Molecular cloning of an ester-forming triterpenoid: UDP-glucose 28-O-glucosyltransferase involved in saponin biosynthesis from the medicinal plant Centella asiatica

De Costa, Fernanda; Barber, Carla J. S.; Kim, Yeon-bok; Reed, Darwin W.; Zhang, Haixia; Fett-Neto, Arthur G.; Covello, Patrick S.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/j.plantsci.2017.05.009
Plant Science, 262, pp. 9-17, 2017-05-23

Molecular cloning of an ester-forming triterpenoid: UDP-glucose 28-*O*-glucosyltransferase involved in saponin biosynthesis from the medicinal plant *Centella asiatica*

Fernanda de Costa[a, b, 1], Carla J. S. Barber[a, 1], Yeon-bok Kim[a, c], Darwin W. Reed[a], Haixia Zhang[a], Arthur G. Fett-Neto[b] and Patrick S. Covello[a, *]

[a] National Research Council of Canada, 110 Gymnasium Pl., Saskatoon, SK S7N 0W9, Canada

[b] Plant Physiology Laboratory, Center for Biotechnology and Department of Botany, Federal University of Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500, Porto Alegre, RS 91501-970, Brazil

[c] Herbal Crop Research Division, Department of Herbal Crop Research, Bisanro 92, Eumseong, Chungbuk, 369-873, Korea

[1] These authors equally contributed to this work.

* Correspondence to:

Patrick S. Covello

National Research Council of Canada, 110 Gymnasium Pl., Saskatoon S7N 0W9, Canada

Phone: +1-306-975-5269

FAX: +1-306-975-4839

E-mail address: *patrick.covello*@*nrc-cnrc.gc.ca*

Fernanda de Costa – *fernandadecosta*@*yahoo.com.br*

Carla J. S. Barber – *carla.barber@nrc-cnrc.gc.ca*

Yeon-bok Kim – *yeonbokkim@daum.net*

Darwin W. Reed – *darwin.reed@sasktel.net*

Haixia Zhang – *haixia.zhang@nrc-cnrc.gc.ca*

Arthur G. Fett-Neto – *fettneto@cbiot.ufrgs.br*

Abstract

Triterpene saponins include bioactive compounds with structures consisting of triterpene aglycones (sapogenins) and one or more sugar moieties linked through acetal or ester glycosidic linkages at one or more sites. *Centella asiatica* (L.) Urban is a medicinal plant that contains bioactive ursane-type saponins, such as madecassoside and asiaticoside. In this work, glucosylation of triterpenoids in *C. asiatica* was investigated starting with plant extracts. An enzyme capable of glucosylating asiatic and madecassic acids was partially purified. Proteomics methods and cDNA sequence data were employed as tools to obtain a full-length cDNA clone encoding a glucosyltransferase. The recombinant gene product, *UGT73AD1*, was functionally expressed in *Escherichia coli* and purified by immobilized metal-affinity chromatography. Purified recombinant UGT73AD1 was found to have a narrow specificity, glucosylating asiatic and madecassic acids at the C28 carboxyl. mRNA accumulated in all tissues tested (leaves, stems, roots and flowers), with highest expression in leaves. Thus, UGT73AD1 was identified as a triterpenoid carboxylic acid: UDP-glucose 28-*O*-glucosyltransferase that appears to be involved in saponin biosynthesis in *C. asiatica*.

Keywords: *Centella asiatica*; Apiaceae; saponin biosynthesis; glucosyltransferase; UGT73AD1

Highlights

- *Centella asiatica* contains bioactive ursane triterpenoid saponins
- *C. asiatica* cDNA encoding a triterpenoid UDP-glucosyltransferase (UGT) was cloned
- The UGT is a triterpenoid carboxylic acid:UDP-glucose 28-*O*-glucosyltransferase
- UGT73AD1 has narrow specificity, being active with asiatic and madecassic acid

Abbreviations

BAS, β-amyrin synthase; BSA, bovine serum albumin; DTT, dithiothreitol; EST, expressed sequence tags; FPP, farnesyl diphosphate; LC-MS, liquid chromatography–mass spectrometry, MeJA, methyl jasmonate; OSC, oxidosqualene cyclase; PSPG, plant secondary product glycosyltransferase; UDP, uridine diphosphate; UGT, UDP-dependent glucosyltransferase.

1. Introduction

Triterpene saponins are a heterogeneous group of bioactive compounds with structures consisting of triterpene aglycones (sapogenins) and one or more sugar moieties linked through an acetal or ester glycosidic linkage at one or more sites. *Centella asiatica* (L.) Urban (Apiaceae), commonly known as Gotu kola or Indian pennywort, is a medicinal plant that contains bioactive ursane-type saponins, such as madecassoside and asiaticoside (see Fig. 1), in its aerial parts (reviewed in [1]). This plant has been used to treat skin diseases, mental illnesses, and leprosy. Various bioactivities have been shown for madecassoside (anti-inflammatory, antioxidant, cardioprotective, neuroprotective) [2-4] and asiaticoside (wound healing, antioxidant, anxiolytic) [5-7].

Despite its biological importance, triterpene saponin biosynthesis has yet to be fully characterized at the molecular level. There is not yet a complete complement of genes identified for synthesis of the backbone nor the subsequent glucosylation steps involved in production of bioactive saponins. However, advances in genomics technologies, such as expressed sequence tags (ESTs) and RNAseq, provide an economical platform for gene discovery in non-model plants such as *C. asiatica* [8]. Identification of the saponin-related genes together with corresponding functional *in vitro* studies could prove valuable for metabolic engineering to increase production of these triterpene saponins in plants and/or microbes.

The triterpenoid saponin skeletons, which include oleanane, ursane, lupane and dammarane types, are synthesized via the isoprenoid pathway through farnesyl diphosphate (FPP) followed by cyclization of 2,3-oxidosqualene by 2,3 oxidosqualene cyclase (OSC). β-amyrin synthase (BAS), a widely distributed OSC, has been functionally characterized from various plants, reviewed in [9]. In *C. asiatica*, genes involved in the main pathway to triterpenoid

formation have been discovered. These include farnesyl diphosphate synthase (CaFPS) [10] squalene synthase (CaSQS) [11] , oxidosqualene synthase (CaOSQs) and the putative β-amyrin synthase (CabAs) [12], an oxidosqualene cyclase, subsequently identified as a dammarenediol synthase (CaDDs) [13] after being cloned and functionally expressed in a lanosterol synthase-deficient yeast mutant.

Modifications to the triterpenoid backbone include oxidation, substitution or glucosylation [14] by enzymes such as P450-dependent monooxygenases, UDP-Glc glucosyltransferases (UGTs) and other enzymes [15-17]. UGTs are pivotal enzymes in the process of glucosylation in plants, contributing to the biosynthesis and stable storage of medically important secondary metabolites [14]. Furthermore, glucosylation has been found to be critical for saponin biological activities [18, 19]. In saponin biosynthesis, UGT's catalyze the transfer of UDP linked sugar moieties to the triterpenoid skeleton. To date, a limited number of UGTs have been uncovered from plants, including UGT73K1 and UGT71G1 [20] and UGT73F3 [16] from *M. truncatula*, UGT74M1 from *S. vaccaria* [21], UGT73P2 and UGT91H4 [22] and UGT73F2 and UGT73F4 from *Glycine max* [23], UGT73C10, UGT73C11, UGT73C12 and UGT73C13 from *Barbarea vulgaris* [24], UGT74H5 from *Avena* spp. [25], UGTPg45 and UGTPg29 from *Panax ginseng* [26] and Pq3-O-UGT2 from *P. quinquefolius* [27]. In this work, glucosylation of triterpenoids in *C. asiatica* was investigated starting with enzyme assays of plant extracts. An enzyme capable of glucosylating asiatic and madecassic acids (see Fig. 1) was partially purified. Proteomic methods and cDNA sequence data were used to obtain a full-length cDNA clone encoding the glucosyltransferase for which the corresponding gene product was characterized.

2. Materials and Methods

2.1. Plant material

Centella asiatica (L.) Urban (Gotu kola) 4-month old plantlets were obtained from Richters Herbs (Goodwood, Canada). The plants were grown under greenhouse conditions at 23°C and 12 h photoperiod.

2.2. Partial purification of a glucosyltransferase from crude extracts of C. asiatica

A crude protein extract from *C. asiatica* leaves was obtained using a protocol previously described [28]. The extraction buffer contained 100 mM of Tris-HCl pH 7.6, 150 mM sorbitol, 12.5 mM β-mercaptoethanol, 1 mM PMSF, 2% (w/v) polyvinylpolypyrrolidone (PVPP) and 0.05% (v/v) plant protease inhibitor cocktail (Sigma®) containing 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), Bestatin, Pepstatin A, E-64, Leupeptin, and 1,10-phenanthroline. Approximately 1 g of leaves was ground with PVPP on ice and stirred with 5 ml of buffer for 5 min. Subsequently, the extract was centrifuged at 4ºC at 14,000 × *g* for 10 min. The recovered supernatant was applied to a PD-10 (Sephadex® G-25 M) for desalting and buffer exchange. The resulting eluate was considered a crude protein extract. The amount of protein in each extract was determined using a modified micro BCA protein assay (Pierce) [29]. Briefly, 10 µl samples were mixed with 90 µl of BCA working solution and incubated at 60 ºC for 30 min, after which optical density at 562 nm was recorded. Glucosyltransferase enzyme assays were carried out in a total volume of 100 µl containing 10 µl of crude protein extract (1.8 to 2.2 mg

139 ml^{-1} of protein), 250 µM UDP-Glc, 1 mM acceptor substrate, 100 mM HEPES pH 7.5, 1 mM
140 DTT and 10 mM of $MgCl_2$. The reaction mixture was incubated at 30 ºC for 1 h and stopped by
141 transferring the reaction to -80 ºC. Samples were lyophilized and resuspended in 20 µl of 70%
142 (v/v) MeOH. Samples were sonicated for 3 min and centrifuged at 12,000 × *g* for 1 min. Samples
143 were then transferred to LC/MS vials for analysis.
144
145 *2.3. Ion trap LC/MS analysis of enzyme assays*
146
147 Enzyme reactions were analyzed using an Agilent 6320 ion trap LC/MS system with
148 electrospray source under default Smart Parameter settings (scanning in the *m/z* range of 50-2200
149 at 8100 mass units s^{-1} with an expected peak width of ≤ 0.35 mass units), with a reverse phase
150 column (Zorbax 80A Extended-C18 column, 5 µm particle size, 2.1 × 150 mm) maintained at 35
151 °C. The binary solvent system consisted of H_2O-acetonitrile (90:10 v/v) containing 0.1% (v/v)
152 formic acid and 0.1% (w/v) ammonium formate (solvent A) and H_2O-acetonitrile (10:90 v/v)
153 containing 0.1% (v/v) formic acid and 0.1% (w/v) ammonium formate (solvent B). The
154 separation gradient was 90:10 A/B to 0:100 A/B over 4.5 ml. The monoglucoside products were
155 detected as $[M+formate]^-$ ions and the substrates madecassic acid and asiatic acid were detected
156 as $[2M]^+$ ions.
157
158 *2.4. Isolation and elucidation of monoglucoside products*
159
160 For isolation of the monoglucosylated products of madecassic and asiatic acid, 500 µl
161 enzyme reactions were performed containing 90 to 110 µg of protein extract (1.8 to 2.2 mg ml^{-1}),

250 µM UDP-Glc, 1 mM substrate (asiatic or madecassic acid), 1 mM DTT, 10 mM $MgCl_2$ and 20 mM Bicine, pH 8.5. The assay was incubated for 4 h at 30 ºC under agitation. Samples were then lyophilized and resuspended in 100 µl of 70% (v/v) MeOH. Separation was performed using a Hewlett Packard 1100 Series HPLC and an Agilent Eclipse XDB-C18 column (5 µm, 4.6 × 150 mm), at a flow rate of 1 ml min^{-1}. A binary solvent system was used, consisting of 100% acetonitrile (solvent A) and 10% acetonitrile (solvent B), and the separation system went from 100% solvent B to 100% solvent A in 20 min and was then maintained at 100% B for 15 min. One ml fractions were collected and analyzed by ion trap LC/MS.

