Characterization of steroidal saponin and glycoalkaloid biosynthesis in plants

Nakayasu, Masaru

(Degree)
博士（農学）

(Date of Degree)
2016-03-25

(Date of Publication)
2018-03-25

(Resource Type)
doctoral thesis

(Report Number)
甲第6660号

(URL)
https://hdl.handle.net/20.500.14094/D1006660

Doctoral Dissertation

Characterization of steroidal saponin and glycoalkaloid biosynthesis in plants

植物におけるステロイドサポニンおよびグリコアルカロイド生合成の解析

Masaru NAKAYASU

January, 2016

Graduate School of Agricultural Science

Kobe University

Contents

General introduction

Saponins are natural surfactants with various biological activities such as hemolytic, cytotoxic, anti-inflammatory, antifungal and antibacterial properties (Sparg et al. 2004), and widely contributed in diverse plants. They are structurally diverse and classified as triterpenoid saponins, steroidal saponins, or steroidal glycoalkaloids (SGAs). The structures of them are consisting of C_{30} triterpenoid, C_{27} steroid, or C_{27} steroidal alkaloid, containing a nitrogen atom in the C_{27} steroid, respectively, bound to oligosaccharide moiety.

The steroidal saponins are widely found in members of the families *Scrophulariaceae*, *Solanaceae*, *Fabaceae*, *Simaroubaceae* and *Zygophyllaceae* in dicots, and *Agavaceae*, *Alliaceae*, *Asparagaceae*, *Liliaceae*, *Amaryllidaceae*, *Bromeliaceae*, *Palmae*, *Poaceae* and *Dioscoreaceae* in monocots (Hoffmann, 2003; Hostettmann & Marston, 2005; Moses et al., 2014).

The edible tubers of *Dioscorea* spp. have been reported to contain high amounts of steroidal saponins (sometimes with a yield >2%) as the functional compounds. In particular, the tubers are known to contain furostanol and spirostanol glycosides such as protodioscin and dioscin (**Figure A**), respectively, which are valuable saponins used for semi-synthetic production of pharmaceutical steroidal drugs such as anti-inflammatory, androgenic, estrogenic, and contraceptive drugs.

On the other hand, SGAs are typically distributed in *Solanaceae* plants (Harrison, 1990; Helmut, 1998; Petersen et al., 1993). In particular, potato (*Solanum tuberosum*) and tomato (*S. lycopersicum*) are known to contain α-solanine and α-chaconine, and α-tomatine (**Figure A**), respectively, which are toxic compounds against fungi, bacteria, insects, animals and humans (Friedman, 2002, 2006). The mechanisms of toxicity include disruption of membranes and inhibition of acetylcholine esterase activity (Roddick, 1989).

α-solanine and α-chaconine are present in most tissues of potato plants and especially contained the greatest levels of the SGAs in floral and tuber sprouts tissues (Kozukue and Mizuno 1985, 1989; Ginzberg et al. 2009). Additionally, these SGAs in the potato tuber are induced by exposure to light, low temperature and mechanical injury (Valkonen et al. 1996). The potato SGAs are distasteful described as bitter, burning, scratchy or acrid tasting (Friedman 2006; Ginzberg et al. 2009; Taylor et al. 2007), and high concentrations of them are harmful to human. Similarly, α-tomatine exists in all plant tissues and richly accumulated in leaves and immature fruits. On the other hand, esculeoside A is stored in the ripe fruit of tomato and its content increases during fruit ripening in contrast to a decrease in α-tomatine (Iijima et al., 2013).

Figure A shows the structures of SGAs in potato and tomato, and steroidal saponins in *Dioscorea* spp. α-solanine and α-chaconine in potato are composed of solatriose and chacotriose, respectively, as oligosaccharide attached to the C-3 hydroxy group of solanidine as aglycone. Similarly, α-tomatine in tomato consists of lycotetraose linked to the C-3 hydroxy group of tomatidine as aglycone. On the other hand, dioscin in yam, the member of the steroidal saponins, is composed of chacotriose bound to the C-3 hydroxy group of diosgenin. Solanidine, tomatidine and diosgenin as aglycones are all six-membered ring heterocycles of C_{27} steroid, although the binding mode of C-5-C-6 and the structures of E- and F-rings are different. In this way, there are some structural similarities between SGAs and steroidal saponins.

The SGAs in potato and tomato are hypothesized to be biosynthesized from cholesterol via the oxidation at C-16, C-22 and C-26 position, the amination at C-26 position, and the glycosylation at the C-3 hydroxy group (Friedman, 2002; Ginzberg et al., 2009; Petersen et al., 1993; Ohyama et al., 2013) (**Figure B**). Similarly, steroidal saponins in *Dioscorea* spp. are biosynthesized from cholesterol via the sequential steps of the oxidation and the glycosylation, except for the amination (Joly et al., 1969a, b; Varma et al., 1969) (**Figure C**). In both steroidal saponins and SGAs biosynthesis in plants, cytochrome P450 monooxygenases (CYPs) are

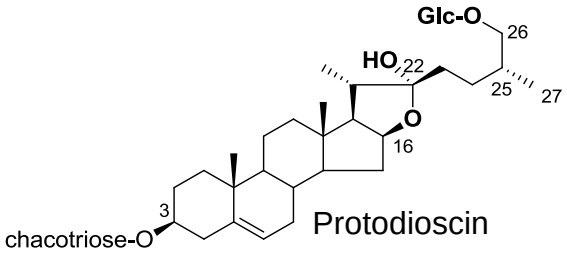
likely involved in the oxygenations at the C-16, C-22 and C-26 positions, and UDP-dependent glycosyltransferases (UGTs) also function in the glycosylation at the C-3, the addition at C-26 in so far as furostanol saponins biosynthesis (**Figure B and C**). However, little is known about enzymes and genes for these biosynthesis, except for some UGTs in the SGAs biosynthesis of potato and tomato (Moehs CP et al., 1997; McCue KF et al., 2005, 2006, 2007; Itkins et al., 2011).

In this study, to elucidate steroidal saponins and SGAs biosynthesis in plants, the biosynthetic genes were identified and the enzymes encoded by these genes were characterized. In chapter 1, comparative transcriptome analysis was performed in order to investigate steroidal saponins biosynthesis in *Dioscorea* spp. by using the tubers of three yam as plant materials, and furostanol glycoside 26-*O*-β-glucosidase from *D. esculenta*, named DeF26G1, which hydrolyzed protodioscin to form dioscin, was identified. In chapter 2, it was revealed that two CYPs, named PGA1 and PGA2, catalyzed the hydroxylation of cholesterol at C-26 and C-22 positon, respectively, and they were involved in SGAs biosynthesis in potato and tomato. Similarly, in chapter 3, 2-oxoglutarate dependent dioxygenase (2OGD), named 16DOX, was identified as the enzyme catalyzing steroid C-16 hydroxylation for SGAs biosynthesis.

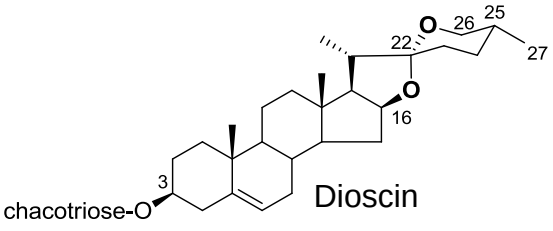
Potato is the world's major food crop and the producers and consumers have called for the removal of SGAs from potato. In previous study, the transgenic potato plants overexpressing a heterologous sterol methyltransferase gene had relatively lower amounts of SGAs (Arnqvist et al. 2003). However, changing the gene expression of the sterol methyltransferase or some glycosyltransferases responsible for potato SGAs biosynthesis (Moehs et al., 1997; McCue et al., 2005, 2006, 2007), does not effectively decrease SGA levels. In Chapter 2 and 3, gene silencing of *PGA1*, *PGA2* and *16DOX* resulted in a significant reduction in SGAs in the transgenic potato plants. Then, the introduction of several genes for dioscin biosynthesis into the transgenic potato plants likely allow the potato plants to accumulate not SGAs but dioscin. Furthermore, the elucidation of steroidal saponins and SGAs biosynthesis may make possible

the metabolic engineering in plants such as the construction of the transgenic plants containing little toxic SGAs and accumulating high amounts of beneficial steroidal saponins for human health.

Dioscorea spp. Steroidal saponins

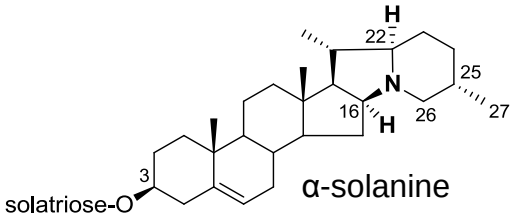


Protodioscin

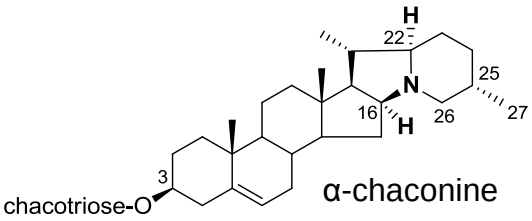


Dioscin

Potato SGAs

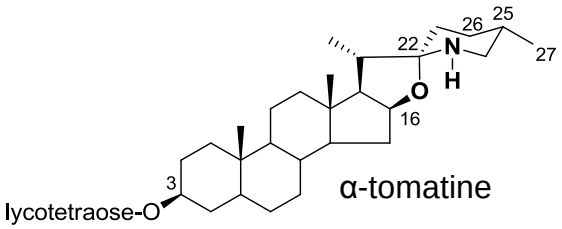


α-solanine



α-chaconine

Tomato SGA



α-tomatine

Oligosaccarides

chacotriose-O- = L-Rhamnose (α1-4) — D-Glucose-O- (β1); D-Glucose-O- (β1) | (α1-2) L-Rhamnose

solatriose-O- = D-Glucose (β1-3) — D-Galactose-O- (β1); D-Galactose-O- (β1) | (α1-2) L-Rhamnose

lycotetraose-O- = Xylose (β1-3) — D-Glucose (β1-4) — D-Galactose-O- (β1); D-Galactose-O- (β1) | (β1-2) D-Glucose

Figure A. The structures of steroidal saponins in *Dioscorea* spp. and steroidal glycoalkaloids (SGAs) in potato and tomato

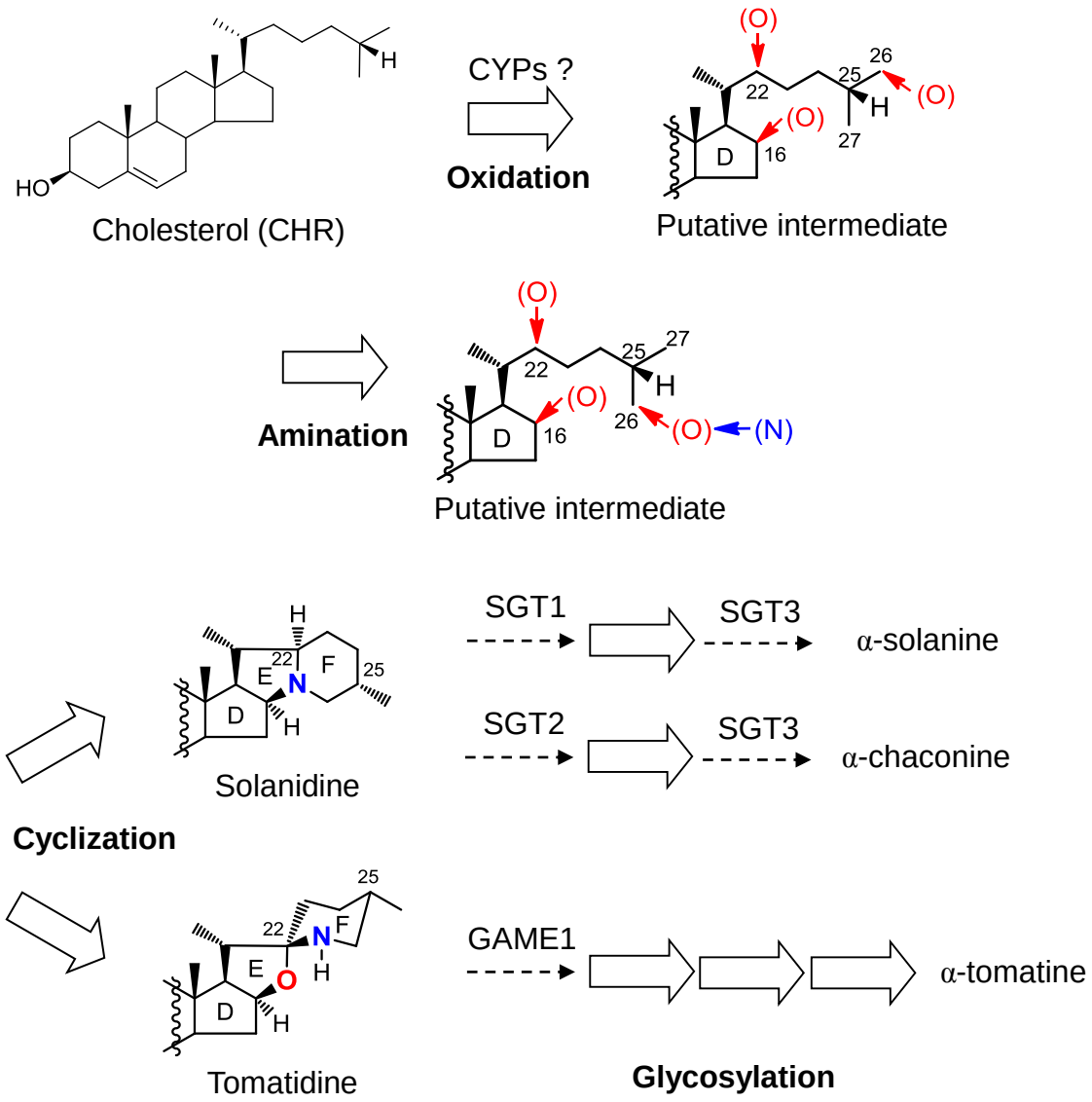


Figure B. The putative biosynthetic pathway of SGAs in potato and tomato. Thick arrows indicate unidentified reaction stages. Dashed arrows indicate the identified reaction stages in previous study.

