemasaponin I (K. Yoshikawa et al., 1991)(**10**), and Gymnemagenin (**11**) (Liu et al., 1992) , Gymnemoside-d and Gymnemosid-e (K. Yoshikawa et al., 1991). The structures of new compounds were elucidated by 2D-NMR, while known compounds were determined using 1D NMR analysis and comparing the physical and spectroscopic data with those in the literature.



Compound	R_3	R_{16}	R_{21}	R_{22}	R_{23}	R_{28}
1	-Ahg	-OH	-Mba	-H	-OH	-Ac
2	-Ahg	-OH	-Tig	-H	-OH	-Ac
3	-Ahg	-OH	-Benz	-H	- OH	-Ac
4	-Ahg	-Ac	-Tig	-Ac	-OH	-OH
5	-Glu	-OH	-H	-H	-H	-glu
6	-GlA	-Ac	-Ac	-Ac	-H	-OH
7	-GlA	-Ac	-Tig	-Ac	-OH	-OH
8	-Ahg	-OH	-Tig	-OH	-OH	-OH
9	GlA	-OH	-Tig	-OH	-OH	-OH
10	-OH	-OH	-OH	-OH	-OH	-glu
11	-OH	-OH	-OH	-OH	-OH	-OH
12	-OH	-OH	-OH	-H	-FerG^6-Glu^6-Xyl	-OH
13	-OH	-OH	-OH	-H	-SynG^6-Glu^6-Xyl	-OH
14	-OH	-OH	-OH	-H	-FerG	-OH
15	-OH	-OH	-H	-H	-FerG	-Glu
16	-OH	-OH	-H	-H	-FerG	-FerG
17	-OH	-OH	- H	-H	-FerG^6-Glu^6-Xyl	-Glu
18	-OH	-OH	-H	-H	-Glu^6-Glu^6-Xyl	-Glu
19	-OH	-OH	-OH	-H	-Glu^6-Glu^6-Xyl	-OH
20	-OH	-OH	-H	-H	-Glu^6-Glu^6-Xyl	-Glu^6-Glu^6

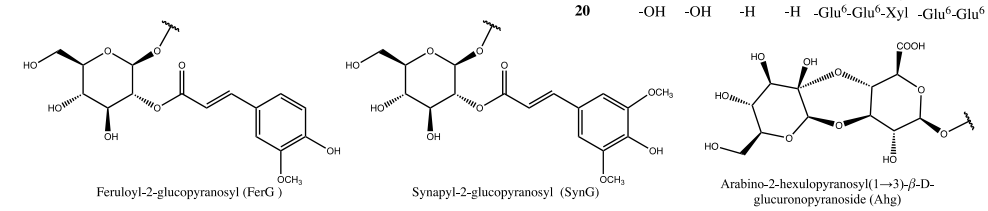


Figure 24. Isolated oleanane triterpenoids from *G. latifolium*

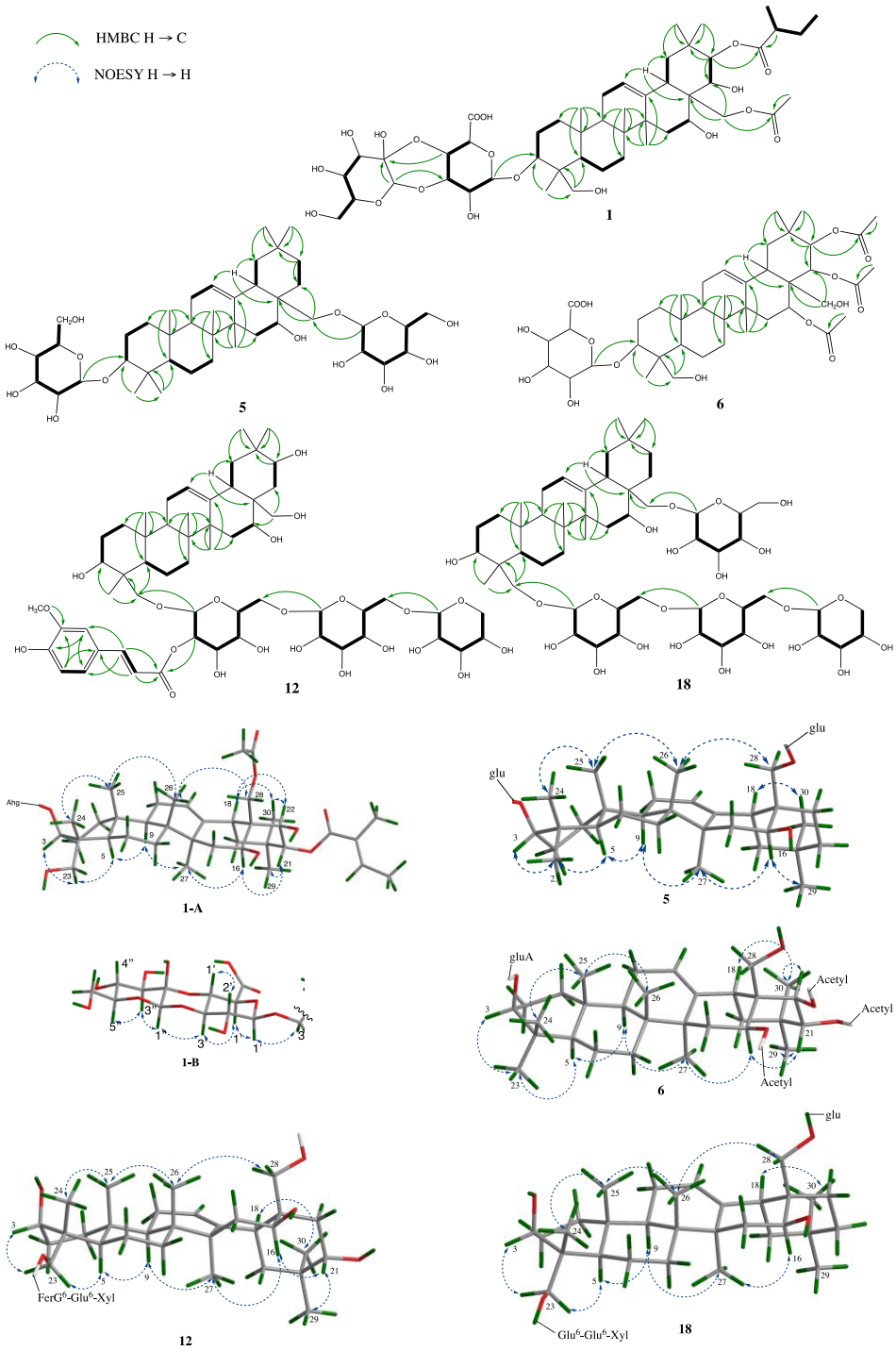


Figure 25. Key HMBC (H→C), COSY (H→H) and NOESY (H→H) of compound **1**, **5**, **6**, **12**, **18** isolated from *G. latifolium*

Table 5. ^{13}C NMR spectroscopic data for new compounds Gymlatinosides GL**1–7**

(in Pyridine-d_5)

	1	**2**	**3**	**4**	**5**	**6**	**7**
position	δ_C, type 125 Mhz	δ_C, type 125 Mhz	δ_C, type 125 Mhz	δ_C, type 150 Mhz	δ_C, type 125 Mhz	δ_C, type 125 Mhz	δ_C, type 150 Mhz
1	39.0, CH_2	39.0, CH_2	39.1, CH_2	39.0, CH_2	39.2, CH_2	39.0, CH_2	39.1, CH_2
2	26.4, CH_2	26.4, CH_2	26.4, CH_2	26.4, CH_2	27.0, CH_2	26.9, CH_2	26.4, CH_2
3	82.1, CH	82.2, CH	82.2, CH	82.2, CH	89.2, CH	89.1, CH	82.2, CH
4	43.8, C	43.9, C	43.9, C	43.9, C	39.9, C	39.8, C	43.9, C
5	47.2, CH	47.3, CH	47.5, CH	47.3, CH	56.1, CH	55.8, CH	47.7, CH
6	18.3, CH_2	18.4, CH_2	18.4, CH_2	18.3, CH_2	18.8, CH_2	18.6, CH_2	18.3, CH_2
7	32.8, CH_2	32.8, CH_2	32.9, CH_2	32.8, CH_2	33.1, CH_2	33.0, CH_2	32.8, CH_2
8	40.5, C	40.6, C	40.6, C	40.6, C	40.5, C	40.5, C	40.6, C
9	47.6, CH	47.6, CH	47.7, CH	47.6, CH	47.3, CH	47.2, CH	47.9, CH
10	36.7, C	36.7, C	36.7, C	37.1, C	37.3, C	36.9, C	36.9, C
11	24.2, CH_2	24.3, CH_2	24.3, CH_2	24.2, CH_2	24.2, CH_2	24.1, CH_2	24.2, CH_2
12	124.9, CH	124.9, CH	125.0, CH	125.0, CH	123.3, CH	125.0, CH	125.0, CH
13	141.5, C	141.6, C	141.6, C	140.9, C	143.8, C	140.8, C	140.9, C
14	43.0, C	43.0, C	43.0, C	43.2, C	44.3, C	43.1, C	43.2, C
15	36.8, CH_2	36.9, CH_2	37.0, CH_2	33.9, CH_2	37.3, CH_2	33.8, CH_2	33.9, CH_2
16	67.8, CH	67.8, CH	67.9, CH	68.7, CH	66.5, CH	68.6, CH	68.7, CH
17	46.0, C	46.0, C	46.0, C	47.9, C	41.7, C	47.8, C	47.4, C
18	42.8, CH	42.8, CH	42.9, CH	42.5, CH	45.1, CH	42.4, CH	42.5, CH
19	46.0, CH_2	46.1, CH_2	46.2, CH_2	46.1, CH_2	47.1, CH_2	46.0, CH_2	46.1, CH_2
20	36.9, C	37.0, C	37.0, C	37.0, C	37.0, C	36.8, C	37.1, C
21	78.6, CH	79.2, CH	80.3, CH	76.7, CH	34.5, CH_2	76.7, CH	76.7, CH
22	71.7, CH	71.9, CH	71.9, CH	72.0, CH	27.5, CH_2	71.9, CH	72.0, CH
23	64.5, CH_2	64.5, CH_2	64.5, CH_2	64.5, CH_2	28.6, CH_2	28.4, CH_2	64.7, CH_2
24	13.9, CH_3	13.9, CH_3	14.0, CH_3	14.0, CH_3	17.4, CH_3	17.2, CH_3	14.0, CH_3
25	16.5, CH_3	16.5, CH_3	16.6, CH_3	16.5, CH_3	16.1, CH_3	15.8, CH_3	16.5, CH_3
26	17.4, CH_3	17.4, CH_3	17.5, CH_3	17.3, CH_3	17.4, CH_3	17.1, CH_3	17.3, CH_3
27	27.7, CH_3	27.8, CH_3	27.8, CH_3	27.7, CH_3	27.5, CH_3	27.6, CH_3	27.7, CH_3
28	62.6, CH_2	62.7, CH_2	62.7, CH_2	58.2, CH_2	78.7, CH_2	58.2, CH_2	58.2, CH_2
29	29.7, CH_3	29.7, CH_3	29.7, CH_3	29.5, CH_3	33.3, CH_3	29.4, CH_3	29.5, CH_3
30	20.9, CH_3	20.1, CH_3	20.2, CH_3	20.1, CH_3	24.4, CH_3	20.0, CH_3	20.1, CH_3
1st sugar	*3-O-GlcA*	*3-O-GlcA*	*3-O-GlcA*	*3-O-GlcA*	*3-O-Glc*	*3-O-GlcA*	*3-O-GlcA*
1′	106.8, CH	106.8, CH	106.9, CH	106.8, CH	107.3, CH	107.5, CH	106.6, CH
2′	72.5, CH	72.5, CH	72.5, CH	72.5, CH	76.2, CH	75.8, CH	75.8, CH
3′	74.2, CH	74.2, CH	74.2, CH	74.2, CH	78.7, CH	78.5, CH	78.5, CH
4′	69.9, CH	69.9, CH	69.9, CH	69.9, CH	72.2, CH	73.8, CH	73.8, CH
5′	75.6, CH	75.6, CH	75.7, CH	75.6, CH	79.1, CH	77.9, CH	78.2, CH
6′	171.6, C	171.6, C	171.6, COOH	171.6, C	63.4, CH_2	173.3, COOH	173.3, COOH
2nd sugar	*3′-O-2-oxoglc*	*3′-O-2-oxoglc*	*3′-O-2-oxoglc*	*3′-O-2-oxoglc*	*28-O-Glc*		
1″	97.4, CH	97.5, CH	97.5, CH	97.5, CH	106.2, CH		
2″	94.1, CH	94.2, CH	94.2, CH	94.2, CH	75.4, CH		
3″	80.2, CH	80.2, CH	80.2, CH	80.2, CH	78.4, CH		
4″	70.1, CH	70.1, CH	70.2, CH	70.1, CH	72.0, CH		
5″	79.9, CH	80.0, CH	80.0, CH	80.0, CH	79.1, CH		
6″	63.2, CH_2	63.2, CH_2	63.3, CH_2	63.2, CH_2	63.1, CH_2		
16-Acyl				*16-Ac*		*16-Ac*	*16-Ac*
1‴				170.2, C		170.5, C	170.3, C
2‴				21.4, CH_3		21.3, CH_3	21.9, CH_3
21-acyl	*21-Mba*	*21-Tig*	*21-Bz*	*21-Tig*		*21-Ac*	*21-Tig*
1	176.8, C	168.4, C	167.2, C	167.6, C		170.3, C	167.6, C
2	42.4, CH	129.9, C	131.9, C	129.3, C		21.8, CH_3	129.3, C
3	27.6, CH_2	137.2, CH	130.4, CH	137.9, CH			137.9, CH
4	12.3, CH_3	12.8, CH_3	129.2, CH	12.6, CH_3			12.6, CH_3
5	17.5, CH_3	14.5, CH_3	133.5, CH	14.6, CH_3			14.6, CH_3
6			129.2, CH				
7			130.4, CH				
22-Acyl				*22-Ac*		*22-Ac*	*22-Ac*
1‴				170.8, C		170.8, C	170.8, C
2‴				21.9, CH_3		21.0, CH_3	21.4, CH_3
28-Acyl	*28-Ac*	*28-Ac*	*28-Ac*				
1‴	171.1, C	171.2, C	170.8, C				
2‴	20.0, CH_3	21.1, CH_3	21.0, CH_3				

Table 6. ^{1}H NMR spectroscopic data for compounds Gymlatinosides GL**1–7** (in Pyridine-d_5)

position	**1** δ_H (J in Hz) 600 Mhz	**2** δ_H (J in Hz) 600 Mhz	**3** δ_H (J in Hz) 500 Mhz	**4** δ_H (J in Hz) 500 Mhz	**5** δ_H (J in Hz) 500 Mhz	**6** δ_H (J in Hz) 500 Mhz	**7** δ_H (J in Hz) 600 Mhz
1	0.88, overlap 1.38, overlap	0.90, overlap 1.42, overlap	0.98, overlap 1.41, overlap	0.89, overlap 1.41, overlap	0.90, ovelap 1.43, overlap	0.88, ovelap 1.41, overlap	0.88, overlap 1.38, overlap
2	1.95, overlap 2.19, overlap	1.96, overlap 2.21, overlap	1.95, overlap 2.21, overlap	1.95, overlap 2.23, overlap	2.24, overlap 1.84, overlap	2.29, overlap 1.86, overlap	1.95, overlap 2.19, overlap
3	4.26, dd (12.3, 4.5)	4.28, dd (12.3, 4.5)	4.26, dd (12.3, 4.5)	4.28, dd (12.3, 4.5)	3.40, dd (11.5, 4.5)	3.41, dd (11.5, 4.5)	4.36, dd (12.3, 4.5)
5	1.64, overlap	1.64, overlap	1.65, overlap	1.64, overlap	0.78, overlap	0.75, overlap	1.64, overlap
6	1.31, overlap 1.65, overlap	1.33, overlap 1.67, overlap	1.32, overlap 1.67, overlap	1.31, overlap 1.73, overlap	1.50, overlap 1.30, overlap	1.51, overlap 1.29, overlap	1.31, overlap 1.65, overlap
7	1.24, overlap 1.63, overlap	1.24, overlap 1.65, overlap	1.22, overlap 1.65, overlap	1.16, overlap 1.56, overlap	1.50, overlap 1.26, overlap	1.50, overlap 1.26, overlap	1.24, overlap 1.63, overlap
9	1.63, overlap	1.63 overlap	1.63, overlap	1.63, overlap	1.56, t (9.0)	1.52, t (9.0)	1.63, overlap
11	1.66, overlap 1.80, overlap	1.66, overlap 1.84, overlap	1.66, overlap 1.84, overlap	1.34, overlap 1.73, overlap	1.84, overlap 1.56, overlap	1.82, overlap 1.56, overlap	1.66, overlap 1.80, overlap
12	5.36, t (3.3)	5.38, t (3.3)	5.37, t (3.3)	5.37, t (3.3)	5.24, t (3.0)	5.39, t (3.0)	5.38, t (3.0)
15	1.50, overlap 2.00, overlap	1.54, overlap 2.02, overlap	1.54, overlap 2.02, overlap	1.41, overlap 1.88, overlap	2.20, overlap 1.74, overlap	1.89, overlap 1.45, overlap	1.88, overlap 1.42, overlap
16	5.09, dd (11.0, 5.2)	5.12, dd (11.0,5.2)	5.12, dd (11.0,5.2)	6.31, dd (11.5,5.5)	4.57, overlap	6.31, overlap	6.30, dd (11.0, 5.5)
18	2.84, dd (13.7, 3.9)	2.87, dd (13.7, 3.9)	2.89, dd (13.7, 3.9)	3.28, dd (14.0, 4.0)	2.30, dd (13.8, 4.0)	3.27, dd (14.0, 4.0)	3.30, dd (14.4, 4.0)
19	1.30, overlap 2.17, overlap	1.32, overlap 2.21, overlap	1.32, overlap 2.21, overlap	1.31, overlap 2.28, overlap	1.88, overlap 1.16, overlap	2.27, overlap 1.34, overlap	2.26, t (14.0) 1.34, overlap
21	5.67, d (10.5)	5.76, d (10.5)	5.77, d (10.5)	5.70, d (11.0)	1.23, overlap 1.63, overlap	5.61, d (11.0)	5.69, d (11.0)
22	4.50, overlap	4.62, overlap 3.72, d	4.6, overlap	6.22, d (11.0) 3.74, d	1.83, overlap 1.35, overlap	5.16, d (11.0)	6.23, d (11.0) 4.36, overlap
23	3.70, d (10.5) 4.32, d (10.5)	(10.5) 4.36, overlap	3.70 d (10.5) 4.18, overlap	(10.5) 4.36, overlap	1.32, s	1.32, s	3.73, d (10.4)
24	0.92, s	0.93, s	0.91, s	0.95, s	1.03, s	0.99, s	0.96, s
25	0.90, s	0.89, s	0.87, s	0.88, s	0.87, s	0.78, s	0.89, s
26	1.05, s	1.06, s	1.05, s	0.90, s	0.98, s	0.86, s	0.90, s
27	1.27, s	1.28, s	1.28, s	1.35, s	1.35, s	1.41, s	1.37, s
28	4.57, overlap 4.99, d (10.5)	4.64, overlap 5.07, d (10.5)	4.62, overlap 5.07, d (10.5)	4.00, overlap 4.02, overlap	4.03, overlap 4.25, overlap	3.99, overlap 4.00, overlap	4.01, s 4.25, overlap
29	1.04, s	1.04, s	1.04, s	0.95, s	0.93, s	0.96, s	0.98, s
30	1.20, s	1.23, s	1.23, s	1.23, s	0.95, s	1.20, s	1.24, s
1st sugar	*3-O-GlcA*	*3-O-GlcA*	*3-O-GlcA*	*3-O-GlcA*	*3-O-Glc*	*3-O-GlcA*	*3-O-GlcA*
1′	5.17, d (7.5)	5.20, d (7.5)	5.15, overlap	5.19, d (7.5)	4.97, d (7.5)	5.19, d (7.5)	5.27, d (7.5)
2′	4.20, overlap	4.20, overlap	4.21, overlap	4.21, overlap	4.01, overlap	4.21, overlap	4.18, t (6.8)
3′	4.89, t (9.5)	4.91, t (9.5)	4.89, t (9.5)	4.91, t (9.5)	4.26, overlap	4.91, t (9.5)	4.26, overlap
4′	5.21, overlap	5.24, t (9.5)	4.36, overlap	5.21, t (9.5)	4.22, overlap	5.21, t (9.5)	4.59, overlap
5′	4.52, overlap	4.54, d (9.5)	4.54, overlap	4.55, d (9.5)	4.02, overlap	4.55, d (9.5)	4.59, overlap
6′					4.40, overlap 4.59, overlap		
2nd sugar	*3′-O-2-oxoglc*	*3′-O-2-oxoglc*	*3′-O-2-oxoglc*	*3′-O-2-oxoglc*	*28-O-Glc*		
1″	5.32, s	5.34, s	5.31, s	5.34, s	4.94, d (7.5)		
2″					4.01, overlap		
3″	4.21, overlap	4.24, overlap	4.22, overlap	4.23, overlap	4.26, overlap		
4″	4.35, overlap	4.38, overlap	5.21, overlap	4.36, overlap	4.22, overlap		
5″	4.02, t (7.5)	4.04, overlap	4.03, t (7.5)	4.02, overlap	4.02, overlap		
6″	4.33, d (10.5) 4.61, d (10.5)	4.36, overlap 4.60, overlap	4.61, overlap	4.34, overlap 4.62, overlap	4.40, overlap 4.59, overlap		
16-Acyl				*16-Ac*		*16-Ac*	*16-Ac*
2				2.02, s		2.11, s	2.03, s
21-acyl	*21-Mba*	*21-Tig*	*21-Bz*	*21-Tig*		*21-Ac*	*21-Tig*
2	2.56, sextet (6.0)					2.17, s	
3	1.53, overlap 1.86, overlap	7.03, q (7.0)	8.27, d (7.5)	7.04, q (7.0)			7.04, q (6.8)
4	0.97, t (7.0)	1.88, s	7.43, t (7.5)	1.90, s			1.90, s
5	1.23, d (7.0)	1.61, d (7.0)	7.51, t (7.5)	1.63, d (7.0)			1.63, d (7.2)
6			7.43, t (7.5)				
7			8.27, d (7.5)				
22-Acyl				*22-Ac*		*22-Ac*	*22-Ac*
2				2.12, s		2.08, s	2.13, s
28-Acyl	*28-Ac*	*28-Ac*	*28-Ac*				
2	1.94, s	2.02, s	1.96, s				

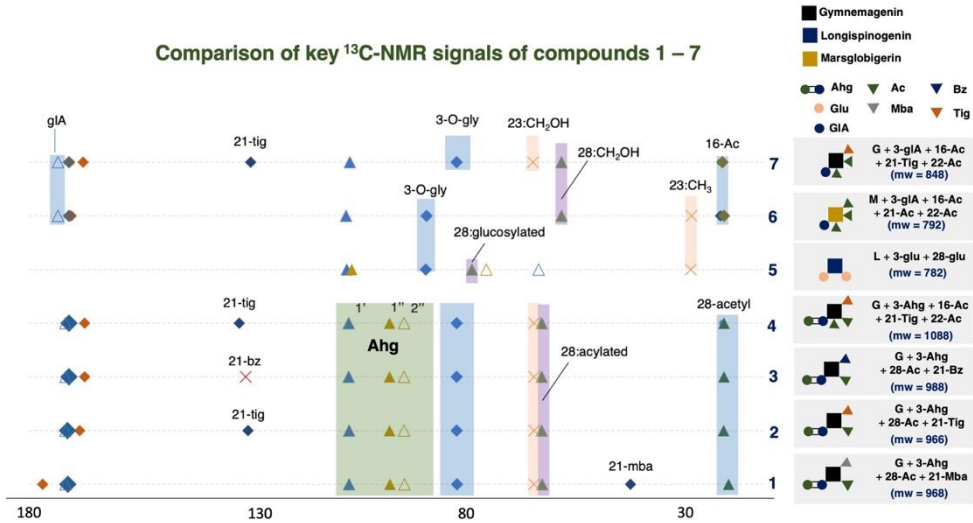


Figure 26. Comparison of key ^{13}C NMR signals of new compound **1–7** isolated from *G. latifolium*

*3.2.1. Gymlatinoside GL1 (**1**)*

Gymlatinoside GL1 (**1**), obtained as a white amorphous powder, with $[\alpha]_D^{25}$ +15.2 (*c* 0.2, MeOH), was found to possess a molecular formula of $C_{49}H_{76}O_{19}$ and and 12 indices of hydrogen deficiency (IHDs) based on the high-resolution electrospray ionization mass spectrometry (HRESIMS) ion peak at *m*/*z* 967.4896 [M − H]$^-$ (calcd for $C_{49}H_{75}O_{19,}$ 967.4903). The broad IR peak at 3389 cm^{-1} indicated the presence of hydroxy groups, and the peaks at 1641 cm^{-1} revealed the existence of carboxylic moieties. The ^{1}H NMR spectrum exhibited seven methyl singlets at δ_H 1.94, 1.27, 1.20, 1.05, 1.04, 0.92 and 0.90 (each 3H). The most downfield resonance (δ_H 1.94) is the signal of an acetoxymethyl, and the others suggested an oleanane backbone. Two other methyls can be observed at δ_H 1.23 (3H, d, J = 7.0 Hz) and δ_H 0.97 (3H, t, J = 7.0 Hz) (**Table 6**). The ^{13}C NMR spectrum showed signals for 49 carbons, including three carboxylic groups at δ_C 176.8, 171.6 and 171.1; two olefinic carbon signals at δ_C 141.5 and 124.9; two anomeric carbons at δ_C 106.8 and 97.4; 15

oxygenated carbons in the range from δ_C 62.6 to 94.1, and others are methine, methylene and methyl signals (**Table 5**). The COSY, HSQC and HMBC spectra indicated a planar structure of oleanane-type triterpene possessing 6 hydroxy groups substituted at C-3, 16, 21, 22, 23 and 28 for the main aglycone of **1** (**Figure 27**). Comparing chemical shifts of **1** with gymnemic acid VIII and gymnemic acid X suggested gymnemagenin to be the main aglycone (Liu et al., 1992). A series of NOESY correlations between H-3 (δ_H 4.26 dd, J = 12.3, 4.5 Hz), H-23 (δ_H 3.70 d, J = 10.5 Hz); δ_H 4.32 d, J = 10.5 Hz), H-5 (δ_H 1.64, overlap), H-9 (δ_H 1.64, overlap), H-27 (δ_H 1.27, s), H-16 [δ_H 5.09, dd, J = 11.0, 5.2 Hz), H-21 [δ_H 5.67, d, J = 10.5 Hz) and H-29 (δ_H 1.04, s) confirmed the α configuration for these protons. Meanwhile, the series of NOESY cross peaks between H-24 (δ_H 0.92, s), H-25 (δ_H 0.92, s), H-26 (δ_H 1.05, s), H-28 [δ_H 4.57, overlap; δ_H 4.99 d, J = 10.5 Hz), H-18 [δ_H 2.84, dd, J = 13.7, 3.9 Hz), H-22 [δ_H 4.50, overlap] and H-30 (δ_H 1.20, s) demonstrated their β configurations (**Figure 27**). Furthermore, the acid hydrolysis of **1** afforded an aglycone which exhibited the same retention time and mass fragmentation with the standard gymnemagenin on LC-MS analysis. Two carboxylic groups were identified to be an (*S*)-2-methylbutyric acid (δ_C 176.8) and an acetyl (δ_C 171.1) substitutions by comparing with those reported in literature (Yoshikawa et al., 1992). Compared with gymnemagenin, the acylation shifts can be observed at C-28 (δ_C 62.6, +4.1 ppm) and C-21 (δ_C 78.6, +1.4 ppm) (Liu et al., 1992). The protons at C-21 and C-28 also showed significant downfield shifts from those of gymnemagenin (H-21: δ_H 4.04 → δ_H 5.67 and H-28: δ_H 4.07 → δ_H 4.57; δ_H 4.71 → δ_H 4.99). The HMBC correlations from H-21 (δ_H 5.67, d, J = 10.5 Hz) to C-$1^{M\text{-}21}$ (δ_C 176.8), and H-28 [δ_H 4.99, d, J = 10.5 Hz); δ_H 4.57, overlap] to C-$1^{A\text{-}28}$ (δ_C 171.1)

confirmed the linkages of the (*S*)-2-methylbutyric acid to C-21 and the acetyl to C-28. The COSY and dimensional (2D) *J*-resolved NMR spectra allowed the detection of two sugar chains, and the precise assignment of the coupling constants of their sugar protons (**Figure 27**). A doublet signal at δ_H 5.18 (d, J = 7.5 Hz) can be assigned to a β-isomer anomeric proton H-1′, corresponding to the anomeric carbon signal at C-1′ (δ_C 106.8). The appearance of proton H-5′ (δ_H 4.51) as a doublet (J = 10.5 Hz) and its HMBC cross peak with C-6′ (δ_C 171.6) suggested that the first sugar moiety is a glucuronic acid. The second anomeric proton H-1″ (δ_H 5.32) which correlated with C-1″ δ_C 97.4 on HSQC spectrum and the signal of the quaternary carbon C-2″ (δ_C 94.1) are the characteristics of a 2-oxo-hexose which is forming an intramolecular hemiacetal with another hydroxy group (Liu et al., 1992). The existence of this 2-oxo-glucose portion was also recognizable by a fragment loss of 160 Da in the positive mass fragmentation. Furthermore, the ^{13}C-NMR data of the sugar portions were almost identical to those of gymnemic acid VIII (Liu et al., 1992). The HMBC correlation from H4′ (δ_H 5.21, t, J = 10.5 Hz) to C-2″, which was not detected by Liu, but finally could be observed in a high resolution HMBC NMR experiment. This HMBC correlation further confirmed the existence dioxane ring between two sugar moieties (**Figure 27**). Coupling constants and NOESY correlations confirmed the β configurations for H-2′ and 4′, and α orientations for H-1′, 3′, 5′, 1″, 3″ and 5″. The ^3J coupling constant (J = 10.5 Hz) observed by dimensional (2D) *J*-resolved NMR indicated the antiperiplanar conformation between H-3″ and H-4″ and thus verified the axial orientation of H-4″. Taken together, compound **1** was deduced as 21-*O*-2(*S*)-methylbutyloyl-28-*O*-acetyl-gymnemagenin 3-*O*-β-D-arabino-2-hexulopyranosyl-(1→3)-β-D-glucuronopyranoside.

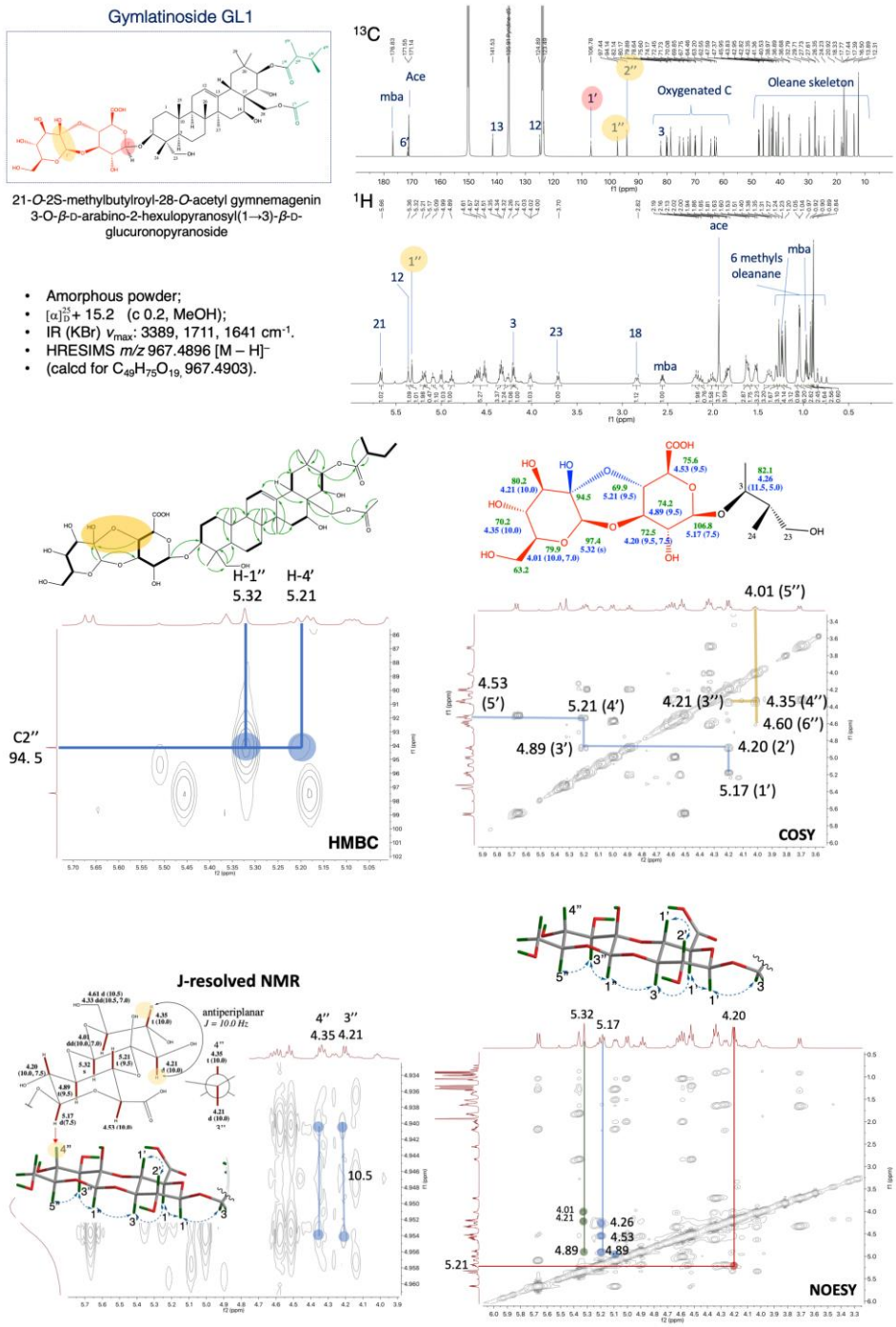


Figure 27. Gymlatinoside GL1

*3.2.2. Gymlatinoside GL2 (**2**)*

Gymlatinoside GL2 (**2**) was obtained as a white amorphous powder, with $[\alpha]_D^{25}$ +9.6 (*c* 0.2, MeOH). Its HRESIMS showed a pseudo molecular ion peak at *m*/*z* 965.4752 $[M - H]^-$ (calcd for $C_{49}H_{73}O_{19}$, 965.4746), indicating a molecular formula of $C_{49}H_{74}O_{19}$ and 13 IHDs. The ^1H and ^{13}C NMR spectroscopic data of **2** (**Tables 5, 6**) showed similar resonances to those of **1**, apart from the signals of the *O*-(S)-2methylbutyloyl moiety which were replaced by the chemical shifts of a tigloyl moiety (Pham et al., 2018) (**Figure 26**). A neutral fragmentation loss of 82 Da observed in positive mode of mass spectrum also established the occurrence of the tigloyl and its linkage to C-21 was demonstrated by the HMBC correlation from H-21 [δ_H 5.76, d(10.5)] to C-$1^{T\text{-}21}$ (δ_C 168.4) (**Figure 28**). Therefore, compound **2** was defined as 21-*O*-tigloyl-28-*O*-acetyl-gymnemagenin 3-*O*-β-D-arabino-2-hexulopyranosyl-(1→3)-β-D-glucuronopyranoside.