For peak identification by NMR, HPLC fractions corresponding to retention times of 9 to 10 min (madecassic acid-derived product) or 10 to 11 min (asiatic acid-derived product) were evaporated to complete dryness. After dissolution in $CDCl_3$, the proton NMR spectrum was recorded on a Bruker Avance DRX 500 MHz spectrometer equipped with a CryoProbe.

2.5. Separation of fractions containing glucosyltransferase activity

For partial purification of the UGT catalyzing triterpenoid glucosylation, 1 ml of crude protein extract, containing ~2 mg of protein, was subjected to anion exchange chromatography using a Mono-Q 5/50 GL column connected to an Agilent 1100 series HPLC equipped with auto injector, a diode array detector and a fraction collector maintained at 5 °C. The column was pre-equilibrated with 20 mM Tris-HCl, pH 8.0. The column was eluted with a mobile phase of 20 mM Tris-HCl, pH 8.0, containing a linear gradient from zero to 1 M NaCl over 30 ml and maintained at 1 M NaCl for 5 ml. One ml samples were collected and assayed for glucosyltransferase activity.

Two fractions from the anion exchange chromatography that catalyzed the formation of madecassic acid monoglucoside - fractions 15 (28 to 30 min) and 16 (30 to 32 min) – were concentrated, mixed with 2x SDS-PAGE loading sample buffer (200 mM Tris pH 6.8, 4% (w/v) SDS, 0.2% (v/v) Bromophenol Blue, 20% (v/v) glycerol, 200 mM DTT) and heated at 95 °C for 5 min. Samples were subjected to SDS-PAGE under denaturing conditions using electrophoresis buffer (25 mM Tris-HCl, pH 7.5, 250 mM glycine, 0.1% (w/v) SDS and a 4% to 15% (w/v) polyacrylamide gradient Ready Gel® (Bio-Rad). Precision Plus Protein™ Unstained Standard (Bio-Rad) was loaded on the same gel. The gel was stained with Oriole™ fluorescent gel stain (Bio-Rad) overnight. The protein bands were visualized with UV light.

2.6. cDNA library construction

Transformed root cultures of *C. asiatica* were obtained by infecting nodes and leaves with *Agrobacterium rhizogenes* strain 15384. After two days of co-cultivation in the dark, explants were transferred to medium devoid of exogenous plant growth regulators containing 3% (w/v) sucrose. The hairy roots were sub-cultured at 3 week-intervals for further growth. For MeJA elicitation, hairy roots were grown at 23 ºC on a rotary shaker (120 rpm) in darkness for four weeks and then treated for 7 days with 100 μM MeJA [30]. Samples were selected and a cDNA library was constructed as described previously [31].

2.7. Isolation of a full-length cDNA clone from C. asiatica

After gel electrophoresis of anion exchange chromatography fractions (see above), twenty-five gel slices from the lanes of two active and the two non-active flanking sample lanes corresponding to a M_r range of 30,000 to 80,000 were excised from an SDS PAGE gel and digested with trypsin using a MassPrep II proteomics work station (Micromass UK Ltd.) following a published procedure [32]. Tryptic peptides from each gel slice were analyzed by Q-TOF LC/MS as described [29], and the resulting data were used for MASCOT searches (version 2.3.02; Matrix Science Ltd., London, UK) against data derived from Roche 454 GS-FLX Titanium and Illumina sequencing of cDNA from MeJA-induced *C. asiatica* root cultures [31].

A contiguous cDNA sequence named *CASRI1PC_rep_c607* was identified by MASCOT searches against a MIRA assembly of Roche 454-derived sequences, as a match to tryptic peptides derived from the two active fractions of SDS-PAGE gel slices (15-5 and 16-5) corresponding to M_r values of ~50,000 (See online supplementary data, Fig. S1). Sequence similarity searches indicated that CASRI1PC_rep_c607 may contain a full-length ORF encoding a UGT.

Platinum® Taq DNA polymerase (Invitrogen) with oligonucleotides (5'-ATGGCATCCA ACATTCAGCA GCTTCACTT-3' and 5'-TCAAAATCGG CTTGAATTTT GCTGCCTCAC AT-3' for the forward and reverse primers, respectively) was used to PCR amplify the gene corresponding to the ORF on *CASRIPC_rep_c607* using cDNA derived from MeJA-induced *C. asiatica* root cultures [31] under the following conditions: denaturation at 95 °C for 4 min; 35 cycles of 94 °C for 20 s, annealing at 55 °C for 30 s, and extension at 72 °C for 2.4 min; followed by 10 min at 72 °C.

The PCR product was gel-purified and cloned into the *E.coli* vector pCR™8/GW/TOPO using ONE SHOT® TOP10 chemically competent *E. coli* cells (Invitrogen) using a TA Cloning

kit (Invitrogen). Resulting transformants were selected on agar plates [Luria Broth (LB), 1% (w/v) agar, spectinomycin (100 μg ml^{-1})]. The resulting plasmids were sequenced with M13 forward and reverse primers. A plasmid containing the *CASRIPC_rep_c607* sequence (an ORF of 1485 bp) was identified and called pCB018. pCB018 was then recombined into the *E. coli* expression vector pDEST^{TM}17 (Invitrogen) giving rise to the plasmid pCB019 which encodes an N-terminal His-tagged fusion protein of 521 amino acids and *M*r of 59,047. pCB019 was used to transform BL21-AI^{TM} *E. coli* competent cells and the resulting clones were selected on agar plates [LB, 1% (w/v) agar, carbenicillin (100 μg ml^{-1})]. DNA sequencing was used to confirm that the construct was in frame for expression of His-tagged UGT73AD1.

2.8. Expression and purification of UGT73AD1

A 4 ml aliquot of an overnight LB (carbenicillin 100 μg ml^{-1}) culture of the BL21-AI/pCB019 strain was used to inoculate 50 ml of Autoinduction Medium [33]. The culture was incubated at 37 °C with agitation until the OD_{600} reached 0.4 at which time L-arabinose [0.2% (w/v)] was added to induce expression from the *araBAD* operon promoter and the culture was shifted to an incubation temperature of 16 °C with agitation overnight. The culture was centrifuged (2,000*g* at 4 °C for 10 min) and the resulting cell pellets were frozen at -80 °C. Pellets were resuspended in B-Per® bacterial protein extraction reagent (Thermo Scientific), following the manufacturer's instructions. The resulting lysates were then centrifuged (15,000 × *g*, 4 °C, 8 min), and the supernatant containing the His-tagged UGT73AD1 was mixed 1:1 (v/v) with metal affinity resin Binding Buffer, containing 200 mM Tris, pH 7.0, 800 mM NaCl and 10 mM imidazole, which was then added to HisPur^{TM} Cobalt Resin (Thermo Scientific) and

incubated for 1 h at 4 °C. His-tagged UGT73AD1 was eluted with 100 mM Tris-HCl pH 7.0, 400 mM NaCl and 150 mM imidazole, following the manufacturer's instructions. The amount of protein applied in the gel was 1.0 µg. For the *C. asiatica* glucosyltransferase, the reduction of pH from 7.5 to 7.0 significantly increases the yield and purity of protein in the elution buffer (data not shown).

Each elution batch was pooled and then concentrated and desalted using spin dialysis with Amicon Ultra Centrifugal Filters Ultracel 30K (Millipore). The protein was quantified using modified micro BCA protein assay (Pierce). Subsequently, the eluate was assayed for glucosyltransferase activity.

2.9. Optimized assay for recombinant UGT73AD1

Unless otherwise stated, glucosyltransferase enzyme reactions were carried out in 100 µl, containing 0.2% (w/v) BSA, 250 µM asiatic acid, 250 µM UDP-glucose, 1 mM DTT and 20 mM Bicine at pH 8.5. One hundred ng purified UGT73AD1 was added to start the reaction, followed by incubation for 1 h at 30 ºC. Reactions were stopped by transferring samples to -80 ºC. Samples were lyophilized and resuspended in 20 µl of 70% (v/v) MeOH. Samples were sonicated for 3 min and centrifuged at 12000 × *g* for 1 min. Samples were then analyzed by ion-trap LC/MS. Controls were performed with the omission of enzyme and the reactions stopped at 0 h.

2.10. Characterization of UGT73AD1 enzyme

BSA concentrations up to 0.3% (w/v) were assayed in 100 µl samples in triplicate. Reactions were performed as described above. The pH optimum of UGT73AD1 was evaluated in triplicate from 6.5 to 10.0 using five different buffers at 20 mM – MOPS sodium salt, pH 6.5 to 7.5; Bicine, pH 7.5 to 8.5; TRIS-HCl, pH 8.0 to 9.5; CHES, pH 8.5 to 9.5 and CAPS, pH 10.

Enzyme assays to determine substrate specificity were performed with the addition of the acceptor substrates to a final concentration of 250 µM in enzyme assay. Substrates were dissolved in absolute ethanol. Acceptor substrates in this study were as follows: oleanane-type sapogenins (β-amyrin, oleanolic acid, echinocystic acid, hederagenin, gypsogenic acid, gypsogenin, 16-OH gypsogenic acid and quillaic acid), ursane-type sapogenins (α-amyrin, ursolic acid, asiatic acid and madecassic acid), lupane-type sapogenins (lupeol, betulinic acid and betulin), polyphenols (quercetin, 2,3-dihydroquercetin and caffeic acid) and the phytosterol stigmasterol. Enzyme assays were also performed in triplicate with various sugar donors: UDP-glucose, UDP-glucuronic acid, UDP-β-L-arabinose and UDP-α-D-xylose at a final concentration of 250 µM.

2.11. Preparation of RNA and first strand cDNA synthesis

Total RNA was isolated from different organs from 6-month old *C. asiatica* plants. An Illustra RNAspin Mini Kit (GE Healthcare) was used for total RNA extraction from leaves, flowers, roots and stems. Total RNA concentration was measured by UV spectrophotometer (Nanodrop, Thermo Scientific) and nucleic acid quality was checked with 2100 Bioanalyzer instrument (Agilent) and by 1% (w/v) agarose gel electrophoresis (data not shown). First-strand cDNA synthesis was performed with the Superscript II Reverse Transcriptase (Invitrogen) using

299 1 µg of total RNA as template and Oligo $d(T)_{18}$ primers, in the presence of RNAse H (Thermo
300 Scientific), according to the manufacturer's instructions. The final cDNA products were diluted
301 100-fold in RNAse-free distilled H_2O prior to use in RT-qPCR.
302
303 *2.12. Selection of C. asiatica reference genes and primer design*
304
305 Six genes were selected to identify the most stable reference genes for RT-qPCR
306 expression analyses of the various organs – actin, two elongation factor protein (EFP) genes,
307 GAPDH, tubulin and ubiquitin. All sequences were named based on similarity to *Arabidopsis*
308 proteins determined via BLASTX searches of transcriptome information from Roche 454 and
309 Illumina sequencing of *C. asiatica* induced root cDNA [31]. Alignments with sequences from
310 other plants for each of these genes were used to design new primer pairs based on the most
311 conserved areas (online supplementary Table S1). Primer pairs for RT-qPCR were designed
312 using Oligo Perfect™ Designer software (Invitrogen) for each gene. The primer specification
313 included: melting temperatures of 59 to 61 ºC, primer length of 20 nt, GC contents of 45 to 55%
314 and PCR amplicon lengths of 100 to 250 bp.
315
316 *2.13. Real-time quantitative PCR analysis*
317
318 Isolation of total RNA from the samples and synthesis of first strand cDNA were
319 performed as described above. The synthesized cDNA was diluted 1:100 with H_2O, and 10 µl of
320 the diluted cDNA was used as a template for quantitative real-time PCR analysis. PCR reactions
321 were carried out in quadruplicate for each sample and performed in a total volume of 20 µl as

described [34]. The RT-qPCR analyses were performed in fast optical 48-well reaction plates 0.1 ml (MicroAmp™ Applied Biosystem) using a StepOne™ Real-Time PCR System (Applied Biosystems) according to the manufacturer instructions. The RT-qPCR program included an initial denaturation step at 95 °C for 5 min, followed by 40 cycles of 15 s at 95 °C, 10 s at 60 °C, and 15 s at 72 °C, followed by a heat dissociation curve from 60 °C to 90 °C. PCR efficiency was calculated with the LinReg PCR program [35] from the raw fluorescence data. Mean efficiencies of the repeated samples were used to adjust Ct values for further analysis. The expression stability of candidate reference genes was evaluated by two different statistical packages: geNorm [36] and NormFinder [37]. All RT-qPCR reactions were normalized using Ct value corresponding to the reference gene. The relative expression levels of target gene were calculated with formula $2^{-\Delta\Delta Ct}$ [38]. Values represented the average of three biological replicates.