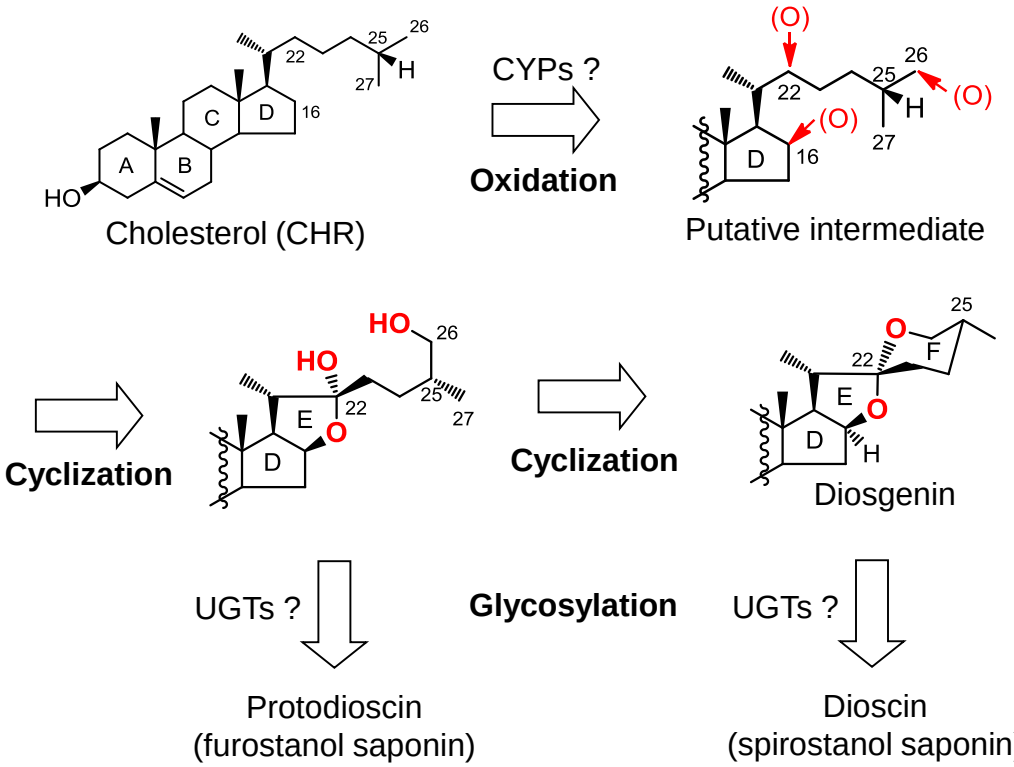


Figure C. The putative biosynthetic pathway of protodioscin and dioscin in *Dioscorea* spp. Thick arrows indicate unidentified reaction stages.

Chapter 1

Identification of furostanol glycoside 26-*O*-β-glucosidase involved in steroidal saponin biosynthesis from *Dioscorea esculenta*

Introduction

As described in General introduction, the tubers of *Dioscorea* spp. are known to contain furostanol and spirostanol glycosides such as protodioscin and dioscin (Figure 1-1), respectively, and these steroidal saponins are hypothesized to be biosynthesized from cholesterol via the sequential modification. However, little is known about enzymes and genes for steroidal saponins biosynthesis in *Dioscorea* spp.

In addition, a β-glucosidase is involved in the conversion of protodioscin to dioscin. Protodioscin, which contains a glucose unit at the C-26 hydroxy group of furostanol, is a precursor of dioscin, and a β-glucosidase, furostanol glycoside 26-*O*-β-glucosidase (F26G), cleaves the glucose unit to form the E/F spiro-ring of dioscin (Figure 1-1).

Purification and cloning of F26G have been reported from several plant species. *Costus speciosus* F26G, which hydrolyzes protogracillin to form gracillin, was purified (Inoue et al. 1996a; Inoue and Ebizuka 1996), and the corresponding *CsF26G* cDNA was isolated (Inoue et al. 1996b). *Avena sativa* avenacosidase, which hydrolyzes avenacosides to 26-degluco-avenacosides, was purified (Grünweller and Kesselmeier 1985; Nisius 1988), and isolation of the cDNA encoding avenacosidase was reported (Gus-Mayer et al. 1994). These monocotyledonous glucosidases belong to glycosyl hydrolase family 1. In contrast, F26G from *Solanum torvum*, which converts furostanol glycosides torvosides to 26-degluco-torvosides, belongs to glycosyl hydrolase family 3 (Arthan et al. 2006). In the case of *Dioscorea* spp., the F26G activity was reported in *D. pseudojaponica* Yamamoto (Yang et al. 2009), and

histochemical analysis of the F26G activity in *D. caucasica* revealed localization of F26G in the thylakoid membrane fraction of the chloroplast (Gurielidze et al. 2004). However, isolation and characterization of the *F26G* gene in *Dioscoreaceae* has so far not been reported.

In this chapter, comparative transcriptome analysis was performed by using the tubers of high saponin producers, *D. esculenta* and *D. cayenensis*, and a low producer, *D. alata*, as plant materials to investigate steroidal saponins biosynthesis in *Dioscorea* spp. and furostanol glycoside 26-*O*-β-glucosidase from *D. esculenta*, named DeF26G1, which hydrolyzes protodioscin to form dioscin, was identified.

Materials and Methods

Plant materials

Dioscorea species used in this chapter (*D. esculenta*, *D. alata* and *D. cayenensis*) were cultivated in the Experiment Station for Agricultural Science of Kagoshima University. Each part of the cultivated plants was separated, frozen in liquid nitrogen and stored at -80°C.

Chemicals

Protodioscin and dioscin were purchased from Funakoshi Co. (Tokyo, Japan) and *p*-nitrophenyl-β-D-glucopyranoside was purchased from NACALAI TESQUE, INC. (Kyoto, Japan).

Extraction of steroidal saponins in plants

Each frozen sample was powdered with a mortar and pestle in the presence of liquid nitrogen. Powdered samples (1 g) were extracted 5 ml of 70% (v/v) aqueous ethanol. After boiled for 1 hour at 80°C, the supernatants were then separated following centrifugation at 5000 rpm for 10

min. The supernatant was diluted by 50 times with 70% (v/v) aqueous ethanol and filtered through a 0.22 μm PVDF membrane (Millipore). The filtrate was analyzed by UPLC-ESI-MS.

Crude enzyme in leaves extraction

Frozen sample of leaves was homogenized with a mortar and pestle in 10 mL of ice-cold extraction buffer of crude enzyme (20 mM sodium citrate buffer (pH 6.0), 30 mM 2-mercaptoethanol, 5 mM EDTA, 5% (w/v) PVPP and 0.5 mM PMSF) per g fresh weight material. After centrifugation at 12,000 *g* for 30 min at 4°C, the supernatant was obtained as crude enzyme.

Enzyme assay

The reaction solution (135 μl) of containing 50 mM sodium citrate buffer (pH 7.0) and 100 μM protodioscin was pre-incubated for 5 min at 37°C, and then 15 μl of crude enzyme from leaves or recombinant proteins was added. The reaction was carried out at 37°C for 60 min. The reaction was stopped after by the addition to 150 μl of n-BuOH saturated with water including 10 μg genistin per mL as an internal standard. The reaction mixture was vortexed for 3 min and centrifuged at 15,000 rpm for 1 min at room temperature, followed by collection of the upper layer. The above extraction was performed by three times. The obtained n-BuOH layer was filtered through a 0.22 μm PTFE membrane (Millipore) and analyzed by UPLC-ESI-MS (Waters). β-glucosidase activity for *p*-nitrophenyl-β-D-glucopyranoside was also measured. The reaction solution (150 μL) containing 50 mM potassium phosphate buffer (pH 6.0) and 5 mM *p*-nitrophenyl-β-D-glucopyranoside was pre-incubated for 5 min at 37°C, and then 50 μl of crude enzyme from leaves of *D. esculenta* or recombinant DeF26G1 was added. The reaction was carried out at 37°C for 15 min. The reaction was stopped after by the addition of 200 μl of 1M Na_2CO_3. The activity was determined spectrophotometrically with liberated *p*-nitrophenol at 405 nm.

RNA extraction

Total RNA of tuber flesh and tuber cortex of *D.esculenta, D. alata* and *D. cayenensis* was prepared using the RNAeasy plant mini-kit (QIAGEN, Hilden, Germany), RNase-Free DNase Set (QIAGEN), and Plant RNA Isolation Reagent (Invitrogen). Each frozen sample was powdered in a mortar with a pestle in the presence of liquid nitrogen. Powdered samples (50 mg) were added to 3 ml of Plant RNA isolation reagent (Invitrogen) and 150 µl of 20% (w/v) PEG. The mixture was incubated at room temperature for 5 min and centrifuged at 15,000 rpm for 10 min at room temperature. The supernatant was added to half volume of 100% EtOH. The mixture was purified by Pink column of RNeasy Plant Mini Kit (QIAGEN) and treated by RNase-Free DNase Set (QIAGEN) according to the instructions from the manufacturer. Total RNA was eluted to microtube with RNase free water. Total RNA of the leaves was extracted from the powdered sample by using TRIzol Reagent (Invitrogen). The obtained RNA was then purified by lithium chloride, followed by sodium acetate.

RT-PCR and TA cloning

Based on *DeF26G1* partial sequence of EST results, the following two primers were designed; DeF26G1_Fw: 5'-GATCACTTTGGTTTGTATGTGGTA-3', DeF26G1_Rv: 5'-AATTTAGAGAACCATTTTGCAGAC-3'. 100 µg of total RNA was used to synthesize the first strand cDNA (10 ng μl^{-1}) using Transcriptor First Strand cDNA Synthesis Kit (TOYOBO). Reverse transcription was carried out at 37°C, and the PCR was undertaken for 2 min at 95°C, followed by 35 cycles of 20 s at 95°C, 30 s at 49°C, and 60 s at 72°C. PCR products were analyzed on a 1% (w/v) agarose gel, and cDNA fragments of expected size (700-bp) were recovered from the agarose gel using a Qiaquick gel extraction kit (Promega) and cloned into a pMD20 vector (TaKaRa) according to the instructions from the manufacturer.

GS-FLX library construction, sequencing, and functional annotation

Total RNA of the tuber fleshs of *D. esculenta* and *D. alata* was extracted using Plant RNA Isolation Reagent (Invitrogen) and the RNAeasy plant mini-kit (QIAGEN, Hilden, Germany), and DNA contamination was eliminated using RNase-Free DNase Set (QIAGEN). Construction of cDNA libraries for the GS-FLX Titanium System (Roche Diagnostic, Tokyo, Japan), sequencing, cleaning-up and assembly of sequences were performed by Dragon Genomics Center, TakaraBio Inc., Mie Japan). The assembled contigs showing significant similarity to steroidal saponin biosynthetic enzymes such as CYPs, UGTs, and GH1 were identified using the EST Viewer software (Dragon Genomics Center, Takara Bio Co. Ltd., Japan) via the BLASTX search with Arabidopsis CYPs, UGTs, and GH1.

Construction of cDNA library of the tuber fleshs of *D. esculenta* and isolation of a full-length *DeF26G1* cDNA

Total RNA of the tuber fleshs of *D. esculenta* was extracted as described above. cDNA library of the tuber fleshs of *D. esculenta* was constructed using SMART cDNA Library Construction Kit (Clontech Laboratories) according to the manufacture's instruction. Approximately 1,000,000 cDNA-containing phages were screened on nylon filters (Hybond-N, Amersham Biosciences) using an alkaline phosphatase-labeled probe based on *DeF26G1* fragment. A total of 30 positive phages, which are longer than 1000 bp, were isolated and converted to pTriplEx2 plasmids according to the manufacture's instruction (Clontech Laboratories). DNA sequence analysis showed that all three clones were identical. A clone, whose insert was 1,985 bp in length, was selected for further analysis.

The nucleotide sequences were determined using ABI 3130 genetic analyzer (Applied Biosystems) and analyzed using BioEdit, a biological sequence alignment editor (http://www.mbio.ncsu.edu).

Real-Time quantitative RT-PCR analysis

Quantitative RT-PCR was performed with LightCycler®Nano (Roche) using THUNDERBIRDTM SYBR® qPCR Mix (TOYOBO) with the following two sets of primers; *DeF26G1* qPCR Fw: 5'-CAAGCTCTTGAGGATGAATATGGAGGC-3', *DeF26G1* qPCR Rv: 5'-CTCCACGGTTCATTCAATGTGATCCAATAC-3', *GAPDH* qPCR Fw: 5'-AATGCTAGCTGCACCACCAACTG-3', *GAPDH* qPCR Rv: 5'-AACTGGCAGCTCTTCCACCTCTC-3'. Cycling was undertaken for 10 min at 95˚C, 45 cycles of 10 sec at 95°C, 10 sec at 60°C, and 15 sec at 72°C for amplification, followed by holding for 30 sec at 95°C and ramping up from 60°C to 95°C at 0.1°C sec^{-1} for melting curve analysis. Three biological repeats were analyzed in duplicate. *DeF26G1* gene expression levels were normalized against the values obtained for the *GAPDH* gene, which was used as an internal reference. Data acquisition and analysis were performed using LightCycler®Nano software (Roche).

Expression of recombinant DeF26G1 in *Escherichia coli*

DeF26G1 cDNA fragments were amplified by PCR using the following three primers containing restriction sites (underlined); EcoRI-full-DeF26G1 Fw: 5'-GAATTCATGGCCTCAATAGTCTCTCA- 3', EcoRI-d80-DeF26G1 Fw: 5'-GAATTCAAGGCCACTGAAGCATTTGT- 3', XhoI-DeF26G1 Rv: 5'-CTCGAGCTAGTTTTGAGGCTTTGG- 3'. The PCR was performed for 2 min at 95°C, followed by 30 cycles 30 sec at 95°C, 30 sec at 59°C, and 2 min at 72°C. The amplified DNA fragments were ligated into the pMD19 vector (TaKaRa) and digested with EcoRI and XhoI. The DNA fragments were ligated into EcoRI-XhoI sites of the pGEX4T-1. *Escherichia coli* strain BL21 (DE3) transformed with constructed plasmid was grown at 37°C in LB medium with 50 μg ml^{-1} ampicillin until its OD_{600} reached appropriate 0.5. The recombinant protein expression was induced by adding IPTG to 0.1 mM and continued for 20 h at 18°C. The culture

was then centrifuged at 3500 rpm for 30 min at 4°C. The cell pellets were resuspended in 5 ml of cold sonication buffer containing 50 mM sodium phosphate (pH 7.4), 300 mM NaCl and 20% (v/v) glycerol, sonicated using a Bandelin Sonopuls HD 2070 ultrasonic homogenizer (Sigma) typeMS73 at a sound intensity of 200 W cm^{-2}, three times for 30 sec each on ice, and centrifuged at 15,000 rpm for 10 min at 4°C. The GST-tagged proteins present in the supernatant were purified using GST Spin Trap columns (GE Healthcare) according to the manufacturer's instructions. After two column washes, the adsorbed proteins were eluted twice in 200 µl of a solution containing 50 mM Tris-HCl (pH 8.0) and 20 mM reduced glutathione, and mixed. The concentration of the purified proteins was determined by Bradford system. The purified recombinant proteins were visualized by SDS–PAGE. The proteins were revealed by staining the gel using Coomassie brilliant blue R-250. The proteins were used for further analyses.