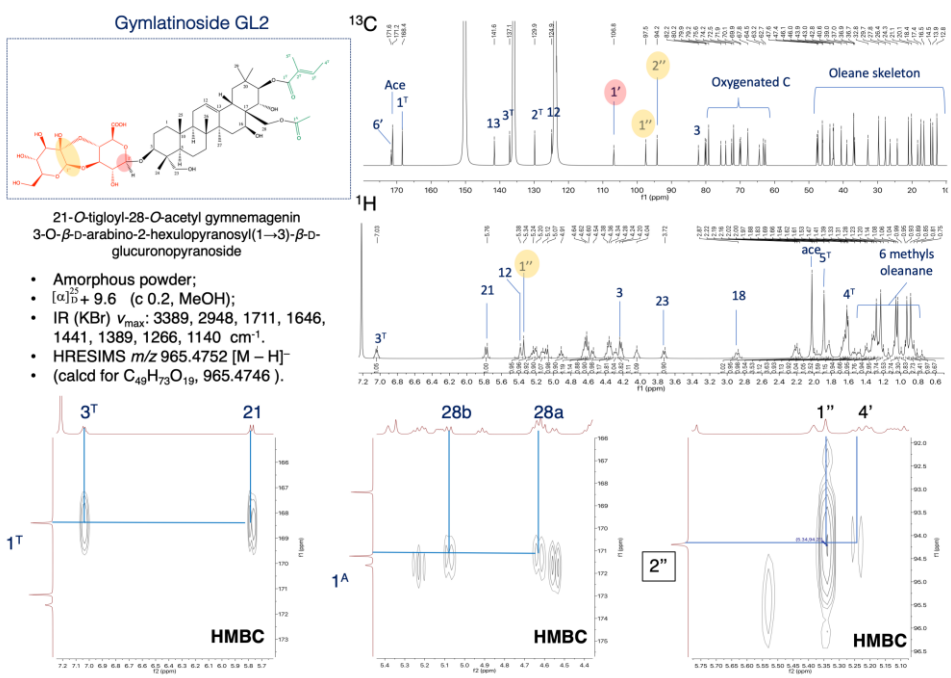


Figure 28. Gymlatinoside GL2

*3.2.3. Gymlatinoside GL3 (**3**)*

Gymlatinoside GL3 (**3**) was obtained as a white amorphous powder, with $[\alpha]_D^{25}$ +10.1 (*c* 0.2, MeOH). Its HRESIMS showed a pseudo molecular ion peak at *m/z* 987.4581 $[M - H]^-$ (calcd for $C_{51}H_{71}O_{19}$, 987.4590), indicating a molecular formula of $C_{51}H_{71}O_{19}$ and 16 indices of hydrogen deficiency. The ^1H and ^{13}C NMR spectroscopic data of **3** (**Tables 5, 6**) exhibited similar resonances to those of **1**, except for the replacement of signals of the *O*-(*S*)-2methylbutyloyl moiety with the chemical shifts of a benzoyl moiety (**Figure 26**). The positive mass fragment loss $\Delta_{m/z}$ = 104 Da observed in **3** confirmed the existence of benzoyl. The downfield shift of C-21 (δ_C 80.3) and the HMBC cross peak from H-21 [δ_H 5.77, d(10.5)] to C-$1^{B\text{-}21}$ (δ_C 167.2) established the connection of the benzoyl to C-21 (**Figure 29**). Therefore, compound **3** was elucidated as 21-*O*-benzoyl-28-*O*-acetyl-gymnemagenin 3-*O*-β-D-arabino-2-hexulopyranosyl-(1→3)-β-D-glucuronopyranoside.

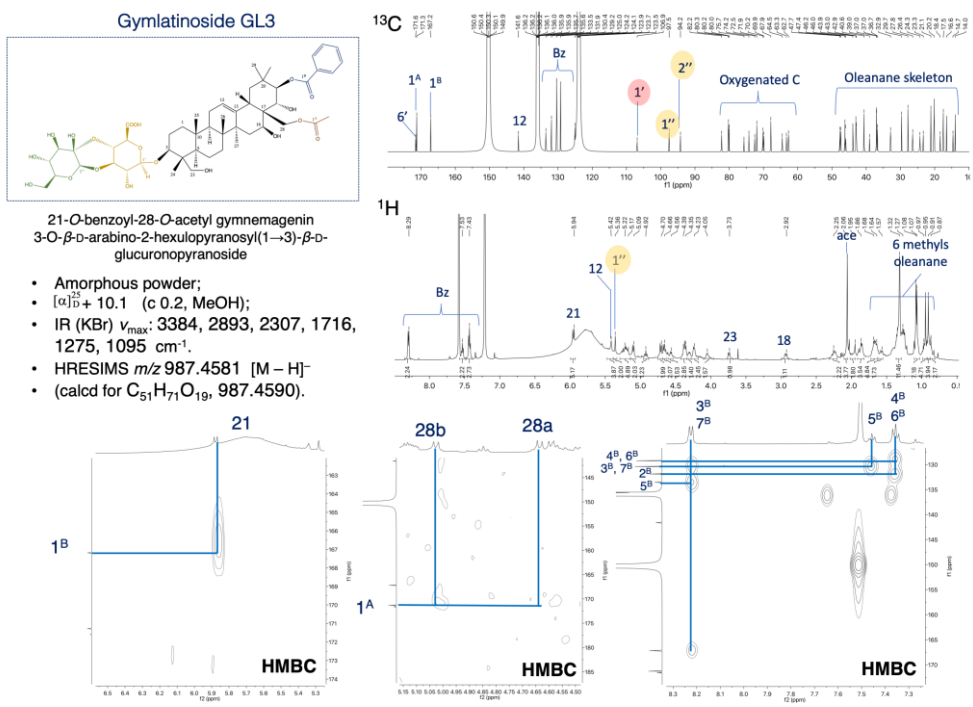


Figure 29. Gymlatinoside GL3

*3.2.4. Gymlatinoside GL4 (**4**)*

Gymlatinoside GL4 (**4**) was obtained as a white amorphous powder, with $[\alpha]_D^{25}$ + 22.1 (*c* 0.2, MeOH), was found to possess a molecular formula of $C_{51}H_{76}O_{20}$ and 14 IHDs based on the high-resolution electrospray ionization mass spectrometry (HRESIMS) ion peak at *m/z* 1007.4826 $[M - H]^-$ (calcd for $C_{51}H_{75}O_{20}$, 1007.4852). Mass fragmentation in the positive mode showed the neutral losses of 2 acetyls (2 × 42 Da) and 1 tigloyl (82 Da). The ^1H and ^{13}C NMR data of **4** (**Tables 5, 6**) showed comparable resonances to those of **2** in the aglycone and glycosyl moieties (**Figure 26)** with some differences occurred in the positions of acyl substitutions. Further investigation of the acylation shifts and HMBC spectra revealed that the tigloyl was attached to C-21 (δ_C 76.7) and two acetyls were substituted at C-16 (δ_C 68.7), C-22 (δ_C 72.0) (**Figure 30**). Therefore, compound **4** was determined as 21-*O*-tigloyl-16,22-*O*-diacetyl gymnemagenin 3-*O*-β-D-arabino-2-hexulopyranosyl-(1→3)-β-D-glucuronopyranoside.

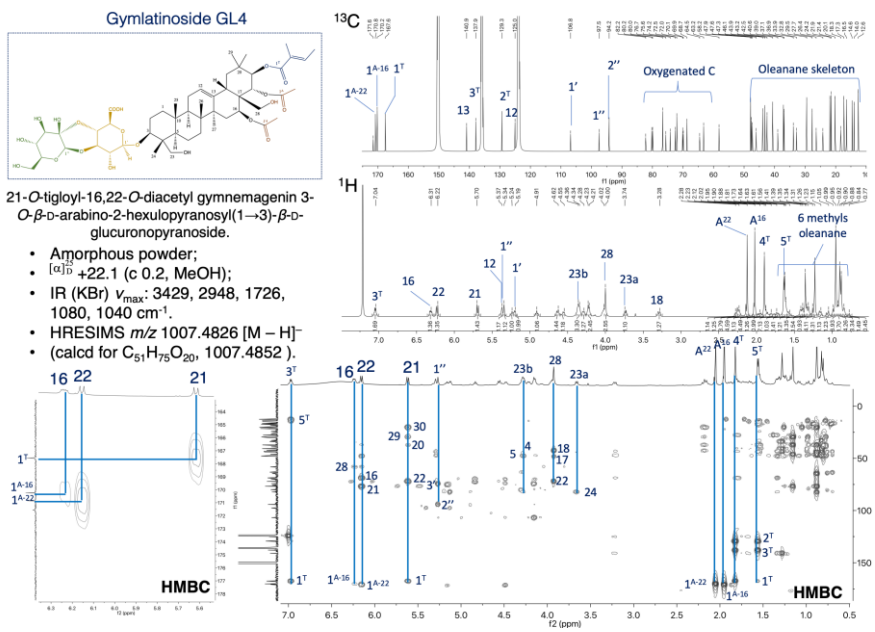


Figure 30. Gymlatinoside GL4

*3.2.5. Gymlatinoside GL5 (**5**)*

Gymlatinoside GL5 (**5**) was obtained as an amorphous powder with $[\alpha]_D^{25}$ +13.4 (c 0.2, MeOH). The molecular formula $C_{42}H_{70}O_{13}$ was determined by a quasimolecular ion peak at *m/z* 781.4746 $[M - H]^-$ (calcd for $C_{42}H_{69}O_{13}$, 781.4738) in HRESIMS. ^1H, ^{13}C, and HSQC NMR spectroscopic data of the aglycone (**Tables 5, 6**) exhibited signals for seven methyl groups: (δ_H 0.87, δ_C 16.1), (δ_H 0.93, δ_C 33.3), (δ_H 0.95, δ_C 24.4), (δ_H 0.98, δ_C 17.4), (δ_H 1.03, δ_C 17.4), (δ_H 1.32, δ_C 28.6), and (δ_H 1.35, δ_C 27.5). The double bond at C_{12-13} was demonstrated by signals of olefinic group [δ_{H-12} 5.24 (br s), δ_{C-12} 123.3] and δ_{C-13} 143.8. Signals of 15 oxygenated carbons can be observed including 2 anomeric carbons at δ_C 107.3, 106.2, 10 other glycosyl carbons and 3 others of the main skeleton. These NMR data suggested the structure of a longispinogenin moiety with two attached sugars (Pham et al., 2018). Sugar analysis and NMR data suggested the sugar type of this structure is glucose which is identified to be in β-configuration by the coupling constant *J*=7.5 Hz of their anomeric protons. Their positions at C-3 and C-28 were evident by the HMBC correlations from H-1′ [4.97, d (7.5)] to C-3 (δ_C 89.2) and H-1″ [4.94, d (7.5)] to C-28 (δ_C 78.7). Consequently, compound **5** was elucidated as 3-*O*-β-D-glucopyranosyl longispinogenin 28-*O*-β-D-glucuronopyranoside.

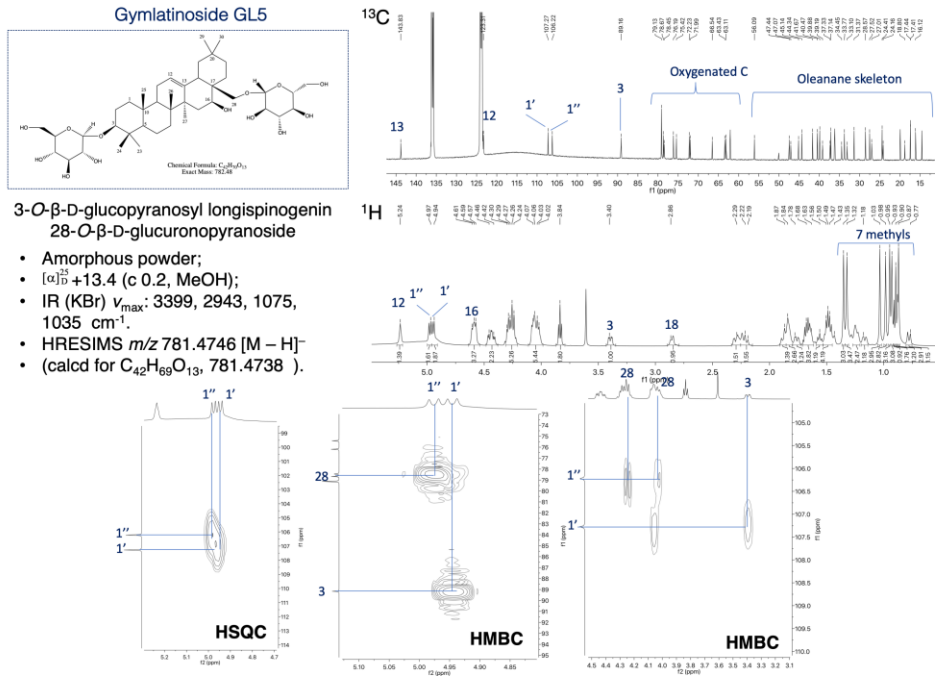


Figure 31. Gymlatinoside GL5

*3.2.6. Gymlatinoside GL6 (**6**)*

Gymlatinoside GL6 (**6**), obtained as an amorphous powder with $[\alpha]_D^{25}$ +5.4 (c 0.2, MeOH). The HRESIMS of this compound revealed an ion peak at *m/z* 791.4225 $[M - H]^-$ (calcd for $C_{42}H_{63}O_{14}$, 791.4218). It suggested a molecular formula $C_{42}H_{64}O_{14}$ and indicated the presence of 11 IHDs. The ^{1}H-NMR spectrum of 6 showed seven methyl groups (δ_H 0.78, 0.86, 0.96, 0.99, 1.20, 1.32 and 1.41), and 12 oxymethine protons (**Table 6**). The signals in the ^{13}C NMR spectrum together with HSQC analysis could be assigned as 11 quaternary carbons (4 carboxylic acids at δ_C 173.3, 170.8, 170.5, 170.3 and one olefinic at δ_C 140.8), 13 tertiary carbons (9 oxygenated methines, one olefinic carbon at δ_C 125.0), 8 secondary carbons (1 oxygenated methylene at δ_C 68.6) and 10 methyl carbons (**Table 5**). Compared its resonances with literature, together with the NOESY experiment (**Figure 32**), suggested that **6** has a Marsglobigerin aglycone which possesses two β-oriented

hydroxyl group substituted at C-16, C-21 and one α-oriented hydroxyl group at C-22 (K. Yoshikawa et al., 1994b). Furthermore, positive fragment ions exhibited a glucuronic fragment loss (176 Da) and 3 acetyl substitutions (3 × 42 Da). The HMBC correlations from H-16 (6.31, overlap) to C-1$^{A\text{-}16}$ (δ_C 179.5), H-21 [δ_H 5.61, d (11.0)] to C-1$^{A\text{-}21}$ (δ_C 170.3), and H-22 [δ_H 5.16, d (11.0)] to C-1$^{A\text{-}22}$ (δ_C 170.8) confirmed the acylated linkage positions at C-16, 21 and 22. The connection of the β glucuronic acid to C-3 was determined by the coupling constant J = 7.5 Hz of the anomeric proton, the glycosylated chemical shift of C-3 (δ_C 89.1, +11.1) and the HMBC correlation from H-1′ (δ_H 4.97) to C-3. Therefore, compound **6** was elucidated as 16,21,22-*O*-triacetyl marsglobigerin 3-*O*-β-D-glucuronopyranoside.

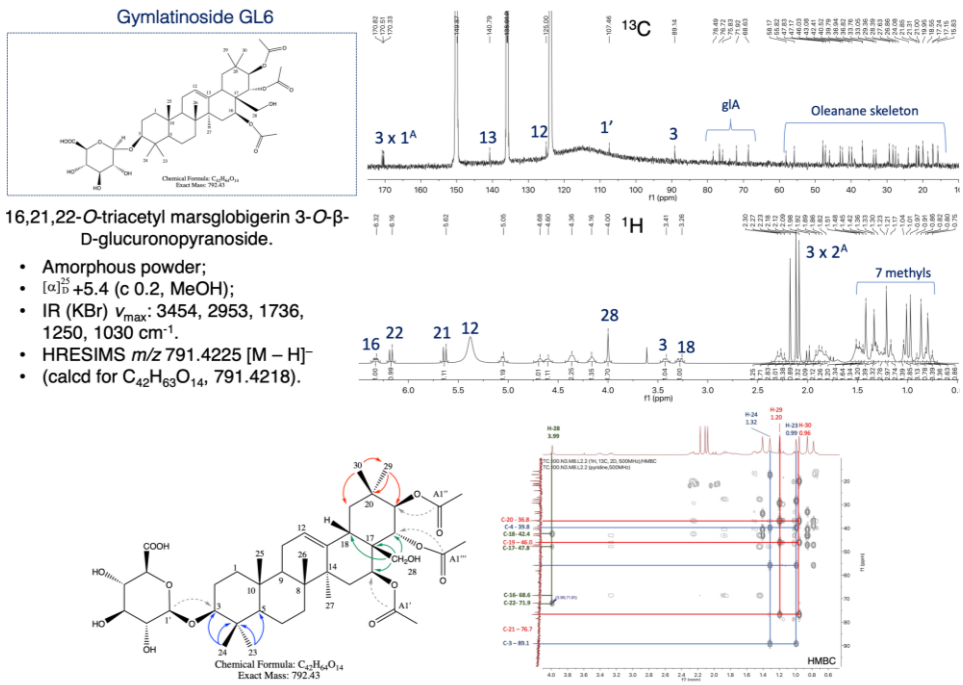


Figure 32. Gymlatinoside GL6

*3.2.7. Gymlatinoside GL7 (**7**)*

Gymlatinoside GL7 (**7**), obtained as an amorphous powder with $[\alpha]_D^{25}$ +12.4 (c 0.2, MeOH), possessed a molecular formula of $C_{45}H_{68}O_{15}$ based on HRESIMS ion peaks at *m/z* 847.4508 $[M - H]^-$ (calcd for $C_{45}H_{67}O_{15}$, 847.4480). The LC-MS experiment in positive mode showed key ions of gymnemagenin (489, 471, 453 and 435) and neutral losses of a glucuronic acid (176 Da), a tigloyl (82 Da) and two acetyls (2 × 42 Da). ^1H, ^{13}C, and HSQC NMR spectroscopic data confirmed the structure of aglycone gymnemagenin similar with compound **1** (**Tables 5, 6**). Meanwhile, its ^{13}C-NMR showed similar acylation chemical shifts for C-16 (δ_C 68.7), C-21 (δ_C 76.7) and C-22 (δ_C 72.0) compared with 4 and the sugar portion showed superimposability with compound **6**. The glycosylated chemical shift of C-3 (δ_C 82.2) and HMBC cross peak from anomeric proton H-1′ [δ_H 5.27, d (7.5)] indicated the presence of a β-D-glucuronic acid substituted at C-3. Accordingly, compound **7** was determined as 21-*O*-tigloyl-16,22-*O*-diacetyl gymnemagenin 3-*O*-β-D-glucuronopyranoside.

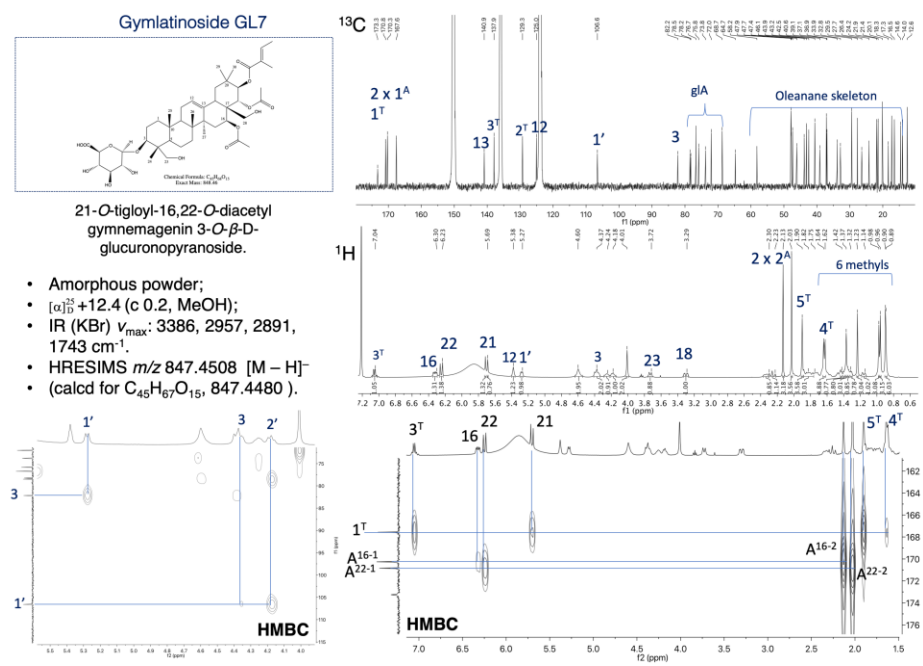


Figure 33. Gymlatinoside GL7

Table 7. ^{13}C NMR spectroscopic data for compounds Gymlatinosides GL**12-18** (in Pyridine-d_5)

	12	**13**	**14**	**15**	**16#**	**17**	**18**
Position	δ_C, type 125 Mhz	δ_C, type 150 Mhz	δ_C, type 150 Mhz	δ_C, type 125 Mhz	δ_C, type 125 Mhz	δ_C, type 125 Mhz	δ_C, , type 150 Mhz
1	38.6, CH_2	38.5, CH_2	38.7, CH_2	38.6, CH_2	38.6, CH_2	38.6, CH_2	39.0, CH_2
2	28.0, CH_2	28.0, CH_2	28.1, CH_2	28.0, CH_2	28.0, CH_2	28.0, CH_2	27.7, CH_2
3	71.5, CH	71.4, CH	71.5, CH	71.4, CH	71.6, CH	71.4, CH	71.4, CH
4	43.6, C	43.6, C	43.6, C	43.6, C	43.5, C	43.6, C	43.2, C
5	47.3, CH	47.3, CH	47.3, CH	47.3, CH	47.2, CH	47.3, CH	47.5, CH
6	18.7, CH_2	18.7, CH_2	18.8, CH_2	18.7, CH_2	18.7, CH_2	18.7, CH_2	18.8, CH_2
7	32.9, CH_2	32.8, CH_2	32.9, CH_2	32.6, CH_2	32.5, CH_2	32.7, CH_2	32.9, CH_2
8	40.5, C	40.5, C	40.5, C	40.4, C	40.5, C	40.5, C	40.4, C
9	47.4, CH	47.3, CH	47.5, CH	47.4, CH	47.2, CH	47.3, CH	47.5, CH
10	37.2, C	37.2, C	37.2, C	37.2, C	36.9, C	37.2, C	37.3, C
11	24.2, CH_2	24.2, CH_2	24.2, CH_2	24.1, CH_2	24.1, CH_2	24.1, CH_2	24.1, CH_2
12	123.1, CH	123.1, CH	123.1, CH	123.4, CH	123.5, CH	123.4, CH	123.2, CH
13	144.0, C	144.2, C	144.0, C	144.2, C	144.6, C	144.2, C	144.0, C
14	44.4, C	44.3, C	44.4, C	44.5, C	44.5, C	44.5, C	44.3, C
15	37.3, CH_2	37.3, CH_2	37.2, CH_2	37.4, CH_2	37.1, CH_2	37.5, CH_2	37.2, CH_2
16	68.0, CH	68.0, CH	68.0, CH	66.5, CH	66.7, CH	66.5, CH	66.5, CH
17	44.3, C	44.0, C	44.4, C	41.7, C	41.7, C	41.4, C	41.6, C
18	44.4, CH	44.2, CH	44.1, CH	45.2, CH	44.5, CH	45.2, CH	45.0, CH
19	48.3, CH_2	48.3, CH_2	48.3, CH_2	47.4, CH_2	47.4, CH_2	47.3, CH_2	47.5, CH_2
20	37.2, C	37.1, C	37.2, C	31.3, C	31.3, C	31.3, C	31.3, C
21	73.1, CH	73.0, CH	73.1, CH	34.4, CH	34.4, CH_2	34.4, CH_2	34.4, CH
22	35.4, CH_2	35.3, CH_2	35.4, CH_2	28.0, CH_2	28.0, CH_2	28.0, CH_2	27.7, CH_2
23	73.5, CH_2	73.4, CH_2	73.5, CH_2	73.4, CH_2	73.7, CH_2	73.4, CH_2	75.3, CH_2
24	13.2, CH_3	13.2, CH_3	13.3, CH_3	13.2, CH_3	13.1, CH_3	13.2, CH_3	13.6, CH_3
25	16.4, CH_3	16.4, CH_3	16.5, CH_3	16.5, CH_3	16.5, CH_3	16.5, CH_3	16.6, CH_3
26	17.2, CH_3	17.2, CH_3	17.3, CH_3	17.4, CH_3	17.1, CH_3	17.4, CH_3	17.5, CH_3
27	28.2, CH_3	28.3, CH_3	28.2, CH_3	28.3, CH_3	28.2, CH_3	28.3, CH_3	27.6, CH_3
28	68.7, CH_2	68.5, CH_2	68.7, CH_2	78.3, CH_2	78.3, CH_2	78.4, CH_2	78.3, CH_2
29	30.4, CH_3	30.4, CH_3	30.5, CH_3	33.8, CH_3	33.7, CH_3	33.8, CH_3	33.7, CH_3
30	18.3, CH_3	18.4, CH_3	18.4, CH_3	24.5, CH_3	24.6, CH_3	24.5, CH_3	24.4, CH_3
A-Glc-1'	104.1, CH	104.1, CH	104.2, CH	104.1, CH	104.0, CH	104.1, CH	105.3, CH
A-Glc-2'	75.4, CH	75.4, CH	75.7, CH	75.7, CH	75.6, CH	75.5, CH	75.4, CH
A-Glc-3'	76.2, CH	76.1, CH	76.4, CH	76.4, CH	76.3, CH	76.2, CH	78.7, CH
A-Glc-4'	71.8, CH	71.8, CH	72.2, CH	72.2, CH	72.1, CH	71.8 CH	71.9, CH
A-Glc-5'	77.5, CH	77.4, CH	79.1, CH	79.1, CH	79.0, CH	77.5, CH	77.4, CH
A-Glc-6'	69.9, CH_2	69.8, CH_2	62.9, CH_2	62.9, CH_2	62.9, CH_2	70.3, CH_2	70.5, CH_2
A-Acyl-1	126.8, C	125.5, C	126.8, C	126.8, C	126.8, C	126.8, C	
A-Acyl-2	124.1, CH	106.9, CH	124.2, CH	124.1, CH	124.1, CH	124.1, CH	
A-Acyl-3	149.3, C	149.6, C	149.3, C	149.3, C	149.2, C	149.2, C	
A-Acyl-4	151.4, C	141.0, C	151.5, C	151.4, C	151.4, C	151.4, C	
A-Acyl-5	117.2, CH	149.6, C	117.2, CH	117.1, CH	117.1, CH	117.1, CH	
A-Acyl-6	111.6, CH	106.9, CH	111.6, CH	111.5, CH	111.6, CH	111.6, CH	
A-Acyl-7	146.2, CH	146.4, CH	146.2, CH	146.2, CH	146.4, CH	146.2, CH	
A-Acyl-8	116.0, CH	116.3, CH	116.0, CH	116.0, CH	115.9, CH	116.0, CH	
A-Acyl-9	167.0, C	166.9, C	167.1, C	167.0, C	167.0, C	167.0, C	
A-Acyl-OCH_3	56.3, CH_3	56.8, 2×OCH_3	56.3, CH_3	56.3, CH_3	56.3, CH_3	56.3, CH_3	
A–glc-1"	105.8, CH	105.8, CH				105.9, CH	105.7, CH
A–glc-2"	75.3, CH	75.3, CH				75.3, CH	75.4, CH
A–glc-3"	78.7, CH	78.7, CH				78.7, CH	78.5, CH
A–glc-4"	71.8, CH	71.8, CH				71.8, CH	71.9, CH
A–glc-5"	77.4, CH	77.3, CH				77.4, CH	77.4, CH
A–glc-6"	70.5, CH_2	70.2, CH_2				69.8, CH_2	70.5, CH_2
A–xyl-1"'	106.5, CH	106.4, CH				106.5, CH	106.3, CH
A–xyl-2"'	75.3, CH	75.2, CH				75.3, CH	75.2, CH

Position	12 δ_C, type 125 Mhz	13 δ_C, type 150 Mhz	14 δ_C, type 150 Mhz	15 δ_C, type 125 Mhz	16# δ_C, type 125 Mhz	17 δ_C, type 125 Mhz	18 δ_C, , type 150 Mhz
A–xyl-3'''	78.5, CH	78.5, CH				78.5, CH	78.3, CH
A–xyl-4'''	71.5, CH	71.5, CH				71.7, CH	71.4, CH
A–xyl-5'''	67.5, CH	67.4, CH				67.5, CH	67.4, CH
B-Glc-I-1'				106.2, CH	103.4, CH	106.2, CH	106.2, CH
B-Glc-I-2'				75.4, CH	75.6, CH	75.3, CH	75.3, CH
B-Glc-I-3'				78.3, CH	76.5, CH	78.4, CH	78.5, CH
B-Glc-I-4'				72.0, CH	72.2, CH	72.0, CH	71.9, CH
B-Glc-I-5'				79.1, CH	78.9, CH	79.1, CH	79.0, CH
B-Glc-I-6'				63.1, CH_2	62.9, CH_2	63.1, CH_2	63.1, CH_2

#Acyl-2: 1^{F-2} (126.8, C); 1^{F-2} (124.2, CH); 3^{F-2} (149.3, C); 4^{F-2} (151.5, C); 5^{F-2} (117.1, CH); 6^{F-2} (111.6, CH); 7^{F-2} (146.4, CH); 8^{F-2} (115.7, CH); 9^{F-2} (167.0, C); O-CH_3 (56.3, C).