3. Results

3.1. Identification of a triterpenoid glucosyltransferase from C. asiatica

With a view towards molecular cloning of the enzymes involved in saponin biosynthesis in *C. asiatica*, crude leaf extracts were tested for glucose transfer to asiatic and madecassic acids and other potential substrates. Assay samples were analyzed with an ion trap liquid chromatography/mass spectrometry (LC/MS). The analysis showed loss of substrate (by single ion monitoring) and formation of products corresponding to MS peaks consistent with asiatic acid monoglucoside (*m/z* = 695.84 $[M + formate]^-$) and madecassic acid monoglucoside (*m/z* = 711.60 $[M + formate]^-$). MS/MS of these ions resulted in the presumptive loss of glucose with

the appearance of diagnostic [$2M^+$] ion for their respective triterpenoids (*m/z* = 1008.68 for madecassic acid and *m/z* = 976.70 for asiatic acid). It was concluded that leaf extracts contain triterpenoid UDP-glucose glucosyltransferase activity. Partial purification of the crude leaf extracts that exhibited glucosyltransferase activity was performed with anion exchange chromatography on a Mono-Q 5/50 GL column (GE HealthCare). Glucosyltransferase enzyme assays were performed on the resulting fractions and the assays were analyzed with madecassic acid as acceptor and LC/MS/MS for detection and quantitation of the resulting monoglucoside. The majority of glucosyltransferase activity was found in fraction 15 (28 to 30 min; see Fig. 2.A, with a lesser amount in fraction 16 (30 to 32 min).

To identify putative UDP-glucosyltransferase candidates in the active chromatography fractions, MS of tryptic digests of the fractions was performed. For tryptic peptide analysis, the active fractions (15 and 16) and neighbouring inactive fractions obtained from anion exchange chromatography were subjected to SDS-PAGE, as shown in online supplementary data (Fig. S1). SDS-PAGE gel slices corresponding to M_r values of 30,000 to 80,000 were excised and treated with trypsin. The resulting peptides from each gel slice were analyzed by quadrupole-time-of-flight (Q-TOF) LC/MS to yield peptide mass and fragment ion information. These data were used in MASCOT searches against the predicted tryptic peptides derived from a MIRA assembly of Roche 454 sequencing of cDNA derived from methyl jasmonate (MeJA)-induced root cultures of *C. asiatica* (called CASRIPC) [31]. MS data from two excised gel bands (15-5 and 16-5; see online supplementary data, Fig. S1) corresponding to a M_r of ~50,000, were found to match three predicted tryptic sequences unique within the *C. asiatica* EST library corresponding to a contig, *CASRIPC_rep_c607,* namely TLGINFFQANEMLREPLEK, GVLVNTFEEMEPR and AIEDGGSSYLYITNLIEDVR. tBLASTn searches of the ORF of *CASRIPC_rep_c607* against

368 plant sequence databases (National Center for Biotechnology Information) revealed a high
369 similarity to UGTs. The corresponding gene product then became the focus for investigation as a
370 UGT involved in saponin biosynthesis. The full-length cDNA with an open reading frame (ORF)
371 of 1485 nucleotides encoding a polypeptide of 494 amino acids with a predicted M_r of 55,953,
372 was cloned and the gene product was named UGT73AD1 [39]. The cDNA sequence was
373 deposited in Genbank as accession no. KP195716.1.

374 Functional characterization of UGT73AD1 was performed with a recombinant N terminal
375 His-tagged enzyme (*M*r 59,047) corresponding to ORF cloned in pDEST^{TM}17 (Invitrogen). The
376 enzyme was purified by metal-affinity batch adsorption. A western blot was performed to
377 monitor an induction expression time course of UGT73AD1 which indicated maximal
378 expression to be ~19 h (data not shown). SDS-PAGE of the resulting purified His-tagged protein
379 UGT73AD1 is shown in Fig. 3. The purified recombinant protein was assayed with *C. asiatica*
380 sapogenins. Extracted ion chromatograms from ion trap LC/MS showed peaks with retention
381 times of 14.3 min or 15.6 min. These peaks were assigned as madecassic acid monoglucoside
382 (Fig. 2.B) and asiatic acid monoglucoside (Fig. 4) respectively, by MS and MS/MS (see
383 Experimental section).

384

385

386 *3.2. Characterization of UGT73AD1*

387

388 For optimum UGT73AD1 activity, the type of buffer, buffer pH, dithiothreitol (DTT)
389 concentration and bovine serum albumin (BSA) concentration were screened to determine
390 optimum reaction conditions. BSA showed an optimum concentration of 0.2% (w/v; online

391 supplementary material, Fig. S2). DTT was also found to be important for optimal UGT activity
392 in the 0.5 to 5 mM range. Five different buffers (20 mM) covering the pH range of 6.5 to 10.0
393 were screened to determine the optimum pH and buffer condition for UGT73AD1 activity. The
394 broad range pH screening indicated that UGT73AD1 was active throughout the range tested and
395 the optimum activity was centered between 8.5 and 9.0 (online supplementary data, Fig. S3). In
396 this optimum pH range, Bicine (pH 8.5) showed maximum product formation. Twenty mM
397 Bicine at pH 8.5 was chosen as optimum buffer and pH for the UGT73AD1 assays.

398 Various donor sugars and acceptor triterpenoid saponin aglycone, polyphenol and
399 phytosterol substrates were screened to determine UGT73AD1 specificity. The purified
400 recombinant enzyme was found to be inactive with both UDP-glucuronic acid and UDP-
401 arabinose when asiatic acid was the triterpene acceptor. However, UDP-xylose appeared to
402 exhibit 10% of activity found with the donor substrate UDP-glucose (online supplementary data,
403 Fig. S4). In reactions in which UDP-Glc was utilized as the donor substrate, UGT73AD1 was
404 found to be active with only asiatic and madecassic acids. No measurable activity was seen with
405 the other triterpenoid saponin aglycones, polyphenols and a phytosterol which were tested as
406 substrates (data not shown; see Experimental section for details).

407 In general, saponins have one or more linear or branched sugar chains attached to the
408 aglycone via a glucosidic ester or ether linkage. The two kinds of products can be distinguished
409 by alkaline hydrolysis [40]. The products obtained from the recombinant enzyme, asiatic acid
410 and madecassic acid monoglucosides, were found to be unstable to treatment with potassium
411 hydroxide 10% (w/v) at 80 ºC for 2 h (data not shown). This procedure cleaves the ester but not
412 the acetal linkages in glucosylated products, indicating that the enzyme is catalyzing an ester
413 formation at the C-28 carboxyl group. To confirm this, ^{1}H NMR spectra of both products were

obtained. By comparison to NMR studies of triterpene glucosides [41, 42], it was possible to assign resonances corresponding to Glc C1 and C6 in the 5.2 to 5.3 ppm and 3.7 to 3.8 ppm ranges, respectively, the first being a characteristic chemical shift signal of glucose ester linkage. Spectra showed characteristic resonances of ursolic acid derivatives - an oleofinic bond at C-12 and a methine at C-18. The chemical shift signals corresponding to C-2, C-3 and C-23 demonstrate that the hydroxyl groups linked to those carbons are not glucosylated (online Supplementary Table S2). Thus, the NMR (and alkaline hydrolysis) data are consistent with the glucosylation of asiatic acid and madecassic acid at the carboxyl group (C-28; see Fig. 1).

Sequence similarity searches indicated that UGT73AD1 protein sequence exhibits highest amino acid sequence identity with UGT from *Daucus carota* (73C6; 64%) (XP_017249880.1), *Panax ginseng* (59%) (AKA44600.1), *Juglans regia* (73C3; 57%) (XP_018840297.1), *Nicotiana attenuata* (73C1; 58%) (XP_019231530.1) and *N. sylvestris* (73C1; 58%) (XP_009769445.1). Inspection of the UGT73AD1 amino acid sequence indicates the presence of a subsequence closely matching (*e.g.,* 72% identity with UGT73D1) with the highly conserved Plant Secondary Product Glycosyltransferase (PSPG) motif [43, 44]. This motif includes amino acids that interact with the sugar donor. The sugar donor specificity of UGT73AD1 for UDP-glucose is consistent with the conserved tryptophan (W-375), glutamic acid (E-396) and glutamine (Q-397) residues in the PSPG motif.*3.3. Tissue-specific expression of UGT73AD1*

Five candidate reference genes were evaluated for use in gene expression analysis by RT-qPCR for *C. asiatica* (See online Supplementary Information). Of these, *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH)* was found to be the most constitutively expressed in the *C. asiatica* tissues tested. As shown in Fig. 5, *UGT73AD1* mRNA accumulated in all tissues tested,

437 but with different relative expression levels. If flower expression is set as the baseline, transcript
438 accumulation was determined to be highest in leaves of *C. asiatica* (38-fold), followed by stems
439 (12-fold) and roots (5.5-fold).
440
441 **4. Discussion**
442
443 As an important step in advancing our understanding of the triterpenoid saponin
444 biosynthesis in *C. asiatica*, a combination of partial enzyme purification, identification of a
445 cDNA candidate, and testing the activity of the corresponding enzyme has led to the elucidation
446 of the first glucosylation step in the biosynthesis of bioactive saponins. Chemical and NMR
447 analyses of UGT73AD1 products demonstrate that this enzyme catalyzes the first glycosylation
448 of asiatic and madecassic acid yielding a C-28 Glc ester (see Fig. 1). Previously, UGTs that
449 glucosylate sapogenins at the C-28 carboxylic group have been identified in only two species: *M.*
450 *truncatula* (UGT73F3) [16] and in *S. vaccaria* (UGT74M1) [17, 21].
451 From an evolutionary point of view, it is notable that of the three known ester-forming
452 UGT from plants, two are from the UGT73 family and one is from the UGT74 family. Although
453 these families are consecutive in number, in fact, they represent genes which diverged in
454 sequence relatively early in the evolution of land plants [45]. This suggests the possibility that
455 ester forming UGTs evolved independently from acetal-forming enzymes after the divergence of
456 the two UGT families.
457 *In vitro* assays showed that UGT73AD1 has a narrow range of substrate specificity, being
458 active only with asiatic acid and madecassic acid out of 18 substrates tested, including ursane,
459 oleanane and lupane triterpenoids, polyphenols and a phytosterol. With regard to sugar donors,
460 the enzyme was active with UDP-glucose and to a lesser extent with UDP-xylose.