LC-MS Analysis of steroidal saponins

LC-MS analyses were performed using a system consisting of an Acquity Ultra Performance Liquid Chromatograph (UPLC) (Waters, Milford, MA) and an Acquity quadruple tandem mass spectrometer (TQ Detector) (Waters), and data acquisition and analysis were performed using MassLynx 4.1 software (Waters). Chromatographic column was a Waters ACQUITY UPLC HSS T3 column (100×2.1 mm, 1.8 µm). The column temperature was set at 30°C. For each sample, 5 µL was injected. The flow rate was set at 0.2 ml min^{-1}. The mobile phases were water with 0.1% (v/v) formic acid (A) and acetonitrile (B), using a gradient elution of 10-55% B at 0-30 min, 55-75% B at 30-35 min (0-30 min and 30-35 min, linear gradient) for analysis of extracts from plants. While, the mobile phases were 50% (v/v) MeOH in H_2O (A) and 100% MeOH (B), using a gradient elution of 0% B at 0-2 min, 0-100% B at 2-12 min, 100% B at 12-16 min (2-12 min, linear gradient) for analysis of enzymatic activities. The mass spectra of steroidal saponins extracted from plants were obtained in positive ESI mode, while the mass

spectra of enzymatic reaction products were detected in negative ESI mode. In ESI conditions, the capillary voltage at 3 kV and sample cone voltage at 80 V were applied. MS scan mode with a mass range of *m/z* 400-1100 was used for enzymatic activities analysis using crude enzymes from leaves and recombinant proteins, while SIM mode with *m/z* 867 corresponding to representative fragment ion $[M-H]^-$ of dioscin was applied for enzymatic kinetics analysis. On the other hand, SIM mode with *m/z* 1031 and *m/z* 869 corresponding to representative fragment ion $[M+H-H_2O]^+$ of protodioscin and $[M+H]^+$ of dioscin, respectively, was applied for quantification of steroidal saponins in extracts from plants. The source and desolvation gas temperature were set at 120°C and 350°C, respectively. The nebulizer and desolvation N_2 gas flows were 50 and 550 L/h, respectively.

Biochemical analysis of recombinant d80-DeF26G1

The pH optimum of d80-DeF26G1 was determined using glycine-HCl buffer, sodium citrate buffer and potassium phosphate buffer for the pH ranges 2.0–3.0, 3.0–6.0 and 6.0–8.0, respectively. The activity for each pH was measured as described above. The kinetics parameters of recombinant d80-DeF26G1 were determined in triplicated assays. The activity was assayed using protodioscin at concentration ranging from 10 to 400 μM. Reaction The reaction was carried out at 37°C for 10 minites. Extraction and LC-MS analysis of the reaction product were performed as described above. Kinetic parameters were determined by non-linear regression with ANEMONA (Hernandez and Ruiz, 1998).

Results

Quantification of Steroidal Saponins in *D. esculenta*, *D. cayenensis* and *D. alata.*

The tubers of *Dioscorea* species are known to contain high amounts of steroidal saponins, and

three species, *D. esculenta*, *D. cayenensis*, and *D. alata* were chosen as plant materials to study steroidal saponin biosynthesis. *D. esculenta* (known as a lesser yam) is widely distributed and cultivated for foods in Okinawa island, parts of southern Asia, and the Pacific, and the tubers are known as a bitter yam probably due to high contents of saponins. The tubers of *D. esculenta* have also been used traditionally as a medicine in the treatment of various diseases. *D. cayenensis* (known as a yellow Guinea yam) is the most popular and economically important yam in West Africa, and *D. cayenensis* has been reported to contain high amounts of steroidal saponins (Sautour et al. 2004a; Sautour et al. 2004b; Sautour et al. 2007). *D. alata* (known as a water yam) is widely cultivated as an important tuber crop in tropical and subtropical regions.

First, the contents of steroidal saponins in the leaves and tubers of *D. esculenta*, *D. cayenensis*, and *D. alata* were analyzed by UPLC-ESI-MS (Figure 1-2). The tubers of *D. esculenta* and *D. cayenensis* contained a spirostanol glycoside dioscin and a furostanol glycoside protodioscin, while their leaves contained only protodioscin but no dioscin was found. In contrast, dioscin and protodioscin were not detected in the leaves and tubers of *D. alata*. Thus, comparative transcriptome analysis of the three *Dioscorea* species was performed to identify the genes responsible for steroidal saponin biosynthesis.

EST Analysis of *Dioscorea* species

The comparative transcriptome among the three species was performed by RNA-seq analysis. Total RNA was isolated from the flesh and cortex of the yam tubers, and a transcriptome dataset for each of the extracted RNAs was prepared via 454 pyrosequencing. After cleanup of the sequences and de novo assembly of the cDNA leads, 67,969 contigs were obtained (Figure 1-3). To identify the candidate genes involved in steroidal saponin biosynthesis, three enzyme superfamilies, CYPs, UGTs, and family 1 glycosidases (GH1s) were surveyed in the transcriptome datasets (Table 1-1). Among the three superfamilies, GH1 showed a clear difference between high- and low-saponin producers. The lead numbers of GH1 were much

higher in the tuber fleshs of *D. esculenta* (2,279 leads/total 578,952 leads) and *D. cayenensis* (1123 leads/121,989 leads) than that in *D. alata* (71 leads/137649 leads). Furthermore, one of the GH1 contigs, which was designated as DeF26G, was highly detected in the tuber fleshs of *D. esculenta* (1526 leads) and *D. cayenensis* (989 leads), indicating that the DeF26G contig account for the vast majority of GH1s expressed in the two species. In contrast, the DeF26G contig was not detected in *D. alata* at all. The DeF26G contig showed the highest similarity in the amino acid sequence to furostanol 26-*O*-β-glucosidase of *Costus speciosus* (CsF26G), which hydrolyzes protogracillin to form gracillin (Inoue et al. 1996b). These results, together with the difference in the saponin contents among the three *Dioscorea* species, suggested that DeF26G is involved in steroidal saponin biosynthesis. Therefore, this study was focused on the characterization of the DeF26G contig.

β-glucosidase activity in crude enzymes from *D. esculenta* and *D. alata*

RNA-seq analysis suggested the existence of the transcript encoding a furostanol 26-*O*-β-glucosidase homolog in *Dioscorea*. To examine the possibility, crude enzymes prepared from the leaves of *D. esculenta* and *D. alata* were incubated with protodioscin, and the reaction products were analyzed by UPLC-ESI-MS (Figure 1-4). After 1h incubation with protodioscin, the crude enzymes from *D. esculenta* gave a product peak with a retention time at 12.23 min. The product is identical to the authentic compound of dioscin in terms of the retention time and the mass spectra. In contrast, the crude enzymes of *D. alata* did not give the corresponding peak. These results suggested that a β-glucosidase hydrolyzing protodioscin to dioscin is present in *D. esculenta* but not in *D. alata*. Thus, the *DeF26G* gene, which is specifically expressed in *D. esculenta,* is likely involved in the hydrolysis of protodioscin.

Isolation of full-length *DeF26G1* cDNA

Based on RNA-seq analysis, a partial cDNA fragment (702-bp) of the *DeF26G* contig was

obtained by PCR, and a *DeF26G* cDNA was isolated from a cDNA library of the *D. esculenta* tubers by using the PCR fragment as a probe. The isolated full-length *DeF26G1* cDNA consists of a 1701-bp open-reading frame and a 270-bp 3'-noncoding region. The deduced DeF26G1 protein is a 566 amino-acid precursor polypeptide consisting of a mature protein of 486 amino acid residues and a putative chloroplast transit peptide of 80 amino acid residues at the *N*-terminus.

The predicted mature polypeptide of DeF26G1 showed the high sequence identity to that of CsF26G (65.5%) and shared 46.5% sequence identity to that of avenacosidase. The phylogenetic analysis of DeF26G1 with other plant GH1s revealed that DeF26G1 was located in the same clade of monocot GH1s including CsF26G and avenacosidase (Figure 1-5). The deduced DeF26G1 contained several sequence motifs that are highly conserved among GH1s (Figure 1-6). The NEP sequence motif, of which the Glu residue is an acid/base catalyst, was found at residues 257–259, and the sequence ITENG, of which the Glu residue is a catalytic nucleophile of β-glucosidases, was also found at residues 468–472 (Jenkins J et al. 1995; Keresztessy Z et al. 1994). The residues involved in the binding of the glycone (glucose) moiety are highly conserved in all GH1s (Barrett et al. 1995; Rye and Withers 2000; Zechel and Withers 2000; Sue et al. 2006; Saino et al. 2014), and these residues were also found at Gln-108, His-212, Asn-257, Glu-258, Glu-526, and Trp-527 in the DeF26G1 sequence (Figure 1-6).

Biochemical characterization of the recombinant DeF26G1

To investigate the potential catalytic activity of DeF26G1, the full-length DeF26G1 protein (full-DeF26G1) and the predicted mature form of DeF26G1 (d80-DeF26G1), of which the 80 amino acid residues at the N-terminus was deleted, were expressed in *E. coli* as a glutathione S-transferase fusion protein (Figure 1-7). The β-glucosidase activity was assayed with protodioscin as a substrate, and the reaction products were analyzed by UPLC-ESI-MS. The

reaction product with the recombinant full-DeF26G1 was detected at Rt 12.2 min (Figure 1-8) and gave an $[M-H]^-$ ion at *m/z* 867 (Figure 1-8). The product is identical to the authentic compound of dioscin in terms of the retention time and the mass spectra. Protodioscin contains the other β-glycosidic bonds of the oligosaccharide moiety attached at the C3-hydroxy position, but these bonds were not hydrolyzed by DeF26G1 because the peak giving an $[M-H]^-$ ion at *m/z* 413 corresponding to the aglycone diosgenin was not detected. The recombinant d80-DeF26G1 essentially gave the same results and showed the better hydrolytic activity than full-DeF26G1. Furthermore, a new product peak at Rt 8.42 min was detected with d80-DeF26G1 (Figure 1-8A (d)). This unknown product gave an $[M-H]^-$ ion at *m/z* 885 (Figure 1-8B (d)), which corresponds to the molecular mass of the 26-degluco form of protodioscin. The results strongly suggested that the product detected at Rt 8.42 min is a hydrolytic product before forming a F-spiro ring and that the 26-degluco form of protodioscin is spontaneously converted to the spirostanol saponin dioscin. Thus, DeF26G1 was able to hydrolyze the 26-*O*-glycosidic bond of protodioscin to form dioscin, clearly indicating that DeF26G1 is a furostanol glycoside 26-*O*-β-glucosidase.

The β-glucosidase activity of the d80-DeF26G1 toward protodioscin was examined within the range of pH2 - pH8, and the optimal pH was determined to be pH 7. The optimal pH for DeF26G1 is higher that that of CsF26G (pH 5-5.5). The apparent K_m value for protodioscin was determined to be 140 ± 0.0760 μM (Table 1-2 and Figure 1-9), and this value is slightly higher than that of CsF26G (50 μM for protogracillin). The d80-DeF26G1 protein did not hydrolyze *p*-nitrophenyl-β-D-glucopyranoside, which is a standard artificial substrate for various β-glucosidases.

Expression of the *DeF26G1* gene in *D. esculenta* and *D. alata*

Quantitive real-time-PCR analysis was performed toward total RNA extracted from the tubers and the leaves of *D. esculenta* and *D. alata* (Figure 1-10). The *DeF26G1* transcript in *D.*

esculenta was detected in the tubers and also 6-fold higher in leaves. In contrast, the *DeF26G1* transcript in *D. alata* was not detected at all. These results are consistent with the results of the RNA-seq analysis and the F26G activity with the leaf crude extracts.

Discussion

In this chapter, comparative transcriptome analysis of three *Dioscorea* species was performed to investigate steroidal saponins biosynthesis in plants. Analysis of endogenous steroidal saponins content in the leaves and tubers revealed a clear contrast between the two *Dioscorea* species, namely, *D. esculenta* as a high saponins producer and *D. alata* as a low saponins producer, and therefore, the transcripts involved in the steroidal saponins biosynthesis were expected to be specifically expressed in *D. esculenta*. The *DeF26G* contig among the candidate genes (*CYPs*, *UGTs*, and *GH1*) was highly detected in the transcriptome datasets of *D. esculenta* but not of *D. alata* (Table 1-1), and biochemical characterization clearly indicated that DeF26G1 is a furostanol glycoside 26-*O*-β-glucosidase.

In the transcriptome datasets, the *DeF26G* contig (1526 leads) accounts for 0.31% of total leads (491,183) in the tuber flesh of *D. esculenta* and 0.92% of total leads (107,342) in the tuber fleshs of *D. cayenensis* (Table 1-1 and Figure 1-3), and furthermore, the *DeF26G* contig occupied 67% of total *GH1* contigs (2279 leads) in the tuber fleshs of *D. esculenta* and 88% of total *GH1* contigs (1123 leads) in the tuber fleshs of *D. cayenensis* (Table 1-1). These results indicate that DeF26G1 is a predominant β-glucosidase in the high saponin producers of *Dioscorea* species. The expression of the *DeF26G1* gene in the leaves of *D. esculenta* was higher than that in the tubers (Figure 1-10). The endogenous contents of protodioscin, which is a natural substrate of DeF26G1, was higher in the leaves than in the tubers (Figure 1-2), and thus, the distribution pattern is well consistent with the *DeF26G1* gene expression. It is noteworthy that the leaves of *D. esculenta* and *D. cayenensis* accumulate only a furostanol-type

protodioscin but not a spirostanol-type dioscin at all. This suggests that the substrate protodioscin and the corresponding degrading enzyme DeF26G1 are stored in a separate compartment such as subcellular localization or different tissue distribution in the leaves. The full-length DeF26G1 protein contains a putative chloroplast transit peptide of 80 amino acid residues at the *N*-terminus, suggesting chloroplast localization of DeF26G1. As supporting to these suggestions, Gurielidze et al. (2004) previously reported that the F26G activity is localized in the membrane fraction of the thylakoids and also that furostanol saponins are localized in idioplasts of the epidermis in the leaves of *D. caucasica*. In this chapter, it has been shown that DeF26G1 belongs to a GH1 family, in which various β-glucosidases are distributed throughout the plant kingdom. Plant GH1 β-glucosidases are involved in defense response against microbials, insects and herbivores, and the pre-toxic glycosides and the degrading GH1s are localized in a separate compartment in plant tissues to prevent poisoning themselves (Morant et al. 2008; Ahn et al. 2004 and 2006). Similarly, it is likely that *Dioscorea* spp. accumulate furostanol saponins and F26G in a separate compartment of the leaves as a defense mechanism and that, upon cell damage, the hydrolysis of 26-*O*-glycosidic bond of furostanol saponins by F26G may immediately produce a bioactive spirostanol saponins. In contrast to the leaves which accumulate only furostanol saponins, the tubers of *D. esculenta* and *D. cayenensis* contain spirostanol saponin dioscin as well as protodioscin. The different distribution of furostanol and spirostanol saponins in the leaves and tubers suggested the presence of unresolved mechanisms by which the organ specific biosynthesis and transport of steroidal saponins are regulated in *Dioscorea* spp. Identification of the biosynthetic genes such as CYPs, UGTs, and transporters and characterization of their tissue specific expression in planta are needed to address the question.