* B-glc-II: 1'' (δC 105.6, CH); 2 (δC 75.5, CH); 3 (δC 78.8, CH); 4 (δC 72.0, CH); 5 (δC 77.5, CH); 6 (δC 62.5, CH);

Table 8. [1]H NMR spectroscopic data for compounds Gymlatinosides GL**12-18** (in Pyridine-d_5)

Position	**12**	**13**	**14**	**15**	**16**[#]	**17**	**18**
	δ_H (J in Hz) 500 Mhz	δ_H (J in Hz) 600 Mhz	δ_H (J in Hz) 600 Mhz	δ_H (J in Hz) 500 Mhz	δ_H (J in Hz) 500 Mhz	δ_H (J in Hz) 500 Mhz	δ_H (J in Hz) 500 Mhz
1	1.36, overlap 1.10, overlap	1.37, overlap 1.08, overlap	1.36, overlap 1.10, overlap	1.36, overlap 1.10, overlap	1.37, overlap 1.08, overlap	1.37, overlap 1.08, overlap	1.36, overlap 1.10, overlap
2	1.82, overlap 1.42, overlap	1.84, overlap 1.40, overlap	1.82, overlap 1.42, overlap	1.82, overlap 1.42, overlap	1.84, overlap 1.40, overlap	1.84, overlap 1.40, overlap	1.82, overlap 1.42, overlap
3	4.08, overlap	4.08, overlap	4.12, dd (11.5, 4.5)	4.07, overlap	4.02, overlap	4.05, overlap	4.18, overlap
5	1.56, overlap	1.54, overlap	1.56, overlap	1.56, overlap	1.56, overlap	1.56, overlap	1.56, overlap
6	1.36, overlap	1.39, overlap	1.36, overlap	1.36, overlap	1.36, overlap	1.36, overlap	1.36, overlap
7	1.94, overlap 1.40, overlap	1.93, overlap 1.40, overlap	1.94, overlap 1.40, overlap	1.94, overlap 1.40, overlap	1.94, overlap 1.40, overlap	1.94, overlap 1.40, overlap	1.94, overlap 1.40, overlap
9	1.79, overlap	1.82, overlap	1.79, overlap	1.79, overlap	1.79, overlap	1.79, overlap	1.79, overlap
11	1.76, overlap	1.77, overlap	1.76, overlap	1.76, overlap	1.76, overlap	1.76, overlap	1.76, overlap
12	5.28, br s	5.28, br s	5.28, br s	5.17, br s	5.19, br s	5.17, br s	5.21, br s
15	2.27, t (11.5)	2.24, t (11.5)	2.27, t (11.5)	2.25, t (11.5)	2.28, t (11.5)	2.22, t (11.5)	2.30, t (11.5)
16	4.81, overlap	4.81, overlap	4.77, dd (11.5, 5.0)	4.58, overlap	4.52, overlap	4.60, overlap	4.52, overlap
18	2.61, dd (13.5, 4.5)	2.63, dd (13.5, 4.5)	2.62, dd (13.5, 4.5)	2.83, dd (13.5, 4.5)	2.61, dd (13.5, 4.5)	2.62, dd (13.5, 4.5)	2.61, dd (13.5, 4.5)
19	1.42, overlap 2.12, overlap	1.42, overlap 2.13, overlap	1.42, overlap 2.13, overlap	1.42, overlap 2.12, overlap	1.42, overlap 2.13, overlap	1.42, overlap 2.12, overlap	1.42, overlap 2.13, overlap
21	4.19, overlap	4.19, overlap	4.19, dd (13.0, 4.5)	4.19, overlap	4.19, overlap	4.19, overlap	4.19, overlap
22	3.30, dd (13.0, 4.5) 2.11, overlap	3.30, dd (13.0, 4.5) 2.13, overlap	3.23, dd (13.0, 4.5) 2.04, t (13.0)	3.30, dd (13.0, 4.5) 2.11, overlap	3.23, dd (13.0, 4.5) 2.04, t (13.0)	3.30, dd (13.0, 4.5) 2.13, overlap	3.30, dd (13.0, 4.5) 2.13, overlap
23	4.03, overlap 3.87, overlap	4.03, overlap 3.95, overlap	4.03, d (10.5) 3.87, d (10.5)	4.01, overlap 3.85, overlap	3.98, overlap 3.90, overlap	4.01, overlap 3.93, overlap	4.09, overlap 3.98, overlap
24	0.84, s	0.84, s	0.86, s	0.85, s	0.87, s	0.84, s	1.02, s
25	0.87, s	0.87, s	0.89, s	0.88, s	0.88, s	0.87, s	0.94, s
26	1.01, s	1.01, s	1.02, s	0.97, s	1.00, s	0.98, s	0.98, s
27	1.68, s	1.68, s	1.65, s	1.61, s	1.53, s	1.64, s	1.22, s
28	4.42, d (10.5) 3.76, d (10.5)	4.42, d (10.5) 3.76, d (10.5)	4.42, d (10.5) 3.76, d (10.5)	4.24, overlap 4.05, overlap	4.24, overlap 4.05, overlap	4.24, overlap 4.05, overlap	4.26, overlap 4.05, overlap
29	1.31, s	1.28, s	1.31, s	0.95, s	0.88, s	0.95, s	0.89, s
30	1.28, s	1.31, s	1.28, s	0.93, s	0.96, s	0.93, s	0.93, s
A-Glc-1'	5.05, d (8.5)	5.05, d (8.5)	5.05, d (8.5)	5.12, d (8.5)	5.09, d (8.5)	5.04, d (8.5)	4.81, d (7.5)
A-Glc-2'	5.68, t (8.5)	5.68, t (8.5)	5.71, t (8.5)	5.77, t (8.5)	5.74, t (8.6)	5.66, t (8.5)	4.03, overlap

Position	12	13	14	15	16[#]	17	18
	δ_H (*J* in Hz) 500 Mhz	δ_H (*J* in Hz) 600 Mhz	δ_H (*J* in Hz) 600 Mhz	δ_H (*J* in Hz) 500 Mhz	δ_H (*J* in Hz) 500 Mhz	δ_H (*J* in Hz) 500 Mhz	δ_H (*J* in Hz) 500 Mhz
A-Glc-3'	4.27, overlap	4.29, overlap	4.29, overlap	4.27, overlap	4.29, overlap	4.27, overlap	4.15, overlap
A-Glc-4'	4.22, overlap	4.21, overlap	4.21, overlap	4.22, overlap	4.21, overlap	4.22, overlap	4.23, overlap
A-Glc-5'	4.07, overlap	4.07, overlap	4.07, overlap	4.07, overlap	4.07, overlap	4.07, overlap	4.00, overlap
A-Glc-6'	4.84, overlap	4.83, overlap	4.83, overlap	4.84, overlap	4.83, overlap	4.84, overlap	4.84, overlap
	4.36, overlap	4.35, overlap	4.35, overlap	4.36, overlap	4.35, overlap	4.36, overlap	4.32, overlap
A-Acyl-2	7.20, overlap	6.98, overlap	7.20, overlap	7.20, overlap	7.20, overlap	7.19, overlap	
A- Acyl -5	7.19, overlap		7.19, overlap	7.19, overlap	7.19, overlap	7.18, overlap	
A- Acyl -6	7.21, overlap	6.98, overlap	7.21, overlap	7.21, overlap	7.20, overlap	7.19, overlap	
A- Acyl -7	7.99, d (16.0)	7.99, d (16.0)	8.01, d (16.0)	7.98, d (16.0)	7.99, d (16.0)	7.97, d (16.0)	
A- Acyl -8	6.62, d (16.0)	6.67, d (16.0)	6.64, d (16.0)	6.64, d (16.0)	6.70, d (16.0)	6.60, d (16.0)	
A- Acyl - OCH_3	3.85, s	3.89, s	3.84, s	3.82, s	3.85, s	3.82, s	
A–glc-1"	4.97, d (7.5)	4.97, d (7.5)				5.00, d (7.5)	4.94, d (7.5)
A–glc-2"	4.05, overlap	4.03, overlap				4.05, overlap	4.03, overlap
A–glc-3"	4.15, overlap	4.15, overlap				4.15, overlap	4.15, overlap
A–glc-4"	4.24, overlap	4.23, overlap				4.24, overlap	4.23, overlap
A–glc-5"	4.00, overlap	4.00, overlap				4.00, overlap	4.00, overlap
A–glc-6"	4.84, overlap	4.84, overlap				4.84, overlap	4.84, overlap
	4.35, overlap	4.32, overlap				4.35, overlap	4.29, overlap
A–xyl-1"'	5.00, d (7.5)	5.00, d (7.5)				4.98, d (7.5)	5.03, d (7.5)
A–xyl-2"'	4.04, overlap	4.02, overlap				4.04, overlap	4.02, overlap
A–xyl-3"'	4.15, overlap	4.15, overlap				4.15, overlap	4.15, overlap
A–xyl-4"'	4.25, overlap	4.23, overlap				4.25, overlap	4.23, overlap
A–xyl-5"'	4.35, overlap	4.35, overlap				4.32, overlap	4.35, overlap
	3.68, t (10.5)	3.67, t (10.5)				3.66, t (10.5)	3.67, t (10.5)
B-Glc-I-1'				4.96, d (7.5)	4.93, d (7.5)	4.97, d (7.5)	4.94, d (7.5)
B-Glc-I-2'				4.05, overlap	5.71, t (8.6)	4.05, overlap	4.05, overlap
B-Glc-I-3'				4.15, overlap	4.15, overlap	4.15, overlap	4.15, overlap
B-Glc-I-4'				4.24, overlap	4.24, overlap	4.24, overlap	4.24, overlap
B-Glc-I-5'				4.00, overlap	4.00, overlap	4.00, overlap	4.00, overlap
B-Glc-I-6'				4.52, overlap	4.59, overlap	4.59, overlap	4.56, overlap
				4.43, dd (12.0, 5.0)	4.44, dd (12.0, 5.0)	4.44, dd (12.0, 5.0)	4.43, dd (12.0, 5.0)

[#] B-Acyl: 2 (7.20, overlap), 5 (7.19, overlap), 6 (7.21, overlap), 7 [8.01, d (16.0)], 8 [6.62, d (16.0)] OCH_3 (3.85, s)

*3.2.8. Gymlatinoside GL8 (**12**)*

Gymlatinoside GL8 (**12**) obtained as a white amorphous powder, with $[\alpha]_D^{25}$ +35.4 (*c* 0.2, MeOH), was found to possess a molecular formula of $C_{57}H_{86}O_{22}$ based on the high-resolution electrospray ionization mass spectrometry (HRESIMS) ion peak at *m/z* 1121.5549 $[M - H]^-$ (calcd for $C_{57}H_{85}O_{22,}$ 1121.5532). The broad IR peak at 3364 cm^{-1} indicated the presence of hydroxy groups. The ^1H NMR spectrum displayed six methyl singlets at δ_H 1.70, 1.30, 1.28, 1.01, 0.86 and 0.84 (each 3H) one olefinic broad singlet at δ_H 5.28, which were characteristics of a oleanane skeleton. Three anomeric proton signals at δ_H 5.07, d (7.5); 4.97, d (7.5) and 5.01, d (7.5) suggested the presences of three sugars moieties. Also, three other signals of overlapped olefinic protons at δ_H 7.19, 7.20, 7.21 revealed by HSQC correlations, together with one methoxy singlet at δ_H 3,89 and two trans olefinic doublets at δ_H 7.99 and 6.66 (both have *J* =16.0 Hz) suggested the existence of a feruloyl substituent. The ^{13}C NMR spectrum showed signals for 57 carbons, including resonances of 1 carboxylic group at δ_C 167.0; 10 aromatic or olefinic carbons; 3 anomeric and 19 oxygenated carbons ranging from δ_C 67.5 to 78.7, and others are methine, methylene and methyl signals. Inspection of COSY, HSQC and HMBC spectra determined a planar structure of oleanane-type triterpene possessing 5 hydroxy groups substituted at C-3,16, 21, 23 and 28 for the main aglycone of **12** (**Figure 34**). Comparing chemical shifts of **12** with gymnemoside-d revealed that they shared an aglycone, gymnestrogenin and a glycosyl chain, -glc^6-glc^6-xyl at C-23 in common (M. Yoshikawa et al., 1997). The additional feruloyl at C-2' was confirmed by the HMBC correlation from H-2' [δ_H 5.69, t (8.6)] to C-9^F (167.0). The relative configuration of the aglycone Gymnestrogenin was analysed by NOESY

experiments which illustrated in **Figure 34**. Consequently, the structure of **12** was identified as Gymnestrogenin 23-*O*-β-D-xylopyranosyl-(1→6)-β-D-glucopyranosyl-(1→6)-(2′-*O*-*trans*-feruloyl)-β-D-glucopyranoside.

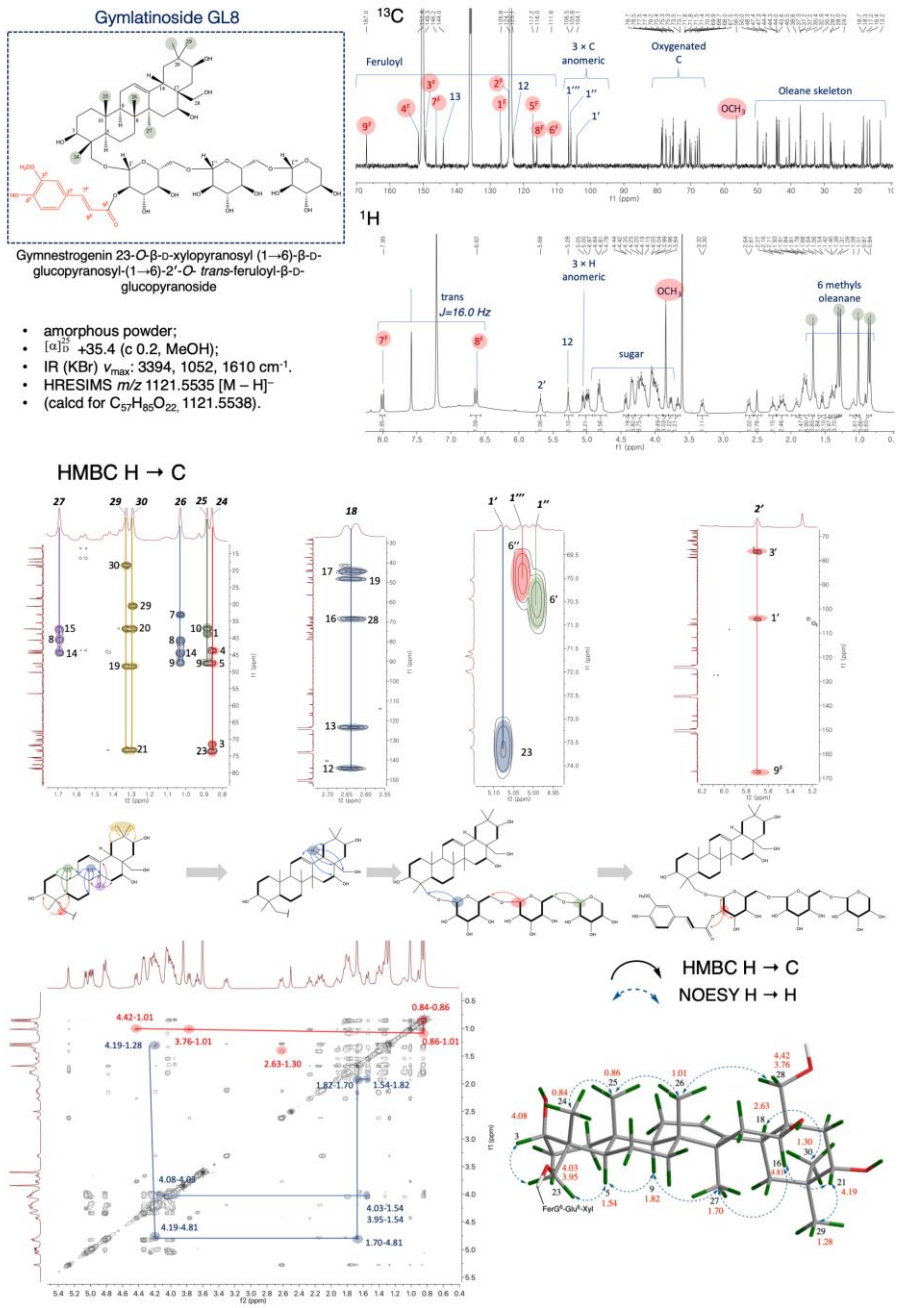


Figure 34. Gymlatinoside GL8

*3.2.9. Gymlatinoside GL9 (**13**)*

Gymlatinoside GL9 (**13**) obtained as a white amorphous powder, with $[\alpha]_D^{25}$ +24.3 (*c* 0.2, MeOH), was found to possess a molecular formula of $C_{58}H_{88}O_{23}$ based on the high-resolution electrospray ionization mass spectrometry (HRESIMS) ion peak at *m*/*z* 1151.5658 $[M - H]^-$ (calcd for $C_{58}H_{87}O_{23,}$ 1151.5638). Comparison of the molecular formula with compound **12** suggested that **13** has one more methoxy (OCH_3) unit. The chemical shifts in 1D-NMR and correlations in the 2D-NMR spectra of **12** and **13** were mostly comparable; the exception was the resonances of the acyl unit connected to C-2′ (**Tables 7, 8**). The existences of two methoxy groups attached to two symmetry-equivalent carbons of the acyl portion were suggested by the proton singlet signal of 6 protons at δ_H 3.89**.** The ^1H and ^{13}C chemical shifts revealed the structure of a sinapyol moiety for the acyl substitution by comparing with literature (Abe et al., 1995) and analysis of HMBC correlations (**Figure 35**). Its trans configuration was identified by the coupling constant *J* =16.0 Hz of two adjacent olefinic doublet protons at δ_H 7.99 and δ_H 6.66. The position of this substituent was determined from the HMBC cross peak correlations from H-2′ [δ_H 5.68, d (8.5)] to C-9^S (δ_C 166.9) and C1' (δ_C 104.1). Accordingly, compound **13** was determined as Gymnestrogenin 23-*O*-β-D-xylopyranosyl-(1→6)-β-D-glucopyranosyl-(1→6)-(2′-*O*-*trans*-sinapoyl)-β-D-glucopyranoside.

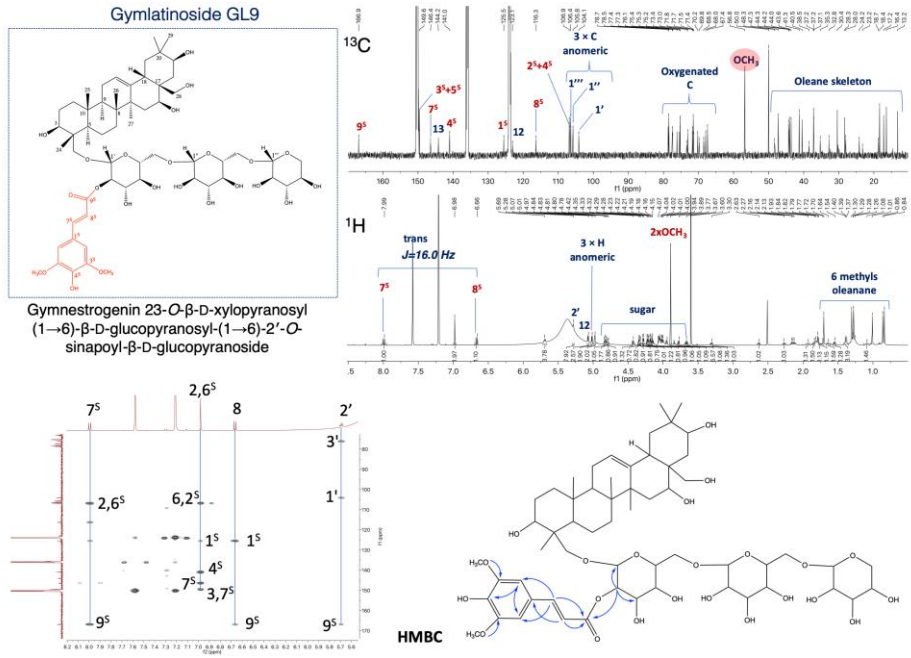


Figure 35. Gymlatinoside GL9

3.2.10. *Gymlatinoside GL10 (**14**)*

Gymlatinoside GL10 (**14**) was obtained as a white amorphous powder, with $[\alpha]_D^{25}$ +10.1 (*c* 0.2, MeOH). Its HRESIMS showed a pseudo molecular ion peak at *m*/*z* 827.4585 $[M - H]^-$ (calcd for $C_{46}H_{67}O_{13}$, 827.4582), indicating a molecular formula of $C_{46}H_{68}O_{13}$. The ^1H, ^{13}C, HSQC and HMBC NMR spectroscopic data of **14** could be discriminated with **12** by the lack of resonances of the xylopyranosyl and the glucopyranosyl units attached to C-6′ and the shielded chemical shift of this carbon C-6′ (δ_C 62.9) (**Tables 7, 8**). HMBC correlations H-2′ [δ_H 5.71, d (8.5)] to C-9^F (δ_C 167.1) and C1' (δ_C 104.2) confirmed the structure of Gymlatinoside GL10 as shown in **Figure 36**. Therefore, compound **14** was elucidated as Gymnestrogenin 23-*O*-(2′-*O*-*trans*-feruoyl)-β-D-glucopyranoside.

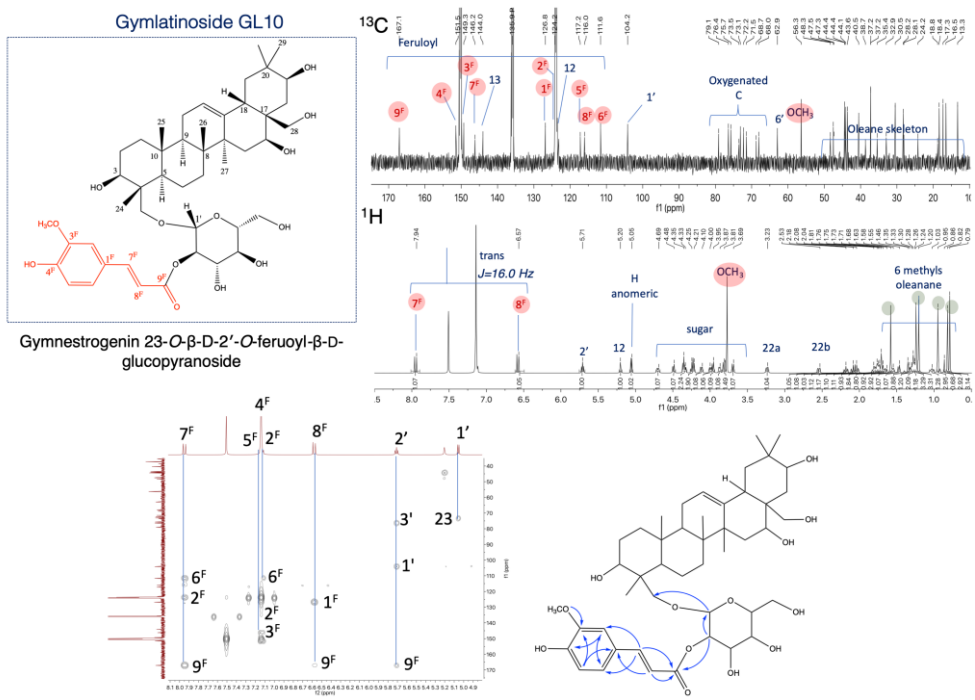


Figure 36. Gymlatinoside GL10

*3.2.11. Gymlatinoside GL11 (**15**)*

Gymlatinoside GL11 (**15**) was obtained as a white and amorphous powder with $[\alpha]_D^{25}$ +26.3 (c 0.2, MeOH). It gave a molecular formula of $C_{52}H_{78}O_{17}$ as determined according to the negative ion peak $[M - H]^-$ at *m/z* 973.5167 (calcd for $C_{52}H_{77}O_{17}$, 973.5161) in the HRESIMS corresponding to 14 IDHs. The ^1H, ^{13}C NMR (**Tables 7, 8**), and HSQC spectroscopic data of **15** displayed 6 methyl singlet resonances at δ_H 0.85, 0.88, 0.97, 1.61, 0.95, 0.93 which corresponding to the carbon resonances at δ_C 13.2, 16.5, 17.4, 28.3, 33.8 and 24.5, respectively. Furthermore, signals of an olefinic proton at δ_H 5.17 (1H, br s) and two olefinic carbon signals at δ_C 144.2 and 123.4 were observed. These resonances revealed the aglycone structure to be 23-hydroxy longispinogenin by referring to NMR data of Gymnemoside-e (K. Yoshikawa et al., 1991). The presence of the additional β-glucopyranosyl unit was

evident by the resonances of 6 oxygenated carbons at C-1″ (δ_C 106.2), C-2″ (δ_C 75.4), C-3″ (δ_C 78.3), C-4″ (δ_C 72.0), C-5″ (δ_C 79.1), and C-6″ (δ_C 63.1). When compared with **14**, a downfield shift of C-28 (δ_C 78.3, +9.6 ppm) and a HMBC correlation between anomeric proton H-1″ (δ_H 4.96) and C-28 (**Figure 37**) were observed, which supported the attachment of the second glucopyranosyl unit to C-28. Accordingly, compound **15** was determined to be 28-*O*-β-D-glucopyranosyl 23-hydroxy-longispinogenin 23-*O*-(2′-*O*-*trans*-feruoyl)-β-D-glucopyranoside.

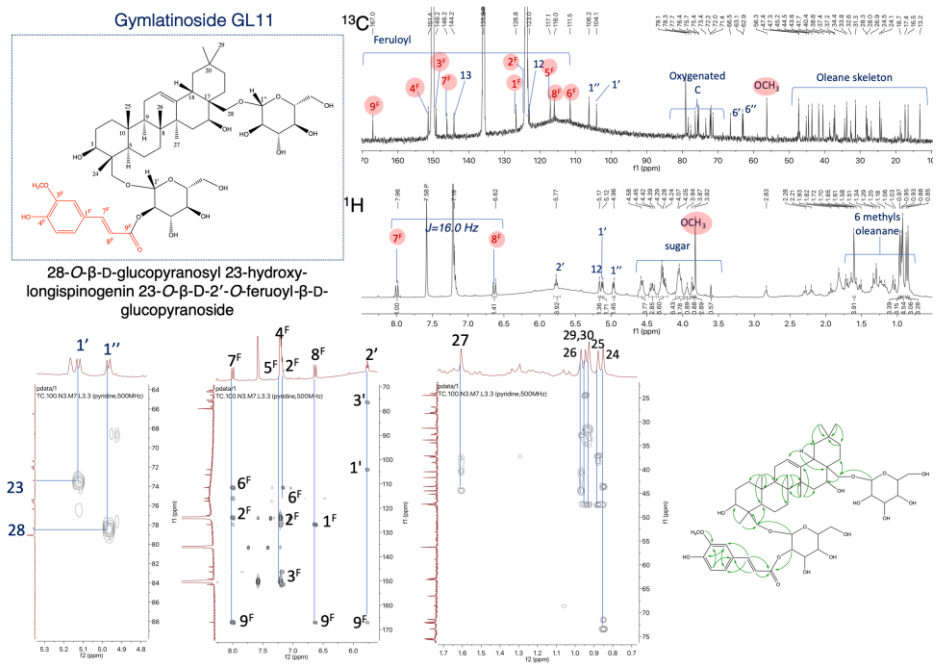


Figure 37. Gymlatinoside GL11

*3.2.12. Gymlatinoside GL12 (**16**)*

Gymlatinoside GL12 (**16**) was obtained as an amorphous powder with $[\alpha]_D^{25}$ +15.2 (c 0.2, MeOH). Its molecular formula was determined to be $C_{62}H_{86}O_{20}$ based on its HRESIMS peak at *m/z* 1149.5629 $[M - H]^-$ (calcd for $C_{62}H_{85}O_{20}$, 1149.5634). The HRMS positive mode showed an extra neutral loss of one feruloyl (176 Da) comparing with **15.** The ^1H and ^{13}C NMR spectroscopic data of **16** (**Tables 7, 8**)

were comparable with those of **15**, except for the addition of resonances of one feruloyl unit which attached to C-2". Further HMBC analysis confirmed the structure of compound **16** as shown in **Figure 38**. Accordingly, compound **16** was elucidated to be 28-*O*-(2′-*O*-*trans*-feruoyl)-β-D-glucopyranosyl 23-hydroxy-longispinogenin 23-*O*-(2′-*O*-feruoyl)-β-D-glucopyranoside.

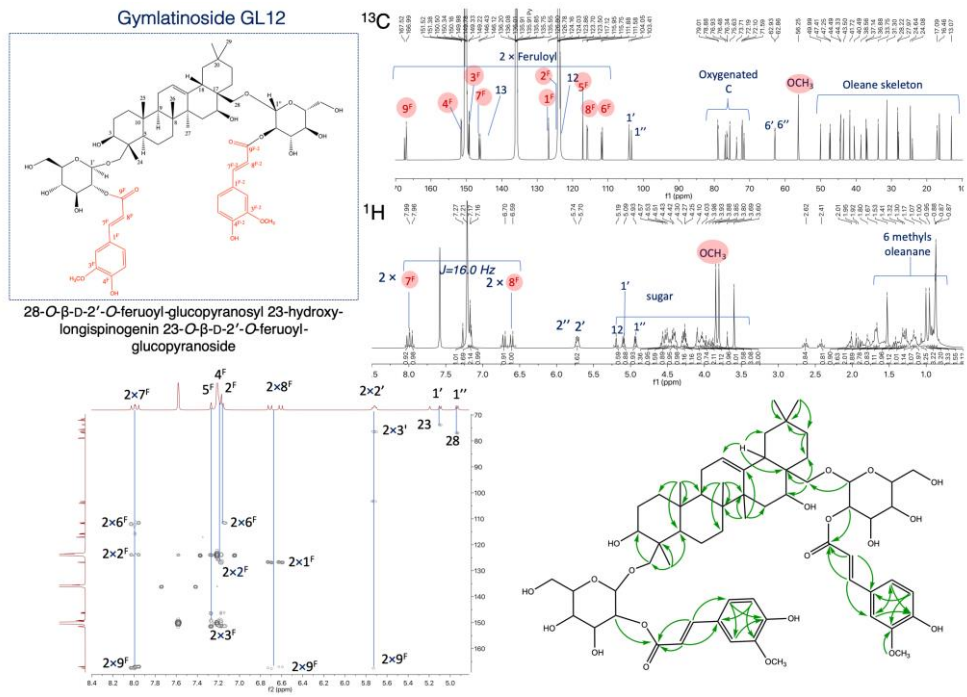


Figure 38. Gymlatinoside GL12

*3.2.13. Gymlatinoside GL13 (**17**)*

Gymlatinoside GL13 (**17**) was isolated as an amorphous powder with $[\alpha]_D^{25}$ +25.9 (*c* 0.2, MeOH). The HRESIMS of this compound revealed an ion peak at *m/z* 1267.6147 $[M - H]^-$ (calcd for $C_{63}H_{95}O_{26}$, 1267.6147). It suggested a molecular formula $C_{63}H_{96}O_{26}$ and indicated the presence of 16 IHDs. NMR data of compound **17** showed partial similarities compared with compounds **13** and **15** (**Tables 7, 8**)**.** The first glycosyl portion at C-23, which was identical to that part of compound **13**, comprised of a feruloyl moiety attached to C-2' of a sugar chain of glc^6-glc^6-xyl. The main aglycone and the second sugar at C-28 were identified to be 23-hydroxy longispinogenin and glucose, respectively by comparing NMR data of this portion with those of compound **15** (**Figure 39**). Finally, the structure of **17** was identified as 28-*O*-β-D-glucopyranosyl 23-hydroxy-longispinogenin 23-*O*-β-D-xylopyranosyl (1→6)-β-D-glucopyranosyl-(1→6)-(2′-*O*-*trans*-feruloyl)-β-D-glucopyranoside.

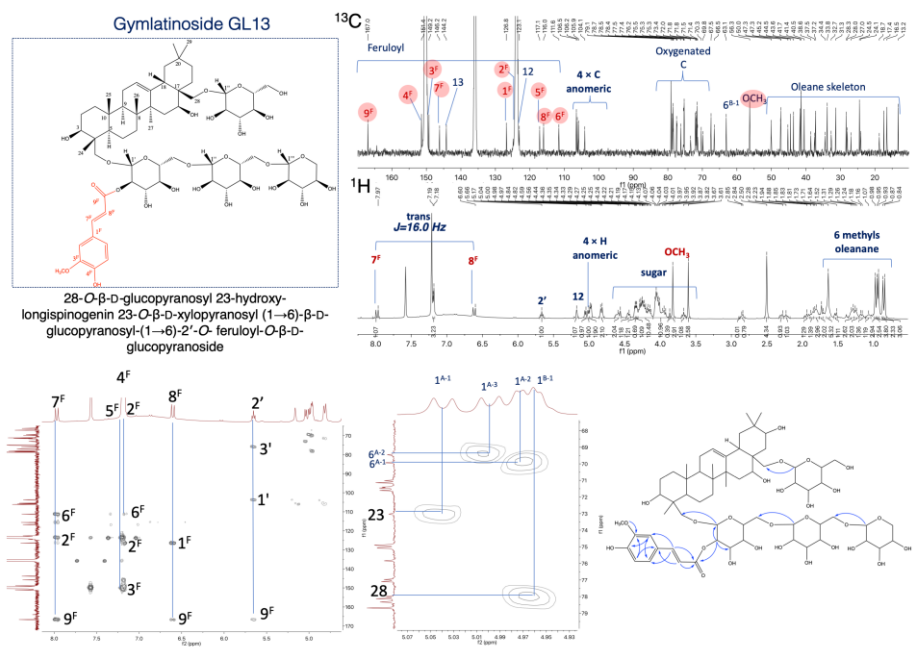


Figure 39. Gymlatinoside GL13

*3.2.14. Gymlatinoside GL14 (**18**)*

Gymlatinoside GL14 (**18**) was purified as an amorphous powder with $[\alpha]_D^{25}$ +12.5 (*c* 0.2, MeOH). Its molecular formula was identified as $C_{53}H_{88}O_{23}$ based on the negative ion peak at *m/z* 1091.5645 $[M - H]^-$ (calcd for $C_{53}H_{87}O_{23}$, 1091.5638). The ^1H and ^{13}C NMR spectroscopic data of **18** (**Tables 7, 8**) showed the same structrure with **17** with the disappearance of the feruloyl moiety. Also, no maximum absorption at 325 on UV spectrum and no mass fragment 176 Da can be observed on possitive mass fragmentation. Further HMBC analysis confirmed the structure of compound **18** as shown in **Figure 40**. Hence, compound **18** was elucidated as 28-*O*-β-D-glucopyranosyl 23-hydroxy-longispinogenin 23-*O*-β-D-xylopyranosyl-(1→6)-β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside.