461 The plant UGTs share a highly conserved motif (PSPG motif) which is an important
462 region for binding UDP-sugars [44]. The two last amino acids in the PSPG motif (Glu-381 and
463 Gln-382) are important in determining sugar specificity and interact with the glucose moiety
464 through forming several hydrogen bonds with O2', O3' and O4' atoms of glucose ring [46]. In
465 addition, the PSPG amino acid Trp360 (22nd amino acid of the PSPG motif) is highly conserved
466 among UGTs and forms a hydrogen bond with the glucose donor through its amide nitrogen
467 atom [47, 48]. In the UGT73AD1 PSPG motif region, the above cited amino acids are also
468 conserved, predicting UDP-Glc as the sugar donor. This is, of course, consistent with our
469 enzyme assays (see Figs. 2 and 4).
470 Results from RT-qPCR showed that all organs tested express the *UGT73AD1* gene, with
471 highest expression in leaves. These results are consistent with studies of *C. asiatica* plantlets in
472 which accumulation of active saponins (asiaticoside and madecassoside) was found to be
473 significantly higher in aerial parts compared to roots [49]. These results are also in good
474 agreement with a previous work [50] in which *C. asiatica* semi-quantitative PCR and RT-qPCR
475 was used to compare the expression profiles of select genes related to secondary metabolic
476 pathways in leaves, stem, seeds and roots tissue samples. All tested genes, i.e. squalene
477 synthases, squalene epoxidases and glucosyltransferases, exhibited highest expression in leaf
478 tissues.
479 In conclusion, the combination of biochemical experiments, transcriptomic and
480 proteomic techniques, as well as heterologous expression in *E. coli,* successfully allowed the
481 elucidation of a glucosyltransferase involved in the biosynthesis of bioactive triterpenoid saponin
482 in *C. asiatica.* In light of the medical attributes of the saponin family, future investigations

should aim at elucidating further steps of the pathway as well as exploring the potential of these saponin UGT's in synthetic applications [51].

Acknowledgements

We are grateful to members of the Genome Canada PhytoMetaSyn project and Simon J. Foote, from National Research Council of Canada (Ottawa, Ontario), for preparing local databases of *C. asiatica* EST library assemblies for MASCOT search compatibility and Dr. Grace Gosmann (Faculty of Pharmacy, UFRGS) for assistance with NMR data analyses. This work was supported by the National Research Council of Canada, CAPES (Coordination for Improvement of Higher Education Personnel, Brazil), PDSE process #13730/12-6, CNPq (National Council for Scientific and Technological Development, Brazil) and the Saskatchewan Agriculture Development Fund. NRCC publication #56094. We would like to thank Dr. Alex Haines (Flinders University, Australia) for naming the UDP-glucosyltransferase identified in this study.

References

[1] R. Azerad, Chemical structures, production and enzymatic transformations of sapogenins and saponins from *Centella asiatica* (L.) Urban, Fitoterapia, 114 (2016) 168-187.

[2] W. Cao, X.Q. Li, X.N. Zhang, Y. Hou, A.G. Zeng, Y.H. Xie, S.W. Wang, Madecassoside suppresses LPS-induced TNF-alpha production in cardiomyocytes through inhibition of ERK, p38, and NF-kappaB activity, Int. Immunopharmacol., 10 (2010) 723-729.

[3] H. Li, X. Gong, L. Zhang, Z. Zhang, F. Luo, Q. Zhou, J. Chen, J. Wan, Madecassoside attenuates inflammatory response on collagen-induced arthritis in DBA/1 mice, Phytomedicine, 16 (2009) 538-546.

[4] C.L. Xu, R. Qu, J. Zhang, L.F. Li, S.P. Ma, Neuroprotective effects of madecassoside in early stage of Parkinson's disease induced by MPTP in rats, Fitoterapia, 90 (2013) 112-118.

[5] J.H. Lee, H.L. Kim, M.H. Lee, K.E. You, B.J. Kwon, H.J. Seo, J.C. Park, Asiaticoside enhances normal human skin cell migration, attachment and growth *in vitro* wound healing model, Phytomedicine, 19 (2012) 1223-1227.

[6] C.L. Xu, Q.Z. Wang, L.M. Sun, X.M. Li, J.M. Deng, L.F. Li, J. Zhang, R. Xu, S.P. Ma, Asiaticoside: attenuation of neurotoxicity induced by MPTP in a rat model of Parkinsonism via maintaining redox balance and up-regulating the ratio of Bcl-2/Bax, Pharmacol. Biochem. Behav., 100 (2012) 413-418.

[7] D.E. Hall, M.M.S. Yuen, S. Jancsik, A.L. Quesada, H.K. Dullat, M. Li, H. Henderson, A. Arango-Velez, N.Y. Liao, R.T. Docking, S.K. Chan, J.E.K. Cooke, C. Breuil, S.J.M. Jones, C.I. Keeling, J. Bohlmann, Transcriptome resources and functional characterization of monoterpene synthases for two host species of the mountain pine beetle, lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*), BMC Plant Biol., 13 (2013) 1-14.

[8] P.J. Facchini, J. Bohlmann, P.S. Covello, V. De Luca, R. Mahadevan, J.E. Page, D.-K. Ro, C.W. Sensen, R. Storms, V.J.J. Martin, Synthetic biosystems for the production of high-value plant metabolites, Trends Biotechnol., 30 (2012) 127-131.

[9] A.C.A. Yendo, F. de Costa, C.T. Da Costa, L.C. Colling, G. Gosmann, A.G. Fett-Neto, Biosynthesis of Plant Triterpenoid Saponins: Genes, Enzymes and their Regulation, Mini Rev. Org. Chem., 11 (2014) 292-306.

[10] O.T. Kim, J.C. Ahn, S.J. Hwang, B. Hwang, Cloning and expression of a farnesyl diphosphate synthase in *Centella asiatica* (L.) Urban, Mol. Cells, 19 (2005) 294-299.

[11] O.-T. Kim, N.-S. Seong, M.-Y. Kim, B. Hwang, Isolation and characterization of squalene synthase cDNA from *Centella asiatica* (L) Urban, J. Plant Biol. (N. Y., NY, U. S.), 48 (2005) 263-269.

[12] O.T. Kim, M.Y. Kim, S.M. Huh, D.G. Bai, J.C. Ahn, B. Hwang, Cloning of a cDNA probably encoding oxidosqualene cyclase associated with asiaticoside biosynthesis from *Centella asiatica* (L.) Urban, Plant Cell Rep., 24 (2005) 304-311.

[13] O.T. Kim, J.W. Lee, K.H. Bang, Y. Chang Kim, D. Yun Hyun, S.W. Cha, Y.E. Choi, M. Lan Jin, B. Hwang, Characterization of a dammarenediol synthase in *Centella asiatica* (L.) Urban, Plant Physiol. Biochem., 47 (2009) 998-1002.

[14] A.C.A. Yendo, F. de Costa, G. Gosmann, A.G. Fett-Neto, Production of plant bioactive triterpenoid saponins: elicitation strategies and target genes to improve yields, Mol. Biotechnol., 46 (2010) 94-104.

[15] K. Haralampidis, M. Trojanowska, A.E. Osbourn, Biosynthesis of triterpenoid saponins in plants, Adv. Biochem. Eng. Biotechnol., 75 (2002) 31-49.

[16] M.A. Naoumkina, L.V. Modolo, D.V. Huhman, E. Urbanczyk-Wochniak, Y. Tang, L.W. Sumner, R.A. Dixon, Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*, Plant Cell, 22 (2010) 850-866.

[17] H. Seki, K. Tamura, T. Muranaka, P450s and UGTs: Key Players in the Structural Diversity of Triterpenoid Saponins, Plant Cell Physiol., 56 (2015) 1463-1471.

[18] V. Kren, L. Martinkova, Glycosides in medicine: "The role of glycosidic residue in biological activity", Curr. Med. Chem., 8 (2001) 1303-1328.

[19] F. de Costa, A.C. Yendo, J.D. Fleck, G. Gosmann, A.G. Fett-Neto, Immunoadjuvant and anti-inflammatory plant saponins: characteristics and biotechnological approaches towards sustainable production, Mini-Rev. Med. Chem., 11 (2011) 857-880.

[20] L. Achnine, D.V. Huhman, M.A. Farag, L.W. Sumner, J.W. Blount, R.A. Dixon, Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*, Plant J., 41 (2005) 875-887.

[21] D. Meesapyodsuk, J. Balsevich, D.W. Reed, P.S. Covello, Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding beta-amyrin synthase and a triterpene carboxylic acid glucosyltransferase, Plant Physiol., 143 (2007) 959-969.

[22] M. Shibuya, K. Nishimura, N. Yasuyama, Y. Ebizuka, Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*, FEBS Lett., 584 (2010) 2258-2264.

[23] T. Sayama, E. Ono, K. Takagi, Y. Takada, M. Horikawa, Y. Nakamoto, A. Hirose, H. Sasama, M. Ohashi, H. Hasegawa, T. Terakawa, A. Kikuchi, S. Kato, N. Tatsuzaki, C. Tsukamoto, M. Ishimoto, The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean, Plant Cell, 24 (2012) 2123-2138.

[24] J.M. Augustin, S. Drok, T. Shinoda, K. Sanmiya, J.K. Nielsen, B. Khakimov, C.E. Olsen, E.H. Hansen, V. Kuzina, C.T. Ekstrøm, T. Hauser, S. Bak, UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance, Plant Physiol., 160 (2012) 1881-1895.

[25] A. Owatworakit, B. Townsend, T. Louveau, H. Jenner, M. Rejzek, R.K. Hughes, G. Saalbach, X. Qi, S. Bakht, A.D. Roy, S.T. Mugford, R.J.M. Goss, R.A. Field, A. Osbourn, Glycosyltransferases from oat (*Avena*) implicated in the acylation of avenacins, J. Biol. Chem., 288 (2013) 3696-3704.

[26] P. Wang, Y. Wei, Y. Fan, Q. Liu, W. Wei, C. Yang, L. Zhang, G. Zhao, J. Yue, X. Yan, Z. Zhou, Production of bioactive ginsenosides Rh2 and Rg3 by metabolically engineered yeasts, Metab. Eng., 29 (2015) 97-105.

[27] C. Lu, S. Zhao, G. Wei, H. Zhao, Q. Qu, Functional regulation of ginsenoside biosynthesis by RNA interferences of a UDP-glycosyltransferase gene in *Panax ginseng* and *Panax quinquefolius*, Plant Physiol. Biochem., 111 (2017) 67-76.

[28] B.R. Madina, L.K. Sharma, P. Chaturvedi, R.S. Sangwan, R. Tuli, Purification and characterization of a novel glucosyltransferase specific to 27beta-hydroxy steroidal lactones from *Withania somnifera* and its role in stress responses, Biochim. Biophys. Acta, 1774 (2007) 1199-1207.

[29] C.J. Barber, P.T. Pujara, D.W. Reed, S. Chiwocha, H. Zhang, P.S. Covello, The two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme, J. Biol. Chem., 288 (2013) 12500-12510.

[30] O.T. Kim, K.H. Bang, Y.S. Shin, M.J. Lee, S.J. Jung, D.Y. Hyun, Y.C. Kim, N.S. Seong, S.W. Cha, B. Hwang, Enhanced production of asiaticoside from hairy root cultures of *Centella asiatica* (L.) Urban elicited by methyl jasmonate, Plant Cell Rep., 26 (2007) 1941-1949.