In conclusion, comparative transcriptome analysis of *Dioscorea* species was performed together with chemical analysis of steroidal saponins content, and DeF26G1 was identified as a furostanol glycoside 26-*O*-β-glucosidase, which is involved in the conversion of protodioscin

to dioscin (Figure 1-1). These approaches are powerful to isolate the candidate genes involved in the steroidal saponins biosynthesis. Several candidate genes (CYPs and UGTs) have been found in the transcriptome datasets used in this chapter (Table 1-1 and Figure 1-1), and their biochemical characterization will unravel the molecular mechanism of the steroidal saponins biosynthesis in plants.

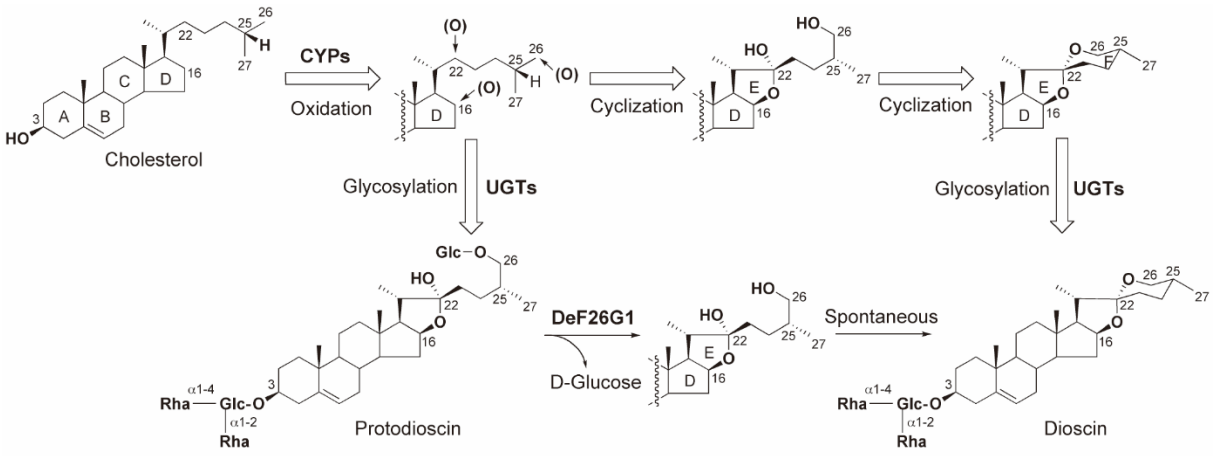


Figure 1-1. The putative biosynthetic pathway of protodioscin and dioscin in *Dioscorea* spp. Thick arrows indicate unidentified reaction stages. Solid arrows mean reaction steps characterized in this work.

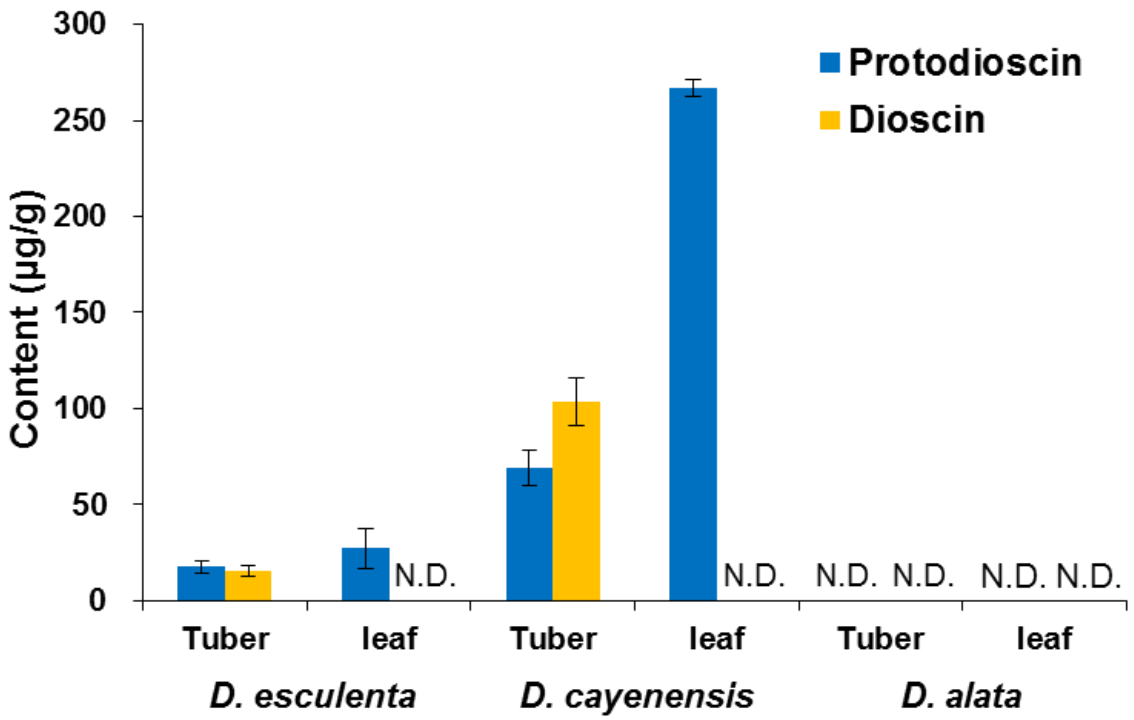


Figure 1-2. Endogenous contents of protodioscin and dioscin in the tubers and the leaves of *D. esculenta*, *D. cayenensis*, and *D. alata*. Bars indicate standard deviation from the mean (n=3). ND indicates not detected.

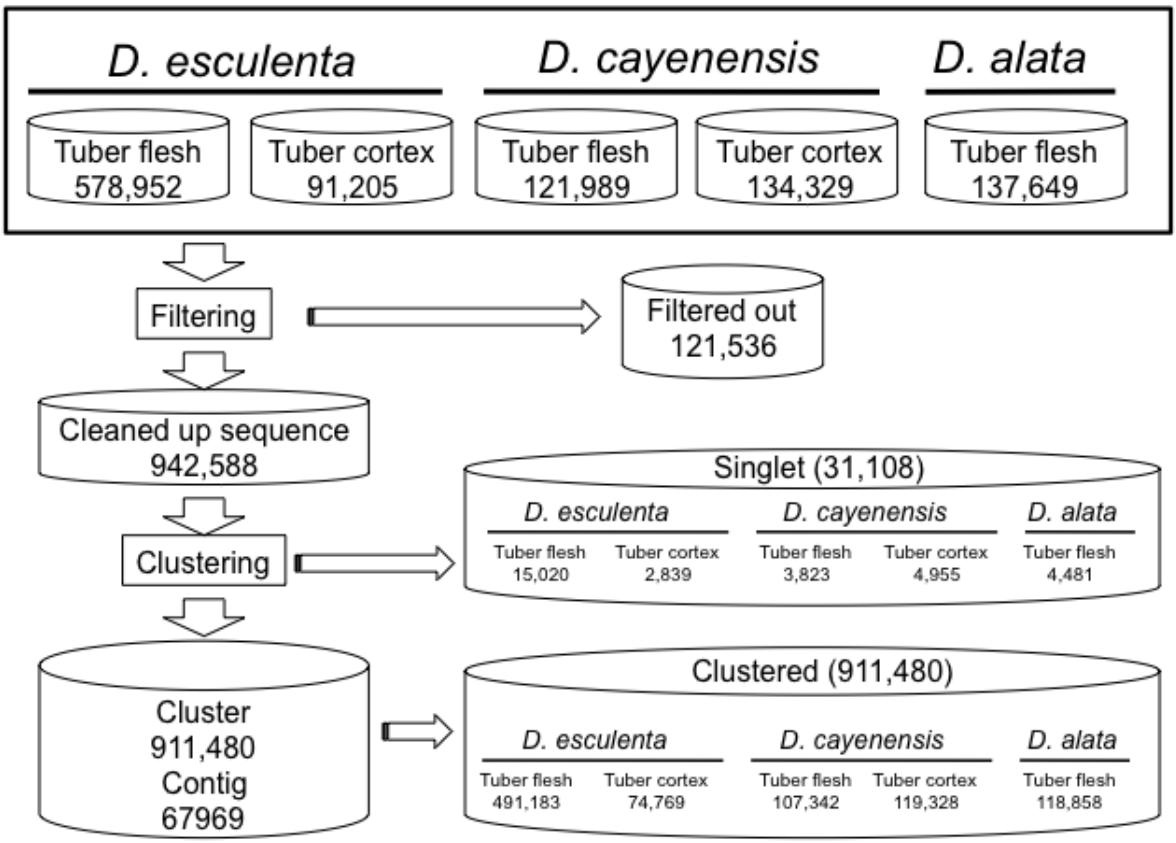


Figure 1-3. Summary of EST libraries construction of *Dioscorea* spp., *D. esculenta*, *D. cayenensis* and *D. alata*, by RNA-seq analysis.

Table 1-1. Number of EST reads of putative three enzyme superfamilies, cytochrome P450 monooxygenases (CYPs), UDP-dependent glycoxyltransferases (UGTs), and family 1 glycosidases (GH1s), involved in steroidal saponin biosynthesis in *Dioscorea* spp.

Gene Family	(contig)	*D. esculenta*		*D. cayenensis*		*D. alata*
		Tuber flesh	Tuber cortex	Tuber flesh	Tuber cortex	Tuber flesh
CYP		4007	722	337	786	299
UGT		1207	243	208	406	232
GH1	(DeF26G)	1526	312	989	27	0
	(other GH1)	753	205	134	159	71

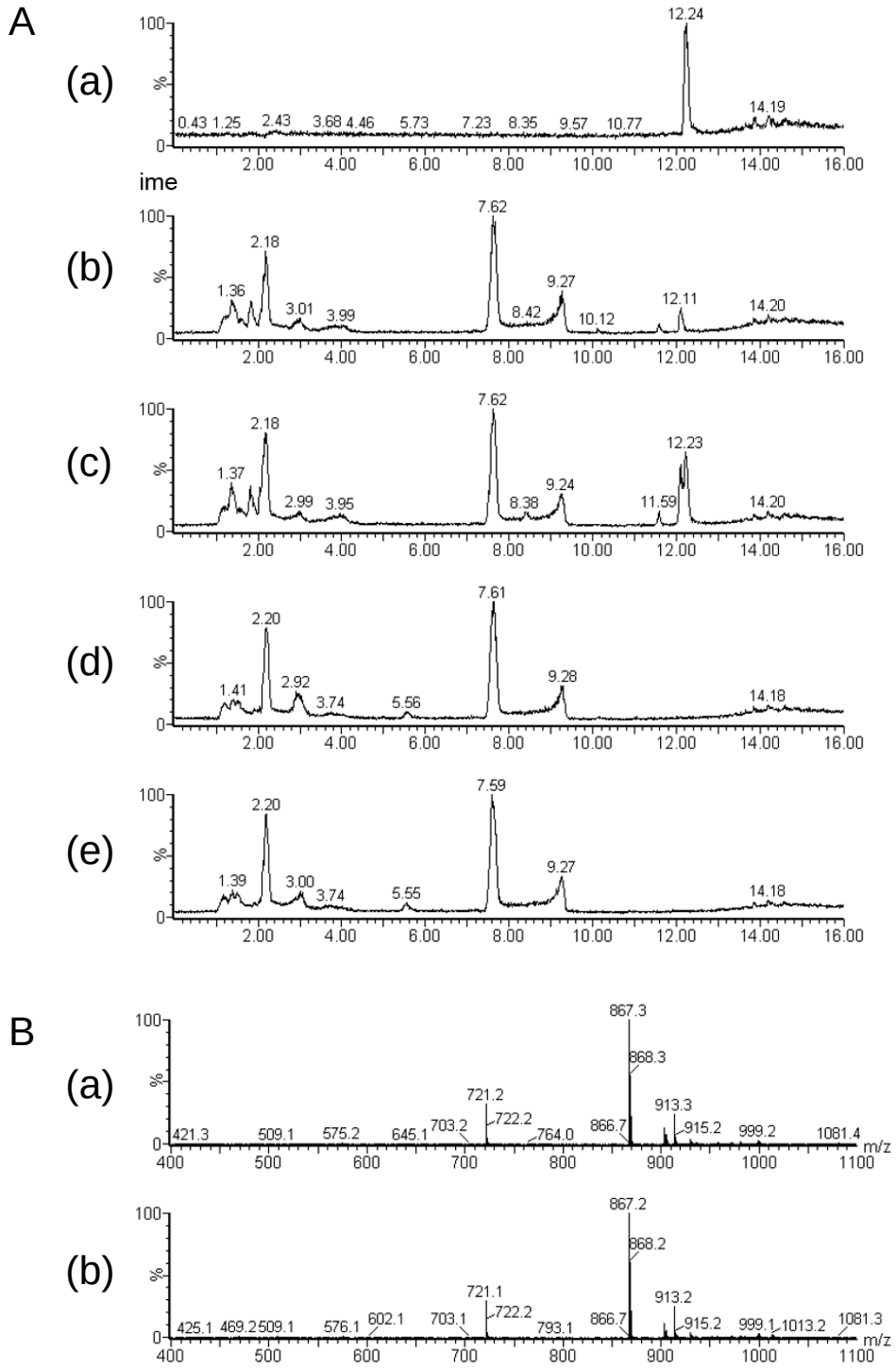


Figure 1-4. UPLC-ESI-MS analysis of the reaction products from crude enzymes prepared from the leaves of *D. esculenta* and *D. alata* with protodioscin as substrate. (A) Total ion chromatogram (TIC) of the authentic compound of dioscin and the reaction products from crude enzymes. (a) authentic standard of dioscin; (b) control (0 min) of *D. esculenta*; (c) reaction products (1 h) of *D. esculenta*; (d) control (0 min) of *D. alata*; (e) reaction products (1 h) of *D. alata*. (B) (a) Mass spectra of the peak with a retention time at 12.24 min from the authentic compound of dioscin; (b) mass spectra of the product peak with a retention time at 12.23 min after 1 h incubation from the crude enzymes from *D. esculenta.*