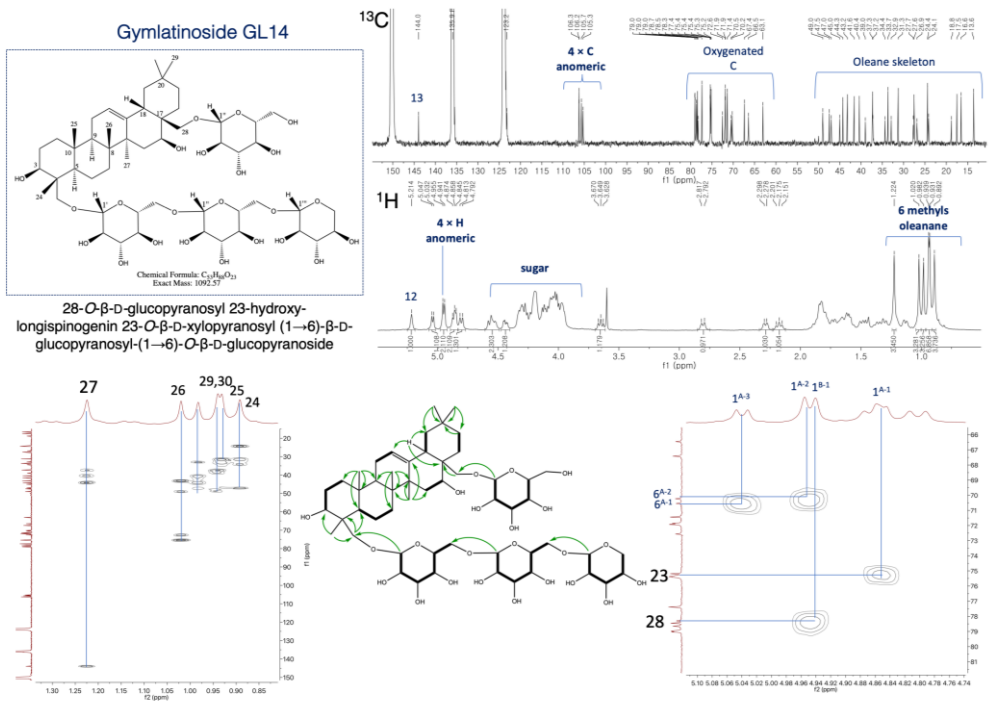


Figure 40. Gymlatinoside GL14

3.3. PTP1B inhibitory activities of isolated compounds

All isolates (**1**–**20**) were evaluated for their PTP1b inhibitory effects and compound **2**, **3** showed considerable inhibitory activities with IC_{50} values of 28.66 ± 2.57 mM and 19.83 ± 0.40 mM, respectively. Based on the inhibitory activities, compounds **2** and **3** were selected for further investigation of their PTP1B enzyme kinetics. Consequently, the compound **2**, **3** were found to be competitive PTP1B inhibitors as shown in Lineweaver-Burk plots (**Figure 41**).

Inhibition effects of compounds **1** - **9** on PTP1B

Compound	IC_{50} values (μM)[a]
1	>30
2	28.66 ± 2.57
3	19.83 ± 0.40
4	>30
5	>30
6	>30
7	>30
8	>30
9	>30
Ursolic acid[b]	1.02 ± 0.20

[a]Values are the means±SD, *n*=3.
[b]Positive control.

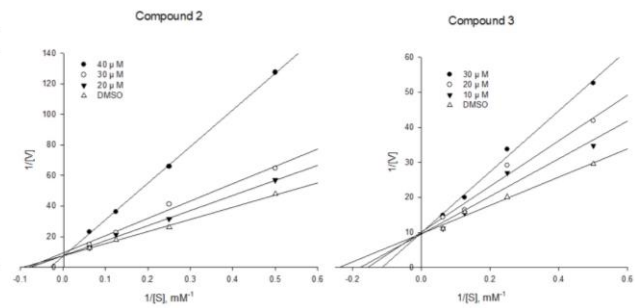


Figure 41. Lineweaver-Burk plots for determination of the type of PTP1B inhibition of compounds **2** and **3** using pNPP assay.

The conditions were as follows: 4 mM substrate, 0.05–0.1 μg/mL of PTP1B enzyme, 50 mM Tris (pH 7.5), at room temperature. In the presence of different concentrations of compounds for lines from bottom to top: A. Compound **2** (20, 30, and 40 μM); B. Compound **3** (10, 20, 30 μM). The data were evaluated in three replicates at each substrate concentration.

3.4. Building block assignments for oleanane glycosides in G.latifolium

3.4.1. 3-β-hydroxy oleanane glycosides

The isolated compounds suggested 3-β-hydroxy oleanane glycosides from *Gymnema latifolium* may be constructed by three main building block components

including a aglycone (gymnemagenin, marsglobiferin, longisopinogenin), a glycosyl chain at C-3 (β-D-arabino-2-hexulopyranosyl-(1→3)-β-D-glucuronopyranosyl, glucuronic acid or glucose) and 1−3 acyl moieties (acyl, tigloyl, Mba, benzoyl) posisted at C-16,21, 22, 28. A list of predicted structures was proposed using the combination of these main building block components following its construction rules. As a result, 75 peaks matched with the masses in the predicted list were extracted successfully from the total extract of *G. latifolium* or its microfractions with mass deviations of less than 5 ppm.

The extracted peaks were then analyzed in the MS^n TOF-positive mode by using the multiple key ion analysis (mKIA) and mass fragmentation analysis (MFA) techniques (**Figure 42**). The sturdy polycyclic structures of various oleanane triterpenoids can be characterized by different sets of dehydrated key MS^n ions during mass fragmentation. Therefore, the mKIA technique utilizing various sets of key positive MS^n ions can be used to effectively discriminate different aglycones in *G. latifolium*. In this study, gymnemagenin and its glycosides showed a set of key ions at *m*/*z* 453, 435, 417, 399, marsglobiferin and its analogs exhibited ions at *m*/*z* 473, 455, 437, 419 and longispinogenin display ions at *m*/*z* 441, 423, 405. On the other hand, the MFA technique was applied in qTOF-positive fragmentation mode to confirm the glycosyl and acyl moieties. Specifically, the β-D-arabino-2-hexulopyranosyl-(1→3)-*β*-D-glucuronopyranosyl (Ahg) portion was defined by a neutral loss of 336 Da, while the two sugars glucuronic acid and glucose were detected by the mass losses of 176 Da and 162 Da, respectively. The acyl moieties can be identified similarly by their neutral losses, such as acetyl (42 Da), tigloyl (82 Da), 2(*S*)-methylbutyloyl (84 Da) and benzoyl (104 Da).

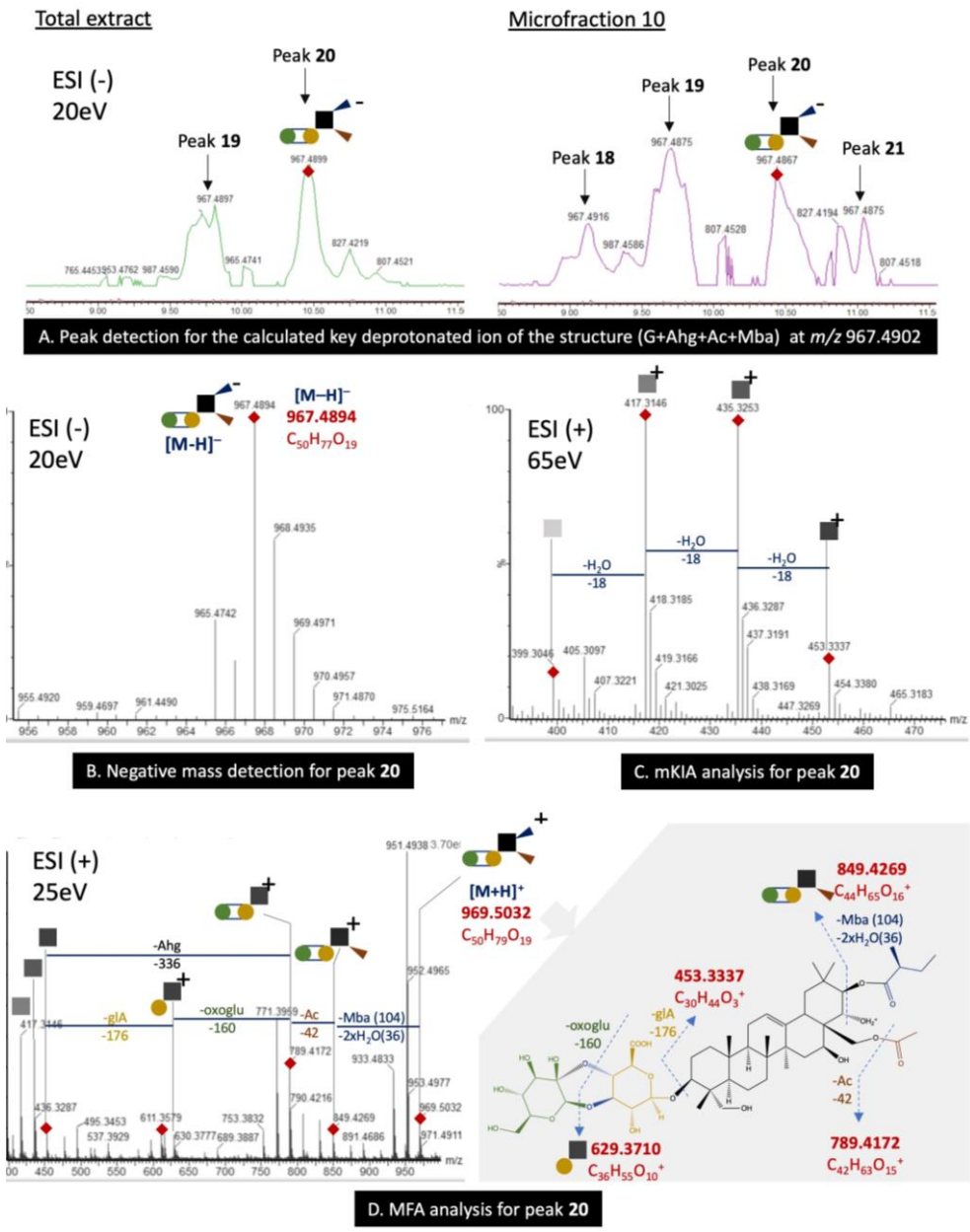


Figure 42. HRMS-based building block strategy to assign constituents in *G. latifolium*.

Figure 42A, B illustrated the application of building block strategy to detect target peaks for the predicted deprotonated ion mass at *m/z* 967.4902 of the molecular formula $C_{50}H_{76}O_{19}$ corresponding to the structure of

Gymenagenin+Ahg+Ac+Mba. The peak peaking resulted in the detection of 2 peaks **16** and **17** from the total extract, and 4 peaks **15–18** in the microfraction 10. The increase of the number of detected peaks and the better mass resolution were two major advantages of the microfractionation applied in this strategy. Structure of peak **20** (compound **1**) was analyzed in detailed by using mKIA analysis (**Figure 42C**) and MFA analysis (**Figure 42D**). 3 other peaks were expected to be structural isomers of compound **1** with the variation of the two acyl moeties at positions 16, 21, 22 and 28.

Finally, 56 of the 76 extracted peaks were confirmed using the same approach. These peaks included 7 previously undescribed compounds which were isolated in this study, 15 peaks can be assigned to reported compounds in the literature, and 34 peaks can be predicted as previously undescribed compounds because no compound indexed in SciFinder can match both of their molecular formulas and mass fragmentation patterns (**Table 9**).

Table 9. Compound assignments for 3-*β*-hydroxy oleanane glycosides *G. latifolium* using the building block strategy

Peak	R_t	*m/z* (-)	Cald. *m/z* (-)	Diff (ppm)	Formula	Exact mass	Building block compound assignment	CAS Number/ identification
1	7.283	505.3527	505.3529	-0.4	$C_{30}H_{49}O_6$	506.3605	G	**22467-07-8**
2	5.17	841.4216	841.4216	-0.7	$C_{42}H_{66}O_{17}$	842.4294	G+Ahg	No report
3	5.862	883.4313	883.4327	-1.6	$C_{44}H_{68}O_{18}$	884.4391	G+Ahg+Ac	No report
4	5.994	883.4352	883.4327	2.8	$C_{44}H_{68}O_{18}$	884.443	G+Ahg+Ac	No report
5	6.529	883.4337	883.4327	1.1	$C_{44}H_{68}O_{18}$	884.4415	G+Ahg+Ac	No report
6	6.801	883.4327	883.4327	1.4	$C_{44}H_{68}O_{18}$	884.4405	G+Ahg+Ac	No report
7	8.362	923.4655	923.464	1.6	$C_{47}H_{71}O_{18}$	924.4733	G+Ahg+Tig	**131653-20-8**
8	8.661	945.4488	945.4484	0.4	$C_{49}H_{70}O_{18}$	946.4566	G+Ahg+Bz	No report
9	8.704	925.4807	925.4797	1.1	$C_{47}H_{74}O_{18}$	926.4885	G+Ahg+Mba	**131653-19-5**
10	8.266	925.4448	925.4433	1.6	$C_{46}H_{70}O_{19}$	926.4526	G+Ahg+Ac+Ac	No report
11	9.204	965.4755	965.4746	0.9	$C_{49}H_{74}O_{19}$	966.4833	G+Ahg+Ac+Tig	No report
12	9.31	965.4757	965.4746	1.1	$C_{49}H_{74}O_{19}$	966.4835	G+Ahg+Ac+Tig	No report
13	9.695	965.4771	965.4746	2.6	$C_{49}H_{74}O_{19}$	966.4849	G+Ahg+Ac+Tig	No report
14	10.028	965.4753	965.4746	0.7	$C_{49}H_{74}O_{19}$	966.4831	G+Ahg+Ac+Tig	Compd. **2**
15	9.528	987.46	987.459	1	$C_{51}H_{72}O_{19}$	988.4678	G+Ahg+Ac+Bz	No report
16	9.695	987.4575	987.459	-1.5	$C_{51}H_{72}O_{19}$	988.4653	G+Ahg+Ac+Bz	No report
17	10.23	987.4595	987.459	0.5	$C_{51}H_{72}O_{19}$	988.4673	G+Ahg+Ac+Bz	Compd. **3**
18	9.362	967.4921	967.4903	1.9	$C_{49}H_{76}O_{19}$	968.4999	G+Ahg+Ac+Mba	No report
19	9.765	967.4899	967.4903	-0.4	$C_{49}H_{76}O_{19}$	968.4977	G+Ahg+Ac+Mba	No report
20	10.362	967.4896	967.4903	-0.7	$C_{49}H_{76}O_{19}$	968.4974	G+Ahg+Ac+Mba	Compd. **1**
21	10.827	967.4894	967.4903	-0.9	$C_{49}H_{76}O_{19}$	968.4972	G+Ahg+Ac+Mba	No report
22	8.792	967.4571	967.4539	3.3	$C_{48}H_{72}O_{20}$	968.4649	G+Ahg+Ac+Ac+Ac	No report
23	10.625	1007.4833	1007.4852	-1.9	$C_{51}H_{76}O_{20}$	1008.4911	G+Ahg+Ac+Ac+Tig	Compd. **4**
24	10.818	1007.4844	1007.4852	1.4	$C_{51}H_{76}O_{20}$	1008.4922	G+Ahg+Ac+Ac+Tig	No report
25	5.284	681.3831	681.385	-2.8	$C_{36}H_{58}O_{12}$	682.3909	G+glcA	**121686-42-8**
26	6.801	723.3965	723.3956	1.2	$C_{38}H_{60}O_{13}$	724.4043	G+glcA+Ac	**147934-05-2**
27	6.968	723.3975	723.3956	2.6	$C_{38}H_{60}O_{13}$	724.4053	G+glcA+Ac	No report
28	8.792	763.4274	763.4269	0.7	$C_{41}H_{63}O_{13}$	764.4352	G+glcA+tig	**121903-96-6, 155023-62-4**
29	9.108	785.4127	785.4158	1.9	$C_{43}H_{62}O_{13}$	786.4205	G+glcA+Bz	**154977-77-2, 154977-76-1**
30	8.959	765.4434	765.4425	1.2	$C_{41}H_{65}O_{13}$	766.4512	G+glcA+Mba	**122074-65-1**
31	9.195	765.4443	765.4425	2.4	$C_{41}H_{65}O_{13}$	766.4521	G+glcA+Mba	**155023-61-3**
32	7.406	765.4057	765.4061	-0.5	$C_{40}H_{62}O_{14}$	766.4135	G+glcA+Ac+Ac	No report
33	8.528	765.4079	765.4061	2.4	$C_{40}H_{62}O_{14}$	766.4157	G+glcA+Ac+Ac	No report
34	9.625	805.4387	805.4374	1.6	$C_{43}H_{66}O_{14}$	806.4465	G+glcA+Ac+Tig	**175033-15-5**
35	9.801	805.4376	805.4374	0.2	$C_{43}H_{66}O_{14}$	806.4454	G+glcA+Ac+Tig	**122168-40-5**
36	10.406	805.4397	805.4374	2.9	$C_{43}H_{66}O_{14}$	806.4475	G+glcA+Ac+Tig	**174232-51-0**
37	9.862	827.4237	827.4218	2.3	$C_{45}H_{64}O_{14}$	828.4315	G+glcA+Ac+Bz	No report
38	10.169	827.4219	827.4218	0.1	$C_{45}H_{64}O_{14}$	828.4297	G+glcA+Ac+Bz	**199618-65-0**
39	10.73	827.4203	827.4218	-1.8	$C_{45}H_{64}O_{14}$	828.4281	G+glcA+Ac+Bz	No report
40	9.897	807.4528	807.4531	-0.4	$C_{43}H_{68}O_{14}$	808.4606	G+glcA+Ac+Mba	No report
41	10.108	807.4529	807.4531	0.2	$C_{43}H_{68}O_{14}$	808.4607	G+glcA+Ac+Mba	**122144-48-3**
42	10.95	807.453	807.4531	-0.1	$C_{43}H_{68}O_{14}$	808.4608	G+glcA+Ac+Mba	No report
43	8.296	807.4177	807.4167	1.2	$C_{42}H_{64}O_{15}$	808.4255	G+glcA+Ac+Ac+Ac	No report
44	9.011	807.4182	807.4167	1.2	$C_{42}H_{64}O_{15}$	808.426	G+glcA+Ac+Ac+Ac	No report
45	9.195	807.4167	807.4167	0	$C_{42}H_{64}O_{15}$	808.4245	G+glcA+Ac+Ac+Ac	No report
46	11.000	847.4492	847.448	1.4	$C_{45}H_{68}O_{15}$	848.457	G+glcA+Ac+Ac+tig	Compd. **7**
47	11.081	869.4357	869.4323	3.9	$C_{47}H_{66}O_{15}$	870.4435	G+glcA+Ac+Ac+Bz	No report
48	11.301	849.4658	849.4636	2.6	$C_{45}H_{70}O_{15}$	850.4736	G+glcA+Ac+Ac+Mba	No report
49	9.537	489.3574	489.358	-1.2	$C_{30}H_{50}O_5$	490.3652	M	**147395-99-1**
50	6.143	665.3894	665.3901	-1.1	$C_{36}H_{58}O_{11}$	666.3972	M+glcA	No report
51	7.415	707.4018	707.4007	1.6	$C_{38}H_{60}O_{12}$	708.4096	M+glcA+Ac	No report
52	7.529	707.3998	707.4007	-1.3	$C_{38}H_{60}O_{12}$	708.4076	M+glcA+Ac	No report
53	8.301	749.4099	749.4112	-1.7	$C_{40}H_{62}O_{13}$	750.4177	M+glcA+Ac+Ac	No report
54	9.072	749.4119	749.4112	0.9	$C_{40}H_{62}O_{13}$	750.4197	M+glcA+Ac+Ac	No report
55	9.862	791.4228	791.4218	1.3	$C_{42}H_{64}O_{14}$	792.4306	M+glcA+Ac+Ac+Ac	Compd. **6**
56	9.318	781.4746	781.4738	1	$C_{42}H_{70}O_{13}$	762.4824	L+glc+glc	Compd. **5**

All of the detected peaks were 56 peaks of 3*β*-hydroxy oleanane triterpenoids were quickly identified and mapped on the plant metabolite fingerprint. A conventional total ion chromatogram (TIC) profiling (**Figure 43-A**) of the total

extract is typically noisy and limited in the number of metabolites (only 37/56 peaks can be detected in the total extract of *G. latifolium*). All the 56 peaks of 3β-hydroxy oleanane triterpenoids were effectively identified and mapped on a multilayer plant metabolite fingerprint (**Figure 43-B**).

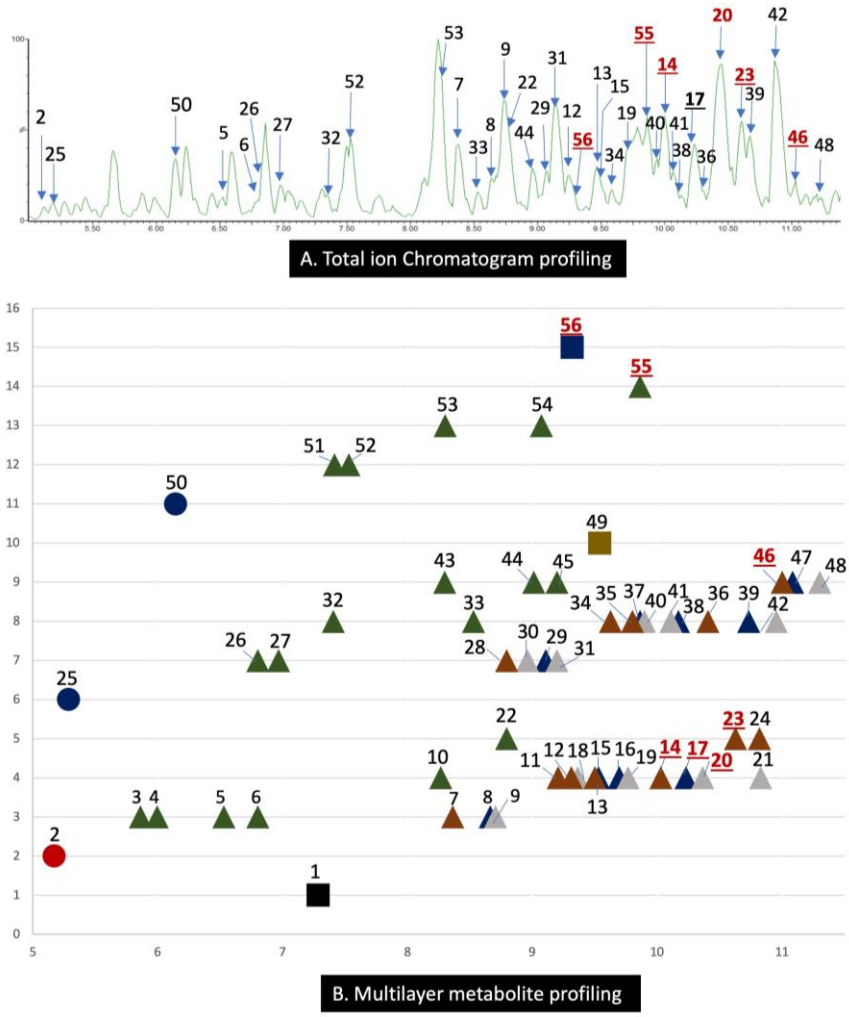


Figure 43. Metabolite profiling for 3β-hydroxy oleanane triterpenoids in *G. latifolium*

3.4.2. 23-hydroxy oleanane glycosides

Similar approach was applied for 23-hydroxy oleanane glycosides led to the assignment of 22 peaks as shown in **Table 10** and 20 peaks were expected to be new

compounds. Among them, the structure of 9 compounds (7 new and 2 known) were isolated and confirmed in this study.

Table 10. Compound assignments for 23-hydroxy oleanane glycosides in *G. latifolium* using the building block strategy

Peak	Compd assignment	R_t	*m/z* (-)	Cald *m/z* (-)	Diff (ppm)	RMD	Formula	Mass
1	GP+2×glc+xyl	6.231	991.5110	991.5114	-0.4	515.4	$C_{47}H_{78}O_{19}$	946.5133
2	GP+3×glc+xyl	5.362	1107.5583	1107.5587	-0.4	504.1	$C_{53}H_{88}O_{24}$	1108.5661
3	GP+glc+fer	7.494	827.4602	827.4582	2.4	556.2	$C_{46}H_{68}O_{13}$	828.4680
4	GP+glc+fer	8.362	827.4585	827.4582	0.4	554.1	$C_{46}H_{68}O_{13}$	828.4663
5	GP+2×glc+fer	6.029	989.5095	989.5110	-1.5	514.9	$C_{52}H_{78}O_{18}$	990.5173
6	GP+2×glc+fer	6.862	989.5118	989.5110	0.8	517.2	$C_{52}H_{78}O_{18}$	990.5196
7	GP+2×glc+fer	7.757	989.5129	989.5110	1.9	518.3	$C_{52}H_{78}O_{18}$	990.5207
8	GP+2×glc+xyl+fer	7.512	1121.5549	1121.5532	1.5	494.8	$C_{57}H_{86}O_{22}$	1122.5627
9	GP+2×glc+xyl+syn	6.485	1151.5658	1151.5638	1.7	491.3	$C_{58}H_{88}O_{23}$	1152.5736
10	GT+3×glc	7.073	959.5219	959.5216	0.3	543.9	$C_{48}H_{80}O_{19}$	960.5297
11	GT+2×glc+xyl	8.450	975.5170	975.5165	0.5	530.0	$C_{47}H_{78}O_{18}$	930.5193
12	GT+3×glc+xyl	6.862	1091.5645	1091.5638	0.6	517.1	$C_{53}H_{88}O_{23}$	1092.5723
13	GT+4×glc+xyl	5.661	1253.6143	1253.6166	-1.8	490.0	$C_{59}H_{98}O_{28}$	1254.6221
14	GT+glc+fer	10.309	811.4618	811.4633	-1.8	569.1	$C_{46}H_{68}O_{12}$	812.4696
15	GT+glc+fer	11.248	811.4628	811.4633	-0.6	570.3	$C_{46}H_{68}O_{12}$	812.4706
16	GT+2×glc+fer	7.503	973.5175	973.5161	1.4	531.6	$C_{52}H_{78}O_{17}$	974.5253
17	GT+2×glc+fer	8.266	973.5176	973.5161	1.5	531.7	$C_{52}H_{78}O_{17}$	974.5254
18	GT+2×glc+fer	8.528	973.5167	973.5161	0.6	530.8	$C_{52}H_{78}O_{17}$	974.5245
19	GT+2×glc+xyl+fer	9.695	1105.5579	1105.5583	-0.4	504.6	$C_{57}H_{86}O_{21}$	1106.5657
20	GT+3×glc+xyl+fer	7.862	1267.6147	1267.6147	-1.8	484.9	$C_{63}H_{96}O_{26}$	1268.6225
21	GT+2×glc+2×fer	9.503	1149.5629	1149.5634	-0.4	489.7	$C_{62}H_{86}O_{20}$	1150.5707
22	GT+2×glc+xyl+syn	8.072	1137.5837	1137.5845	-0.7	513.1	$C_{58}H_{90}O_{22}$	1138.5915

4. Conclusions

To the best of our knowledge, this is the first report on chemical composition and bioactivities of *Gymnema latifolium* Wall. ex Wight. Consequently, the authentication using conventional morphological technique together with DNA sequencing using the universally accepted sequence ITS1-5.8S-ITS2 will help the use of this new plant material correctly in the future studies. 20 oleanane analogs isolated from *G. latifolium* highlighted the chemical similarities of species in the same genus. Also, all the known compounds (including 6 isolated and 18 assigned structures) found in this study were reported from the genus *Gymnema*. The building block strategy using UPLC-MS^n qTOP also applied to identify quickly 78 structures, including 14 new compounds isolated in this study and 40 other oleanane glycosides which can be assigned as previously undescribed structures. Among them, the total extract of *G. latifolium* and compounds Gymlatinosides GL2 and GL3 showed considerable PTP1B inhibitory effect and can be study further for their antidiabetic activity.

Part 3: 12, 23-Dione dammarane triterpenoids from *Gynostemma longipes* and their muscle cell proliferation activities via activation of the AMPK pathway.

1. Introduction

One of the most distinctive effects of aging is the involuntary loss of muscle mass, strength and function, which is termed sarcopenia (Charlier et al., 2016). Skeletal muscle decreases approximately 3-8% per decade after 30 years of age, and muscle loss rapidly increases by about 25-30% after 60 years of age (English and Paddon-Jones, 2010). This loss of muscle mass, strength and function is the key cause of many diseases in elderly people, and the increased risk of injury and falls due to sarcopenia can also lead to continuous functional dependence and disability. The decreased muscle mass is also associated with a continuous increase in fat mass and consequently a change in body composition, and this in turn is associated with a higher incidence of insulin resistance in the elderly people (Jura and Kozak, 2016; Kalyani et al., 2014). Therefore, an increase of muscle mass in the elderly population is a promising approach to alleviate many of the symptoms of geriatric diseases (Dutt et al., 2015).

Recent studies suggested that muscle loss can be prevented by exercise in both young and older people (Denison et al., 2015). Conversely, increased physical inactivity and obesity due to a sedentary lifestyle are associated with attenuated muscle mass and reduced AMP-activated protein kinase (AMPK) activity (Fu et al., 2016; 2015). AMPK plays an important role in the regulation of cellular energy homeostasis, and the activation of AMPK promotes glucose uptake, fatty acid

oxidation, mitochondrial biogenesis and insulin sensitivity. In eukaryotes, there are three subunits of the AMPK, including the catalytic α-subunit (α1 and α2), regulatory β-subunit (β1 and β2) and γ-subunit (γ1, γ2 and γ3) (Ruderman et al., 2013) (Hardie et al., 2012). The regenerative process in skeletal muscle tissue generally involves a coordinated sequence of the activation of satellite cells, which proliferate to produce a population of myogenic progenitors (termed myoblasts) (Fu et al., 2015) (Asakura et al., 2002; Chen et al., 2007). Especially, the AMPKα1 isoform has an important role in enhancing Warburg-like glycolysis through non-canonical sonic hedgehog (Shh) signaling in myogenic cells during muscle regeneration. AMPKα1 knock-out in satellite cells reduces their activation and also inhibit the proliferation and differentiation of myogenic progenitor cells (Fu et al., 2015). Taken together, the activation of AMPK (especially AMPKα1) is a key mediator linking obesity and impaired muscle regeneration, providing an effective drug target to facilitate muscle regeneration in obese and elderly populations (Fu et al., 2016; 2015; Ruderman et al., 2013).

Our ethnopharmacological investigation of medicinal plants with AMPK activation activity resulted in the selection of *Gynostemma longipes*, commonly called “That diep dom”, a Vietnamese traditional herb used widely by the local community for tonic, treatment of diabetes and health strengthening. A fraction from this plant was then found to show potential activity in accelerating the muscle proliferation in the myoblast C2C12 cell model. The main ingredients of *G. longipes* are known as dammarane triterpenes (Anh et al., 2015; Guo et al., 1997) similar to other species of *Gynostemma* genus (C. Lee et al., 2015; Nguyen et al., 2011). Dammarane triterpenes have been also reported to play a significant role in the

development of natural drugs for the treatment of different metabolic diseases such as diabetes mellitus, metabolic syndrome, aging and neurodegenerative diseases (Cao et al., 2015). In this study, bioassay-guided fractionation and isolation of *G. longipes* were performed, and resulted in the purification of nine 12,23-dione dammarane triterpenes (**1**–**9**), including 6 new compounds longipenosides A1-A6, longipengenol, 3-dehydro longipengenol and a known compound gypentonoside A. All isolates were evaluated for their enhancement effects on muscle proliferation through activation of the AMPK pathway using the C2C12 myoblast cell model.

Purpose of research

- Authenticate the scientific name of medicinal plant.
- Identify the the chemical composition of *Gynostemma longipes VK1*.
- Evaluate the muscle cell proliferation activities via activation of the AMPK pathway of the isolates.

2. Materials and methods

2.1. Plant materials

Plant material was collected in the Ha Giang province of Vietnam in March 2016. The sample was identified as *G. longipes* C. Y. Wu based on the morphological characteristics by Prof. Tran Van On, Head of Department of Botany, Hanoi University of Pharmacy, Hanoi, Vietnam (Figure 44-A). A voucher specimen was deposited in the Medicinal Herbarium of Hanoi Univeristy of Pharmacy (HNIP) with the accession number HNIP.18500/16.

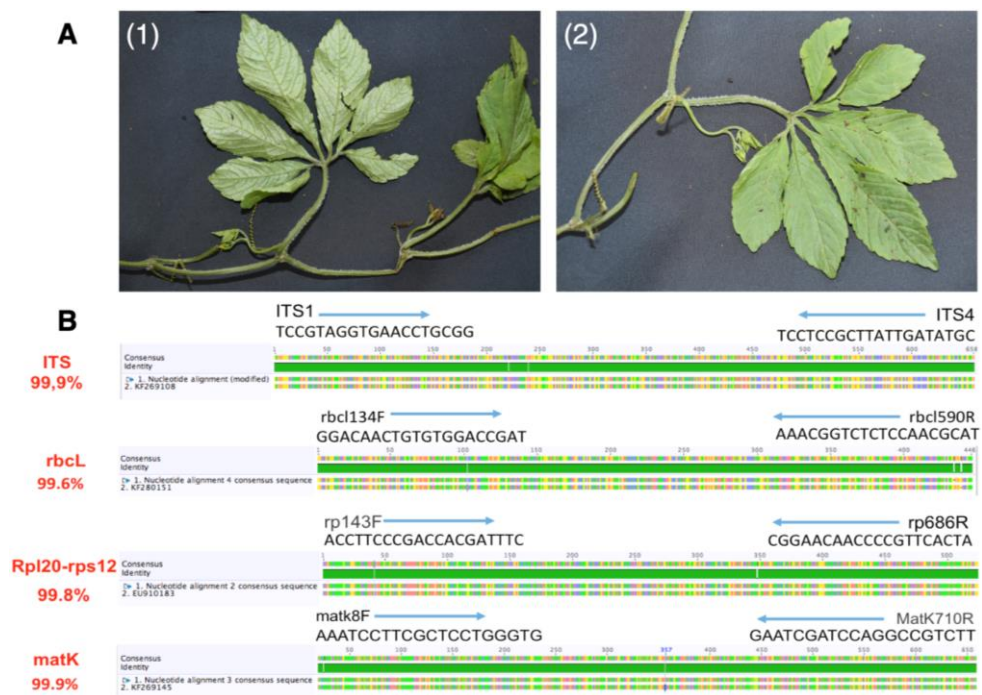


Figure 44. Morphological and DNA authentication of *G. longipes*.

(A) Authentication of *G. longipes* based on morphology (1) Whole plant with abaxial leaves, (2) adaxial leaf. (B) DNA analysis of *G. longpipes* with 4 DNA sequences ITS, rbcL, rpl20-rps12 and matK. All sequences of *Gynostemma longipes* were found to be at least 99.5% identical to sequences registered on Genebank.