[31] M. Xiao, Y. Zhang, X. Chen, E.-J. Lee, C.J.S. Barber, R. Chakrabarty, I. Desgagné-Penix, T.M. Haslam, Y.-B. Kim, E. Liu, G. MacNevin, S. Masada-Atsumi, D.W. Reed, J.M. Stout, P. Zerbe, Y. Zhang, J. Bohlmann, P.S. Covello, V. De Luca, J.E. Page, D.-K. Ro, V.J.J. Martin, P.J. Facchini, C.W. Sensen, Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest, J. Biotechnol., 166 (2013) 122-134.

[32] I.S. Sheoran, D.J. Olson, A.R. Ross, V.K. Sawhney, Proteome analysis of embryo and endosperm from germinating tomato seeds, Proteomics, 5 (2005) 3752-3764.

[33] F.W. Studier, Protein production by auto-induction in high density shaking cultures, Protein Expression Purif., 41 (2005) 207-234.

[34] M. de Almeida, C. Ruedell, F. Ricachenevsky, R. Sperotto, G. Pasquali, A. Fett-Neto, Reference gene selection for quantitative reverse transcription-polymerase chain reaction normalization during *in vitro* adventitious rooting in *Eucalyptus globulus* Labill, BMC Mol. Biol., 11 (2010) 73.

[35] C. Ramakers, J.M. Ruijter, R.H. Deprez, A.F. Moorman, Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data, Neurosci. Lett., 339 (2003) 62-66.

[36] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol., 3 (2002) RESEARCH0034.

[37] C.L. Andersen, J.L. Jensen, T.F. Orntoft, Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets, Cancer Res., 64 (2004) 5245-5250.

[38] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, Methods (Amsterdam, Neth.), 25 (2001) 402-408.

[39] P.I. Mackenzie, I.S. Owens, B. Burchell, K.W. Bock, A. Bairoch, A. Belanger, S. Fournel-Gigleux, M. Green, D.W. Hum, T. Iyanagi, D. Lancet, P. Louisot, J. Magdalou, J.R. Chowdhury, J.K. Ritter, H. Schachter, T.R. Tephly, K.F. Tipton, D.W. Nebert, The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence, Pharmacogenetics, 7 (1997) 255-269.

[40] K.A. Hostettmann, A. Marston, Saponins. Chemistry and Pharmacology of Natural Products, Cambridge University Press, Cambridge, UK, 1995.

[41] C.S. Rumalla, Z. Ali, A.D. Weerasooriya, T.J. Smillie, I.A. Khan, A new triterpene glycoside from *Centella erecta*, Fitoterapia, 81 (2010) 751-754.

[42] Q.-L. Yu, H.-Q. Duan, Y. Takaishi, W.-Y. Gao, A novel triterpene from *Centella asiatica*, Molecules, 11 (2006) 661-665.

[43] A. Noguchi, M. Horikawa, Y. Fukui, M. Fukuchi-Mizutani, A. Iuchi-Okada, M. Ishiguro, Y. Kiso, T. Nakayama, E. Ono, Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in Lamiales, Plant Cell, 21 (2009) 1556-1572.

[44] S.A. Osmani, S. Bak, B.L. Møller, Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling, Phytochemistry, 70 (2009) 325-347.

[45] K. Yonekura-Sakakibara, K. Hanada, An evolutionary view of functional diversity in family 1 glycosyltransferases, Plant J., 66 (2011) 182-193.

[46] H. Shao, X. He, L. Achnine, J.W. Blount, R.A. Dixon, X. Wang, Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*, Plant Cell, 17 (2005) 3141-3154.

[47] X.Z. He, X. Wang, R.A. Dixon, Mutational analysis of the *Medicago* glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation, J. Biol. Chem., 281 (2006) 34441-34447.

[48] X. Wang, Structure, mechanism and engineering of plant natural product glycosyltransferases, FEBS Lett., 583 (2009) 3303-3309.

[49] S. Mangas, M. Bonfill, L. Osuna, E. Moyano, J. Tortoriello, R.M. Cusido, M.T. Pinol, J. Palazon, The effect of methyl jasmonate on triterpene and sterol metabolisms of *Centella asiatica*, *Ruscus aculeatus* and *Galphimia glauca* cultured plants, Phytochemistry, 67 (2006) 2041-2049.

[50] R.S. Sangwan, S. Tripathi, J. Singh, L.K. Narnoliya, N.S. Sangwan, *De novo* sequencing and assembly of *Centella asiatica* leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism, Gene, 525 (2013) 58-76.

[51] T. Moses, J. Pollier, J.M. Thevelein, A. Goossens, Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*, New Phytol., 200 (2013) 27-43.

[52] L. Gutierrez, M. Mauriat, S. Guenin, J. Pelloux, J.F. Lefebvre, R. Louvet, C. Rusterucci, T. Moritz, F. Guerineau, C. Bellini, O. Van Wuytswinkel, The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants, Plant Biotechnol. J., 6 (2008) 609-618.

[53] J. Hellemans, G. Mortier, A. De Paepe, F. Speleman, J. Vandesompele, qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data, Genome Biol., 8 (2007) R19.

Figure legends

Figure 1. Proposed biosynthetic pathway for *C. asiatica* saponins. Solid arrows indicate one-step reaction, dashed arrows indicate uncharacterized steps.

Figure 2. Madecassic acid glucoside produced from madecassic acid after 60 min assay of A) anion exchange chromatography fraction 15 (see text) and B) HIS-purified UGT73AD1. Insets: total ion chromatograms from LC-MS analysis were each extracted for two ions, a) substrate (madecassic acid dimer *m/z* 1008) and b) product (madecassic acid glucoside *m/z* 712) at i. time = 0, ii. time = 60 min. The assay with the fraction 15 contained approximately 3 µg of total protein (A), while the HIS-purified UGT73AD1 assay was performed with 100 ng protein (B). Assay A was performed with 1 mM acceptor substrate while assay B used an optimized protocol containing 250 µM of the acceptor substrate (section 2.9). Both assays were performed with 250 µM of sugar donor (UDP-Glc).

Figure 3. Purification of *E. coli* expressed recombinant *C. asiatica* UGT73AD1 by metal affinity chromatography. The amount of protein applied in the gel was 1.0 µg. The Precision Plus Protein™ M_r standards and the purified UGT73AD1 (arrow) were subjected to SDS-PAGE under denaturing conditions using 4-15% (w/v) polyacrylamide gel, and stained with fluorescent Oriole^{TM} stain. The mobility of M_r standards of 25,000; 50,000 and 75,000 is shown on the right.

Figure 4. Asiatic acid glucoside produced from asiatic acid after 60 min assay of HIS-purified UGT73AD1. Total ion chromatograms from LC-MS analysis were each extracted for two ions,

689 a) substrate (asiatic acid dimer *m/z* 976) and b) product (asiatic acid glucoside *m/z* 695) at i. time
690 = 0 and ii. time = 60 min.
691
692 **Figure 5.** Expression of *UGT73AD1* in different *C. asiatica* organs. The *GAPDH* gene was used
693 to as reference to normalize qPCR data for *UGT73AD1* for the organs shown. Expression is
694 relative to flowers. Vertical bars represent standard error of three biological replicates.
695
696 **Supplementary Material**
697
698 ***Determination of reference genes in C. asiatica***
699
700 Two different approaches, established by the GeNorm and NormFinder software, were
701 used to determine the expression stability of candidate reference genes in the plant parts
702 evaluated. The GeNorm algorithm calculates the average expression stability (*M* value), which
703 gives the average pairwise variation of expression of a particular gene in relation to all candidate
704 reference genes. Using an *M* value below the threshold of 1.5 was initially recommended [36],
705 although a maximum value of 0.5 has been proposed for more accurate results [52, 53]. Genes
706 with the lowest *M* values have the most stable expression. Our results indicated that expression
707 levels of the *Act* and *GAPDH* genes were the most stable (i.e., with the lowest *M* value = 0.439),
708 whereas *EFP2* was the least (*M* value = 0.946) (online supplementary Table S3).
709 NormFinder is another algorithm used to identify the optimal reference gene in a group of
710 candidate genes, taking into account intra- and intergroup variations for calculation, and genes
711 with lowest values are considered the most stable. As shown on supplementary Table S3, the

712 lowest value, obtained for *GAPDH*, indicated the most stably expressed gene (0.131), followed
713 by *Act* and *Ubiq* (0.137).

714 Both GeNorm and NormFinder identified essentially the same two most stably expressed
715 genes, with minor variations in the ranking of the other genes evaluated.

716

717

718 **Supplementary figure legends**

719

720 **Figure S1.** SDS-PAGE of anion exchange fractions which were active [15 (c) and 16 (d)] and
721 inactive [14 (b) and 17 (e)] in madecassic acid glucosyltransferase assays. Lane "a" represents
722 the protein standards. Rectangles indicate gel slices which were excised and subjected to
723 trypsinization and Q-TOF LC/MS. Asterisks indicate gel slices fractions (15-5 and 16-5) from
724 which UGT73AD1 was identified. The mobility of M_r standards of 25,000; 50,000 and 75,000 is
725 shown on the left.

726

727 **Figure S2.** The effect of increasing BSA concentration on enzyme activity. Recombinant
728 UGT73AD1 was assayed with asiatic acid at the BSA concentrations indicated. The means and
729 standard deviations (n = 3) of the rate of formation of the monoglucoside product for 1 h assays
730 are indicated.

731

732 **Figure S3.** Effect of pH on UGT73AD1 activity. Recombinant *C. asiatica* was assayed with
733 asiatic acid at the pH values and buffers indicated. The means and standard deviations (n = 3) of
734 the rate of formation of monoglucoside product for 1 h assays are indicated.

735

736 **Figure S4.** Asiatic acid monoglucoside produced from asiatic acid after a 60 min assay with
737 either A) UDP-Glucose; B) UDP-α-D-xylose; C) UDP-glucuronic acid or D) UDP-β-L-arabinose.
738 Enzyme assays were performed in triplicate at a final concentration of 250 µM for sugar donors
739 and 250 µM for substrate acceptor. Assay A produced asiatic acid + glucose (*m/z* 695) and assay
740 B produced asiatic acid + xylose (*m/z* 665). Assays C and D yielded negative results, with
741 absence of monoglucosylated product.

Supplementary Table S1. Primer sequences for RT-qPCR amplification and product sizes for each of the five reference gene candidates of *C. asiatica.*

Supplementary Table S2. Assignments of selected ^{1}H-NMR chemical shifts (given in parts per million) and coupling constant *J* (Hz) for the asiatic acid and madecassic acid monoglucoside products of UGT73AD1.