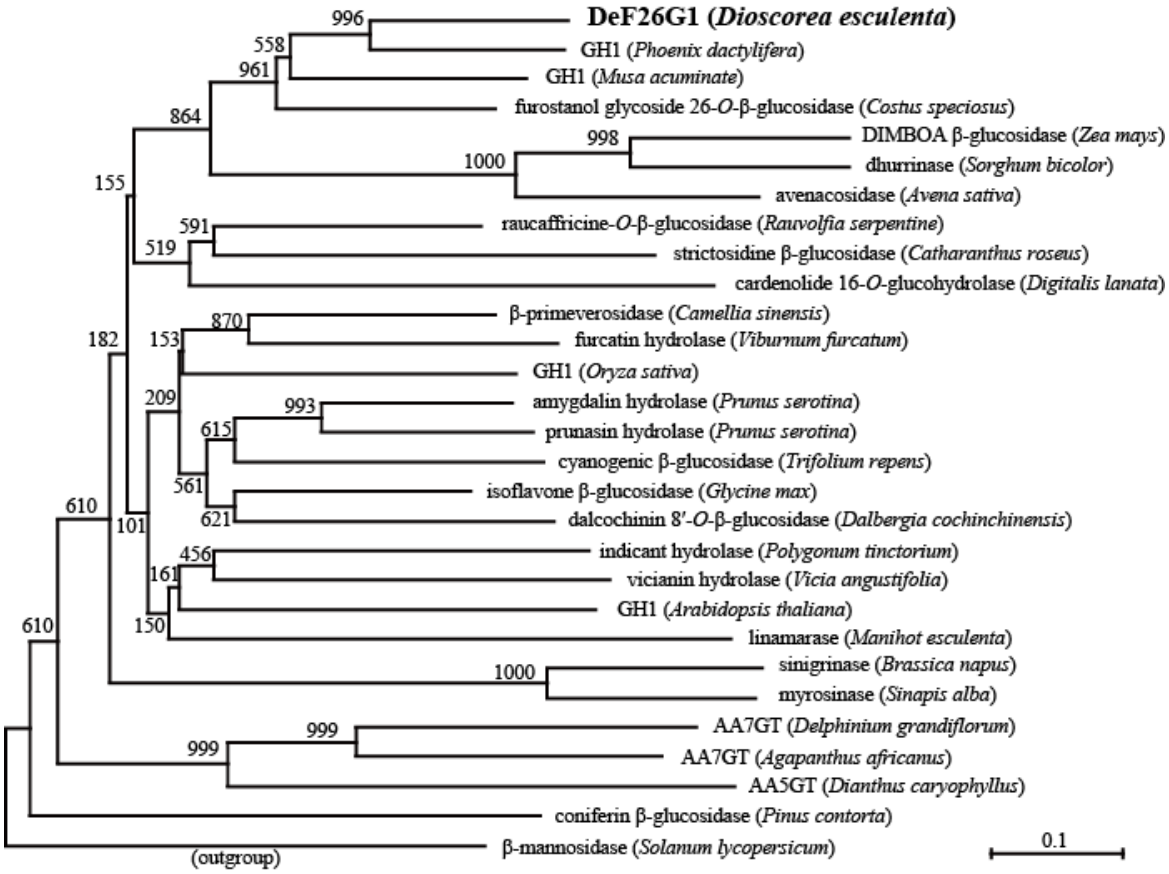


Figure 1-5. A neighbor-joining phylogenetic tree of glycosyl hydrolase family 1 (GH1) members from various plants. The entire amino acid sequences and their functions of the GH1 members were obtained from the published studies. The Genbank ID numbers of the GH1 members are as follows: GH1 (*Phoenix dactylifera*), XP_008799933; GH1 (*Musa acuminate*), XP_009416645; β-primeverosidase (*Camellia sinensis*), Q7X9A9; furcatin hydrolase (*Viburnum furcatum*), Q75W17; AA5GT (*Dianthus caryophyllus*), BAJ33501; AA7GT (*Delphinium grandiflorum*) BAJ33502; AA7GT (*Agapanthus africanus*), BAM29304; vicianin hydrolase (*Vicia angustifolia*), A2SY66; GH1 (Oryza-sativa), Q7XKV5; amygdalin hydrolase (*Prunus serotina*), Q40984; prunasin hydrolase (*Prunus serotina*), Q9M5X4; indicant hydrolase (*Polygonum tinctorium*), Q9XJ67; GH1 (*Arabidopsis thaliana*), Q8GY78; DIMBOA β-glucosidase (*Zea mays*), P49235; dhurrinase (*Sorghum bicolor*), Q41290; isoflavone β-glucosidase (*Glycine max*), BAF34333; coniferin β-glucosidase (*Pinus contorta*), Q9ZT64, dalcochinin 8′-*O*-β-glucosidase (*Dalbergia cochinchinensis*), Q9SPK3; raucaffricine-*O*-β-glucosidase (*Rauvolfia serpentine*), Q9SPP9; strictosidine β-glucosidase (*Catharanthus roseus*), Q9M7N7; linamarase (Manihot esculenta), Q41172; cyanogenic β-glucosidase *(Trifolium repens)*, P26205; β-mannosidase *(Solanum lycopersicum),* AAL37714; sinigrinase (*Brassica napus*), Q00326; myrosinase (*Sinapis alba*), P29736; cardenolide 16-*O*-glucohydrolase *(Digitalis lanata)*, CAB38854. The statistical significance of the NJ tree topology was evaluated by bootstrap analysis with 1000 iterative tree constructions. The scale indicates the evolutionary distances of the base substitution per site, estimated by the Kimura's two-parameter method.

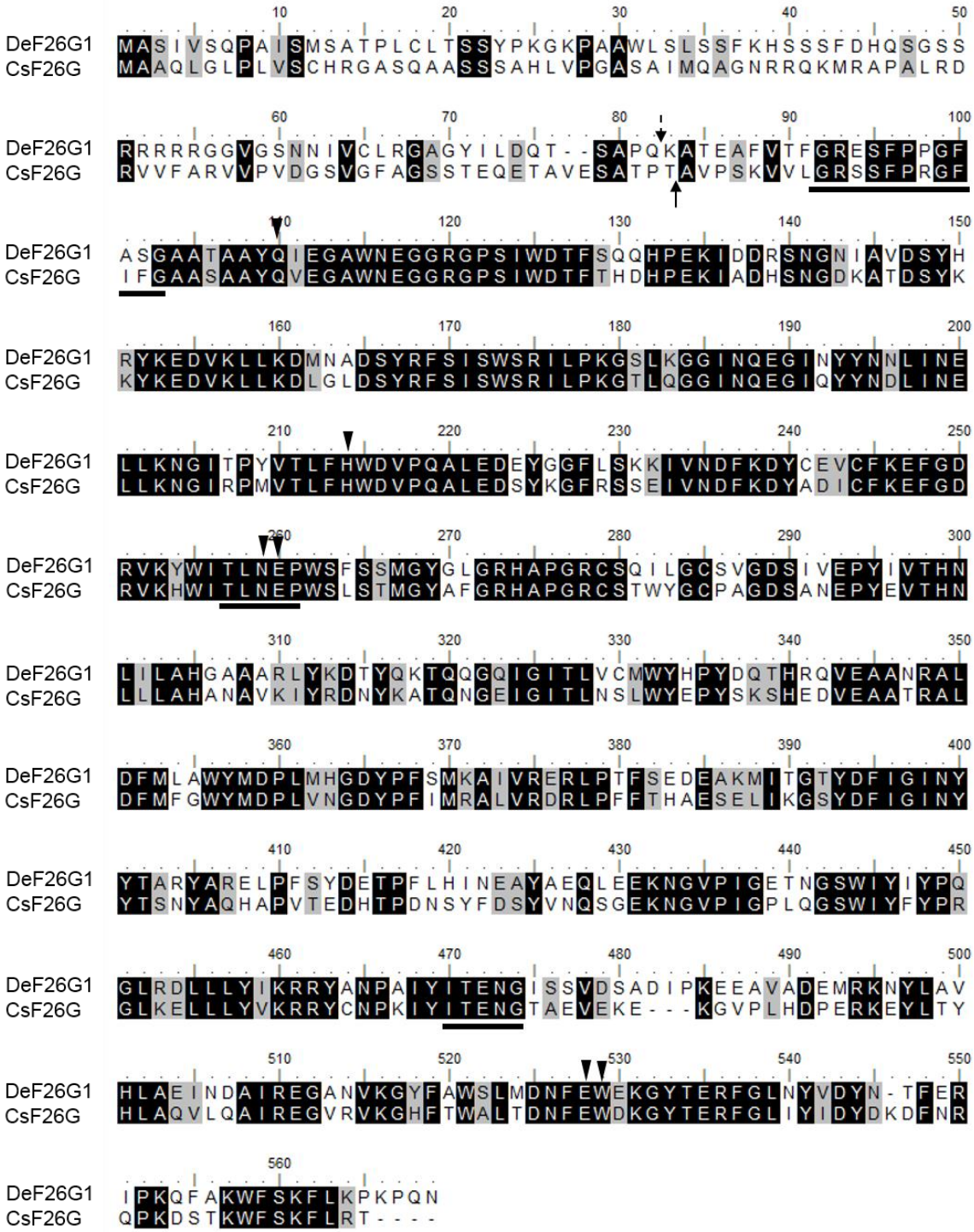


Figure 1-6. Sequence alignment of amino acid of DeF26G1 and CsF26G. Multiple sequence alignment was performed using the ClustralW multiple alignment analysis tool of BioEdit. Identical and similar amino acid residues are shaded in black and gray respectively. Conserved sequence motifs among family 1 glucosidase (GH1) are underlined. Six amino acids for glucoside recognition are indicated with filled triangles on the sequence. The predicted transit peptide cleavage site of CsF26G and DeF26G1 are indicated with solid arrows. DeF26G1 arrow is the site predicted by the PSORT program (psort.nibb.ac.jp).

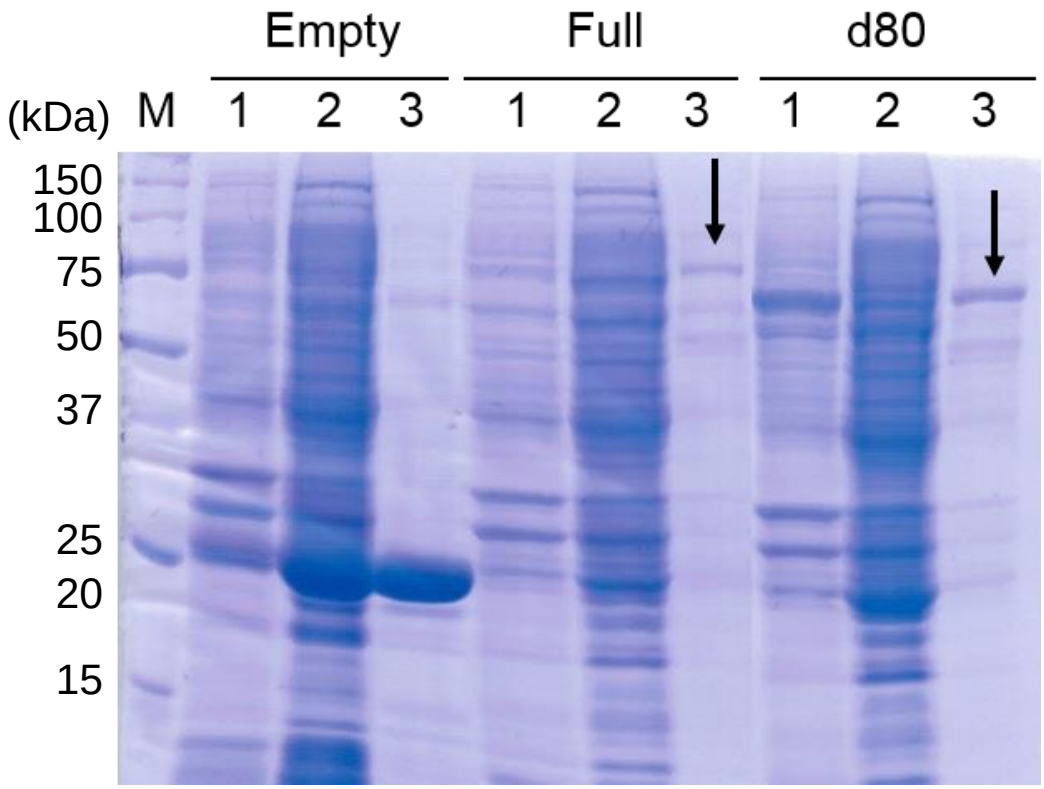


Figure 1-7. Expression of the recombinant DeF26G1 protein in *E. coli*. *E. coli* BLR cell extracts were analyzed by 10%-SDS-PAGE, and the proteins were visualized by Coomassie Brilliant Blue R-250 staining. Lane M, molecular size markers; lane 1, crude precipitation from the cells containing expression vector; lane 2, crude soluble proteins from the cells; lane 3, the purified GST fusion protein from the crude soluble proteins; Empty; pGEX4T-1 (empty vector), Full and d80; pGEX4T-1 containing Full- and d80-DeF26G1, respectively. Full- and d80-DeF26G1 proteins are indicated with solid arrows.