Total DNA was extracted from 200 mg of fresh plant leaves using the Dneasy Plant Mini Kit (QIAGEN, Germany) with some modifications. PCR amplification and sequencing were performed by Macrogen Inc. (Seoul, Korea) using 4 pairs of primers: ITS1-F and ITS4-R for ITS1-5.8S-ITS2 region; matK8-F and matK710-R for partial maturase K gene, rbcL134-F and rbcL590-R for partial ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, rp143-F and rp686-R for rpl20-rps12 intergenic spacer region. The sequences of all the primers were shown in **Table 11**.

Table 11: Primers used to amplify and sequence 4 gene regions in *G. longipes*

No.	Gene sequence	Primers sequences
1	ITS1-5.8S-ITS2	ITS1-F (5′-TCCGTAGGTGAACCTGCGG-3′) ITS4-R (5′- TCCTCCGCTTATTGATATGC-3′)
2	Maturase K	matK8-F (5′-AAATCCTTCGCTCCTGGGTG-3′), matK710-R (5′-GAATCGATCCAGGCCGTCTT-3′)
3	Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	rbcL134-F (5′-GGACAACTGTGTGGACCGAT-3′) rbcL590-R (5′-AAACGGTCTCTCCAACGCAT-3′)
4	Intergenic spacer region	rp143-F (5′-ACCTTCCCGACCACGATTTC-3′) rp686-R (5′-CGGAACAACCCCGTTCACTA-3′)

All nucleotide sequence data were recorded in the GenBank database, National Center for Biotechnology Information (NCBI), with the accession numbers KX619478, KX619479, KX619480 and KX619481 respectively. DNA sequencing data were analyzed using the Geneious DNA sequencing analysis software (ver. 8.1.8, Biomatters Ltd, New Zealand). DNA sequences were compared to published sequences available in Genbank using the Basic Local Alignment Search Tool

(BLAST). All analyzed sequences were matched with *Gynostemma longipes* at pairwise identity 99.9% (ITS1-5.8S-ITS2 region), 99.6% (rbcL gene), 99.8% (rpl20-rps12 intergenic spacer region) and 99.9% (matK gene). This additional genetic information supported morphological authentication of *Gynostemma longipes* C.Y.Wu (**Figure 44-B**).

2.2. Isolation and structure elucidation of compounds

2.2.1. General experimental procedures

Optical rotations were measured in a JASCO P-2000 polarimeter (JASCO International Co. Ltd., Tokyo, Japan). IR data were recorded on a Nicolet 6700 FT-IR spectrometer (Thermo Electron Corp., Waltham, MA, USA). NMR data were analyzed using an AVANCE 500 MHz spectrometer (Bruker, Germany) or a JNM-ECA 600 MHz spectrometer (Jeol, Japan). HRESIMS values were analyzed using an Agilent Technologies 6130 Quadrupole LC/MS spectrometer equipped with an Agilent Technologies 1260 Infinity LC system (Agilent Technologies, Inc., Santa Clara, CA, USA) and an INNO C18 column (4.6 × 150 mm, 5 μm particle size, 12 nm, J.K.Shah & Company, Korea). Silica gel (particle size: 63–200 μm) and RP-C_{18} (particle size: 40–63 μm) were purchased from Merck (Darmstadt, Germany) and Sephadex LH-20 from Sigma-Aldrich (St. Louis, MO, USA) were used for column chromatography (CC). Silica gel 60 F_{254} and RP-18 F_{254} TLC plates were obtained from Merck (Darmstadt, Germany). A Gilson HPLC purification system, equipped with an Optima Pak C_{18} column (10 × 250 mm, 10 μm particle size; RS Tech, Seoul, Korea), was used at a flow rate of 2 mL/min and UV detection at 205 and 254 nm.

2.2.2. Isolation scheme

The aerial part of *G. longipes* (1.5 kg) was powdered, sonicated with 70% EtOH (3 × 3L), filtered and the solvent was evaporated in vacuo. The crude extract obtained (262 g) was suspended in water, absorbed onto Sephabeads SP70 resin and washed with water, 50% EtOH, 95% EtOH, 100% EtOH and acetone in a sequential elution process. The 95% EtOH fraction (75 g) was subjected to silica gel column chromatography (10 × 35 cm; 63–200 μm particle size) using a gradient of EtOAc:MeOH:H_2O (from 20:1:0.1 to 4:1:0.1) to give 9 sub-fractions (C1-C9) based on the thin-layer chromatography profile. Compound **1** (50.5 g) was obtained by direct crystallization from sub-fraction C7.

Sub-fraction 8 was chromatographed on C18- reversed phase silica gel (RP-C18) column chromatography eluting with MeOH:H_2O (from 1:1 to 1:0) and HPLC (Optima Pak C_{18}, CH_3CN:H_2O (isocratic 2:3), flow rate 2 mL/min) to yield compounds **9** (16.0 mg) and **3** (21.3 mg).

Sub-fraction C9 was chromatographed on RP-C18 column chromatography [CH_3OH:H_2O (from 1:1 to 1:0)] to yield 8 subfractions C9-1 to C9-8. Fraction C9.R3 by preparative HPLC (Optima Pak C_{18}, CH_3CN:H_2O (1:2), flow rate 2 mL/min) gave compounds **5** (10.5 mg) and **8** (18.5 mg). Fraction C9.R4 followed by HPLC (Optima Pak C_{18}, CH_3CN:H_2O (2:3), flow rate 2ml/min) led to the isolation of compounds **2** (12.5 mg), and **4** (20.4 mg).

The 100% EtOH fraction was applied sequentially to Sephadex LH-20 using 100% MeOH and HPLC (Optima Pak C_{18}, CH_3CN:H_2O (7:3), flow rate 2 mL/min) to yield compounds **6** (22.2 mg) and **7** (17.1 mg).

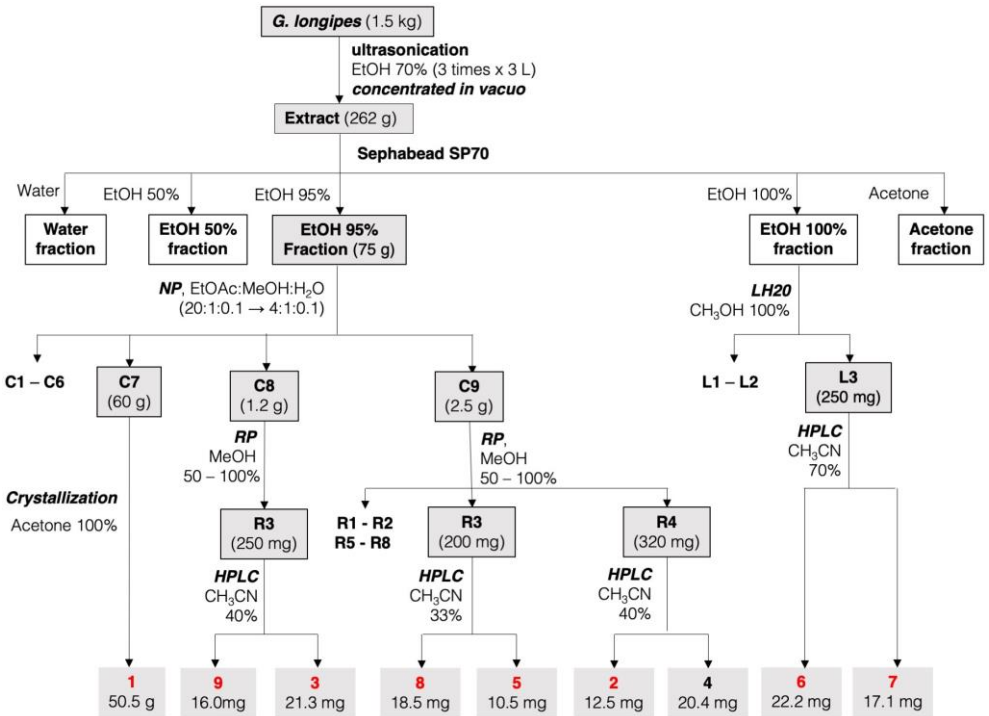


Figure 45. Isolation scheme of compounds **1** – **9** from of *G. longipes*.

2.2.3. Physical and chemical characteristics of isolated compounds

*Longipenoside A1 (**1**)*: white crystal; $[\alpha]_D^{25}$ -32.0 (*c* 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 244 nm (3.8); IR (KBr) ν_{max} 3404, 2973, 2938, 1697, 1616, 1073, 1055, 1010 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS *m/z* 925.5157 $[M-H]^-$ (calcd for $C_{48}H_{77}O_{17}$, 925.5166).

*Longipenoside A2 (**2**)*: white, amorphous powder; $[\alpha]_D^{25}$ -37.5 (*c* 0.1, MeOH); UV (MeOH); λ_{max} (log ε) 244 nm (3.8); IR (KBr) ν_{max} 3400, 2970, 2938, 1698, 1617, 1069, 1053, 1008 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS *m/z* 1087.5682 $[M-H]^-$ (calcd for $C_{54}H_{87}O_{22}$, 1087.5694).

*Longipenoside A3 (**3**)*: white, amorphous powder; $[\alpha]_D^{25}$ -35.2 (*c* 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 244 nm (3.8); IR (KBr) ν_{max} 3402, 2971, 2938, 1702, 1616,

1060, 1054, 1010 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS *m/z* 967.5276 $[M-H]^-$ (calcd for $C_{50}H_{79}O_{18}$, 967.5272).

*Longipenoside A4 (**4**)*: white, amorphous powder; $[\alpha]_D^{25}$ -33.0 (*c* 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 244 nm (3.7); IR (KBr) ν_{max} 3403, 2972, 2940, 1702, 1616, 1060, 1054, 1010 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS *m/z* 779.4597 $[M-H]^-$ (calcd for $C_{42}H_{67}O_{13}$, 779.4587).

*Longipengenol (**6**)*: white, amorphous powder; $[\alpha]_D^{25}$ -33.0 (*c* 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 244 nm (3.9); IR (KBr) ν_{max} 3403, 2972, 2940, 1702, 1616, 1060, 1010 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS *m/z* 471.3393 $[M-H]^-$ (calcd for $C_{30}H_{47}O_4$, 471.3480).

*3-dehydro longipengenol (**7**)*: white, amorphous powder, $[\alpha]_D^{25}$ -38.0 (*c* 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 244 nm (3.8); IR (KBr) ν_{max} 3403, 2972, 2940, 1702, 1616, 1060, 1054, 1010 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS *m/z* 471.3452 $[M+H]^+$ (calcd for $C_{30}H_{47}O_4$, 471.3469).

*Longipenoside A5 (**8**)*: white, amorphous powder, $[\alpha]_D^{25}$ -32.5 (*c* 0.1, MeOH); IR (KBr) ν_{max} 3403, 2972, 2940, 1702, 1616, 1060, 1054, 1010 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS *m/z* 943.5374 $[M-H]^-$ (calcd for $C_{48}H_{79}O_{18}$, 943.5380).

*Longipenoside A6 (**9**)*: White amorphous powder, $[\alpha]_D^{25}$ -33.0 (*c* 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 275nm (3.6); IR (KBr) ν_{max} 3403, 2972, 2940, 1702, 1616, 1060, 1054, 1010 cm^{-1}; ^{13}C and ^{1}H NMR data, Tables 12 and 13; HRESIMS data at *m/z* 907.5066 $[M-H]^-$ (calcd for $C_{48}H_{75}O_{16}$, 907.5061).

2.2.4. Acid hydrolysis

20 mg of Longipenoside A1 (Compounds **1**) was hydrolyzed with 1N HCl (MeOH 70%, 1 mL) at 90 °C for 1 hours. The solution was neutralized by NaOH 10%, dried and suspended in H_2O and partitioned with EtOAc. The residual H_2O layer was concentrated and dissolved in pyridine (1.0 mL) and 5.0 mg of L-cysteine methyl ester hydrochloride was added. The mixture was kept for 1 hours at 60 °C, and then 4.4 μL of benzylisothiocyanate was added. This solution was filtered through a 0.2-μm Whatman hydrophilic membrane filter into an HPLC sample vials immediately before LC-MS analysis. Analysis was performed using an Agilent 1200 HPLC system (Agilent Technologies, Palo Alto, CA, USA) with INNO C18 (4.6 × 250 mm inner diameter, 5 μm particle size; Young Jin Bio Chrom Co., Ltd) at a column temperature of 30 °C. Chromatographic condition was carried out using a mobile phase of 27% CH_3CN/H_2O isocratic elution at a flow rate 0.6 mL/min in 60 minutes. The sugar derivatives showed the retention time identical to the trimethylsilyl-L-cysteine derivatives of authentic D-glucose and L-rhamnopyranose, respectively.

*2.2.5. Quantitative analysis of Longipenoside A1 (compound **1**)*

The plant extracts were prepared using different extraction solvents with water, 30% EtOH, 50% EtOH, 70% EtOH, 100% EtOH and MeOH, respectively. After the extracts were dried *in vacuo* and lyophilized respectively, the extraction yield of each extract was calculated by the percentage of each extract in 1 g of raw material. For preparation of stock solutions, plant extracts were dissolved in extracted solvents at concentrations of 1 mg/mL. This solution was filtered through a 0.2-μm Whatman hydrophilic membrane filter into an HPLC sample vials immediately before LC-MS

analysis. Analysis and validation were performed using an Agilent 1200 HPLC system (Agilent Technologies, Palo Alto, CA, USA) with INNO C18 (4.6 × 250 mm inner diameter, 5 μm particle size; Young Jin Bio Chrom Co., Ltd) at a column temperature of 30°C. All chromatographic separations were carried out using a mobile phase of 10% to 100% CH_3CN/H_2O gradient elution at a flow rate 0.6 mL/min. A calibration curve was drawn using the crystal form of compound Longipenoside A1 (**1**) at concentrations ranging from 0.025 mg/mL to 0.2 mg/mL. The regression equation is expressed as y = 11333x + 55.578 (correlation coefficient R^2 = 0.9990), where y and x correspond to the peak area and concentration, respectively. The relative amount of compound in each of the extracts (mg/g of the extract) were calculated using this equation.

2.3. Muscle regeneration activity of G.longipes extract and isolated compounds

2.3.1. Cell proliferation assay

C2C12 cell lines (mouse skeletal myoblasts) were maintained in growth medium consisting of Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Gibco, NY, USA), and cells were incubated at 37 °C and 5% CO_2. The cell proliferation assay was carried out using the (3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) (Sigma-Aldrich, St. Louis, MO, USA) or the WST-1 cell proliferation reagent (Roche, Mannheim, Germany). Briefly, C2C12 cells were seeded onto a 96-well plate at 4,000 cells/well and incubated for 8 hours. The cells were then treated with different concentrations of compounds or fraction dissolved in low-serum medium containing 5% FBS. After 24 hours or 48 hours of incubation, 2 mg/mL MTT solution (20 μL) was added to each well and incubated for 4 hours. Formazan was

dissolved in 100 μL of DMSO and then absorbance was measured with an Absorbance Microplate Reader (VersaMax^{TM}, Randor, PA, USA) at 550 nm. In addition, 10 μL of this reagent was added to each well to evaluate cell viability using the WST-1 cell proliferation reagent, and incubated for 2 hours, and absorbance was measured at 440 nm. For the cell counting assay, experiments were performed using 36 mm cell culture dishes. After 48 hours of incubation with the tested compound, cells were harvested and evaluated with a Countess^{TM} automated cell counter (Invitrogen, Eugene, OR, USA). For measuring the cell viability of cancer cell lines, MCF-7 and MDA-MB 231 cells were seeded onto 96-well plate at 4,000 cells/well and incubated for 8 hours. Next, cells were cultured in DMEM containing various compounds. After 2 days of incubation, cell viability was evaluated by the above procedure.

2.3.2. Western blot analysis

C2C12 myoblasts were seeded into 6-well plates and incubated until a confluence of 70-80%. To prepare C2C12 myotubes, cells were incubated with DMEM supplemented with 2% horse serum (Gibco, NY, USA). The differentiation medium was changed every 2 days until myotubes were formed. For Western blot analysis, myotubes or myoblasts were incubated with various concentrations of compounds or fractions for 30-60 minutes. The cells were washed with cold PBS and lysed using the lysis buffer [50 mM Tris-HCl (pH 7.6), 120 mM NaCl, 1 mM EDTA, 0.5% NP-40, 50 mM NaF] and centrifuged at 12,000 rpm for 20 minutes. The protein concentrations were determined using a protein assay kit (Bio-Rad Laboratories, Inc., CA, USA). Aliquots of lysates were electrophoresed on 8% or 12% SDS-polyacrylamide gels and then transferred electronically to

polyvinylidene fluoride (PVDF) membranes (PVDF 0.45 μm, Immobilon-P, USA). Membranes were then incubated with primary antibodies for *p*-AMPKα Thr^{172}, AMPKα, *p*-ACC Ser^{79}, ACC (Cell signaling Technology, Inc., Beverly, MA, USA) or mouse monoclonal actin (Abcam, Cambridge, UK). After incubation with secondary antibodies, membranes were detected using an enhanced chemiluminescence Western blot detection kit (Thermo sci., Rockford, IL, USA).

2.3.3. Immunochemical staining with BrdU antibody

C2C12 myoblasts were grown on sterile glass coverslips and the cells were treated with test samples. After incubation for 8 hours, 50 μM BrdU labeling solution (Sigma-Aldrich, St. Louis, MO, USA) was added to the culture cells followed by incubation at 37 °C for 2 hours. The cells were then washed three times with PBS and fixed with 4% formaldehyde in PBS for 15 minutes at room temperature. After washing three times with PBS, the cells were permeabilized with Triton X-100 buffer (Sigma-Aldrich, St. Louis, MO, USA) for 20 minutes at room temperature. Next, the cells were incubated on ice for 20 minutes with 2N HCl solution. After removing the acid solution, the cells were incubated with phosphate/citric acid buffer (pH 7.4) for 10 minutes at room temperature. Cells were washed three times with Triton X-100 permeabilization buffer and incubated overnight with anti-BrdU monoclonal antibody (Santa Cruz, CA, USA) diluted 1:50 in PBS (pH 7.4). Cells were continuously incubated with FITC-conjugated goat anti-mouse lgG (Life technologies, Eugene, OR, USA) for 1 hour and then stained with 500 nM DAPI solution (Life technologies, Eugene, OR, USA) for 10 minutes at room temperature. The slides were fixed with a mounting medium for fluorescence detection (Vectashield®, Vector Lab, Inc., USA) and images were observed using fluorescence

microscopy (Olympus ix70 Fluorescence Microscope, Olympus Corporation, Tokyo, Japan).

2.3.4. Flow cytometry analysis for BrdU and PI staining

For flow cytometry with anti-BrdU antibody staining, C2C12 myoblasts were maintained in 36-mm culture dishes. After treatment with different concentrations of the test compounds, the cells were incubated for 8 hours. Next, 50 μM BrdU labeling solution was added to the culture medium and incubated for 2 hours. After removing BrdU medium, the cells were washed with PBS and harvested by trypsin solution. Cells were permeabilized using precooled 70% ethanol solution and further incubated at 4°C for 1 hour. After centrifugation at 1000 rpm for 2 minutes, the pellet cells were continually added to a 2N HCl solution containing 0.5% Triton X-100 reagent and incubated at room temperature for 30 minutes. Cells were then washed with PBS containing 1% BSA (Sigma-Aldrich, St. Louis, MO, USA) and resuspended in PBS containing 0.5% Tween 20 and 1% BSA. Cells were incubated with anti-BrdU monoclonal antibody (Santa Cruz, CA, USA) for 4 hours at room temperature. Cells were resuspended in PBS containing 20 μg/mL propidium iodide (PI) (Merck Millipore, MA, USA) and kept in the dark for 30 minutes. Then, double-stained cells were analyzed using a flow cytometer (BD, FACSCalibur, San Jose, CA, USA) and at least 10,000 cells were counted for each analysis.

2.3.5. Flow cytometry of cell cycle status

C2C12 myoblasts were maintained in 36-mm culture dishes and cells were treated with the test compounds. After 12 or 24 hour - incubation, cells were harvested using trypsinization and washed twice with pre-cooled PBS. Cells were

then fixed with pre-cooled 70% ethanol solution and incubated overnight at 4 °C. After washing with PBS, cells were stained sequentially with 20 μg/mL of PI in PBS solution and incubated at room temperature for 30 minutes. Cellular DNA content was analyzed using a flow cytometer (BD, FACSCalibur, San Jose, CA, USA). At least 10,000 cells were counted for each assay. At each phase of the cell cycle, cell distribution was displayed using a histogram.

2.3.6. Glucose uptake assay

C2C12 myoblasts were seeded into black 96-well plates at 10^4 cells/well using glucose-free medium containing 10% FBS. After 1 day of incubation, cells were treated with test samples or insulin (as a positive control) in the presence or absence of 2-NBDG (Invitrogen, Eugene, OR, USA). The cells were continually incubated for 1 hour at 37°C and they were then washed with cold PBS. The fluorescence intensity was measured with excitation/emission at 450/535 nm using a fluorescence microplate reader (Spectra Max Gemini XPS, Molecular Devices, San Jose, CA, USA).

2.3.7. Measurement of ATP level

Intracellular ATP production was evaluated using ATP bioluminescence determination assay kit (Invitrogen, Eugene, OR, USA). C2C12 myoblasts were grown on 24-well plates overnight and the cells were then exposed to test samples from 0 to 70 minutes. After that, the cells were lysed using lysis buffer (100 μL). The cell lysates and ATP standard dilutions (20 μL) were transferred to white 96-well plates and then luciferase mixture (80 μL) was added to each well. ATP levels were measured using a luminescence microplate reader (Spectra Max L, Molecular Devices, San Jose, CA, USA).

2.3.8. Statistical analysis

All data were expressed as the mean ± SD of 2-3 independent experiments. The difference between group mean values was calculated by analysis of variance (ANOVA) followed by Duncan's or Tukey's post hoc test as appropriate, conducting in SPSS Statistics 23 (SPSS, Inc., Chicago, IL, USA). Statistical significance was accepted at * $p < 0.05$, and ** $p < 0.01$, *** $p < 0.001$.

3. Results and discussion

3.1. Structural determination of new compounds

Bioassay-guided fractionation and isolation of the bioactive fraction from the leaves of *G. longipes* afforded eight new dammarane triterpenes (**1**–**3** and **5**–**9**), along with one known compound gypentonoside A (**4**) (**Figure 46**).

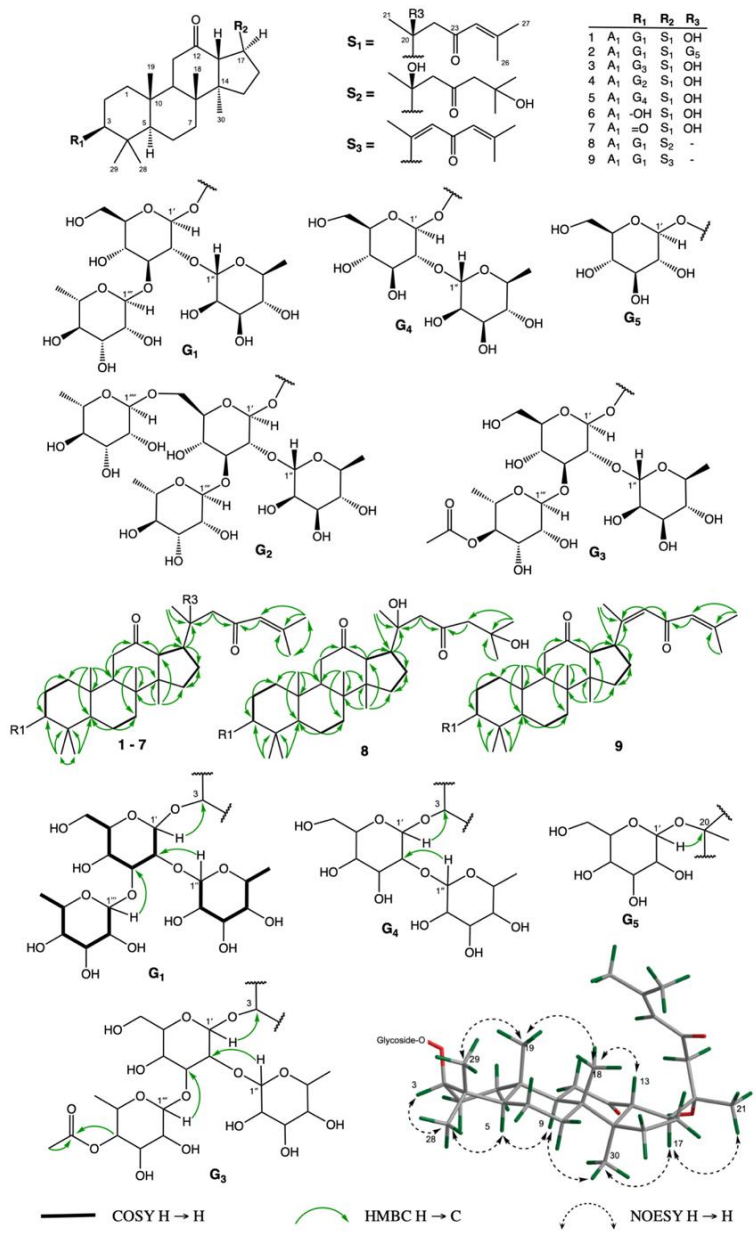


Figure 46. Nine compounds (**1**–**9**) isolated from *Gynostemma longipes*.

Table 12. ^{13}C NMR spectroscopic data for new compounds **1**–**3** and **5**–**9** (in Pyridine-d_5).

	1	**2**[#]	**3**[##]	**5**	**6**	**7**	**8**	**9**
position	δ_C, type 125 MHz	δ_C, type 150 MHz	δ_C, type 150 MHz	δ_C, type 125 MHz	δ_C, type 150 MHz	δ_C, type 125 MHz	δ_C, type 125 MHz	δ_C, type 125 MHz
1	39.3, CH_2	39.4, CH_2	39.4, CH_2	39.4, CH_2	39.3, CH_2	39.5, CH_2	39.4, CH_2	39.4, CH_2
2	26.9, CH_2	26.9, CH_2	27.0, CH_2	27.1, CH_2	29.0, CH_2	34.4, CH_2	26.9, CH_2	28.7, CH_2
3	88.5, CH	88.6, CH	88.6, CH	88.8, CH	78.1, CH	216.2, C	88.7, CH	88.7, CH
4	39.9, C	39.9, C	39.9, C	40.0, C	39.9, C	47.7, C	40.0, C	39.8, C
5	56.6, CH	56.6, CH	56.7, CH	56.7, CH	56.7, CH	55.3, CH	56.7, CH	56.6, CH
6	18.8, CH_2	18.8, CH_2	18.8, CH_2	18.9, CH_2	19.2, CH_2	20.3, CH_2	18.9, CH_2	19.0, CH_2
7	34.8, CH_2	35.0, CH_2	34.8, CH_2	34.9, CH_2	34.9, CH_2	34.1, CH_2	34.9, CH_2	34.8, CH_2
8	40.9, C	41.1, C	41.0, C	41.0, C	41.0, C	40.9, C	41.1, C	41.0, C
9	54.6, CH	54.9, CH	54.8, CH	54.7, CH	54.9, CH	54.0, CH	54.7, CH	54.1, CH
10	37.7, C	37.7, C	37.7, C	37.7, C	38.2, C	37.7, C	37.8, C	37.7, C
11	40.1, CH_2	40.3, CH	40.1, CH_2	40.2, CH	40.3, CH_2	40.3, CH_2	40.2, CH	39.8, CH_2
12	211.7, C	211.4, C	211.8, C	211.8, C	211.9, C	211.2, C	211.9, C	209.9, C
13	56.2, CH	56.5, CH	56.3, CH	56.3, CH	56.3, CH	56.8, CH	56.9, CH	59.0, CH
14	56.5, C	56.5, C	56.5, C	56.7, C	56.7, C	56.3, C	56.3, C	55.3, C
15	32.2, CH_2	32.6, CH_2	32.2, CH_2	32.3, CH_2	32.3, CH_2	32.3, CH_2	32.2, CH_2	32.2, CH_2
16	25.2, CH_2	25.3, CH_2	25.2, CH_2	25.2, CH_2	25.2, CH_2	25.2, CH_2	25.2, CH_2	28.7, CH_2
17	45.2, CH	43.5, CH	45.2, CH	45.3, CH	45.3, CH	45.2, CH	45.4, CH	45.7, CH
18	16.0, CH_3	16.2, CH_3	16.0, CH_3	16.1, CH_3	16.1, CH_3	15.7, CH_3	16.0, CH_3	16.1, CH_3
19	16.5, CH_3	16.6, CH_3	16.5, CH_3	16.5, CH_3	16.5, CH_3	16.0, CH_3	16.5, CH_3	16.4, CH_3
20	74.0, C	80.7, C	74.1, C	74.1, C	74.1, C	74.1, C	73.9, C	159.7, C
21	27.2, CH_3	23.7, CH_3	27.2, CH_3	27.3, CH_3	27.2, CH_3	27.3, CH_3	27.0, CH_3	17.7, CH_3
22	54.8, CH_2	55.1, CH_2	55.0, CH_2	55.0, CH_2	55.0, CH_2	55.0, CH_2	55.1, CH_2	126.3, CH
23	201.7, C	199.8, C	201.7, C	201.7, C	201.8, C	201.7, C	212.9, C	191.7, C
24	126.4, CH	126.6, CH	126.5, CH	126.6, CH	126.6, CH	126.5, CH	54.8, CH	127.5, CH
25	154.9, C	154.5, C	154.9, C	154.9, C	154.9, C	155.0, C	70.0, C	154.1, C
26	20.9, CH_3	20.8, CH_3	21.0, CH_3	21.0, CH_3	21.0, CH_3	21.0, CH_3	30.8, CH_3	20.8, CH_3
27	27.7, CH_3	27.6, CH_3	27.7, CH_3	27.7, CH_3	27.7, CH_3	27.7, CH_3	30.4, CH_3	27.7, CH_3
28	28.1, CH_3	28.1, CH_3	28.2, CH_3	28.3, CH_3	28.4, CH_3	27.0, CH_3	28.2, CH_3	28.2, CH_3
29	16.9, CH_3	16.9, CH_3	16.8, CH_3	17.2, CH_3	16.6, CH_3	21.5, CH_3	17.0, CH_3	17.1, CH_3
30	17.4, CH_3	17.3, CH_3	17.5, CH_3	17.5, CH_3	17.5, CH_3	17.3, CH_3	17.5.CH_3	17.7, CH_3
1′	105.3, CH	105.3, CH	105.1, CH	105.8, CH			105.3, CH	105.4, CH
2′	78.3, CH	78.5, CH	78.6, CH	78.1, CH			78.4, CH	78.5, CH
3′	87.5, CH	87.6, CH	87.8, CH	80.3, CH			87.8, CH	87.7, CH
4′	70.6, CH	70.4, CH	70.6, CH	72.5, CH			70.8, CH	70.8, CH
5′	78.3, CH	78.3, CH	78.6, CH	78.7, CH			78.4, CH	78.4, CH
6′	62.8, CH_2	62.8, CH	62.8, CH_2	63.3, CH_2			63.0, CH_2	63.0, CH_2
1″	102.5, CH	102.4, CH	101.9, CH	102.1, CH			102.5, CH	102.6, CH
2″	72.2, CH	71.0, CH	70.3, CH	72.8, CH			72.4, CH	72.5, CH
3″	72.7, CH	72.7, CH	72.3, CH	72.9, CH			72.8, CH	72.9, CH
4″	73.9, CH	73.9, CH	76.0, CH	74.5, CH			74.1, CH	74.1, CH
5″	70.5, CH	70.6, CH	67.6, CH	69.9, CH			70.5, CH	70.6, CH
6″	18.9, CH_3	18.9, CH_3	18.2, CH_3	19.1, CH_3			18.9, CH_3	18.9, CH_3
1‴	103.9, CH	104.0, CH	104.1, CH				104.0, CH	104.1, CH
2‴	72.7, CH	72.6, CH	72.7, CH				72.9, CH	72.7, CH
3‴	72.9, CH	73.0, CH	72.9, CH				73.0, CH	73.0, CH
4‴	73.8, CH	73.8, CH	73.8, CH				73.9, CH	74.0, CH
5‴	71.0, CH	71.0, CH	71.1, CH				71.1, CH	71.2, CH
6‴	18.7, CH_3	18.7, CH_3	18.7, CH_3				18.7, CH_3	18.8, CH_3

[#] Chemical shift of *20-0-glc*: C-1‴ (98.8, CH), C-2‴ (73.0, CH), C-3‴ (72.9, CH), C-4‴ (74.0, CH), C-5‴ (78.2, CH), C-6‴ (63.3, CH_2); [##] Chemical shift of *3″-Acetyl*: 171.9 (C=0); 21.4 (CH_3).

Table 13. ^{1}H NMR spectroscopic data for compounds **1**–**3** and **5**–**9**.