Supplementary Table S3. Ranking of candidate reference genes for *C. asiatica* in decreasing order of expression stability calculated by *geNorm* and *NormFinder.*

752 **Supplementary Table S1.** Primer sequences for RT-qPCR amplification and product sizes for
753 each of the five reference gene candidates of *C. asiatica.*
754

Abbreviation	**Gene name**	***Arabidopsis* homologue locus**	**Primer sequence (5'-3') forward/reverse**	**Amplicon length (bp)**
Act	Actin	NP_196543.1	TGCAGATCGTATGAGCAAGG/ GTGCACAATTGATGGACCAG	187
EFP1	Elongation factor protein	NP_849818.1	GGACAGGTGCTTTCTTGAGC/ CCAAACTTGGAGGCGTACAT	222
EFP2	Elongation factor protein	NP_175135.1	CTCCTGGTCACGTGGATTTT/ CTGACTCTGCACACCTCCAA	101
GAPDH	Glyceraldehyde 3-phosphate Dehydrogenase	NP_187062.1	GATCGTTGAGGGGCTTATGA/ CAGTGCTGCTCGGAATGATA	128
Tub	Tubulin	NP_197478.1	GCGAGACTGGTTCTGGAAAG/ GATTAGCTGTTCCGGATGGA	119
Ubiq	Ubiquitin	NP_849292.1	TTCACTTGGTGCTGAGGTTG/ TCCATCTTCCAATTGCTTCC	188

755
756

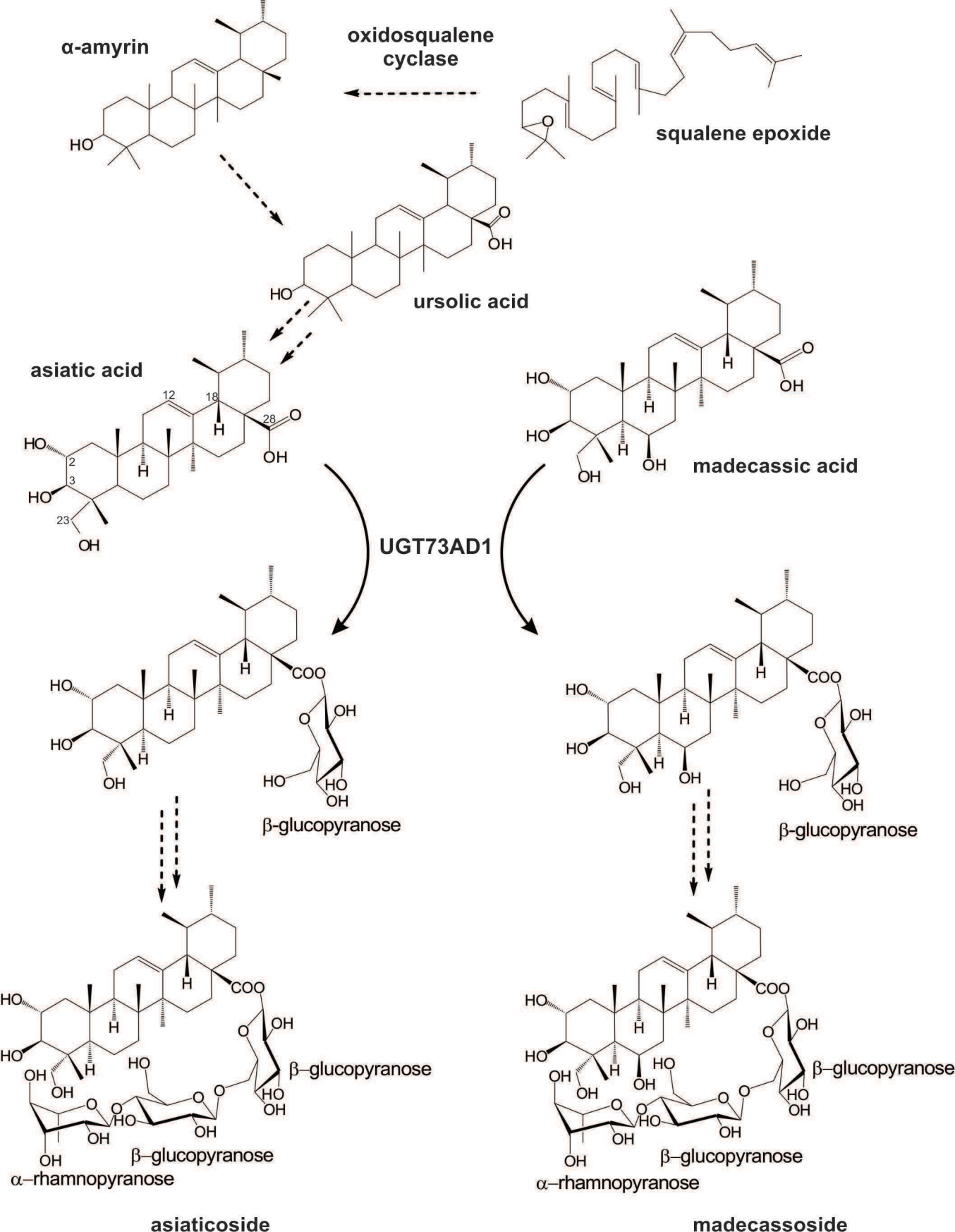
Supplementary Table S2. Assignments of selected ^{1}H-NMR chemical shifts (given in parts per million) and coupling constant *J* (Hz) for the asiatic acid and madecassic acid monoglucoside products of UGT73AD1.

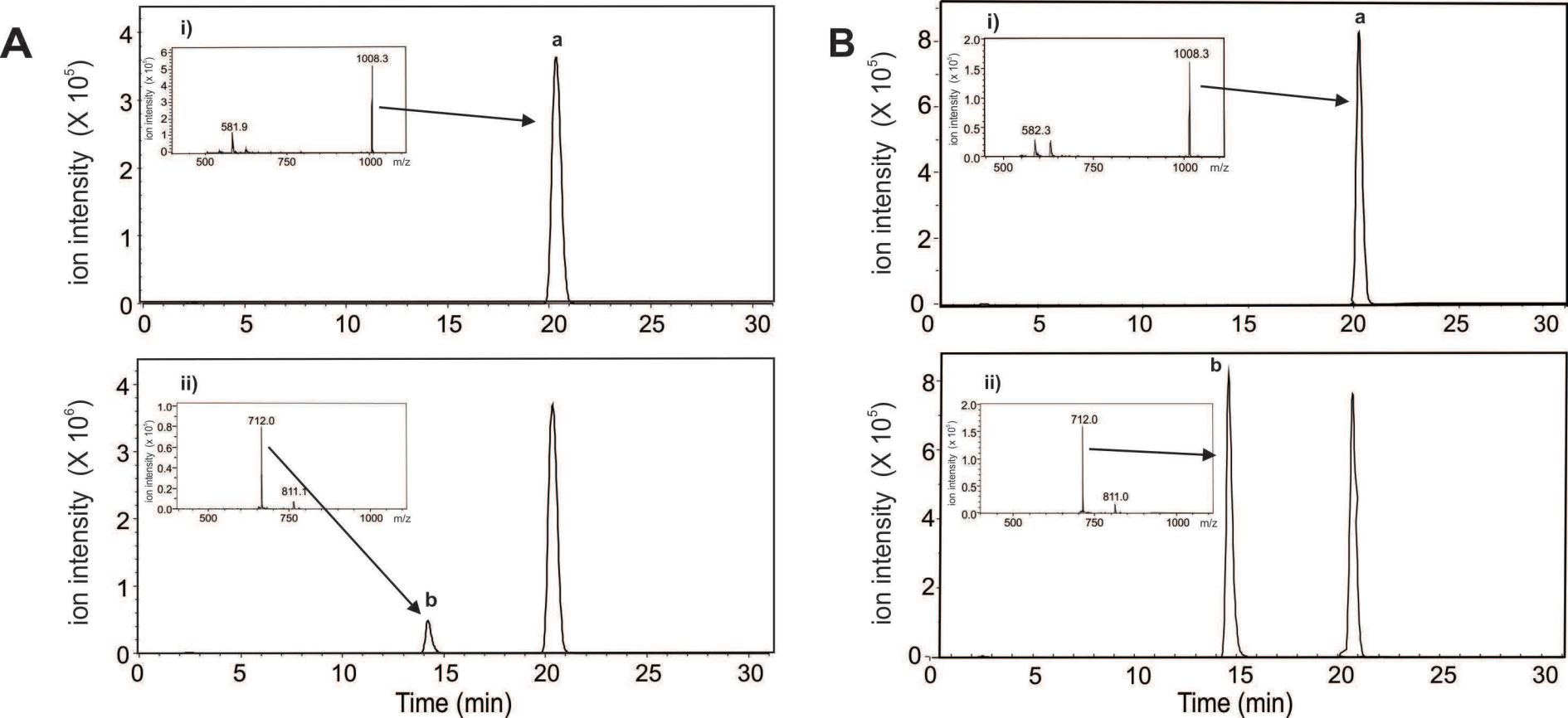
Atom No.	Asiatic Acid	Asiatic acid monoglucoside	Madecassic acid	Madecassic acid monoglucoside
2	3.68 (1H, *m*)	3.68 (1H, *m*)	3.72 (1H, *m*)	3.72 (1H, *m*)
3	3.34 (1H, *d*, 10 Hz)	3.37 (1H, *d*, 10 Hz)	3.27 (1H, *m*)	3.27 (1H, *m*)
12	5.22 (1H, *t*, 3.5 Hz)	5.24 (1H, *t*, 3.5 Hz)	5.26 (1H, *t*, 3.5 Hz)	5.29 (1H, *t*, 3.5 Hz)
18	2.19 (1H, *d*, 11 Hz)	2.20 (1H, *d*, 11 Hz)	2.20 (1H, *d*, 11 Hz)	2.23 (1H, *d*, 11 Hz)
23	3.25 (1H, *d*, 11 Hz) and 3.49 (1H, *d*, 11 Hz)	3.26 (1H, *d*, 11 Hz) and 3.49 (1H, *d*, 11 Hz)	3.43 (1H, *d*, 11 Hz) and 3.57 (1H, *d*, 11 Hz)	3.43 (1H, *d*, 11 Hz) and 3.55 (1H, *d*, 11 Hz)
Glc-1	-	5.32 (1H, *d*, 8 Hz)	-	5.33 (1H, *d*, 8.5 Hz)
Glc-6	-	3.77 (1H, *d*, 2.5 Hz) and 3.80 (1H, *d*, 2.5 Hz)	-	3.78 (1H, *d*, 2 Hz) and 3.80 (1H, *d*, 2.5 Hz)

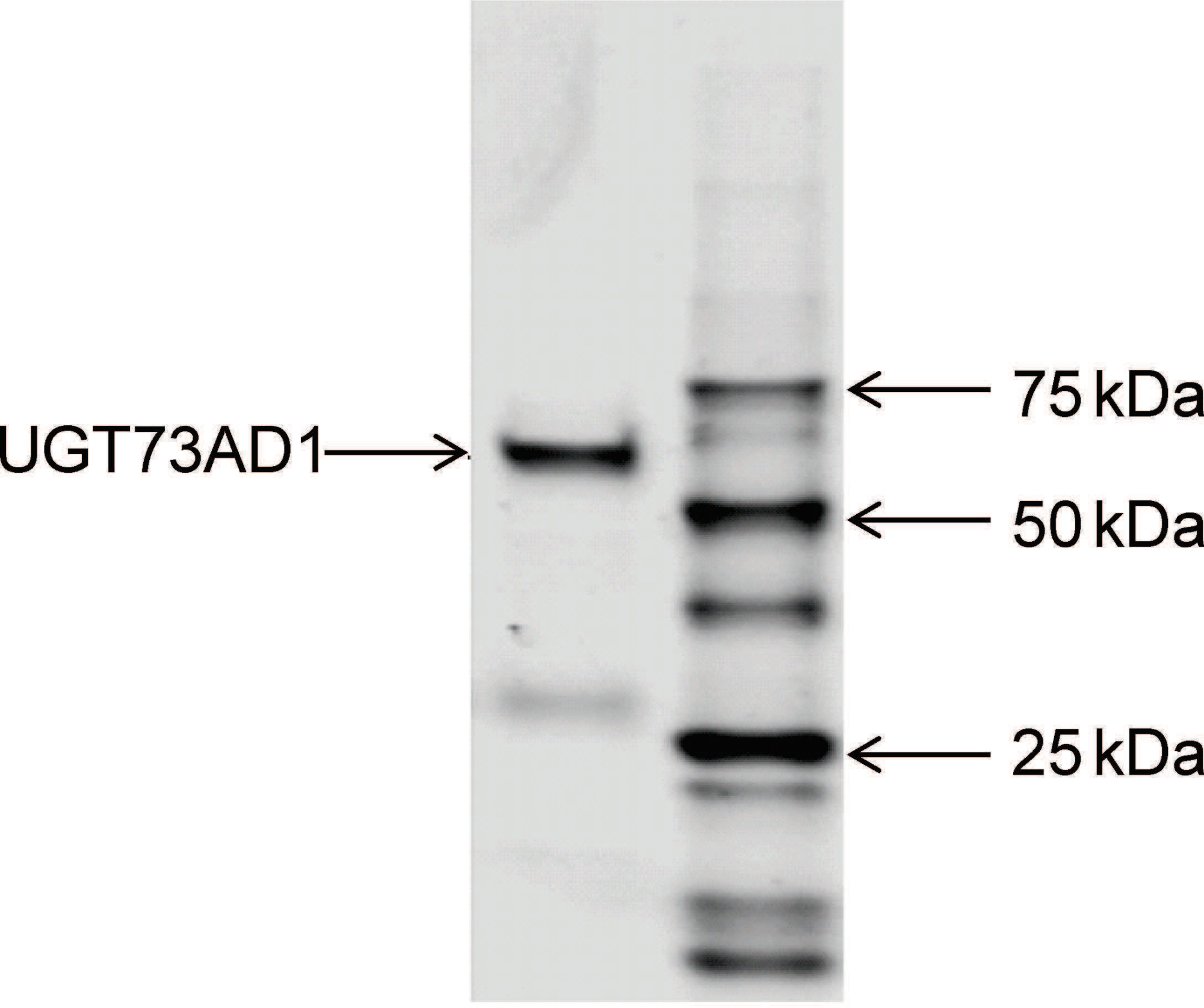
765 **Supplementary Table S3.** Ranking of candidate reference genes for *C. asiatica* in decreasing
766 order of expression stability calculated by *geNorm* and *NormFinder*
767