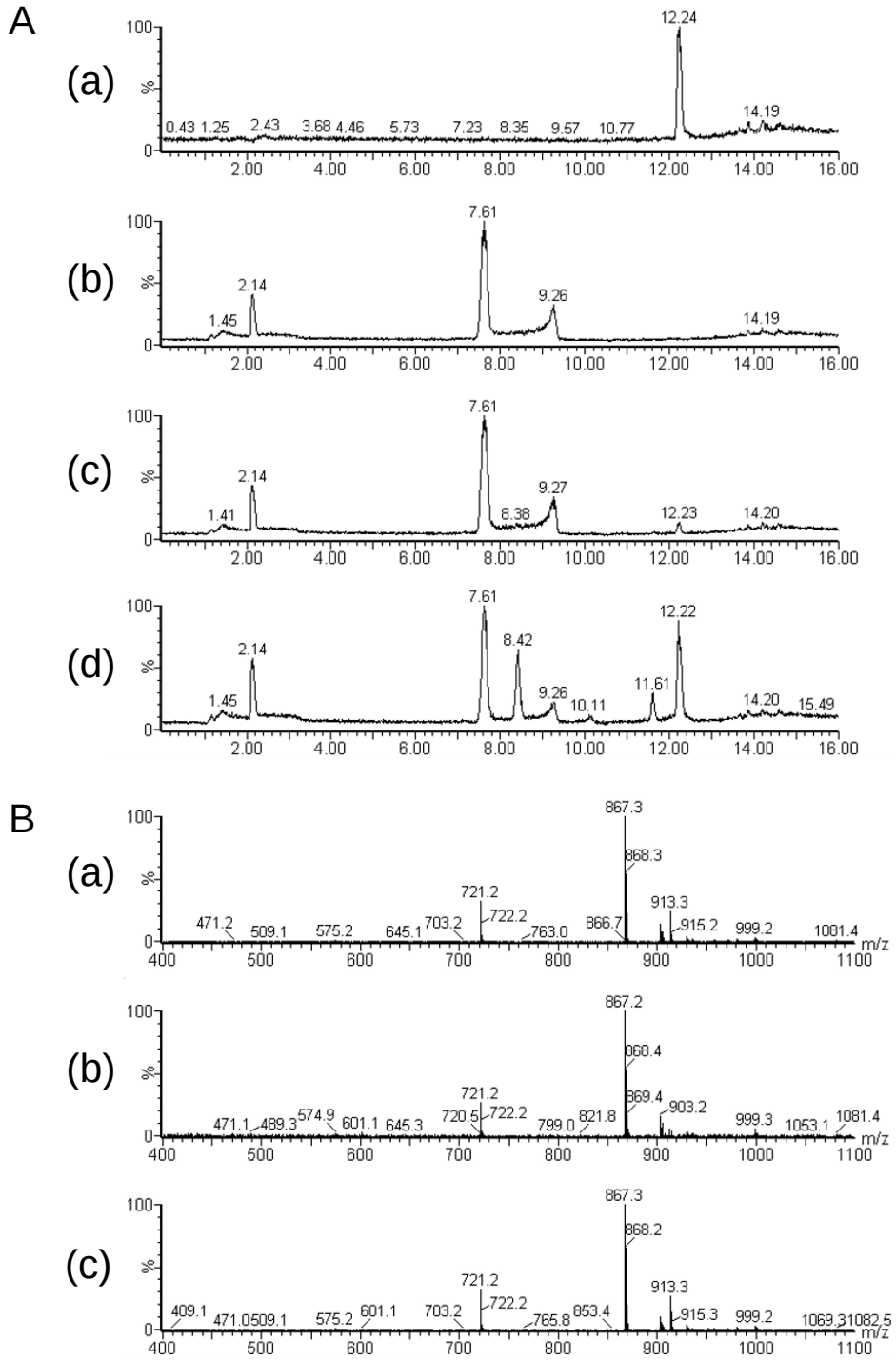


Figure 1-8. UPLC-ESI-MS analysis of the reaction products from the recombinant GST fusion proteins with protodioscin as substrate. (A) Total ion chromatogram (TIC) of the authentic compound of dioscin and the reaction products from the recombinants. (a) authentic standard of dioscin; (b) the reaction products with the empty vector; (c) the reaction products with the full-length DeF26G1; (d) the reaction products with d80-DeF26G1. (B) (a) Mass spectra of the peak with a retention time at 12.24 min from the authentic compound of dioscin; (b) mass spectra of the product peak with a retention time at 12.2 min from the full-length DeF26G1; (c) mass spectra of the product peak with a retention time at 12.2 min from d80-DeF26G1; (d) mass spectra of the product peak with a retention time at 8.4 min from d80-DeF26G1; (e) mass spectra of the product peak with a retention time at 11.6 min from d80-DeF26G1.

Table 1-2. The biochemical properties of d80-DeF26G1

Parameter	value
Optimal pH	7.0
K_m [μM]	124 ± 0.934
k_{cat} [s^{-1}]	0.161 ± 0.00427
k_{cat}/K_m [$mM^{-1}s^{-1}$]	1.30 ± 0.0287

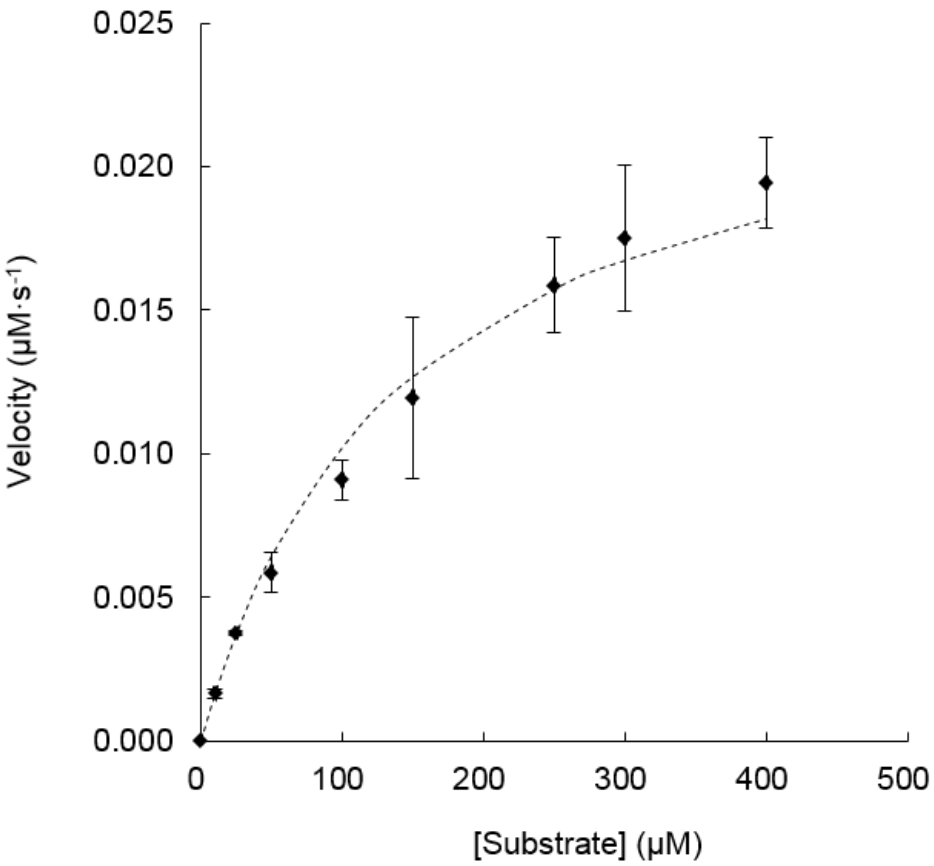


Figure 1-9. d80-DeF26G1 recombinant enzyme activity curves. Enzyme activities were measured with substrate concentrations up to 400 μM protodioscin. Michaelis–Menten curve (featuring *K*m of 140 μM and *V*max of 0.0245 μM·s-1) was fitted to the values obtained. Kinetic parameters were determined by non-linear regression with ANEMONA (Hernandez and Ruiz, 1998).

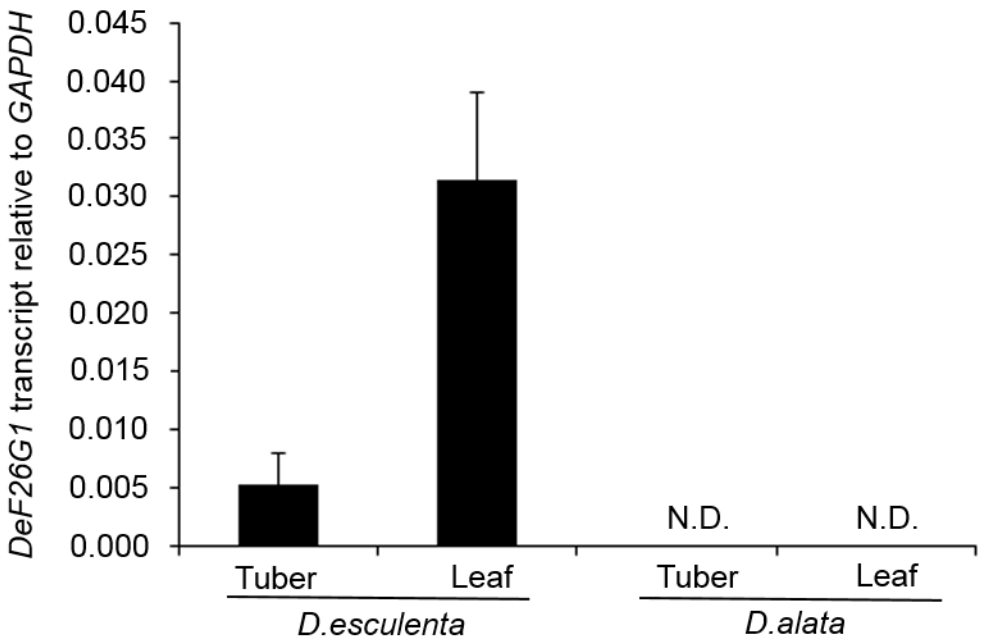


Figure 1-10. Quantitative RT-PCR analysis of *DeF26G1* gene in different organs of *D. esculenta* and *D. alata*. Values represent the ration between each *DeF26G1* and the corresponding *GAPDH* levels. Bars indicate standard deviation from the mean (*n*=3). N.D. indicates not detected.

Chapter 2

Identification of two cytochrome P450s catalyzing the hydroxylation of cholesterol for steroidal glycoalkaloids biosynthesis in potato

Introduction

As described in General introduction, only three UGT genes involved in the glycosylation for α-solanine and α-chaconine, potato SGAs, have previously been identified (Moehs et al., 1997; McCue et al., 2005, 2006, 2007). To elucidate the SGAs biosynthesis in potato, the modification steps from cholesterol as the precursor to solanidine as the aglycone were investigated. The solanidine biosynthetic pathway is hypothesized to require the three oxidation steps at C-16, C-22 and C-26 position and the amination at C-26 as described in General introduction. Recently, the later step was shown to be the amination reaction (Ohyama et al. 2013).

In plants, numerous CYPs have been known to be involved in the biosynthesis or the catabolism of triterpenoids and steroids. The CYPs associated with the triterpenoids biosynthesis, which modify β-amyrin backbone at C-11, 12, 13, 16, 22, 23, 28 or 30 position, dammarenediol-II backbone at C-6, 12 or 28 position, and lupeol at C-20 position, have been identified (Moses et al., 2014). Additionally, the CYPs for the brassinosteroid, a member of plant hormones, biosynthesis and catabolism, which oxidize steroid backbone at C-2, 3, 6, 22, 23 or 26 position, have also been characterized (Ohnishi et al., 2009).

It is suggested that several CYPs are likely involved in the oxidation at the C-16, C-22, and C-26, followed by the amination at C-26 in SGAs biosynthesis, similarly to that of triterpenoids and brassinosteroid as described above. Then, this chapter was focused on CYPs superfamily as candidate genes involved in the oxidation steps of cholesterol in the potato SGAs

biosynthesis.

In this chapter, two CYPs, PGA1 and PGA2 were identified as the enzymes catalyzing the two hydroxylation steps at C-26 and C-22 position, respectively. Additionally, the *PGA1*- and *PGA2*-silenced transgenic plants contained a significantly low amounts of the SGAs. Furthermore, the silenced plants had novel phenotypes, including the suppression of flower development and tuber sprouting. Since, the sprouting process reduces the quality and yield of potato tubers, the suppression of that is remarkably benefit to the industry for the long-term storage of tubers.

Materials and Methods

Chemicals

Authentic samples of α-solanine, α-chaconine, α-tomatine and the two 22-hydroxycholesterols were purchased from Sigma-Aldrich. Cholesterol was purchased from Tama Biochemical Co. Authentic compounds of the two 26-hydroxycholesterols and (22S, 25S)-22, 26-dihydroxycholesterol were kindly provided by Bunta Watanabe.

Cloning of *PGA1 and PGA2* cDNAs.

The potato cDNA template was prepared from mRNA isolated from sprouts of *S. tuberosum* cv. Sassy using an RNeasy Plant Mini Kit (Qiagen) and SuperScript First-Strand Synthesis System for RT-PCR (Life Technologies). The transcripts of *PGA1 and PGA2* genes were PCR amplified with primer sets designed from the potato unigene sequences (TC135549 and TC141445 in the Potato Gene Index database, respectively); U841: 5’- GCTTGCTCTGTTCTTGTACATCTC-3’ / U842: 5’- TGAAAAGCAGAATTAGCAGCA-3’ and U875: 5’- CCAAGGGACAGAGCAATCAA-3’ / U876: 5’- TGATGTGAACTTGAGATTGGTG-3’,

respectively. The PCR products of *PGA1* and *PGA2* were cloned into the pCR4-TOPO plasmid (Life technologies).

Expression of recombinant PGA1 and PGA2 in insect cells

The in vitro enzyme activity assays with homogenates of insect cells expressing recombinant PGA1 and PGA2 proteins were conducted as previously reported (Ohnishi et al. 2006a). *PGA1* and *PGA2* cDNAs were cloned into the pDEST8 vector (Life Technologies) and were then used to generate the corresponding recombinant Bacmid DNAs by transforming *Escherichia coli* DH10Bac (Life Technologies). Preparation of the recombinant baculovirus DNAs that contained *PGA1* and *PGA2* cDNAs and transfection of *Spodoptera frugiperda* 9 insect cells were carried out according to the manufacturer's instructions (Life Technologies). Heterologous expression of PGA1 and PGA2 proteins in the insect cells and spectrometric analyses were performed as described (Saito et al. 2004). Microsomal fractions of the insect cells expressing PGA1 and PGA2 were obtained from the infected cells (100 mL of suspension cultured cells). Infected cells were washed with PBS buffer and suspended in buffer A consisting of 20 mM potassium phosphate (pH 7.25), 20% (v/v) glycerol, 1 mM EDTA, and 1 mM DTT. The cells were sonicated using a Bandelin Sonopuls HD 2070 ultrasonic homogenizer (Sigma) typeMS73 at a sound intensity of 200 W cm^{-2}, three times for 30 sec each on ice, and cell debris was removed by centrifugation at 15,000 rpm for 60 min at 4°C. The pellet was homogenized with buffer A to provide the microsomal fractions. The microsomal fractions were stored at –80 ºC before the enzyme assays.

In vitro enzyme activity assay

PGA1 and PGA2 activities were reconstituted by mixing the PGA1 and PGA2-containing microsomes with purified Arabidopsis NADPH-P450 reductase (Mizutani and Ohta 1998). The reaction mixture (100 μL) consisted of 100 mM potassium phosphate, pH 7.25, 50 pmol/mL

recombinant P450 protein, 0.1 unit/mL NADPH-P450 reductase, 1 mM NADPH, and 20 mM of substrate sterols. Reactions were initiated by addition of NADPH and were performed at 30˚C for 3 h. The reaction products were extracted three times with an equal volume of ethyl acetate. The organic phase was collected and evaporated. The residue was trimethylsilylated with TMS-HT (=HMDS and TMCS in Anhydrous Pyridine) [for Gas Chromatography] (Tokyo Chemical Industry Co., Ltd.) at 80˚C for 45 min. GC-MS was conducted using a GC-MS-QP2010 Ultra (Shimadzu) with a DB-1MS (30 m × 0.25 mm, 0.25 μm film thickness; J&W Scientific) capillary column to analyze the PGA1 reaction products and with a DB-5MS (30 m × 0.25 mm, 0.25 μm film thickness; J&W Scientific) capillary column to analyze the PGA2 reaction products. The injection temperature was 250˚C. The column temperature program for analysis of the PGA1 reaction products was as follows: 80˚C for 1 min, followed by a rise to 300˚C at a rate of 20˚C min^{-1}, and a hold at 300˚C for 20 min. The column temperature program for analysis of the PGA2 reaction products was as follows: 180˚C for 1 min, followed by a rise to 280˚C at a rate of 20˚C min^{-1}, followed by a rise to 300˚C at a rate of 2˚C min^{-1}, and a hold at 300˚C for 20 min. The carrier gas was He, and the flow rate was 1.0 ml min^{-1}; the interface temperature was 280˚C with a splitless injection.