	1	**2**[#]	**3**[##]	**5**	**6**	**7**	**8**	**9**
position	$\delta_{H,}$ (*J* in Hz) 500 MHz	$\delta_{H,}$ (*J* in Hz) 600 MHz	$\delta_{H,}$ (*J* in Hz) 600 MHz	$\delta_{H,}$ (*J* in Hz) 500 MHz	$\delta_{H,}$ (*J* in Hz) 600 MHz	$\delta_{H,}$ (*J* in Hz) 500 MHz	$\delta_{H,}$ (*J* in Hz) 500 MHz	$\delta_{H,}$ (*J* in Hz) 500 MHz
1.	1.20, overlap	1.18, overlap	1.20, overlap	1.22, overlap	1.20, overlap	1.58, m	1.24, m	1.22, overlap
	0.69, overlap	0.69, overlap	0.69, overlap	0.73, overlap	0.69, overlap	1.23, m	0.75, m	0.73, overlap
2.	2.11, overlap	2.12, overlap	2.11, overlap	2.23, overlap	2.10, overlap	2.48, m	2.15, m	2.04, m
	1.70, overlap	1.71, overlap	1.70, overlap	1.80, overlap	1.70, overlap	2.42, m	1.75, m	1.56, overlap
3.	3.25, overlap	3.26, dd (11.2, 4.4)	3.32, dd (11.2, 4.4)	3.32, dd (11.6, 4.0)	3.40, dd (10.8, 5.4)	3.28, dd (11.7, 4.1)	3.28, dd (11.7, 4.1)	3.30, dd (11.7, 4.2)
5.	0.66, overlap	0.65, overlap	0.66, m	0.67, d (9.6)	0.80, overlap	1.35, m	0.69, m	0.69, overlap
6.	1.50, overlap	1.49, overlap	1.50, m	1.49, overlap	1.50, overlap	1.50, m	1.54, m	1.52, m
	1.39, overlap	1.40, overlap	1.39, m	1.41, overlap	1.39, overlap	1.41, m	1.41, m	1.42, m
7.	1.39, overlap	1.38, overlap	1.39, m	1.42, overlap	1.39, overlap	1.43, m	1.41, m	1.45, m
	1.25, overlap	1.25, overlap	1.25, m	1.26, overlap	1.25, overlap	1.28, m	1.27, m	1.27, m
9.	1.62, overlap	1.79, overlap	1.62, m	1.65 (dd, 13.0, 4.0)	1.62, overlap	1.78, m	1.66, overlap	1.62, m
11.	2.24, overlap	2.24, overlap	2.24, overlap	2.23, overlap (2H)	2.34, overlap	2.37, overlap	2.24, overlap	2.27, overlap
	2.22, dd (10.5, 3.1)	2.22, overlap	2.22, dd (10.5, 3.1)		2.32, overlap	2.28, dd (10.5, 3.8)	2.22, overlap	2.21, overlap
13.	3.26, d (9.6)	3.66, d (9.4)	3.28, d (9.6)	3.27, d (10.0)	3.31, d (9.6)	3.31, d (9.6)	3.20, d (9.7)	3.04, d (10.5)
15.	1.15, overlap	1.12, overlap	1.15, overlap	1.16, m, overlap	1.17, overlap	1.17, m	1.80, m	1.45, m
	2.07, overlap	1.97, overlap	2.07, overlap	1.84, m, overlap	1.88, overlap	1.87, m	1.14, m	1.27, m
16.	2.07, overlap	2.30, overlap	2.07, overlap	2.10, m, overlap	2.10, overlap	2.12, m	2.02, m	2.07, m
	1.91, overlap	2.00, overlap	1.91, overlap	1.93, m, overlap	1.92, overlap	1.93, m	1,88, m	1.91, m
17.	2.77, td (10.0, 5.1)	3.12, td (10.0, 5.1)	2.80, td (10.0, 5.0)	2.80, td (10.0, 5.1)	2.80, td (10.0, 5.1)	2.8, td (5.2, 9.6)	2.78, td (10.1, 5.3)	3.20, ddd (17.0, 10.6, 6.7)
18.	1.12, s	1.28, s	1.11, s	1.13, s	1.19, s	1.19, s	1.12, s	1.15, s
19.	0.76, s	0.76, s	0.82, s	0.79, s	0.85, s	0.83, s	0.78, s	0.81, s
21.	1.50, s	1.80, s	1.53, s	1.53, s	1.54, s	1.54, s	1.53, s	2.34, s
22.	2.90, d (14.2)	3.38, d (14.2)	2.90, d (14.2)	2.93, d (14.2)	2.94, d (14.2)	2.94, d (14.2)	3.11, d (15.6)	6.47, s
	2.72, d (14.2)	3.21, d (14.2)	2.72, d (14.2)	2.74, d (14.2)	2.74, d (14.2)	2.75, d (14.2)	2.83, d (15.6)	
24.	6.28, s	6.22, s	6.30, s	6.31, brs	6.32, s	6.31, s	2.96, d (14.8)	6.18, s
							2.95, d (14.8)	
26.	2.15, s	2.14, s	2.17, s	2.17, s	2.17, s	2.17, s	1.46, s	2.23, s
27.	1.70, s	1.67, s	1.71, s	1.71, s	1.72, s	1.72, s	1.46, s	1.70, s
28	1.17, s	1.17, s	1.19, s	1.24, s	1.23, s	1.14, s	1.19, s	1.22, s
29	1.11, s	1.11, s	1.11, s	1.19, s	1.04, s	1.05, s	1.12, s	1.16, s

	1	**2**[#]	**3**[##]	**5**	**6**	**7**	**8**	**9**
position	$\delta_{H,}$ (J in Hz) 500 MHz	$\delta_{H,}$ (J in Hz) 600 MHz	$\delta_{H,}$ (J in Hz) 600 MHz	$\delta_{H,}$ (J in Hz) 500 MHz	$\delta_{H,}$ (J in Hz) 600 MHz	$\delta_{H,}$ (J in Hz) 500 MHz	$\delta_{H,}$ (J in Hz) 500 MHz	$\delta_{H,}$ (J in Hz) 500 MHz
30.	0.84, s	0.83, s	0.86, s	0.86, s	0.88, s	0.89, s	0.83, s	0.89, s
1′	4.80, d (7.7)	4.82, d (7.5)	4.84, d (7.7)	4.96, d (7.7)			4.82, d (7.4)	4.86, d (7.7)
2′	4.01, overlap	4.01, overlap	4.06, overlap	4.30, overlap			4.03, overlap	4.07, overlap
3′	4.11, overlap	4.11, overlap	4.11, overlap	4.29, overlap			4.12, t (8.0)	4.11, t (8.0)
4′	4.01, overlap	4.01, overlap	4.01, overlap	4.16, t (9.1)			4.02, overlap	4.07, overlap
5′	3.85, overlap	3.87, overlap	3.89, overlap	3.98, overlap			3.85, m	3.89, m
6′	4.48, overlap	4.53, overlap	4.61, overlap	4.61, dd (12.0, 2.4)			4.46, dd (11.8, 3.4)	4.60, overlap
	4.33, overlap	4.43, overlap	4.47, overlap	4.41, dd (12.0, 6.6)			4.32, dd (11.8, 5.4)	4.52, overlap
1″	5.92, br s	5.95, br s	5.99, br s	6.59, brs			5.94, br s	6.00, br s
2″	4.66, overlap	4.66, overlap	4.72, overlap	4.88, m			4.66, overlap	4.72, overlap
3″	4.45, overlap	4.45, overlap	4.48, overlap	4.70, dd (9.4, 3.4)			4.45, overlap	4.50, overlap
4″	4.26, overlap	4.28, overlap	5.79, t (10.0)	4.35, overlap			4.25, t (9.4)	4.31, overlap
5″	4.61, overlap	4.62, overlap	4.74, overlap	4.81, dq (9.5, 6.0)			4.66, overlap	4.68, overlap
6″	1.64, d (6.0)	1.65, d (6.0)	1.41, d (6.0)	1.73, d (6.0)			1.66, d (6.0)	1.70, d (6.0)
1‴	5.70, br s	5.72, br s	5.72, br s				5.70, br s	5.77, br s
2‴	4.87, br s	4.87, br s	4.84, br s				4.87, br s	4.93, br s
3‴	4.53, overlap	4.52, overlap	4.53, overlap				4.53, dd (9.5, 3.1)	4.59, overlap
4‴	4.28, overlap	4.30, overlap	4.30, overlap				4.26, t (9.4)	4.32, overlap
5‴	4.70, overlap	4.69, overlap	4.72, overlap				4.66, overlap	4.75, overlap
6‴	1.60, d (6.0)	1.62, d (6.0)	1.63, d (6.0)				1.61, d (6.0)	1.66, d (6.0)

*3.1.1. Longipenoside A1 (**1**)*

Longipenoside A1 (**1**) was obtained as a white crystal with $[\alpha]_D^{25}$ -32.0 (*c* 0.1, MeOH). The molecular formula of compound 1 was deduced as $C_{48}H_{78}O_{17}$ from the $[M - H]^-$ high resolution electrospray ionization mass spectrometry (HRESIMS) ion peak observed at *m*/*z* 925.5157 (calcd for $C_{48}H_{77}O_{17}$, 925.5166). Infrared (IR) absorption at 1616 and 1697 cm^{-1} and ultraviolet–visible (UV) peak at 254 nm suggested the existence of an *α*, *β*-unsaturated carbonyl group. The IR broad peak at 3404 cm^{-1} appeared in companion with the peak at 1035 cm^{-1} indicated the presence of a hydroxyl group. The proton nuclear magnetic resonance (^{1}H NMR) spectrum demonstrated oxygenated methine proton at δ_H 3.25 (1H, H-3) and eight methyl singlets at δ_H 2.15, 1.71, 1.50, 1.17, 1.12, 1.11, 0.84, and 0.76 (each 3H). Its carbon-13 nuclear magnetic resonance (^{13}C NMR) spectrum revealed two ketone carbons (δ_C 211.7 and 201.7) and one oxygenated quaternary carbon (δ_C 74.0) (**Figure 47**). The above observations strongly suggested that compound **1** is a dammarane triterpenoid with one hydroxyl and two ketone groups. The position of hydroxyl group was confirmed at C-20 by heteronuclear multiple bond correlation (HMBC) correlations from H-17 (δ_H 2.77), H-21 (δ_H 1.50) and H-22 (δ_H 2.90 and 2.72) to C-20 (δ_C 74.0). One ketone group was located at C-12 from the HMBC experiment showing cross-peaks from H-11 (δ_H 2.24 and 2.22), H-13 (δ_H 3.26, d, J = 9.6 Hz), and H-17 (δ_H 2.32) to C-12 (δ_C 211.7). As suggested by UV and IR spectra, an *α*,*β*-unsaturated ketone moiety was further confirmed at C-23 (C=O) and C-24/25 (CH=C) by HMBC cross peaks from H-22 (δ_H 2.90, 2.72), H-24 (δ_H 6.28) to C-23 (δ_C 201.7), from H-22 (δ_H 2.90, 2.72), H-26 (δ_H 2.15), and H-27 (δ_H 1.71) to C-24 (δ_C 126.4) (**Figure 47**). The nuclear overhauser effect spectroscopy (NOESY) correlation of H-3 (δ_H 3.25) and H-5 (δ_H 0.66) demonstrated an *α* orientation for H-

3. Both H-17 and Me-21 were also determined as α orientation based on the coupling constant of H-13 (J = 9.6 Hz) and NOESY correlations from H-17 (δ_H 2.77) to H-21 (δ_H 1.50) and H-30 (δ_H 0.84).

The chemical shifts of the side chain of **1** are comparable with those of gypentonoside A, and the chemical shift of C-21 (27.1 ppm) is close to that of the typical chemical shift value of 20(S) dammarane derivatives. Also, proton H-17 and the methyl group at C-21 were determined as α oriented based on the coupling constant of H-13 (J = 9.7 Hz) and NOESY correlations from H-17 (δH 2.79) to H-21 (δH 1.52) and H-30 (δH 0.86). Therefore, the aglycone of **1** was deduced as 3β,20(S)-dihydroxydammar-24-en-12,23-dione. The diagnostic NMR data suggested the existence of three sugar units in compound **1**. Acid hydrolysis of **1** with HCl afforded D-glucopyranose and L-rhamnopyranose, based on liquid chromatography analysis following treatment with L-cysteine methyl ester hydrochloride and benzyl isothiocyanate derivatization. The coupling patterns of anomeric protons indicated β configuration for the glucopyranosyl (δ_H 4.80, J = 7.7 Hz) unit and α configuration for the two rhamnopyranosyl (δ_H 5.92, 5.70, br s) units. The HPLC-MS experiment at the positive mode showed fragment ions at 909 $[M + H - H_2O (18)]^+$, 763 $[M + H - H_2O$ – rha (146)$]^+$, 617 $[M + H - H_2O$ – rha – rha (146)$]^+$, 455 $[M + H - H_2O$ – 2 × rha – glc (162)$]^+$ corresponding to the presence of a sugar chain consisting of 2 rhamnoses and 1 glucose. The linkage patterns of three sugar units were analyzed based on the HMBC experiment (**Figure 47**) which also constructed the connection between H-1' (δ_H 4.80) and C-3 (δ_C 88.5), H-1" (δ_H 5.92) and C-2' (δ_C 78.3), H-1'" (δ_H 5.70) and C-3' (δ_C 87.5). Therefore, compound **1** was

elucidated as 3β,20(*S*)-dihydroxydammar-24-en-12,23-dione-3-*O*-α-L-rhamnopyranosyl- (1→2)-[α-L-rhamnopyranosyl-(1→3)]-β-D-glucopyranoside.

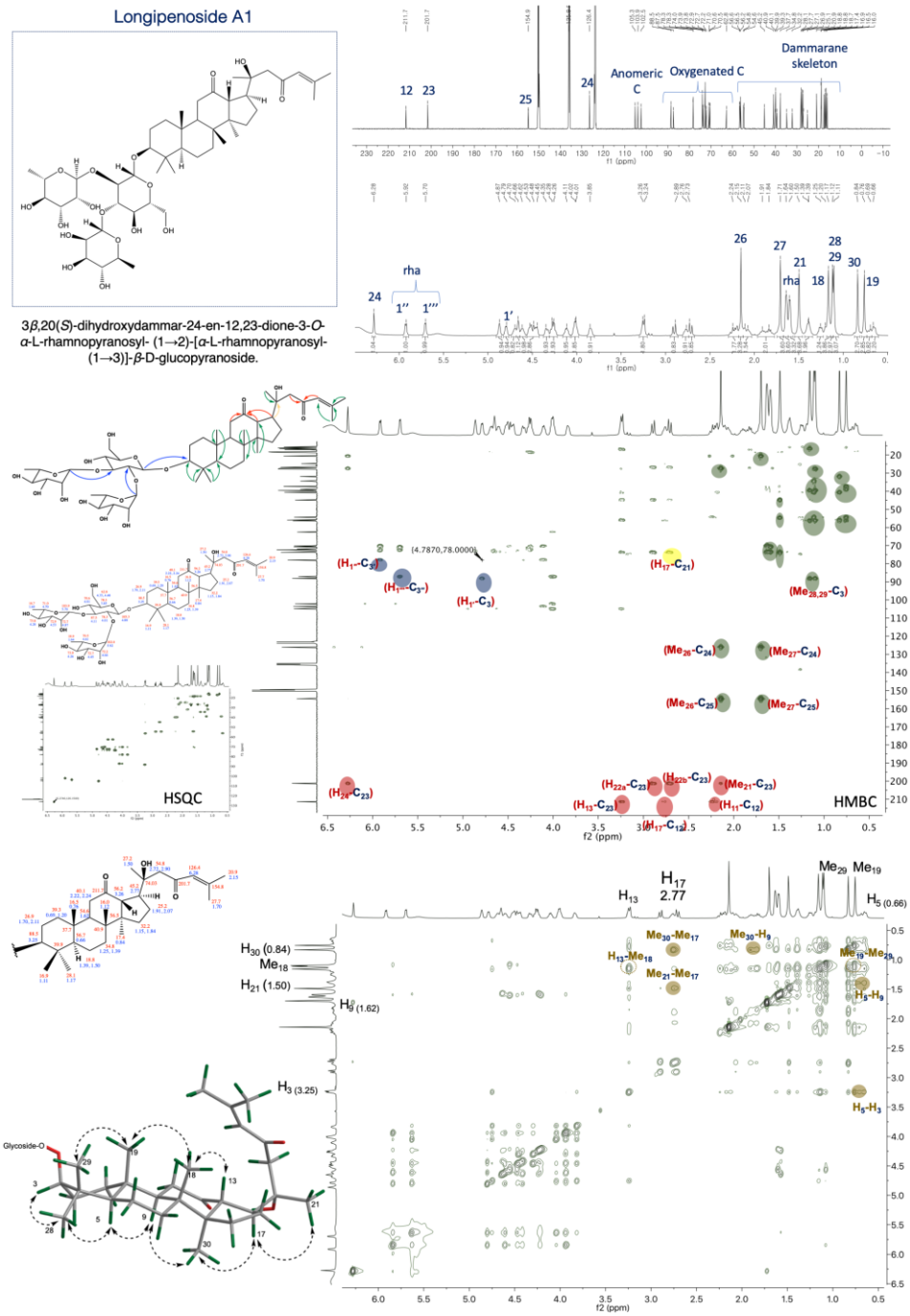


Figure 47. Longipenoside A1

*3.1.2. Longipenoside A2 (**2**)*

Compound **2** was obtained as a white and amorphous powder. Its molecular formula of $C_{54}H_{88}O_{22}$ was evident by the negative HRESIMS $[M - H]^-$ ion peak at *m/z* 1087.5682 (calcd for $C_{54}H_{87}O_{22}$, 1087.5694). The high-performance liquid chromatography-mass spectrometry (HPLC-MS) experiment in the positive mode showed a similar fragmentation with compound **1** with extra neutral loss of one hexose (162). The ^{1}H, ^{13}C NMR data of **2** were similar in resonance to those of **1** except for an additional *β*-glucopyranosyl unit which can be proved by the resonance of anomeric proton H-1′′′′ (δ_H 5.16, d, J = 7.7 Hz) and the addition of 6 carbon resonances at C-1′′′′ (δ_C 98.8), C-2′′′′ (δ_C 73.0), C-3′′′′ (δ_C 72.9), C-4′′′′ (δ_C 74.0), C-5′′′′ (δ_C 78.2), and C-6′′′′ (δ_C 63.3). When compared with 1, a downfield shift of C-20 (δ_C 80.7, +6.7 ppm) and an upfield shift of C-21 (δ_C 23.7, −3.5 ppm) were observed, which supported the attachment of the glucopyranosyl unit to C-20. Further HMBC analysis also constructed a link between H-1′′′′ (δ_H 5.16) and C-20 (**Figure 48**). Accordingly, compound **2** was determined to be 3*β*,20(*S*)-dihydroxydammar-24-en-12,23-dione-3-*O*-*α*-L-rhamnopyranosyl-(1→2)-[*α*-L-rhamnopyranosyl-(1→3)]-*β*-D-glucopyranosyl-20-*O*-*β*-D-glucopyranoside.

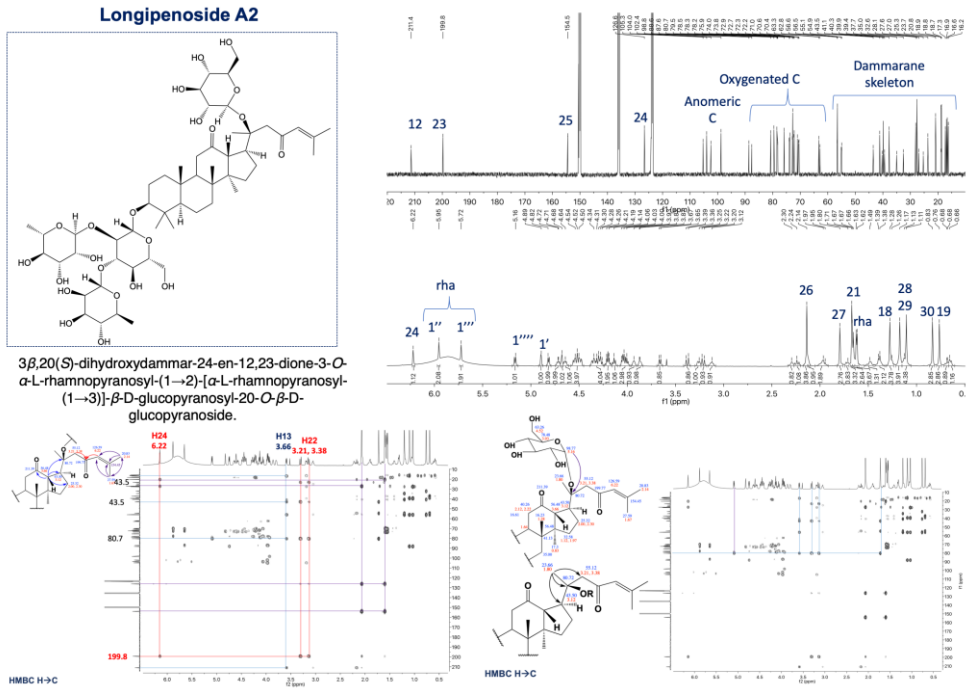


Figure 48. Longipenoside A2

*3.1.3. Longipenoside A3 (**3**)*

Compound **3** was obtained as a white and amorphous powder. The negative HRESIMS showed the deprotonated molecular ion peak at *m/z* 967.5276 $[M - H]^-$ (calcd for $C_{50}H_{79}O_{18}$, 967.5272) indicating that the molecular formula is $C_{50}H_{80}O_{18}$. HPLC-MS experiments at the positive mode showed the presence of an oligosaccharide chain consisted of 2 rhamnoses and 1 glucose similar with compound **1** together with an acetyl substitution. The ^{1}H, ^{13}C, and HSQC NMR spectroscopic data of **3** also showed similar resonances to those of **1**, apart from an additional acetyl unit which can be deduced by the addition of 2 carbon resonances of one ketone (δ_C 171.9) and a methyl singlet (δ_C 21.4; δ_H 2.10). When compared to 1, a downfield shift of C-4″ (δ_C 76.0, +2.1 ppm) and an upfield shift of C-5″ (δ_C 67.6, −2.9 ppm) demonstrated the substitution of an acetyl moiety to C-4″. The HMBC spectrum showed a cross peak from H-4″ (δ_H 5.79) to acetyl carbonyl (δ_C 171.9)

confirming the linkage (**Figure 49**). Accordingly, compound **3** was determined as 3*β*,20(*S*)-dihydroxydammar-24-en-12,23-dione-3-*O*-*α*-L-rhamnopyranosyl-(1→2)-[(4-*O*-acetyl)-*α*-L-rhamnopyranosyl-(1→3)]-*β*-D-glucopyranoside.

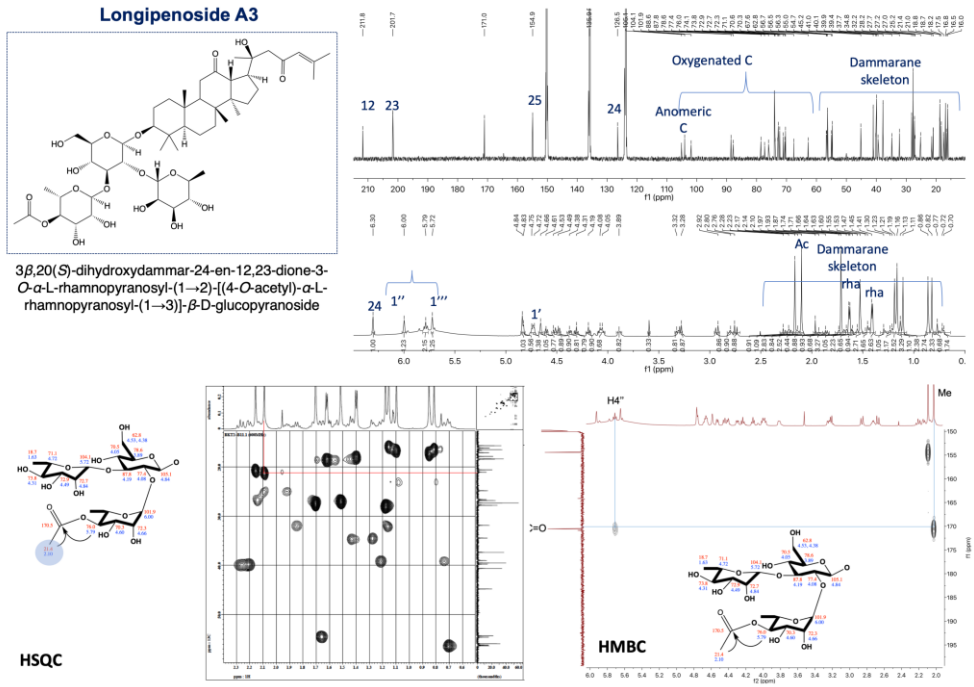


Figure 49. Longipenoside A3

*3.1.4. Longipenoside A4 (**5**)*

Compound **5** was obtained as a white and amorphous powder. The negative HRESIMS showed a deprotonated molecular ion peak at *m/z* 779.4597 $[M - H]^-$ (calcd for $C_{42}H_{67}O_{13}$, 779.4587) indicating the molecular formula to be $C_{42}H_{68}O_{13}$. The HPLC-MS experiment at positive mode showed fragment ions at 763 $[M + H - H_2O\ (18)]^+$, 617 $[M + H - H_2O - \text{rha}\ (146)]^+$, 455 $[M + H - H_2O - \text{rha} - \text{glc}\ (162)]^+$ corresponding to the existence of a oligosaccharide chain [-glc-rha]. The ^1H, ^{13}C, HSQC and HMBC NMR spectroscopic data of **5** could be discriminated with **1** by the lack of resonances of the rhamnopyranosyl unit attached to C-3′ and the shielded

chemical shift of this carbon (δ_C 80.3). Accordingly, compound **5** was determined as 3*β*,20(*S*)-dihydroxydammar-24-en-12,23-dione-3-*O*-*α*-L-rhamnopyranosyl-(1→2)-*β*-D-glucopyranoside.

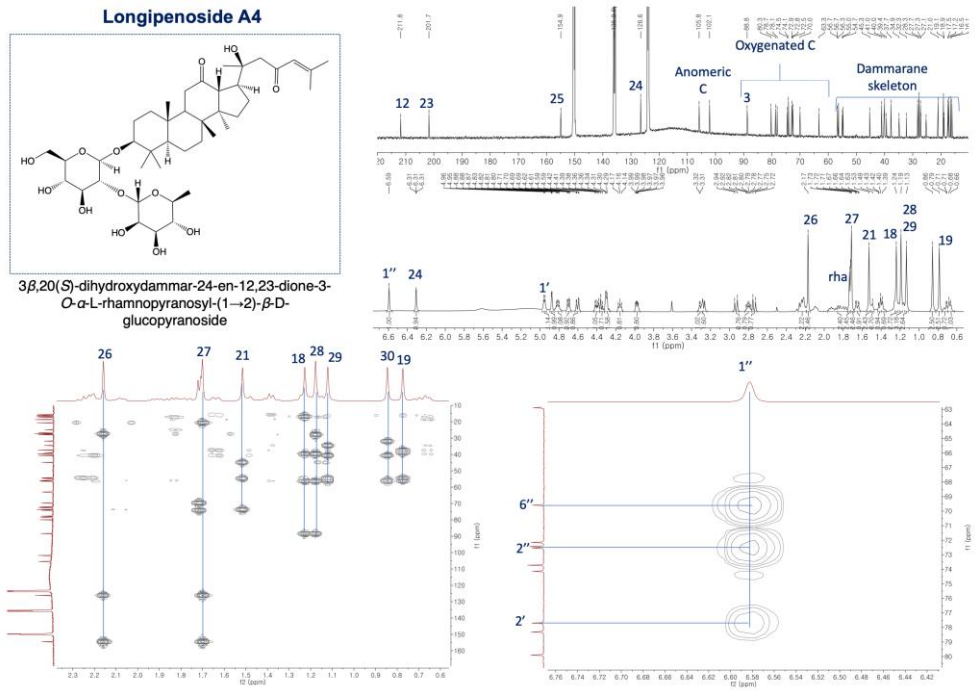


Figure 50. Longipenoside A4

*3.1.5. Longipengenol (**6**)*

Compound **6** was obtained as a white and amorphous powder. The negative HRESIMS showed a deprotonated molecular ion peak at *m/z* 471.3393 $[M - H]^-$ (calcd $C_{30}H_{47}O_4$, 471.3480), representing the molecular formula $C_{30}H_{48}O_4$. The ^1H, ^{13}C NMR, HSQC and HMBC spectroscopic data shown that **6** is the aglycone of compounds **1**−**4**. The NOESY correlations from H-3 (δ_H 3.40) to H-5 (δ_H 0.80) and H-28 (δ_H 0.77) suggested OH-3 to be *β*-oriented, which was also supported by the coupling pattern of H-3 (dd, J = 9.2 and 2.6 Hz). As a consequence, compound **6** was identified as 3*β*,20(*S*)-dihydroxydammar-24-en-12,23-dione.

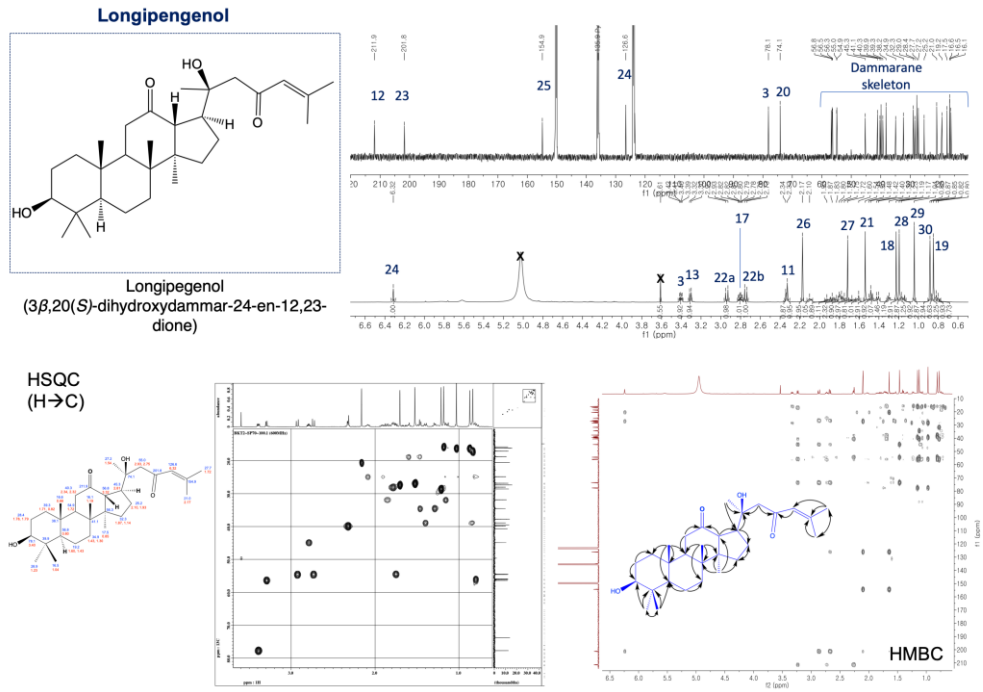


Figure 51. Longipengenol

*3.1.6. 3-dehydro longipengenol (**7**)*

Compound **7** was obtained as a white and amorphous powder. The negative HRESIMS showed a deprotonated molecular ion peak at *m/z* 471.3452 $[M + H]^+$ (calcd for $C_{30}H_{47}O_4$, 471.3469), which represents the molecular formula $C_{30}H_{46}O_4$. The ^1H, ^{13}C, HSQC, and HMBC NMR spectroscopic data of **7** showed similar resonances to those of **6** except for the presence of the third ketone (δ_C 216.2) and the disappearance of H-3 resonance, together with the deshielded chemical shift of C-2 (δ_C 34.4, +5.4 ppm) and C-4 (δ_C 47.7, +7.8 ppm) suggested ketone oxidation of OH-3. Thus, **7** was identified as 20(*S*)-hydroxydammar-24-en-3,12,23-trione.

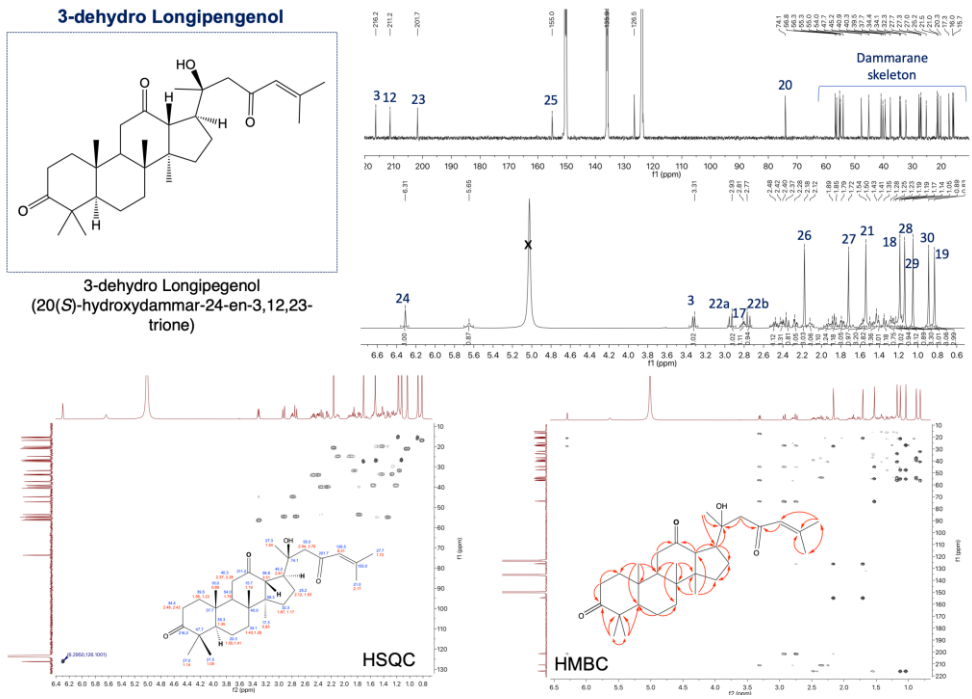


Figure 52. 3-dehydro longipengenol

*3.1.7. Longipenoside A5 (**8**)*

Compound **8** was obtained as an white, amorphous powder, and its molecular formula of $C_{48}H_{80}O_{18}$ was evident on the basis of a negative $[M-H]^-$ HRESIMS ion peak at *m/z* 943.5374 (calcd for $C_{48}H_{79}O_{18}$, 943.5380). HPLC-MS experiments at positive mode suggested the existence of a sugar chain similar with compound **1**, together with the presence of two hydroxyl groups substituted on the its aglycone. The ^{1}H, ^{13}C NMR, HSQC spectroscopic data of **8** (**Tables 11, 12**) showed similar resonances to those of **1** except for the hydration of double bonds at C-24 and C-25. When compared with **1**, an upfield shift of C-24 (δ_C 54.8, -71.6 ppm), C-25 (δ_C 70.0, -84.8 ppm) and a downfield shift of C-23 (δ_C 212.9, +11.2 ppm) were observed, which supported the attachment a hydroxyl group at C-25. Further HMBC analysis also established the connection between Me-26, Me-27 (each δ_H 1.46, s), and C-25

(δ_C 70.0). Thus, compound **8** was determined to be 3β,20(*S*),25-trihydroxydammaran-12,23-dione 3-*O*-α-L-rhamnopyranosyl-(1→2)-[α-L-rhamnopyranosyl-(1→3)]-β-D-glucopyranoside.