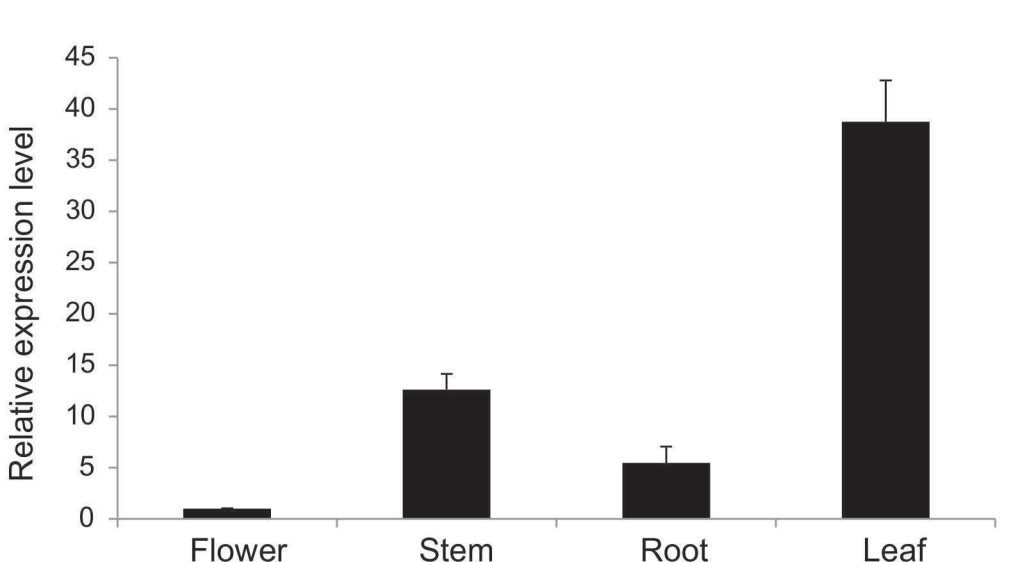
Ranking order	*geNorm* (*M* value)	*NormFinder* (stability value)
1	*Act/GAPDH* 0.439	*GAPDH* (0.131 ± 0.017)
2		*Act* (0.137 ± 0.022)
3	*Ubiq* 0.666	*Ubiq* (0.137 ± 0.029)
4	*Tubu* 0.718	*EFP1* (0.157 ± 0.014)
5	*EFP1* 0.859	*Tubu* (0.164 ± 0.014)
6	*EFP2* 0.946	*EFP2* (0.184 ± 0.038)

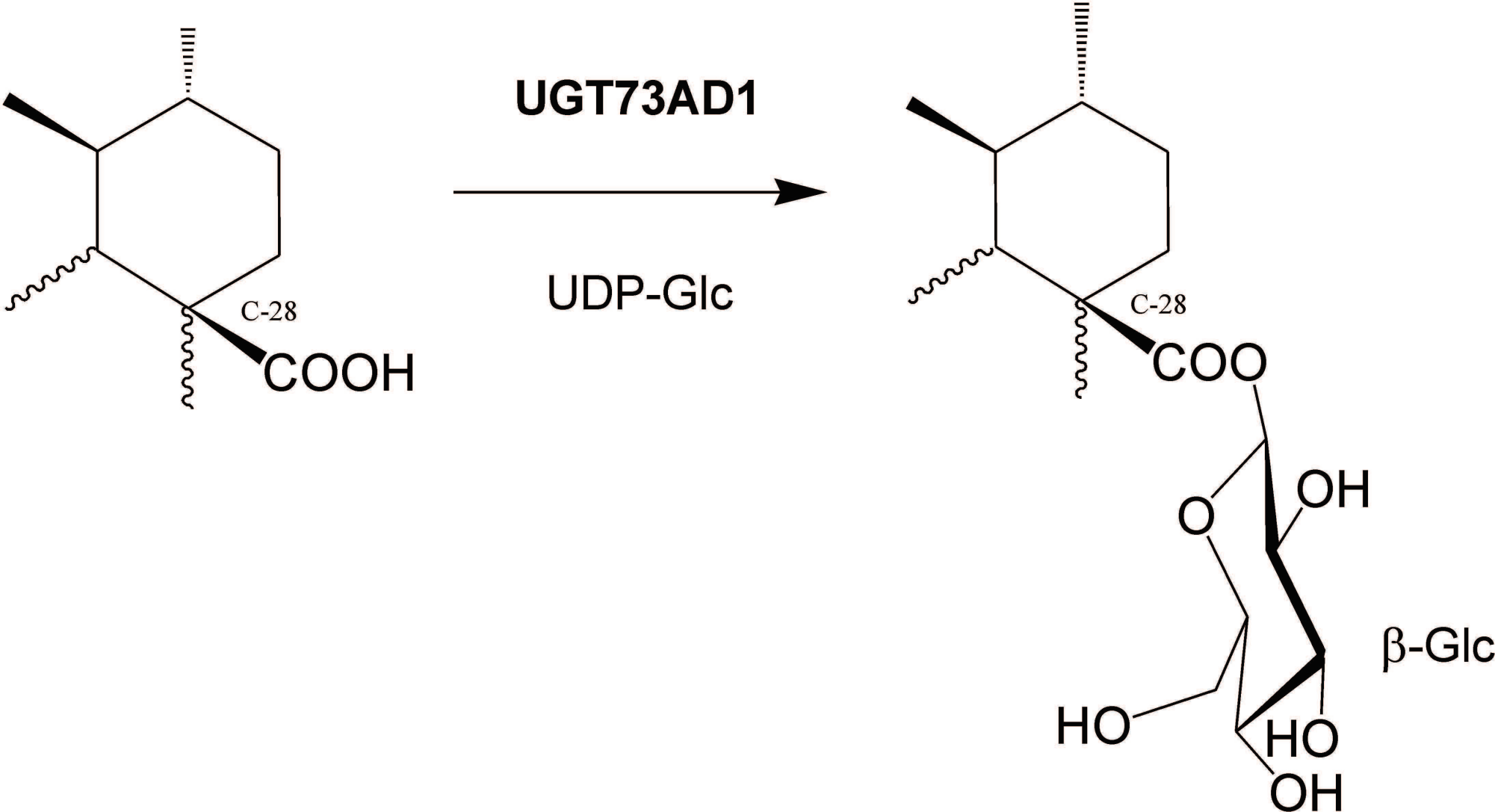
768
769
770
771

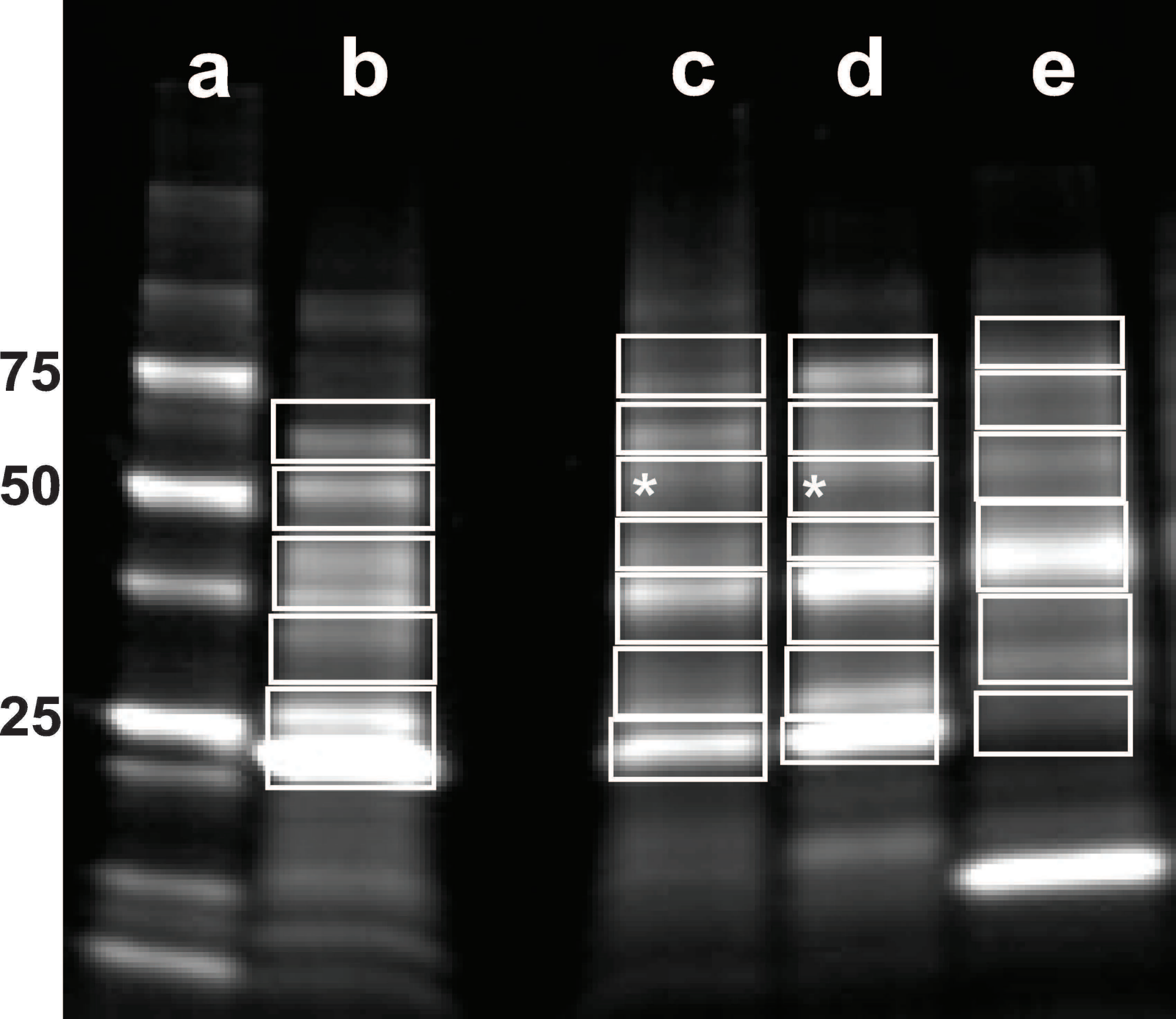

α-amyrin
oxidosqualene cyclase
squalene epoxide
ursolic acid
asiatic acid
madecassic acid
UGT73AD1
β-glucopyranose
β-glucopyranose
β–glucopyranose
β–glucopyranose
α–rhamnopyranose
asiaticoside
β–glucopyranose
β–glucopyranose
α–rhamnopyranose
madecassoside