Generation of transformation vectors, plant transformation and growth conditions of *PGA1*- and *PGA2*-silenced potato plants

PGA1- and *PGA2*-silenced transgenic potato plants were generated by RNAi with *Agrobacterium*-mediated transformation of tuber discs obtained from *S. tuberosum* cv. Sassy (Momma 1990). The control vector pKT19 for plant transformation was generated by inserting the GUS gene into the pBin19 (Frisch et al. 1995)-based binary vector (Umemoto et al. 2001) between the Cauliflower mosaic virus 35S (CaMV35S) promoter and the nopaline synthase (NOS) terminator. To silence the *PGA1* and *PGA2* transcripts, two different regions of *PGA1* gene and a region of *PGA2* gene were designed. A 372 bp fragments of *PGA1* cDNA was PCR

amplified using primer sets; U724: 5’- GAGCTCTAGAGAAGCAAAGAAAACACC-3’ / U725: 5’- GGATCCATATGCTAACCAATTCCTCCCATC-3’. The amplification product was cloned into the pCR4-TOPO plasmid. An RNAi binary vector pKT226 targeting *PGA1* gene was constructed from the binary vector pKT19 (Umemoto et al., 2001), substituting for the GUS gene, by locating two *PGA1* fragments in opposite interposing the third intron of the *Arabidopsis thaliana At4g14210* gene under the control of the cauliflower mosaic virus 35S (CaMV35S) promoter in the T-DNA region (Figure 2-1). Two RNAi binary vectors, pKT249 targeting another region of the *PGA1* gene and pKT227 targeting PGA2, were constructed similarly. A 402 bp fragment of PGA1 cDNA was PCR amplified using primer set; U869: 5’-GAGCTCTAGACCACAGCTTTGCTCTCTTG-3’ / U870: 5’-GGATCCATATGCATCGTCTCCCCATACT-3’. A 325 bp fragment of PGA2 cDNA was PCR amplified using primer set; U726: 5’-GAGCTCTAGAGGTTAAGAGTTTGTGCCAACG-3’ / U727: 5’-GGATCCATATGGCTTTCTCTTGCCAATCTG-3’. Transgenic potato plants carrying the pKT19, pKT226, pKT249, and pKT226 constructs were generated using *A. tumefaciens* GV3101 pMP90 cells. In vitro grown plants were cultured at 20˚C under a 16 h light/8 h dark condition. The obtained transformants were individually selected by genomic PCR of the shoots with the primer set; NP2: 5’-TAAAGCACGAGGAAGCGGT-3’ and NP3: 5’-GCACAACAGACAATCGGCT-3’ targeting the kanamycin resistance gene on the T-DNA region integrated into the potato genome. RT-PCR analysis of *PGA1*, *PGA2* and the *elongation factor 1α* (*EF1α*) control (Nicot et al. 2005) was carried out using primer sets U724 / U840: 5’-GAGCTCTAGAGAAGCAAAGAAAACACC-3’, U1057: 5’-GGATCCATATGCTAACCAATTCCTCCCATC-3’ /U879: 5’-GGATCCATATGCTAACCAATTCCTCCCATC-3’ and U785: 5’-ATTGGAAACGGATATGCTCCA-3’ / U786: 5’-TCCTTACCTGAACGCCTGTCA-3’, respectively. Total RNA prepared from stems of five independent lines of in vitro cultured plants. For a quantitative comparison of the transcript abundance, the number of PCR reaction cycles

were optimized to ensure both clear visualization of the amplified products on an agarose gel and amplification in the exponential phase. The optimal cycles were 25 for PGA1, 25 for PGA2 and 20 for EF1α. Plants were grown in a greenhouse under long-day conditions to analyze their flowering and tuberization.

LC-MS analysis of SGAs in transgenic potato plants

Extractions and LC-MS analyses of the plant materials were performed with the similar method as previously described (Sasaki 2011). Fresh plant materials (100 mg) were homogenized with a mixer mill at 4°C in a 1 mL solution containing 80% (v/v) methanol and 0.1% (v/v) formic acid. For analyses of the levels of α-solanine and α-chaconine in the stems of in vitro-grown *PGA1*- and *PGA2*-silenced plants, 10 mg brassinolide was added as an internal standard. After centrifugation, 25 mL of supernatant was diluted with 475 mL 0.1% (v/v) formic acid solution and filtered with a MultiScreen Solvinert (Millipore). An aliquot (10 mL) was analyzed by liquid chromatography-mass spectrometry (LC-MS) using 10 mM ammonium hydrogen carbonate in water (pH 10): acetonitrile (2:3, v/v) as eluent at a flow rate of 0.2 mL min^{-1} at 40˚C. LC-MS was performed with a Shimadzu LCMS-2010EV apparatus operating in ESI mode attached to an XBridge Shield RP18-5 column (150 mm × 2.1 mm i.d.; Waters). Quantifications of α-solanine and α-chaconine were calculated from the ratio of peak area at m/z 868 and 852 from positive ion scans using a calibration curve of authentic samples (with both coefficients of determination: $r^2 > 0.999$), respectively.

GC-MS analysis of metabolites in transgenic potato plants

GC-MS analyses were conducted with the same conditions as described before (Seki et al., 2008). The steroidal compounds that were fluctuated in the transgenic potato plants were analyzed. Peaks were identified by comparing the retention times and mass spectra with those of the authentic standards.

Results

Identification of candidate PGA1 and PGA2 of potato

Some cytochrome P450 monooxygenases (CYPs) are associated with steroid oxidization (Ohnishi et al. 2009). In order to identify the candidate genes involved in the oxidative modification of cholesterol for SGA biosynthesis, *CYP* superfamily, which are present in potato EST databases of the Gene Index project (http://compbio.dfci.harvard.edu/tgi/), were surveyed. Several *CYP*s were selected based on the correlation between the read numbers of the EST contigs and the tissue specific accumulation of SGAs in potato. Two CYP candidates, designated as *PGA1* and *PGA2*, were identified among 400 CYP genes in potato genome. The EST databases show that *PGA1* and *PGA2* are highly expressed in flowers and tuber sprouts, where the high levels of SGAs are accumulated (Kozukue and Mizuno 1985, 1989; Ginzberg et al. 2009). These results suggest *PGA1* and *PGA2* as candidate genes involved in oxidation of cholesterol for SGA biosynthesis. PGA1 and PGA2, encoding CYP72A208 and CYP72A188, respectively (http://drnelson.uthsc.edu/cytochromeP450.html), share 52% amino acid sequence identity and have about 45% amino acid identity to licorice CYP72A154 that is associated with the biosynthesis of glycyrrhizin, a triterpenoid saponin (Seki et al. 2011). CYP72A208 and CYP72A188, designated as PGA1 and PGA2, are located on chromosomes 6 (PGSC0003DMG400026586 and PGSC0003DMG400026596) and on chromosomes 7 (PGSC0003DMG400011750) of *S. tuberosum* Group Phureja DM1-3 (Potato Genome Sequencing Consortium 2011), respectively.

In vitro functional analysis of the recombinant PGA1 and PGA2 proteins

To investigate the catalytic functions of PGA1 and PGA2, the recombinant PGA1 and PGA2

proteins were prepared with baculovirus-mediated expression system in insect cells, and in vitro enzyme assay was performed with cholesterol, 22*S*-hydroxycholesterol, and 22*R*-hydroxycholesterol as substrates. The PGA1 protein metabolized 22*S*-hydroxycholesterol and 22*R*-hydroxycholesterol but not cholesterol to products with retention times at 21.3 min and 21.4 min, respectively (Figure 2-2). The former product is identical to the authentic compound (22*S*, 25*S*)-22, 26-dihydroxycholesterol in terms of the retention time and the mass spectra and the latter product is presumed to be (22*R*, 25*S*)-22, 26-dihydroxycholesterol. On the other hand, The PGA2 protein metabolized cholesterol to two products with retention times at 20.9 min and 21.2 min, which are identical to 22*S*-hydroxycholesterol and 22*R*-hydroxycholesterol, respectively (Figure 2-3). These results strongly suggest that PGA1 catalyzes the 26-hydroxylation of 22-hydroxycholesterols and PGA2 catalyzes the 22-hydroxylation of cholesterol in the SGA biosynthesis.

Construction and SGA analysis of *PGA1- and PGA2*-silenced transgenic potato plants

To confirm the contribution of *PGA1 and PGA2* to SGA biosynthesis in potato plants, potato plants were transformed with an RNA interference vector to create *PGA1*- and *PGA2*-silenced transgenic potato plants (pga1-1 and pga2) (Figure 2-1). Of the 28 *PGA1*-silenced transgenic lines, the in vitro shoots of four independent lines (#20, #35, #45 and #67 in Figure 2-4A-E) had the significantly reduced *PGA1* transcript levels than the control (Figure 2-4B), and consistently had a much lower SGA content (Figure 2-4A). Similarly, of the 27 *PGA2*-silenced transgenic lines, the in vitro shoots of four independent lines (#9, #28, #45 and #59 in Figure 2-4F-J) had the significantly reduced *PGA1* transcript levels than the control (Figure 2-4G), and consistently had a much lower SGA content (Figure 2-4F). All the silenced lines grew in a greenhouse, and the potato tubers were harvested. The SGA contents in tuber peel and tuber cortex of all the silenced lines were significantly lower than the control in both dark and light conditions (Figure 2-4C, D, H and I), although the transgenic tuber peels also accumulated

normal levels of chlorophyll and anthocyanin (Figure 2-5). Most of the *PGA1*- and *PGA2*-silenced plants had similar tuber yields as compared to the control (Figure 2-4E and J), although two lines of the *PGA2*-silenced plants had lower tuber yields than the control. To confirm the results, potato plants were constructed with the vector that used a different region of the *PGA1* gene (pga1-2) (Figure 2-1). Of the 27 *PGA1*-silenced transgenic lines (pga1-2), three independent lines also had lower levels of SGAs and similar vegetative growth as compared to the control and the *PGA1*-silenced transgenic plants (pga1-1). These results indicate the involvement of *PGA1* and *PGA2* in SGA biosynthesis in potato plants.

Metabolite analyses of *PGA1*- and *PGA2*-silenced transgenic potato plants

To examine the effects of silencing of the *PGA1* and *PGA2* expression on the endogenous metabolites, the steroidal compounds that were fluctuated in the *PGA1*- and *PGA2*-silenced plants (pga1-1 and pga2, respectively) were analyzed by GC-MS. The *PGA1*-silenced plants had four new peaks in the extracted ion chromatogram (EIC) of 173 *m/z* steroids (Figure 2-6Aa, peaks A-D), as compared to non-transgenic plants and vector control plants. The retention times and mass spectra of peak A and B were identical to those of authentic 22*S*-hydroxycholesterol and 22*R*-hydroxycholesterol, respectively (Figure 2-6Aa and 2-6Ba). From the mass spectra of peaks C and D, it was inferred that the two peaks corresponded to 22-dihydroxycholesterol derivatives which are presumed to be enantiomers of 16, 22-dihydroxycholesterols (Figure 2-6Ba and 2-7). On the other hand, the *PGA2*-silenced plants accumulated several to 10 times more cholesterol (CHR) than control plants (Peak E as shown in Figure 2-6Ab and 2-6Bb) and had a new minor peak in the extracted ion chromatogram (EIC) of 456 *m/z* steroids (Figure 2-6Ab, peaks F), as compared to non-transgenic plants and vector control plants. Therefore, it was hypothesized that PGA2, a C-22 oxygenase, oxidizes cholesterol to 22-OH-CHR and that PGA1, a C-26 oxygenase, metabolize 22-OH-CHR to 22, 26-diOH-CHR (Figure 2-8).

Phenotypes of *PGA1*- and *PGA2*-silenced transgenic potato plants

The *PGA1*- and *PGA2*- silenced transgenic plants demonstrated two distinct phenotypes. First, the *PGA1*-silenced plants (pga1-1; #20, #35, #45 and #67) and the *PGA2* -knockdown plants (pga2; #28, #45 and #59) produced abnormal flowers that were sterile (Figure 2-9A-C). Flower initiation in these plants seemed normal, but development stopped and the flowers withered. The transgenic plants produced no pollen and were, thus, infertile. The alternative series of *PGA1*-silenced lines constructed with another silencing region from the *PGA1* gene (pga1-2; #11 and #20, Figure 2-1) showed the same phenotype. The second distinctive phenotype was that the tubers of the *PGA1*- and *PGA2*-silenced transgenic plants could not sprout without treatment. Sprouting initiates normally, but the sprout clusters did not grow even after more than three months at 20 ºC or for a one year at 4 ºC after the cessation of usual dormancy in the control plants (Figure 2-9D-H). Interestingly, the sprout clusters could grow after the tubers were planted in soil (Figure 2-9G) but not in water. Two independent series of *PGA1*-silenced lines constructed with two different regions in the *PGA1* gene (pga1-1 lines and pga1-2 lines, Figure 2-1) showed the same phenotype. The reproducibility of the non-sprouting tuber phenotype was confirmed in triplicate for the *PGA1*-silenced plants (pga1-1), and twice for the *PGA2*-silenced plants (pga1-2), during different growing seasons. After placing excised sprout tips on tissue culture media, the sprouts could also grow (Figure 2-10), suggesting that the sprout clusters were still alive and had only stopped growing.

Tomato plants have a *PGA1* homolog (*PGA1H*) and a *PGA2* homolog (*PGA2H*). In this chapter, *PGA1H*- and *PGA2H*-silenced transgenic tomato plants were obtained and these plants had lower levels of the predominant tomato SGA, α-tomatine (Figure 2-11). Both of the silenced tomato lines exhibited retarded growth, dwarfing, abnormal flowers and sterility (Figure 2-11A-C). The transgenic tomato plants did not produce any fruit.

Discussion

SGAs are biosynthesized from cholesterol via oxidation, amination and glycosylation (Friedman, 2002; Ginzberg et al., 2009; Petersen et al., 1993). Recently, the cholesterol synthase gene named sterol side chain reductase 2 (SSR2) have been identified and found to be involved in SGAs biosynthesis (Sawai et al., 2014).