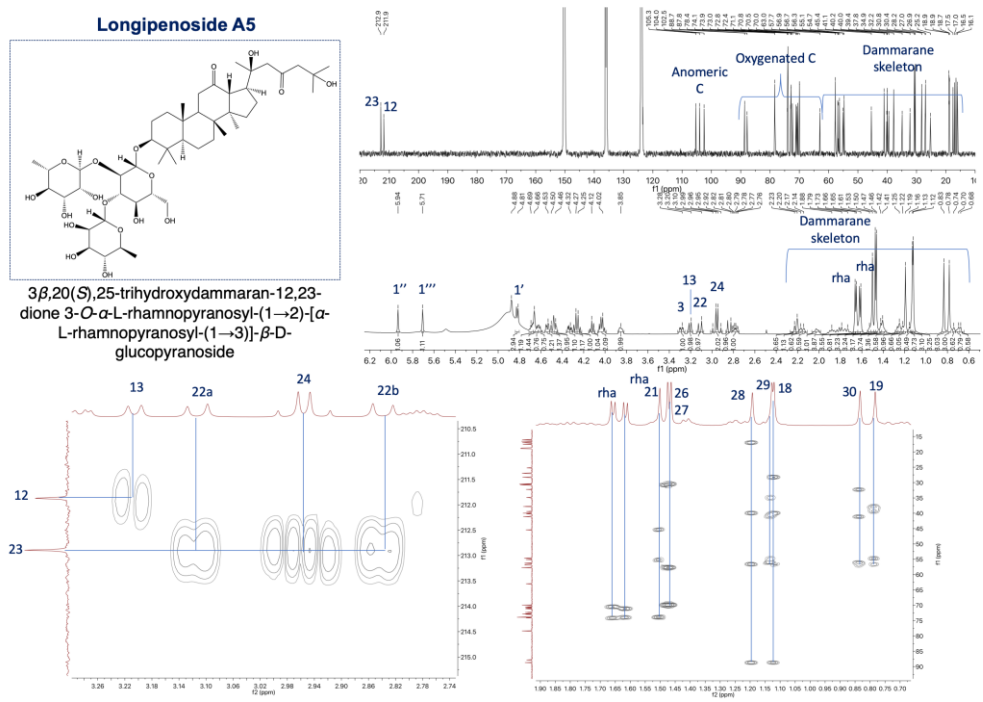


Figure 53. Longipenoside A5

*3.1.7. Longipenoside A6 (**9**)*

Compound **9** was obtained as a white and amorphous powder. The molecular formula of compound **9** was deduced as $C_{48}H_{76}O_{16}$ from the $[M-H]^-$ HRESIMS ion peak observed at *m/z* 907.5066 (calcd for $C_{48}H_{75}O_{16}$, 907.5061). The HPLC-MS experiment at positive mode also demonstrated the presence of a glycoside chain identical with compound **1**. The ^1H, ^{13}C, and HSQC NMR spectroscopic data of **9** (**Tables 11, 12**) showed a similar resonance to **1** with the lack of a hydroxy group signal at C-20 and contained a olefinic linkage resonance. Moreover, when compared with **1**, an upfield shift of C-21 (δ_C 17.7, -9.5 ppm), and downfield shifts

of C-20 (δ_C 159.7, +85.7 ppm) and C-22 (δ_C 126.3, +71.5 ppm) were observed, suggesting the dehydration of this compound at C-20. The HMBC spectrum exhibited cross peaks from H-21 and H-17 to C-20 and C-22, respectively (CHARGÉ and Rudnicki, 2004; Hawke and Garry, 2001), further confirming the proposed structure. Thus, compound **9** was determined as 3*β*-hydroxydammar-20,24-dien-12,23-dione-3-*O*-*α*-L-rhamnopyranosyl-(1→3)-[*α*-L-rhamnopyranosyl-(1→2)]-*β*-D-glucopyranoside.

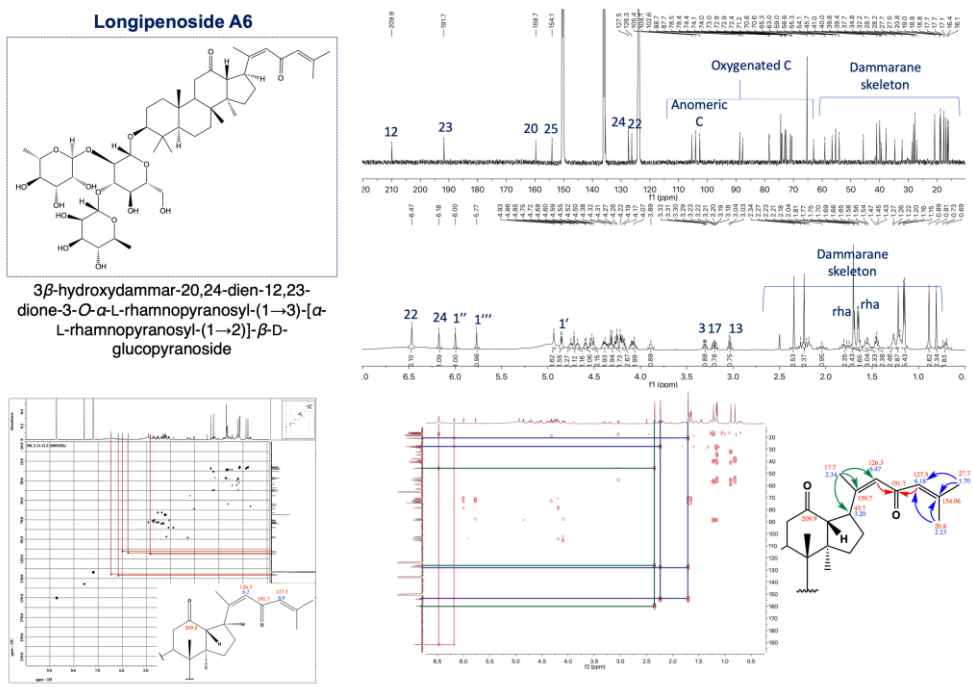


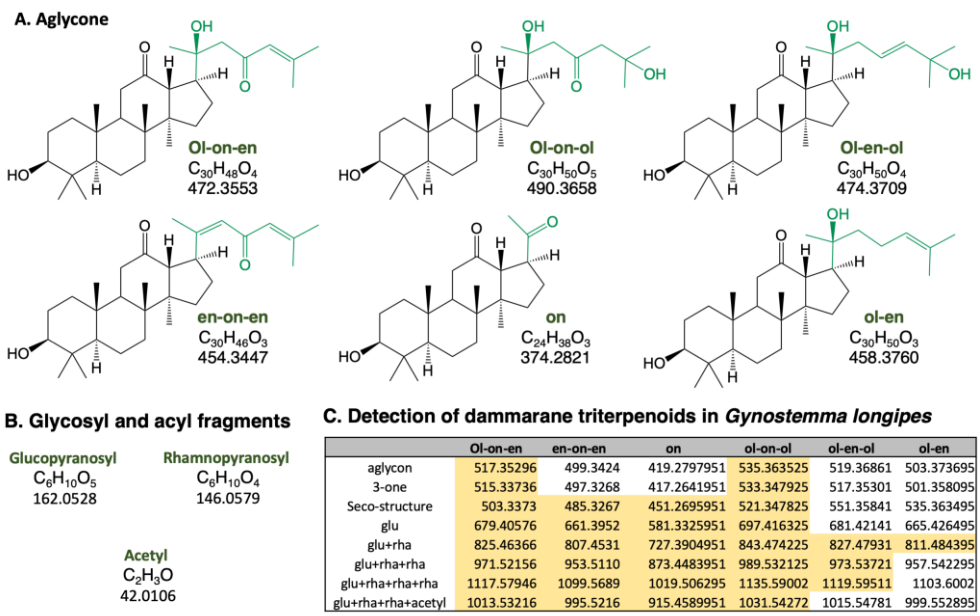
Figure 54. Longipenoside A6

3.2. Annotations of dammarane triterpenoids in G.latifolium using building block strategy

A total of 19 compounds have been isolated from the sample of G. longipes, including 9 compounds isolated in this study and 10 compounds reported in our

previous publication (Pham et al., 2018). By analysis the structures of these compounds,

six aglycones (**Figure 55-A**), two sugar types and one acyl mass fragments (**Figure 55-B)** were identified to be building blocks of dammarane triterpenoids in G. longipes. Main construction rules can be proposed were oxidation of 3-hydoxy to ketone, ring cleavage at $C_{3\text{-}4}$, and elongation of glycosyl chain at C-3. In the negative mode of MS^1, masses of precursor ions $[M - H]^-$ or $[M + HCOO]^-$ of all predicted structures were calculated. However signals of $[M + HCOO]^-$ could be detected with higher sensitivity for the triterpene compounds in G. longipes as shown in **Figure 55-C**.



C. Detection of dammarane triterpenoids in *Gynostemma longipes*

	Ol-on-en	en-on-en	on	ol-on-ol	ol-en-ol	ol-en
aglycon	517.35296	499.3424	419.2797951	535.363525	519.36861	503.373695
3-one	515.33736	497.3268	417.2641951	533.347925	517.35301	501.358095
Seco-structure	503.3373	485.3267	451.2695951	521.347825	551.35841	535.363495
glu	679.40576	661.3952	581.3325951	697.416325	681.42141	665.426495
glu+rha	825.46366	807.4531	727.3904951	843.474225	827.47931	811.484395
glu+rha+rha	971.52156	953.5110	873.4483951	989.532125	973.53721	957.542295
glu+rha+rha+rha	1117.57946	1099.5689	1019.506295	1135.59002	1119.59511	1103.6002
glu+rha+rha+acetyl	1013.53216	995.5216	915.4589951	1031.54272	1015.54781	999.552895

Figure 55. Building blocks of dammarane triterpenoids and the detection of targeted mass in the extract

All the detected masses were analyzed by KIA and MFA techniques to consolidate the annotation of their structures. Consequently, a total of 45 peaks were detected and assigned. Besides 1 known compounds, 10 new compounds previously reported and 8 new compounds

Table 14: Compound annotation for peaks detected in *Gynostemma longipes*

Peak	R_t	*m/z* (-)	Cald. *m/z* (-)	Diff (ppm)	Formula	Ion type	Building block compound assignment
1.	14.123	517.3546	517.3529	3.3	$C_{30}H_{47}O_4$	$[M+H]^-$	[Ol-on-en]
2.	14.861	515.3384	515.3373	2.1	$C_{31}H_{47}O_6$	$[M+HCOO]^-$	[Ol-on-en] – on
3.	12.178	503.3387	503.3373	2.8	$C_{30}H_{47}O_6$	$[M+H]^-$	[Ol-on-en] – seco
4.	10.362	679.4069	679.4057	1.8	$C_{37}H_{59}O_{11}$	$[M+HCOO]^-$	[Ol-on-en]+glc
5.	9.862	825.4659	825.4636	2.8	$C_{43}H_{69}O_{15}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha
6.	9.414	825.4661	825.4636	3.0	$C_{43}H_{69}O_{15}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha
7.	9.528	971.5241	971.5216	2.6	$C_{49}H_{79}O_{19}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha
8.	10.038	971.5219	971.5216	0.3	$C_{49}H_{79}O_{19}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha
9.	8.441	1117.5793	1117.5795	-0.2	$C_{55}H_{89}O_{23}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha+rha
10.	8.941	1117.5787	1117.5795	-0.7	$C_{55}H_{89}O_{23}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha+rha
11.	9.397	1117.5795	1117.5795	1.0	$C_{55}H_{89}O_{23}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha+rha
12.	10.695	1013.5336	1013.5321	1.5	$C_{51}H_{81}O_{20}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha+ace
13.	10.774	1013.5338	1013.5321	1.7	$C_{51}H_{81}O_{20}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha+ace
14.	11.362	1013.5327	1013.5321	0.6	$C_{51}H_{81}O_{20}$	$[M+HCOO]^-$	[Ol-on-en]+glc+rha+rha+ace
15.	14.291	485.3275	485.3267	1.6	$C_{30}H_{45}O_5$	$[M+H]^-$	[En-on-en] – seco
16.	12.256	661.3958	661.3952	0.9	$C_{37}H_{57}O_{10}$	$[M+HCOO]^-$	[En-on-en]+glc
17.	11.090	807.4534	807.4531	0.4	$C_{43}H_{67}O_{14}$	$[M+HCOO]^-$	[En-on-en]+glc+rha
18.	11.572	807.4542	807.4531	1.4	$C_{43}H_{67}O_{14}$	$[M+HCOO]^-$	[En-on-en]+glc+rha
19.	12.081	807.4535	807.4531	0.5	$C_{43}H_{67}O_{14}$	$[M+HCOO]^-$	[En-on-en]+glc+rha
20.	11.563	953.5114	953.5110	0.4	$C_{49}H_{77}O_{18}$	$[M+HCOO]^-$	[En-on-en]+glc+rha+rha
21.	10.748	1099.5682	1099.5689	-0.6	$C_{55}H_{87}O_{22}$	$[M+HCOO]^-$	[En-on-en]+glc+rha+rha+rha
22.	11.861	995.5212	995.5216	-0.4	$C_{51}H_{79}O_{19}$	$[M+HCOO]^-$	[En-on-en]+glc+rha+rha+ace
23.	12.361	995.5219	995.5216	0.3	$C_{51}H_{79}O_{19}$	$[M+HCOO]^-$	[En-on-en]+glc+rha+rha+ace
24.	11.861	995.5212	995.5216	-0.4	$C_{51}H_{79}O_{19}$	$[M+HCOO]^-$	[En-on-en]+glc+rha+rha+ace
25.	11.300	535.3658	535.3635	4.3	$C_{31}H_{51}O_7$	$[M+HCOO]^-$	[Ol-on-ol]
26.	12.178	533.3487	533.3478	1.7	$C_{31}H_{51}O_7$	$[M+HCOO]^-$	[Ol-on-ol] – on
27.	9.491	521.3496	521.3484	2.3	$C_{30}H_{49}O_7$	$[M+HCOO]^-$	[Ol-on-ol] – seco
28.	8.195	697.4163	697.4163	0.0	$C_{37}H_{61}O_{12}$	$[M+HCOO]^-$	[Ol-on-ol]+glc
29.	7.029	989.5342	989.5321	2.1	$C_{49}H_{81}O_{20}$	$[M+HCOO]^-$	[Ol-on-ol]+glc+rha+rha
30.	7.529	989.5341	989.5321	2.0	$C_{49}H_{81}O_{20}$	$[M+HCOO]^-$	[Ol-on-ol]+glc+rha+rha
31.	7.195	1135.5922	1135.5900	1.9	$C_{55}H_{91}O_{24}$	$[M+HCOO]^-$	[Ol-on-ol]+glc+rha+rha+rha
32.	6.099	1135.5928	1135.5900	2.5	$C_{55}H_{91}O_{24}$	$[M+HCOO]^-$	[Ol-on-ol]+glc+rha+rha+rha
33.	8.195	1031.5419	1031.5427	-0.8	$C_{51}H_{83}O_{21}$	$[M+HCOO]^-$	[Ol-on-ol]+glc+rha+rha+ace
34.	8.643	1031.5424	1031.5427	-0.3	$C_{51}H_{83}O_{21}$	$[M+HCOO]^-$	[Ol-on-ol]+glc+rha+rha+ace
35.	9.862	405.2524	405.2641	-1.7	$C_{24}H_{37}O_5$	$[M+H]^-$	[On] – seco
36.	7.695	681.3843	681.3850	-1.0	$C_{36}H_{57}O_{12}$	$[M+H]^-$	[On]+glc+rha
37.	8.143	681.3849	681.3850	-0.1	$C_{36}H_{57}O_{12}$	$[M+H]^-$	[On]+glc+rha
38.	7.950	827.4439	827.4429	1.2	$C_{42}H_{67}O_{16}$	$[M+H]^-$	[On]+glc+rha+rha
39.	7.503	1019.5089	1019.5063	2.6	$C_{49}H_{79}O_{22}$	$[M+HCOO]^-$	[On]+glc+rha+rha
40.	7.695	827.4801	827.4793	1.0	$C_{43}H_{71}O_{15}$	$[M+HCOO]^-$	[Ol-en-ol]+glc+rha
41.	7.397	973.5375	973.5372	0.3	$C_{49}H_{81}O_{19}$	$[M+HCOO]^-$	[Ol-en-ol]+glc+rha+rha
42.	7.862	973.5386	973.5372	1.4	$C_{49}H_{81}O_{19}$	$[M+HCOO]^-$	[Ol-en-ol]+glc+rha+rha
43.	8.028	973.5380	973.5372	0.8	$C_{49}H_{81}O_{19}$	$[M+HCOO]^-$	[Ol-en-ol]+glc+rha+rha
44.	7.634	1119.5970	1119.5951	1.7	$C_{55}H_{91}O_{23}$	$[M+HCOO]^-$	[Ol-en-ol]+glc+rha+rha+rha
45.	8.766	811.4844	811.4844	0.0	$C_{43}H_{71}O_{14}$	$[M+HCOO]^-$	[Ol-en]+glc+rha

3.3. Dammarane triterpenes enhanced muscle proliferation through activating AMPK

Regeneration of damaged skeletal muscles depends on satellite cells (known as quiescent muscle precursor cells), which play an important role in the proliferation and differentiation of myoblasts to form or repair muscle fibers (Fu et al., 2015; Li et al., 2012). Recently, several studies have demonstrated that decreased proliferation of myoblasts and cytotoxicity can reduce the number of muscle fibers (Li et al., 2012; Tews et al., 1997; Yen et al., 2012). Therefore, mouse C2C12 cells, also known as myoblastic cells, were chosen in this research because they are a valid model to study muscle cell proliferation (Fu et al., 2015; Li et al., 2012). In a search for new bioactive natural products with promising activity for muscle proliferation, we found that the 95% EtOH eluted fraction on SP70 resin of the total extract could potentially be a 'hit' using an *in vitro* screening method. This fraction stimulated an increase in proliferation of C2C12 myoblast cell about 20-30% compared to the negative control during 24 and 48 hours of treatment using 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) and 4-[3-(4-iodophenyl)-2-(4-nitrophenyl)-2H-5-tetrazolio]-1,3-benzene disulfonate (WST-1) methods (**Figure 56-A,B**). The proliferation also expressed dose-dependently using a cell counting method. As AMPKα is an important regulator during the process of muscle proliferation, we tested the effect of the *G. longipes* fraction on *p*-AMPKα stimulation in C2C12 myotubes and myoblasts. As a result, this fraction stimulated the phosphorylation of AMPK significantly in a concentration-dependent manner (**Figure 56-C,D**). On the other hand, when the cells were pre-incubated with compound C (dorsomorphin – a reversible and potent AMPK inhibitor), the *G. longipes* fraction did not affect the expression of *p*-AMPKα, similarly to the vehicle-treated cells.

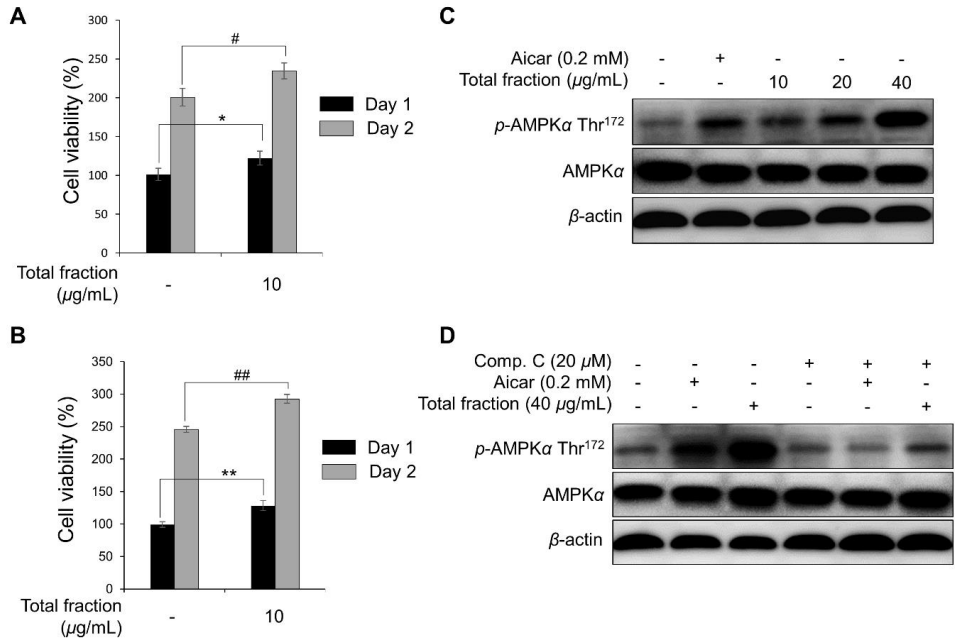


Figure 56. Effect of the SP70-EtOH 95% fraction from *G*. *longipes* on C2C12 myoblast cell proliferation using the MTT method (A) or WST method (B) after 1 - 2 days incubation.

Data are presented as the mean ± SD (n = 3), *, # $p < 0.05$ and **, ## $p < 0.01$, compared to the negative control group. (C) *Gynostemma longipes* induced phosphorylation (Thr172) of AMPKα in a concentration-dependent manner when C2C12 myotubes were incubated with the SP70-EtOH 95% fraction for 30 minutes. (D) C2C12 myotubes were preincubated for 15 minutes with or without compound C before treatment with fractions or Aicar as a positive control. After 30 minutes of incubation, cells were harvested and Western blot performed to evaluate the expression of *p*-AMPK.

Bioassay-guided isolation from *G*. *longipes* resulted in the isolation of nine compounds (**1**–**9**) and all of the isolates were evaluated for their enhancing effect on muscle proliferation in C2C12 myoblast cells. After 48 hours of treatment, cell viability was assessed using the MTT and WST-1 methods. Seven compounds (**1**

and **3**–**8**) increased cell viability by about 20-30% (**Figure 57-A**). Notably, compounds 6 and 7 exhibited the highest induction of cell proliferation at a value from 131.95 ± 2.54% to 136.19 ± 2.73%. Since compound **1** was identified as a major constituent of this plant (over 2.08% of the material), it was selected to examine the effects on the proliferation of two cancer cell lines, MCF-7 and MDA-MB-231 breast cancer cells. Cell viability was not changed when the cells were incubated with a fraction of *G. longipes* (10 μg/mL) or compound **1** (10 μM) for 2 days. These results suggested that the active fraction and major compound of *G. longipes* stimulate muscle cell proliferation selectively and do not induce the outgrowth of harmful cancer cell lines.

A number of studies demonstrated that dammarane triterpenoids are potent classes of AMPK-activating agents (Cao et al., 2015; H. M. Lee et al., 2012; Nguyen et al., 2011). In order to evaluate the potential activities of compounds from *G. longipes*, nine isolates (**1**–**9**) were measured the stimulatory effects on the *p*-AMPK and *p*-ACC expressions using mouse C2C12 myotubes. Compounds **5**, **6**, and **7** increased the phosphorylation of AMPK and ACC proteins strongly, while compounds **1**, **3**, **4**, and **8** exhibited less efficient. Therefore, potential candidates (**1**, **5**, **6**, and **7**) were further confirmed the activation effect on *p*-AMPK in C2C12 myoblasts. These compounds clearly increased the phosphorylation of the target protein compared to Aicar as a positive control.

Because compound **1** is the major constituent of the active fraction, it was selected for further evaluating the activation of *p*-AMPK in a time-dependent manner using the mouse C2C12 myoblast model. Remarkably, compound **1** up-regulated the phosphorylation of AMPK from 30 to 90 minutes (**Figure 57-B**).

Besides, compound **1** also induced the expression of AMPK slightly at 24 hours after treatment. These results suggested that triterpenoids from *G. longipes* could significantly stimulate the *p*-AMPK expression in both mouse C2C12 myotubes and myoblasts.

Considering the structures and activities of all nine isolates, the core structure of the dammarane skeleton might contribute mainly to the AMPK activation effect. Further examination of structure-activity relationships (SAR) revealed that compounds with a longer saccharide side chain at C-3 showed lower AMPK inhibition effects: **4** (4 sugar units) < **1** (3 sugar units) < **5** (2 sugar units) < **6** and **7** (no sugar). It was also noteworthy that with the same number of sugar units, the attachment of a sugar unit to C-20 remarkably reduced the activity (**2** < **4**). Furthermore, the hydration of olefinic bond at C-20 and C-24 slightly increased the effect in the following order: **9** (two double bonds) < **1** (one double bond) < **8** (no double bond). The change of hydroxy (**6**) and ketone substitutions (**7**) at C-3 did not cause any difference in activity.

Finally, to assess the effects of compound C during muscle proliferation, the cell growth and expression of *p*-AMPK were analyzed using the mouse C2C12 myoblast model. Interestingly, co-treatment of compound C with the active fraction or compound **1** decreased the percentage of cell viability after 48 hours of incubation, compared to the number of C2C12 cells exposed to the fraction or compound **1** alone (Figure 57-C). The decline of the *p*-AMPK expression was also observed after 1 hour of incubation. Similarly, the stimulation effect of compound 1 on *p*-AMPK was down-regulated by co-treatment with compound C (**Figure 57-D**). Consequently, the activation of the AMPK signaling pathway by *G. longipes*'s fraction and

compound **1** contributes relatively to muscle proliferation by inducing the growth of C2C12 myoblasts.

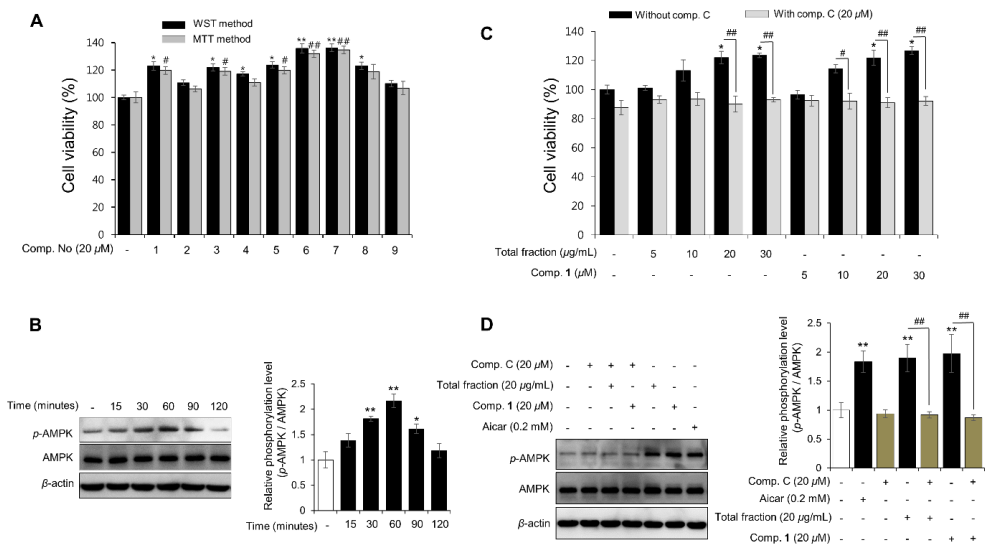


Figure 57. (A) Effect of isolated compounds (**1**–**9**) on the proliferation of myoblast cells. C2C12 cells were incubated with tested compounds for 48 hours and cell viability was assessed using the MTT or WST method.

Each value is expressed as the mean ± SD (n=3); * $p < 0.05$ and ** $p < 0.01$, compared to negative control group using the WST method; while $^{\#}p < 0.05$ and $^{\#\#}p < 0.01$, compared to the vehicle group using the MTT method. (B) C2C12 myoblast cells were exposed to compound **1** (20 μM) and incubated for 15 to 120 minutes. Phosphorylation of AMPK protein in the cells was assessed by Western blot analysis. Values are expressed as the mean ± SD (n = 3), * $p < 0.05$ and ** $p < 0.01$, compared to the negative control group. (C) The effect of co-treatment compound C with the active fraction and compound **1** on the cell viability. C2C12 myoblasts were re-treated with compound C (20 μM) for 15 minutes and then the cells were continuously incubated with test samples for 2 days. After that, the percentage of

cell viability was evaluated using MTT method. Results are presented as the mean ± SD (n=3); * $p < 0.05$ and ** $p < 0.01$, compared to negative control group (without compound C treatment); while $^{\#}p < 0.05$ and $^{\#\#}p < 0.01$ compared to co-treatment with compound C and test samples group, respectively. (D) Down-regulatory effect of compound C on p-AMPK (Thr^{172}) when co-treatment with the active fraction or compound **1**. After 15 minutes re-incubated with compound C (20 μM), mouse C2C12 myoblasts were then exposed to test samples for 1 hour. Phosphorylation of AMPK protein was measured by Western blotting. Data were calculated as the mean ± SD (n=3); ** $p < 0.01$ compared to the negative control, while $^{\#\#}p < 0.01$ compared to co-treatment with compound C and test samples group, respectively.

3.4. Effects of dammarane triterpenes on DNA synthesis during cell proliferation.

It has been previously reported that 5-bromo-2′-deoxyuridine (BrdU) can be used to label the DNA of cells in the S-phase. This method is useful for quantifying the degree of DNA synthesis during cell proliferation (Staszkiewicz et al., 2009). Therefore, in order to confirm whether compound **1** from *G. longipes* induces cell proliferation, immunochemical staining analysis with a BrdU antibody was performed. In the first experiment, C2C12 myoblast cells were treated with compound **1** (20 μM) or the SP70-EtOH 95% fraction (20 μg/mL) and incubated for 8 hours. Next, the cells were incubated for 2 hours with or without BrdU. The BrdU incorporation into the nucleus was assessed using a fluorescein isothiocyanate (FITC)-conjugated anti-BrdU antibody and DAPI staining. As shown in **Figure 58-A**, the number of BrdU-positive cells was significantly increased in the group treated with the active fraction of *G. longipes* or compound **1** compared to DMSO treatment alone. We also evaluated the BrdU-labeled cells using a flow cytometer for BrdU

and PI staining. When C2C12 cells were incubated with or without BrdU treatment, the fluorescence intensity was significantly increased from 33.46 ± 0.58% to 55.41 ± 0.88%. In order to confirm the concentration dependence in the cells, compound 1 was treated for 8 hours at different concentrations (20, 10, and 5 μM) and then incubated with BrdU for 2 hours. The BrdU positive cell population increased in a concentration-dependent manner to 65.22 ± 3.43%, 59.90 ± 0.84% and 56.00 ± 0.81%, respectively (**Figure 58-B**).

Moreover, to better understand the effects of co-treatment compound C with test samples on DNA synthesis during cell proliferation, immunofluorescence detection and flow cytometry analysis for BrdU staining were carried out. For qualitative analysis, the number of red stained cells notably reduced in the groups co-treated with compound C compared to the active fraction or compound **1** alone (**Figure 58-A**). To further examine the fluorescence intensity of each group, the BrdU-labeled cells were determined using flow cytometry for BrdU staining only. This result indicated that co-treatment of compound C with the active fraction or compound **1** failed to increase the number of BrdU-positive cells.

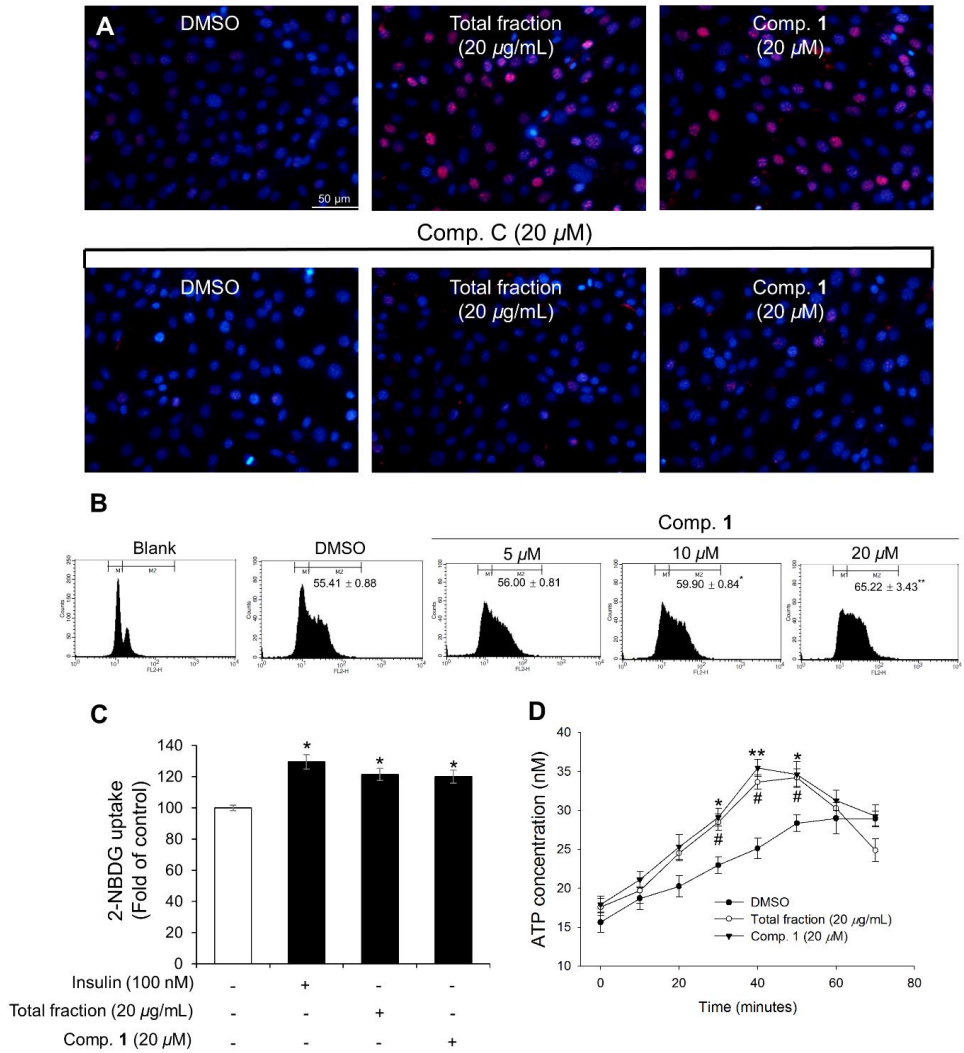


Figure 58. Increased DNA synthesis in C2C12 myoblasts cells by dammarane triterpenes from *G. longipes*. (A) C2C12 myoblast cells were re-treated with compound C (20 μM) for 15 minutes and the cells were then incubated with the active fraction (20 μg/mL) or compound **1** (20 μM) for 8 hours.