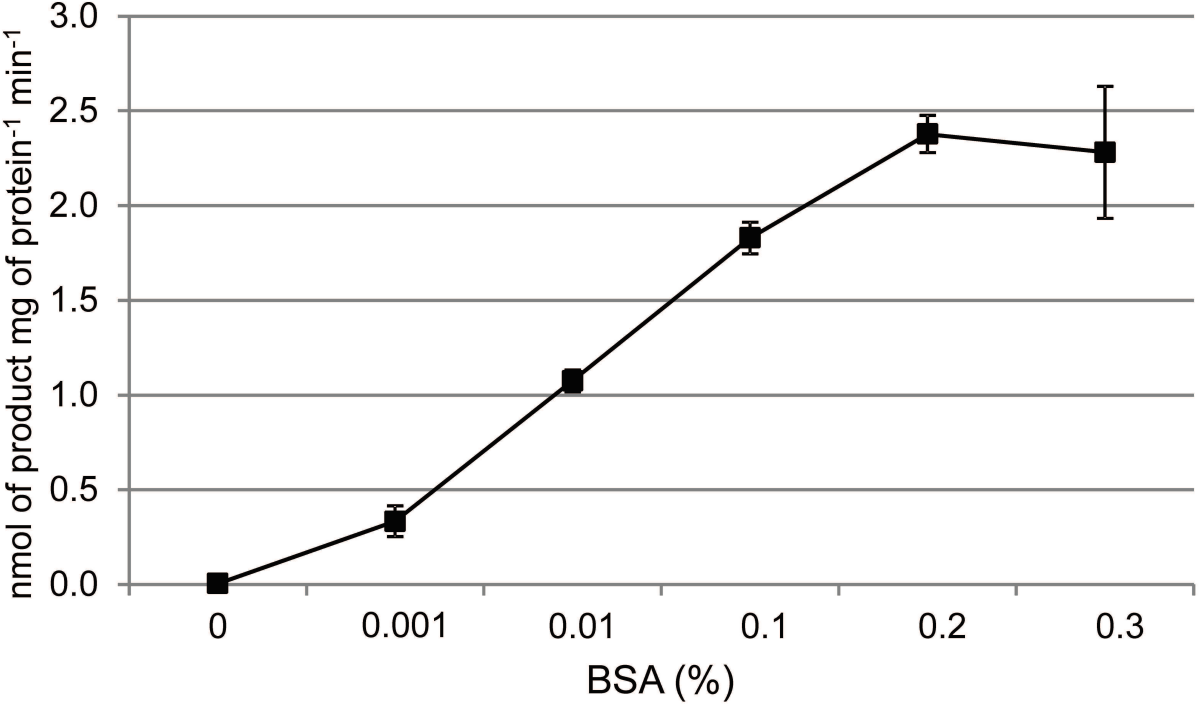

A
B
i)
ii)
a
b
1008.3
581.9
582.3
712.0
811.1
811.0
ion intensity (X 10^5)
ion intensity (X 10^6)
ion intensity (x 10^5)
m/z
Time (min)

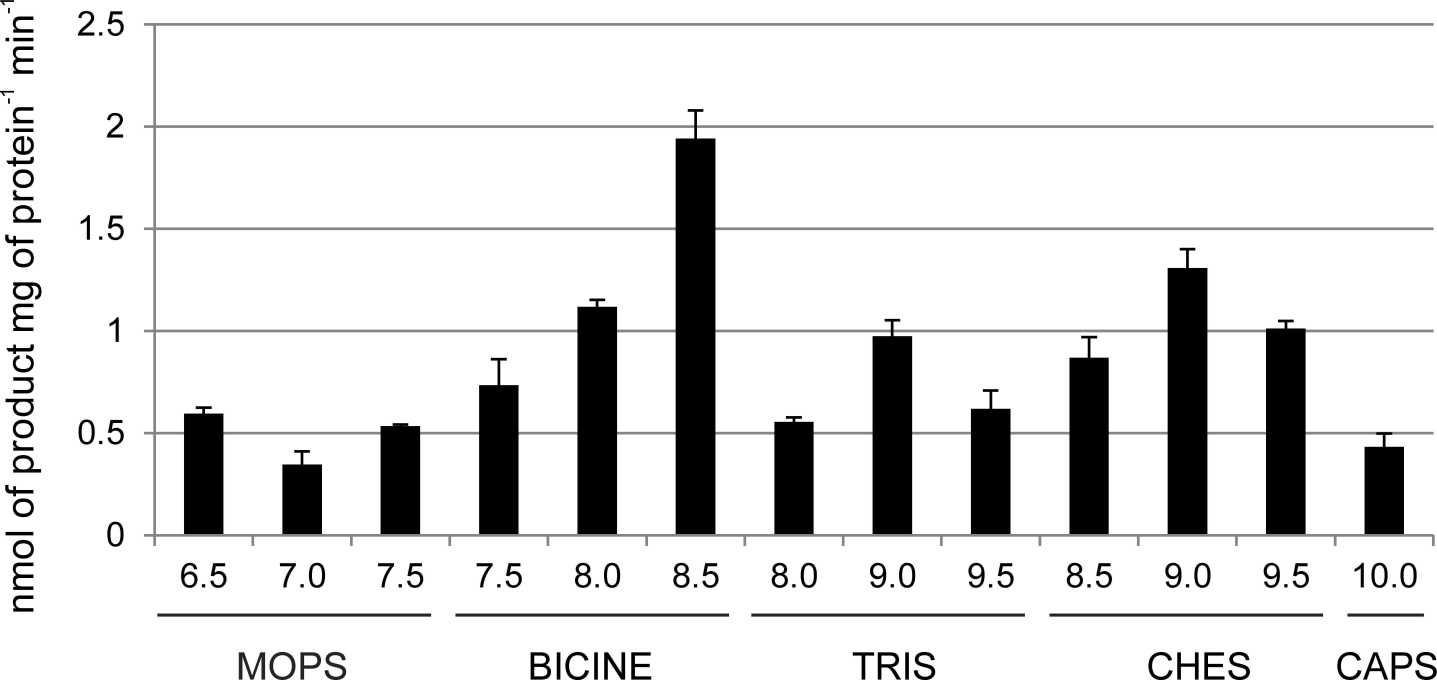

UGT73AD1
75 kDa
50 kDa
25 kDa

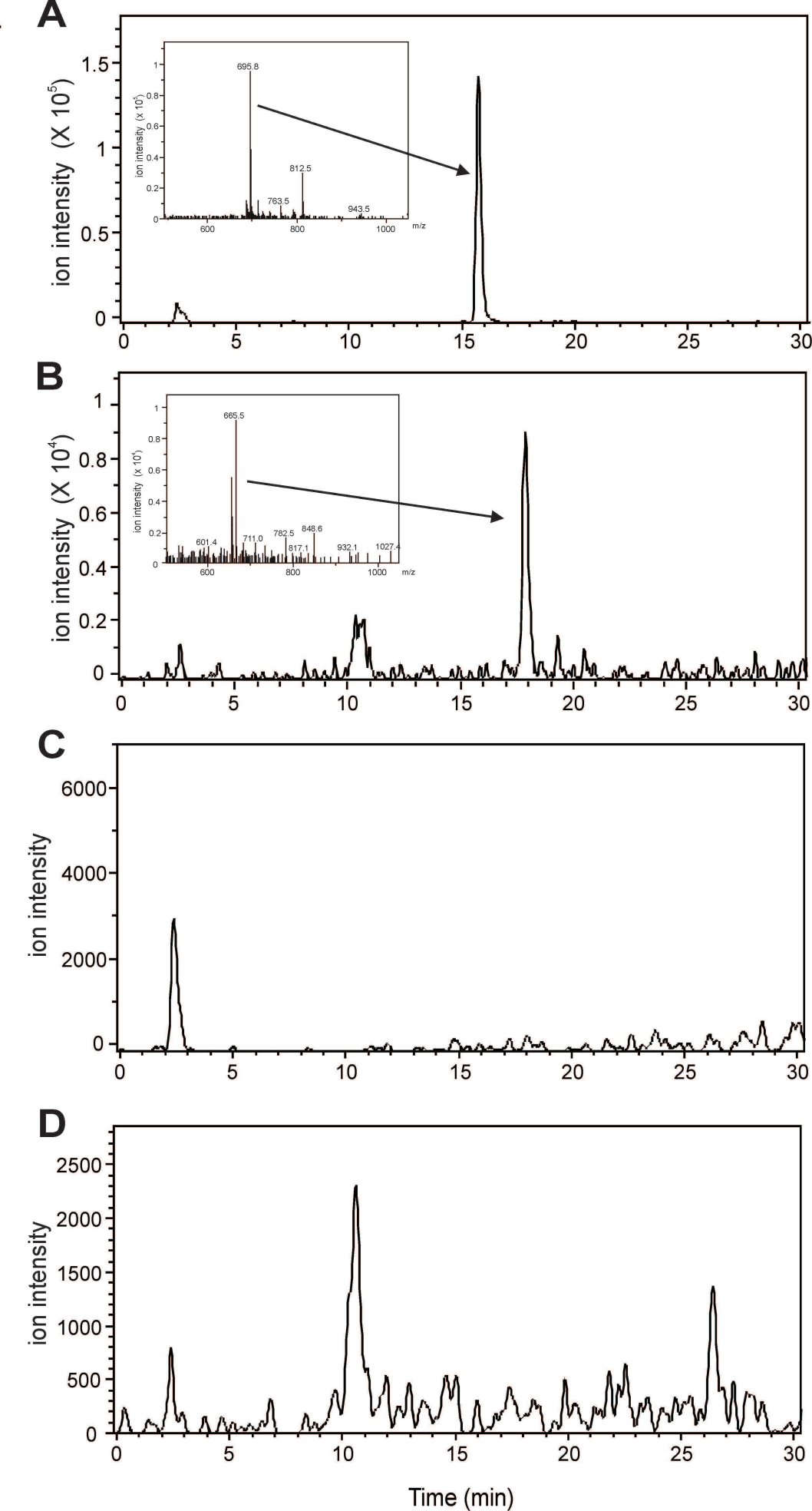

Relative expression level
45
40
35
30
25
20
15
10
5
0
Flower
Stem
Root
Leaf


UGT73AD1
UDP-Glc
C-28
COOH
C-28
COO
O
OH
β-Glc
HO
HO
OH


a
b
c
d
e
75
50
25
*
*


3.0
2.5
2.0
1.5
1.0
0.5
0.0
nmol of product mg of protein-1 min-1
0
0.001
0.01
0.1
0.2
0.3
BSA (%)


nmol of product mg of protein-1 min-1
2.5
2
1.5
1
0.5
0
6.5
7.0
7.5
7.5
8.0
8.5
8.0
9.0
9.5
8.5
9.0
9.5
10.0
MOPS
BICINE
TRIS
CHES
CAPS


A
ion intensity (X 10^5)
695.8
812.5
763.5
943.5
ion intensity (x 10^5)
m/z
B
ion intensity (X 10^4)
665.5
601.4
711.0
782.5
848.6
817.1
932.1
1027.4
ion intensity (x 10^4)
m/z
C
ion intensity
D
ion intensity
Time (min)