Cells were then incubated for 2 hours in the presence or absence of BrdU. After fixation and permeabilization, the cells were incubated with antibodies and stained with the DAPI solution. Cells images were observed by fluorescence microscopy. (B) C2C12 myoblast cells were incubated with compound **1** at different concentrations for 8 hours. After incubation for 2 hours with or without BrdU, the

cells were harvested and flow cytometric analysis for BrdU and PI staining was performed. The cell distribution was displayed using a histogram. Data presented as the mean ± SD (n = 2), * $p < 0.05$ and ** $p < 0.01$, compared to the DMSO group. (C) Stimulation effects of the active fraction or compound **1** on glucose uptake in C2C12 myoblasts using the 2-NBDG probe. The cells were incubated with test samples or insulin (as a positive control) for 1 hour. The fluorescent intensity was observed at ex/em = 450/535 nm using a fluorescence microplate reader. Each value is represented as the mean ± SD (n = 3), * $p < 0.05$ compared to negative control. (D) Dammarane triterpenes from *G. longipes* increased intracellular ATP level of C2C12 myoblasts. The cells were treated with the active fraction (20 μg/mL) or compound **1** (20 μM) from 0 to 70 minutes. Mitochondrial ATP synthesis was measured using ATP bioluminescence determination assay. Data were calculated as the mean ± SD (n = 3); the p value for compound **1** (* $p < 0.05$, ** $p < 0.01$) and total fraction (# $p < 0.05$) compared to negative control.

3.5. Effects of dammarane triterpenes on glucose uptake and ATP levels

Recent studies suggested the importance of AMPKα enhances glucose uptake and ATP production during the process of muscle proliferation (Wagatsuma and Sakuma, 2013). Considering that the glucose uptake level was evaluated using a fluorescent-tagged glucose analogue, 2-[*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino]-2-deoxy-D-glucose (2-NBDG), to monitor intracellular glucose uptake. As illustrated in **Figure 58C**, the active fraction and compound **1** as well as insulin (as a positive control) increased glucose uptake in mouse C2C12 myoblasts.

The effects of dammarane triterpenes from *G. longipes* on the mitochondrial ATP synthesis was examined by treatment mouse C2C12 myoblasts with the active fraction and compound **1** in approximately 70 minutes. The ATP levels of the active fraction and compound **1** increased drastically from 30 to 50 minutes comparison

with the vehicle group (**Figure 58-D**). However, the rate of ATP production reduced rapidly after 1 hour of incubation with these samples. During the proliferation of myoblasts, the production of ATP is an essential downstream target of glycolysis, and cells used these ATP (approximately 30%) which is supplied by oxidative phosphorylation (OXPHOS) as the major source of metabolic energy (Medema and Macůrek, 2012). Taken together, the role of triterpenoids from *G. longipes*, which regulate muscle proliferation by increasing glucose uptake and ATP production, could be suggested.

3.6. Enhancement of cell proliferation by dammarane triterpenes through cell cycle regulation

Several previous studies have proposed that ginseng root extract promotes glucose uptake in muscle cells and also induces the proliferation of β-cells, and ginsenoside Rg_1 from *Panax ginseng* enhances myoblast differentiation and myotube growth (Cha et al., 2010; Go et al., 2017). In this study, the effect of dammarane triterpenoid from *G. longipes* on cell proliferation through regulation of cell cycle was also evaluated using propidium iodide (PI) staining method (Medema and Macůrek, 2012). C2C12 myoblasts were exposed to compound 1 (20 μM) for 12 and 24 hours. The cell cycle distributions at different phases were significantly changed, compared to the vehicle group. After 12 hours of incubation with compound **1**, the cells with 2N DNA content (G_0/G_1 phases) decreased remarkably from 72.07 ± 0.63% to 65.09 ± 2.49%, while the cells with >2N DNA content (S or G_2/M phases) increased about 8% compared with negative control. Moreover, the proportion of G_2/M phases increased significantly from 13.64 ± 1.47% to 22.41 ± 0.81% when the

cells were incubated with compound **1** for 24 hours in comparison with the group treated with DMSO.

4. Conclusion

In this study, we reported the findings of eight new 12,23-dione dammarane triterpenes from *G. longipes*, a traditional Vietnamese medicinal plant. Seven of the compounds (**1** and **3**–**8**) activated significantly AMPK phosphorylation in mouse C2C12 cell lines. The result of AMPK activation by compound **1** could stimulate glucose uptake into the cells and increase mitochondrial ATP synthesis. The effect of dammarane triterpenes on DNA synthesis during cell proliferation also showed that compound **1** increased the BrdU positive cell population in a dose-dependent manner. Analysis of cell cycle using flow cytometry displayed clearly that compound **1** reduced the 2N DNA content (G_0/G_1 phases) while the ratio of G_2/M phases was significantly increased. These results suggested strongly that compound **1** enhances significantly the proliferation of C2C12 myoblast cells (**Figure 59**). Furthermore, a considerable amount of active compound **l** (over 2.08% per dried raw plant) in *G. longipes* suggested that it may be a promising candidate for development of functional food or botanical drug. These results also indicated that new dammarane-type compounds are promising candidates for muscle proliferation via activation of AMPK signaling pathways and could be further studied and developed as therapeutics for geriatric diseases.

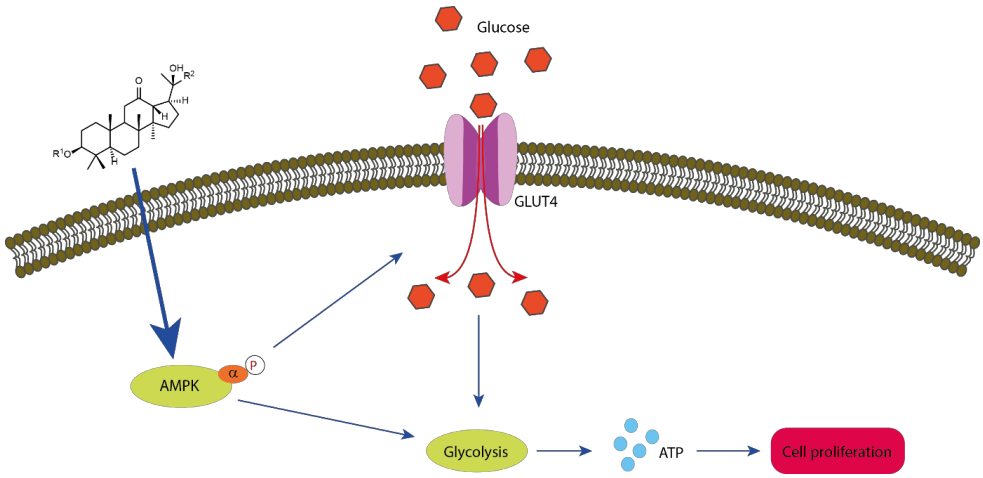


Figure 59. The proposed mechanism for enhancing cell proliferation by dammarane triterpenoids from *G*. *longipes* through AMPK pathway.

REFERENCES

Abe, F., Iwase, Y., Yamauchi, T., Yahara, S., T.N., 1995. Flavonol sinapoyl glycosides from leaves of *Thevetia peruviana*. Phytochemistry, 40, 577–581.

Alabdul Magid, A., Voutquenne, L., Moretti, C., Long, C., Lavaud, C., 2006. Triterpenoid saponins from the fruits of *Caryocar glabrum*. J. Nat. Prod. 69, 196–205.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anh, P.T., Ky, P.T., Cue, N.T., Nhiem, N.X., Yen, P.H., Ngoc, T.M., Anh, H.L.T., Tai, B.H., Trang, D.T., Minh, C.V., Kiem, P.V., 2015. Damarane-type saponins from *Gynostemma longipes* and their cytotoxic activity. Nat Prod Commun 10, 1351–1352.

Asakura, A., Seale, P., Girgis-Gabardo, A., Rudnicki, M.A., 2002. Myogenic specification of side population cells in skeletal muscle. J. Cell Biol. 159, 123–134.

Bohni, N., Cordero-Maldonado, M.L., Maes, J., Siverio-Mota, D., Marcourt, L., Munck, S., Kamuhabwa, A.R., Moshi, M.J., Esguerra, C.V., de Witte, P.A.M., Crawford, A.D., Wolfender, J.-L., 2013. Integration of microfractionation, qNMR and zebrafish screening for the in vivo bioassay-guided isolation and quantitative bioactivity analysis of natural products. PLoS ONE 8, e64006.

Cao, J., Zhang, X., Qu, F., Guo, Z., Zhao, Y., 2015. Dammarane triterpenoids for pharmaceutical use: a patent review (2005 – 2014). Expert Opinion on Therapeutic Patents 25, 805–817.

Cha, J.-Y., Park, E.-Y., Kim, H.-J., Park, S.-U., Nam, K.-Y., Choi, J.-E., Jun, H.-S., 2010. Effect of white, taegeuk, and Red Ginseng root extracts on insulin-stimulated glucose uptake in muscle cells and proliferation of β-cells. Journal of Ginseng Research 34, 192–197.

Chargé, S.B.P., Rudnicki, M.A., 2004. Cellular and molecular regulation of muscle regeneration. Physiological Reviews 84, 209–238.

Charlier, R., Knaeps, S., Mertens, E., Van Roie, E., Delecluse, C., Lefevre, J., Thomis, M., 2016. Age-related decline in muscle mass and muscle function in Flemish Caucasians: a 10-year follow-up. Age (Dordr) 38, 36.

Chen, S.-E., Jin, B., Li, Y.-P., 2007. TNF-alpha regulates myogenesis and muscle regeneration by activating p38 MAPK. Am. J. Physiol., Cell Physiol. 292, C1660–71.

Cho, H., 2013. Protein tyrosine phosphatase 1B (PTP1B) and obesity. Vitam. Horm. 91, 405–424.

Coquerel, D., Neviere, R., Delile, E., Mulder, P., Marechal, X., Montaigne, D., Renet, S., Remy-Jouet, I., Gomez, E., Henry, J.-P., do Rego, J.-C., Richard, V., Tamion, F., 2014. Gene deletion of protein tyrosine phosphatase 1B protects against sepsis-induced cardiovascular dysfunction and mortality. Arterioscler. Thromb. Vasc. Biol. 34, 1032–1044.

DeFronzo, R.A., Ferrannini, E., Groop, L., Henry, R.R., Herman, W.H., Holst, J.J., Hu, F.B., Kahn, C.R., Raz, I., Shulman, G.I., Simonson, D.C., Testa, M.A., Weiss, R., 2015. Type 2 diabetes mellitus. Nat Rev Dis Primers 1, 15019.

Denison, H.J., Cooper, C., Sayer, A.A., Robinson, S.M., 2015. Prevention and optimal management of sarcopenia: a review of combined exercise and nutrition

interventions to improve muscle outcomes in older people. Clin Interv Aging 10, 859–869.

Dewick, P.M., 2009. Secondary metabolism: The building blocks and construction mechanisms, in: medicinal natural products. John Wiley & Sons, Ltd, Chichester, UK, pp. 7–38.

Dutt, V., Gupta, S., Dabur, R., Injeti, E., Mittal, A., 2015. Skeletal muscle atrophy: Potential therapeutic agents and their mechanisms of action. Pharmacol. Res. 99, 86–100.

Ekanayaka, E.A.P., Celiz, M.D., Jones, A.D., 2015. Relative mass defect filtering of mass spectra: a path to discovery of plant specialized metabolites. Plant Physiol. 167, 1221–1232.

English, K.L., Paddon-Jones, D., 2010. Protecting muscle mass and function in older adults during bed rest. Curr Opin Clin Nutr Metab Care 13, 34–39.

Fischer, E.H., Charbonneau, H., Tonks, N.K., 1991. Protein tyrosine phosphatases: a diverse family of intracellular and transmembrane enzymes. Science 253, 401–406.

Franck, B., 1979. Key building blocks of natural product biosynthesis and their significance in chemistry and medicine. Angew. Chem. Int. Ed. Engl. 18, 429–439.

Frank, R.A., Mize, S.J.S., Kennedy, L.M., de los Santos, H.C., Green, S.J., 1992. The effect of *Gymnema sylvestre* extracts on the sweetness of eight sweeteners. Chem. Senses 17, 461–479.

Fu, X., Zhu, M., Zhang, S., Foretz, M., Viollet, B., Du, M., 2016. Obesity impairs skeletal muscle regeneration through inhibition of AMPK. Diabetes 65, 188–200.

Fu, X., Zhu, M.-J., Dodson, M.V., Du, M., 2015. AMP-activated protein kinase stimulates Warburg-like glycolysis and activation of satellite cells during muscle regeneration. J. Biol. Chem. 290, 26445–26456.

Gent, J.F., Hettinger, T.P., Frank, M.E., Marks, L.E., 1999. Taste confusions following gymnemic acid rinse. Chem. Senses 24, 393–403.

Go, G.-Y., Lee, S.-J., Jo, A., Lee, J., Seo, D.-W., Kang, J.-S., Kim, S.-K., Kim, S.-N., Kim, Y.K., Bae, G.-U., 2017. Ginsenoside Rg1 from *Panax ginseng* enhances myoblast differentiation and myotube growth. Journal of Ginseng Research 41, 608–614.

Guo, X.L., Wang, T.J., Bian, B.L., 1997. Studies on the chemical constituents of Gynostemma longipes C.Y. Wu. Yao Xue Xue Bao 32, 524–529.

HaiLong, L., 1998. Flora of Taiwan, 2nd ed. Editorial Committee of the Flora of Taiwan, Department of Botany, National Taiwan University.

Hardie, D.G., Ross, F.A., Hawley, S.A., 2012. AMPK: a nutrient and energy sensor that maintains energy homeostasis. Nat. Rev. Mol. Cell Biol. 13, 251–262.

Hawke, T.J., Garry, D.J., 2001. Myogenic satellite cells: physiology to molecular biology. J Appl Physiol 91, 534–551.

He, R.-J., Yu, Z.-H., Zhang, R.-Y., Zhang, Z.-Y., 2014. Protein tyrosine phosphatases as potential therapeutic targets. Acta Pharmacol. Sin. 35, 1227–1246.

Hossain, P., Kawar, B., Nahas, El, M., 2007. Obesity and diabetes in the developing world--a growing challenge. N. Engl. J. Med. 356, 213–215.

Hung, T.M., Hoang, D.M., Kim, J.C., Jang, H.-S., Ahn, J.S., Min, B.S., 2009. Protein tyrosine phosphatase 1B inhibitory by dammaranes from Vietnamese Giao-Co-Lam tea. J Ethnopharmacol 124, 240–245.

Ito, A., Chai, H.-B., Kardono, L.B.S., Setowati, F.M., Afriastini, J.J., Riswan, S., Farnsworth, N.R., Cordell, G.A., Pezzuto, J.M., Swanson, S.M., Kinghorn, A.D., 2004. Saponins from the bark of *Nephelium maingayi*. J. Nat. Prod. 67, 201–205.

Jiang, C.-S., Liang, L.-F., Guo, Y.-W., 2012. Natural products possessing protein tyrosine phosphatase 1B (PTP1B) inhibitory activity found in the last decades. Acta Pharmacol. Sin. 33, 1217–1245.

Jura, M., Kozak, L.P., 2016. Obesity and related consequences to ageing. AGE 38, 23.

Jürgens, A., Dötterl, S., 2004. Chemical composition of anther volatiles in Ranunculaceae: genera-specific profiles in Anemone, Aquilegia, Caltha, Pulsatilla, Ranunculus, and Trollius species. Am. J. Bot. 91, 1969–1980. doi:10.3732/ajb.91.12.1969

Kalyani, R.R., Corriere, M., Ferrucci, L., 2014. Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. The Lancet Diabetes & Endocrinology 2, 819–829.

Kashima, H., Eguchi, K., Miyamoto, K., Fujimoto, M., Endo, M.Y., Aso-Someya, N., Kobayashi, T., Hayashi, N., Fukuba, Y., 2017. Suppression of oral sweet taste sensation with *Gymnema sylvestre* affects postprandial gastrointestinal blood flow and gastric emptying in humans. Chem. Senses 42, 295–302.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Leach, M.J., 2007. *Gymnema sylvestre* for diabetes mellitus: a systematic review. J Altern Complement Med 13, 977–983.

Lee, C., Lee, J.W., Jin, Q., Jang, H., Jang, H.-J., Rho, M.-C., Lee, M.K., Lee, C.K., Lee, M.K., Hwang, B.Y., 2015. Isolation and characterization of dammarane-type saponins from *Gynostemma pentaphyllum* and their inhibitory effects on IL-6-induced STAT3 activation. J. Nat. Prod. 78, 971–976.

Lee, H.M., Lee, O.H., Kim, K.J., Lee, B.Y., 2012. Ginsenoside Rg1 promotes glucose uptake through activated AMPK pathway in Insulin-resistant muscle cells. Phytotherapy Research 26, 1017–1022. doi:10.1002/ptr.3686

Li, W.-X., Chen, S.-F., Chen, L.-P., Yang, G.-Y., Li, J.-T., Liu, H.-Z., Zhu, W., 2012. Thimerosal-induced apoptosis in mouse C2C12 myoblast cells occurs through suppression of the PI3K/Akt/Survivin pathway. PLoS ONE 7, e49064.

LIU, H.-M., KIUCHI, F., TSUDA, Y., 1992. Isolation and Structure Elucidation of Gymnemic Acids, Antisweet Principles of *Gymnema sylvestre*. Chem. Pharm. Bull. 40, 1366–1375.

Liu, X., Ye, W., Xu, D., Zhang, Q., Che, Z., Zhao, S., n.d. Chemical study on triterpene and saponin from *Gymnema sylvestre*. Zhongguo Yaoke Daxue Xuebao 30, 174–176.

Liu, Y., Xu, T.-H., Zhang, M.-Q., Li, X., Xu, Y.-J., Jiang, H.-Y., Liu, T.-H., Xu, D.-M., 2014. Chemical constituents from the stems of *Gymnema sylvestre*. Chin J Nat Med 12, 300–304.

Marcovecchio, M.L., Dalton, R.N., Daneman, D., Deanfield, J., Jones, T.W., Neil, H.A.W., Dunger, D.B., Adolescent type 1 Diabetes cardio-renal Intervention Trial (AdDIT) study group, 2019. A new strategy for vascular complications in young people with type 1 diabetes mellitus. Nat Rev Endocrinol 376, 1419.

Medema, R.H., Macůrek, L., 2012. Checkpoint control and cancer. Oncogene 31, 2601–2613.

Meve, U., Alejandro, G.J., 2011. Taxonomy of the Asian *Gymnema latifolium* (Apocynaceae, Asclepiadoideae, Marsdenieae), including lectotypification of the synonymous *G. khandalense*. Will 43, 81–86. doi:10.3372/wi.43.43108

Mohn, T., Plitzko, I., Hamburger, M., 2009. A comprehensive metabolite profiling of *Isatis tinctoria* leaf extracts. Phytochemistry 70, 924–934.

Na, M., Cui, L., Min, B.S., Bae, K., Yoo, J.K., Kim, B.Y., Oh, W.K., Ahn, J.S., 2006a. Protein tyrosine phosphatase 1B inhibitory activity of triterpenes isolated from *Astilbe koreana*. Bioorg. Med. Chem. Lett. 16, 3273–3276.

Na, M., Yang, S., He, L., Oh, H., Kim, B.S., Oh, W.K., Kim, B.Y., Ahn, J.S., 2006b. Inhibition of protein tyrosine phosphatase 1B by ursane-type triterpenes isolated from *Symplocos paniculata*. Planta Med. 72, 261–263.

Nguyen, P.H., Gauhar, R., Hwang, S.L., Dao, T.T., Park, D.C., Kim, J.E., Song, H., Huh, T.L., Oh, W.K., 2011. New dammarane-type glucosides as potential activators of AMP-activated protein kinase (AMPK) from *Gynostemma pentaphyllum*. Bioorg. Med. Chem. 19, 6254–6260.

Pham, H.T.T., Hoang, M.C., Ha, T.K.Q., Dang, L.H., Tran, V.O., Nguyen, T.B.T., Lee, C.H., Oh, W.K., 2018. Discrimination of different geographic varieties of *Gymnema sylvestre*, an anti-sweet plant used for the treatment of type 2 diabetes. Phytochemistry 150, 12–22.

Pothuraju, R., Sharma, R.K., Chagalamarri, J., Jangra, S., Kumar Kavadi, P., 2014. A systematic review of *Gymnema sylvestre* in obesity and diabetes management. J. Sci. Food Agric. 94, 834–840.

Qi, L.-W., Wang, H.-Y., Zhang, H., Wang, C.-Z., Li, P., Yuan, C.-S., 2012. Diagnostic ion filtering to characterize ginseng saponins by rapid liquid chromatography with time-of-flight mass spectrometry. J Chromatogr A 1230, 93–99.

Qian, S., Li, H., Chen, Y., Zhang, W., Yang, S., Wu, Y., 2010. Synthesis and biological evaluation of oleanolic acid derivatives as inhibitors of protein tyrosine phosphatase 1B. J. Nat. Prod. 73, 1743–1750.

Qiao, X., Li, R., Song, W., Miao, W.-J., Liu, J., Chen, H.-B., Guo, D.-A., Ye, M., 2016. A targeted strategy to analyze untargeted mass spectral data: Rapid chemical profiling of *Scutellaria baicalensis* using ultra-high performance liquid chromatography coupled with hybrid quadrupole orbitrap mass spectrometry and key ion filtering. J Chromatogr A 1441, 83–95.

Ruderman, N.B., Carling, D., Prentki, M., Cacicedo, J.M., 2013. AMPK, insulin resistance, and the metabolic syndrome. J. Clin. Invest. 123, 2764–2772.

Shobana, S., Kumari, S.R.U., Malleshi, N.G., Ali, S.Z., 2007. Glycemic response of rice, wheat and finger millet based diabetic food formulations in normoglycemic subjects. Int J Food Sci Nutr 58, 363–372.

Staszkiewicz, J., Gimble, J., Cain, C., Dietrich, M., Burk, D., Kirk-Ballard, H., Gawronska-Kozak, B., 2009. Flow cytometric and immunohistochemical detection of in vivo BrdU-labeled cells in mouse fat depots. Biochem. Biophys. Res. Commun. 378, 539–544.

Tews, D.S., Goebel, H.H., Schneider, I., Gunkel, A., Stennert, E., Neiss, W.F., 1997. DNA-fragmentation and expression of apoptosis-related proteins in experimentally denervated and reinnervated rat facial muscle. Neuropathology and Applied Neurobiology 23, 141–149.

Tiwari, P., Ahmad, K., Baig, M.H., 2017. *Gymnema sylvestre* for diabetes: From traditional herb to future's therapeutic. Curr. Pharm. Des. 23, 1667–1676.

Tiwari, P., Mishra, B.N., Sangwan, N.S., 2014. Phytochemical and pharmacological properties of *Gymnema sylvestre*: an important medicinal plant. Biomed Res Int 2014, 830285–18.

Wagatsuma, A., Sakuma, K., 2013. Mitochondria as a potential regulator of myogenesis. The Scientific World Journal 2013, 1–9.

Warren, R.M., Pfaffmann, C., 1959. Suppression of sweet sensitivity by potassium gymnemate. J Appl Physiol 14, 40–42.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in: PCR Protocols. Elsevier, pp. 315–322.

Wolfender, J.-L., Marti, G., Thomas, A., Bertrand, S., 2015. Current approaches and challenges for the metabolite profiling of complex natural extracts. J Chromatogr A 1382, 136–164.

Wu, Z.Y., Raven, P.H., 1995. *Flora of China*. *Vol*. *16 (Gentianaceae through Boraginaceae)*. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis.

Xie, T., Liang, Y., Hao, H., A, J., Xie, L., Gong, P., Dai, C., Liu, L., Kang, A., Zheng, X., Wang, G., 2012. Rapid identification of ophiopogonins and ophiopogonones in *Ophiopogon japonicus* extract with a practical technique of mass defect filtering based on high resolution mass spectrometry. J Chromatogr A 1227, 234–244.

Xing, R., Zhou, L., Xie, L., Hao, K., Rao, T., Wang, Q., Ye, W., Fu, H., Wang, X., Wang, G., Liang, Y., 2015. Development of a systematic approach to rapid

classification and identification of notoginsenosides and metabolites in rat feces based on liquid chromatography coupled triple time-of-flight mass spectrometry. Analytica Chimica Acta 867, 56–66.

Yang, B., Li, H., Ruan, Q.-F., Xue, Y.-Y., Di Cao, Zhou, X.-H., Jiang, S.-Q., Yi, T., Jin, J., Zhao, Z.-X., 2018. A facile and selective approach to the qualitative and quantitative analysis of triterpenoids and phenylpropanoids by UPLC/Q-TOF-MS/MS for the quality control of *Ilex rotunda*. Journal of Pharmaceutical and Biomedical Analysis 157, 44–58.

Yang, J.L., Ha, T.K.Q., Lee, B.W., Kim, J., Oh, W.K., 2017. PTP1B inhibitors from the seeds of *Iris sanguinea* and their insulin mimetic activities via AMPK and ACC phosphorylation. Bioorg. Med. Chem. Lett.

Ye, W., Liu, X., Zhang, Q., Che, C.T., Zhao, S., 2001a. Antisweet saponins from *Gymnema sylvestre*. J. Nat. Prod. 64, 232–235.

Ye, W.-C., Liu, X., Zhao, S.-X., Che, C.-T., 2001b. Triterpenes from *Gymnema sylvestre* growing in China. Biochemical Systematics and Ecology 29, 1193–1195.

Ye, W.-C., Zhang, Q.-W., Liu, X., Che, C.-T., Zhao, S.-X., 2000a. Oleanane saponins from *Gymnema sylvestre*. Phytochemistry 53, 893–899.

Ye, W.-C., Zhang, Q.-W., Liu, X., Che, C.-T., Zhao, S.-X., 2000b. Oleanane saponins from *Gymnema sylvestre*. Phytochemistry 53, 893–899.

Yen, Y.-P., Tsai, K.-S., Chen, Y.-W., Huang, C.-F., Yang, R.-S., Liu, S.-H., 2012. Arsenic induces apoptosis in myoblasts through a reactive oxygen species-induced endoplasmic reticulum stress and mitochondrial dysfunction pathway. Arch Toxicol 86, 923–933.

Yoshikawa, K., Amimoto, K., Arihara, S., Matsuura, K., 1989. Structure studies of new antisweet constituents from *Gymnema sylvestre*. Tetrahedron Letters 30, 1103–1106.

Yoshikawa, K., Arihara, S., Matsuura, K., 1991. A new type of antisweet principles occurring in *Gymnema sylvestre*. Tetrahedron Letters 32, 789–792.

Yoshikawa, K., Nakagawa, M., Yamamoto, R., Arihara, S., Matsuura, K., 1992. Antisweet Natural Products. V. Structures of Gymnemic acids VIII-XII from *Gymnema sylvestre* R. BR. Chem. Pharm. Bull. 40, 1779–1782.

Yoshikawa, K., OGATA, H., Arihara, S., Chang, H.-C., Wang, J.-D., 1998. Antisweet Natural Products. XIII. Structures of Alternosides I-X from *Gymnema alternifolium*. Chem. Pharm. Bull. 46, 1102–1107.

Yoshikawa, K., Takahashi, K., Matsuchika, K., Arihara, S., Chang, H.-C., Wang, J.-D., 1999. Antisweet Natural Products. XIV. Structures of Alternosides XI-XIX from *Gymnema alternifolium*. Chem. Pharm. Bull. 47, 1598–1603.

Yoshikawa, K., Taninaka, H., Kan, Y., Arihara, S., 1994a. Antisweet natural products. XI. Structures of sitakisosides VI-X from *Stephanotis lutchuensis* Koidz. var. *japonica*. Chem. Pharm. Bull. 42, 2455–2460.

Yoshikawa, K., Taninaka, H., Kan, Y., Arihara, S., 1994b. Antisweet natural products. X. Structures of sitakisosides I-V from *Stephanotis lutchuensis* Koidz. var. japonica. Chem. Pharm. Bull. 42, 2023–2027.

Yoshikawa, M., Murakami, T., Matsuda, H., 1997. Medicinal foodstuffs. X. Structures of new triterpene glycosides, gymnemosides-c, -d, -e, and -f, from the leaves of *Gymnema sylvestre* R. Br.: influence of gymnema glycosides on glucose uptake in rat small intestinal fragments. Chem. Pharm. Bull. 45, 2034–2038.

Zabolotny, J.M., Bence-Hanulec, K.K., Stricker-Krongrad, A., Haj, F., Wang, Y., Minokoshi, Y., Kim, Y.-B., Elmquist, J.K., Tartaglia, L.A., Kahn, B.B., Neel, B.G., 2002. PTP1B regulates leptin signal transduction in vivo. Dev. Cell 2, 489–495.

Zhang, M.-Q., Liu, Y., Xie, S.-X., Xu, T.-H., Liu, T.-H., Xu, Y.-J., Xu, D.-M., 2012. A new triterpenoid saponin from *Gymnema sylvestre*. J Asian Nat Prod Res 14, 1186–1190.

Zhang, Y.-N., Zhang, W., Hong, D., Shi, L., Shen, Q., Li, J.-Y., Li, J., Hu, L.-H., 2008. Oleanolic acid and its derivatives: new inhibitor of protein tyrosine phosphatase 1B with cellular activities. Bioorg. Med. Chem. 16, 8697–8705.

Zhu, X.-M., Xie, P., Di, Y.-T., Peng, S.-L., Ding, L.-S., Wang, M.-K., 2008. Two new triterpenoid saponins from *Gymnema sylvestre*. J Integr Plant Biol 50, 589–592.

국문초록

Part 1.

*Gymnema sylvestre*는 전통적으로 아유르베다 의학에서 당뇨치료에 사용되어왔다. 본 연구에서는 베트남과 인도의 *Gymnema sylvestre*을 형태학적으로 구분하였고 베트남의 *Gymnema sylvestre*에서 분리한 6개의 신규화합물(1-6)의 조각이온 패턴을 화학성분 분석에 활용하였다. 본 식물의 분리 및 동정을 통하여 이미 알고 있는 조각이온의 패턴과 이 식물의 oleanane triterpenoid 의 생합성 과정을 토대로 building block 전략을 활용하여 각각의 피크를 동정하였다. 이 과정에서 소분획, mass 조각이온 패턴, relative mass defect filtering, reference와 비교 등의 방법이 적용되었다. 총 119개의 피크가 oleanane triterpenoid 계열로 분류되었으며 그 중 77개의 피크가 신규화합물임을 확인하였다.

Part 2.

천연물은 PTP1B억제를 통한 항당뇨 효과 후보물질의 훌륭한 원천이며 이에 따라 선별된 Korea Bioactive Natural Material Bank의 식물추출물들에 대한 스크리닝 작업을 실시하였으며 그 결과 *Gymnema latifolium* 70% 에탄올

추출물이 PTP1B 억제 효능을 보였다. 이는 현재 베트남에서 항당뇨 치료제로 *Gymnema latifolium*이 활용되고 있는 결과와도 일치하였다. 이에 본 연구는 14종의 신규 oleanane triterpene인 Gymlatinosides (1-14)와 6개의 기지 물질을 분리 동정하였다. Gymlatinoside GL2와 GL3는 생리활성 결과 각각 IC50 값 28.66±2.57 과 19.83±0.40을 나타내어 PTP1B를 유의미하게 억제하였음을 확인하였다. 또한 building block 전략을 이용하여 54개의 신규 피크를 추정하였다.

Part 3.

Gynostemma longipes 에서는 8종의 신규 12,23-dione dammarane triterpenoid와 1종의 gypetonoside가 분리되었으며 building block 전략을 이용하여 32개의 피크가 신규물질로 추정됨을 확인하였다. 분리한 9종의 화합물에 대해서는 AMPK 활성을 검사하였으며 C2C12 근육세포에 7종의 화합물 (1, 3-8)을 처치하였을 경우 AMPK 활성을 증가시키는 것으로 확인하였다. 화합물 1의 경우 식물 건조량 대비 2.08%의 함량을 가지는 것으로 평가되어 향후 기능성식품으로의 발전 가능성을 제시하였다. 본 연구에서는 *Gynostemma longipes*의 AMPK 활성을 통한 근육세포 증식을 확인하였다.

주요어: *Gynostemma longipes*, *Gymnema sylvestre*, *Gymnema latifolium*, oleanane triterpenes, dammarane triterpenes, building block 전략, PTP1B, 근육세포증식.

학번: 2015-30877