In this chapter, in vitro functional analysis of the recombinant proteins demonstrated that PGA1 and PGA2, potato two CYPs, catalyze the hydroxylation step at the C-26 and C-22 position, respectively (Figure 2-2 and Figure 2-3). The results were clearly supported by analysis of the metabolites that accumulated in the gene silencing transgenic plants (Figure 2-6). Since metabolites accumulated in *PGA1*- and *PGA2*-silenced plants are 22-hydroxycholesterols and 16, 22-dihydroxycholesterols (Figure 2-6Aa and Figure 2-6Ba), and cholesterol and a little 26-hydroxycholesterol (Figure 2-6Ab and Figure 2-6Bb-c), respectively, PGA1 and PGA2 catalyze early hydroxylation steps of cholesterol in the SGAs biosynthesis.

In brassinosteroid biosynthesis and catabolism, sterol C-22 oxidase and C-26 oxidase are encoded by different genes, DWF4 and BAS1 in *Arabidopsis*, which belong to CYP90B and CYP734A subfamilies, respectively (Ohnishi et al. 2009). These orthologs in potato were found as PGSC0003DMG400014902 in CYP90B and PGSC0003DMG400001060 in CYP734A, respectively, in the potato genome (Potato Genome Sequencing Consortium 2011). Using Cytochrome P450 Homepage (http://drnelson.uthsc.edu/cytochromeP450.html), PGA1 and PGA2 were found to belong to the CYP72A subfamily. The CYP72A subfamily includes some oxidases involved in the triterpenoid saponins biosynthesis; CYP72A154 from licorice (*Glycyrrhiza* plants), which oxidizes 11-oxo-β–amyrin at C-30 position, CYP72A61v2 and CYP72A68v2 from *Medicago truncatula*, which hydroxylate 24-hydroxy-β-amyrin at C-22 position and oxidize oleanolic acid at C-23 position, respectively (Seki et al., 2011; Fukushima et al., 2013).

In previous study, the transgenic potato plants overexpressing a heterologous sterol methyltransferase gene had relatively lower amounts of SGAs (Arnqvist et al. 2003). However, changing the expression of the sterol methyltransferase or some glycosyltransferases, which have been previously identified as potato SGAs biosynthetic genes (Moehs et al., 1997; McCue et al., 2005, 2006, 2007), does not effectively decrease SGA levels. Whereas, the transgenic potato plants in which either the *PGA1* or *PGA2* genes were silenced, contained significantly low amounts of SGAs (Figure 2-4A, C, D, F, H and I). Additionally, the transgenic plants grew normally in the vegetative stages and produced yields of tubers in the greenhouse similar to control plants (Figure 2-4 H and J). These results suggest that SGAs may be dispensable for the growth of potato, at least under unstressed conditions. The results are coincident with that of *SSR2*-silenced and *SSR2*-disrupted potato plants (Sawai et al., 2014). Thus, the results can provide the metabolic engineering techniques such as the generation of SGAs-free potato.

The above described reports about *SSR2*-suppressed or disrupted potato plants indicated that these plants were phenotypically identical to control plants, although the plants accumulated significantly lower SGAs contents. These findings imply that SGAs are dispensable for flowering and sprouting morphogenesis. However, the *PGA1*- and *PGA2*-silenced transgenic potato plants were sterile and the tuber sprouts did not grow (Figure 2-9A-F). For the *PGA1*-silenced plants, two independent series of silenced lines (pga1-1 and pga1-2) were constructed using different silencing regions (Figure 2-1) and both lines had the same phenotype, further supporting the hypothesis that SGAs biosynthesis are associated with the sprouting of the tuber by an unknown mechanism. Furthermore, it was demonstrated that the *PGA1H*- or *PGA2H*-silenced transgenic tomato plants had the phenotypes of retarded growth, dwarfing, abnormal flowers and sterility (Figure 2-11A-C). Tomato *GAME1* encodes tomatidine galactosyltransferase in α-tomatine biosynthesis and is the ortholog of potato *SGT1*, solanidine galactosyltransferase (Itkin et al. 2011). These more severe growth phenotypes of *PGA1H*- and *PGA2H*-silenceed tomato plants are similar to those of *GAME1*-suppressed

transgenic tomato plants. The *GAME1*-silenced plants formed small flower buds, suggesting a potential influence of the modified SGAs metabolism to the unusual phenotypes. The *PGA1*- and *PGA2*-silenced plants might accumulate morphogenesis inhibitors that can be disabled by planting tubers in soil (Figure 2-9G). The hypothesis is supported by the observation that floral and tuber sprouts tissues contain the greatest levels of SGAs (Kozukue and Mizuno 1985, 1989; Ginzberg et al. 2009).

Finally, there are no reported relationships between SGAs biosynthesis and tuber sprouting. The new phenotype indicated in this chapter is not understood completely, but offers the possibility that potato storage could be controlled without the use of postharvest chemicals by gene silencing and genome editing technologies.

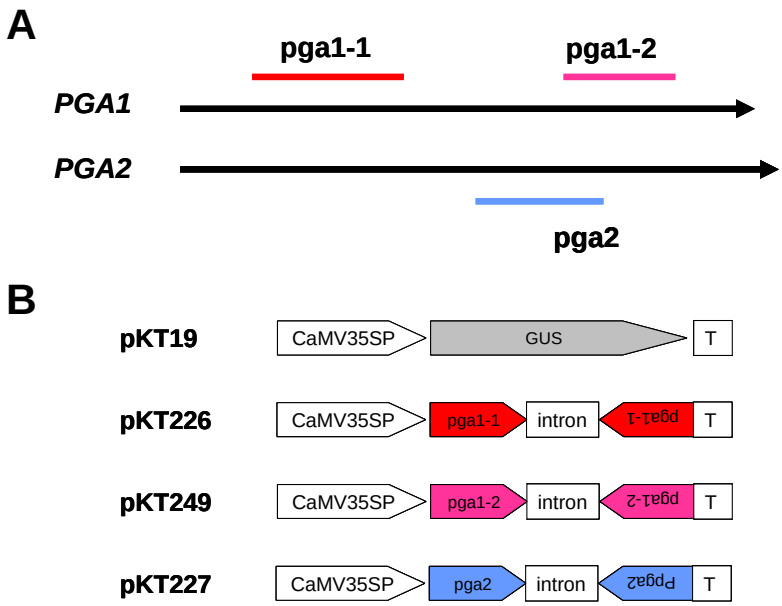


Figure 2-1. *PGA1* and *PGA2*-knockdown vectors. (A) Regions used in constructing the knockdown vectors (pKT226: red line, pKT249: orange line and pKT227: blue line). (B) Structure of control (pKT19) and knockdown vectors and the number of transgenic lines obtained. CaMV35SP, Cauliflower mosaic virus 35S promoter; GUS, β-glucuronidase gene; intron, third intron of the Arabidopsis *phytoene desaturase* gene (*PDS*, At4g14210); T, terminator.

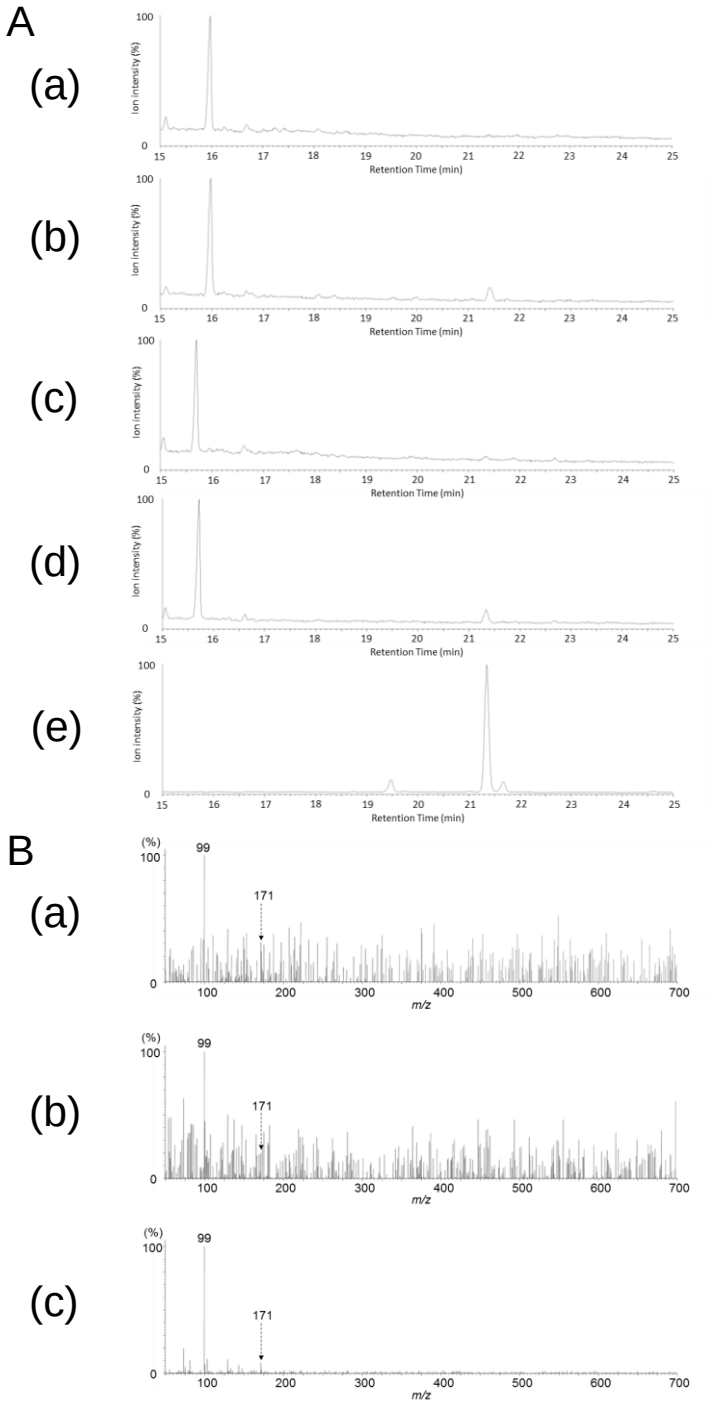


Figure 2-2. GC-MS analysis of the reaction products from the recombinant proteins with 22-hydroxycholesterols as substrate. (A) The extracted ion chromatogram (EIC) of *m/z* 99 of the authentic compounds and the reaction products from the recombinants. (a) the reaction products for (22*R*)-22-hydroxycholesterol with the empty vector; (b) the reaction products (22*R*)-22-hydroxycholesterol with the PGA1; (c) the reaction products for (22*S*)-22-hydroxycholesterol with the empty vector; (d) the reaction products (22*S*)-22-hydroxycholesterol with the PGA1; (e) the authentic compound of (22*S*, 25*S*)-22, 26-dihydroxycholesterol. (B) (a) Mass spectra of the product peak with a retention time at 21.4 min from the PGA1 for (22*R*)-22-hydroxycholesterol; (b) mass spectra of the product peak with a retention time at 21.3 min from the PGA1 for (22*S*)-22-hydroxycholesterol; (c) mass spectra of the peak with a retention time at 21.3 min from the authentic compound of (22*S*, 25*S*)-22, 26-dihydroxycholesterol

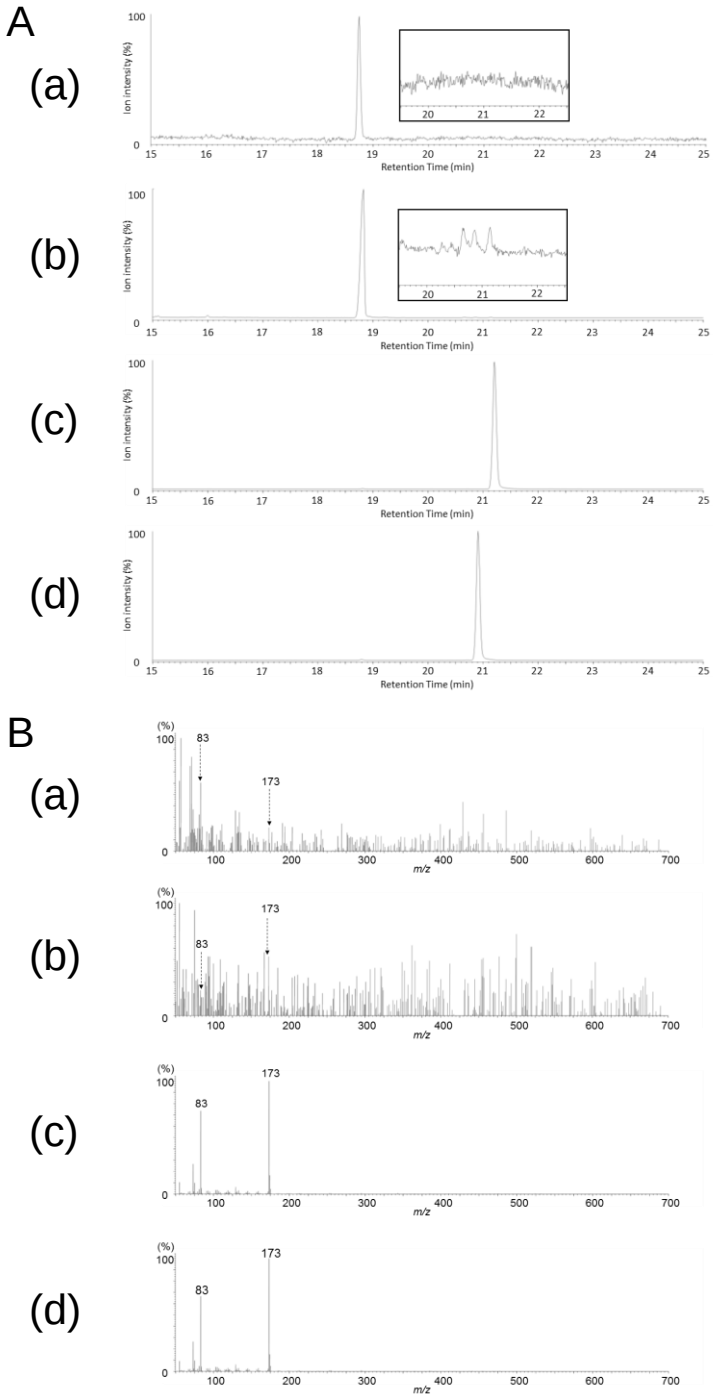


Figure 2-3. GC-MS analysis of the reaction products from the recombinant proteins with cholesterols as substrate. (A) The extracted ion chromatogram (EIC) of *m/z* 173 of the authentic compounds and the reaction products from the recombinants. (a) the reaction products with the empty vector; (b) the reaction products with the PGA2; (c) the authentic compound of (22*R*)-22-hydroxycholesterol; (d) the authentic compound of (22*S*)-22-hydroxycholesterol. (B) (a) Mass spectra of the product peak with a retention time at 20.9 min from the PGA2; (b) mass spectra of the product peak with a retention time at 21.2 min from the PGA2; (c) mass spectra of the peak with a retention time at 21.2 min from the authentic compound of (22*R*)-22-hydroxycholesterol; (c) mass spectra of the peak with a retention time at 21.9 min from the authentic compound of (22*S*)-22-hydroxycholesterol

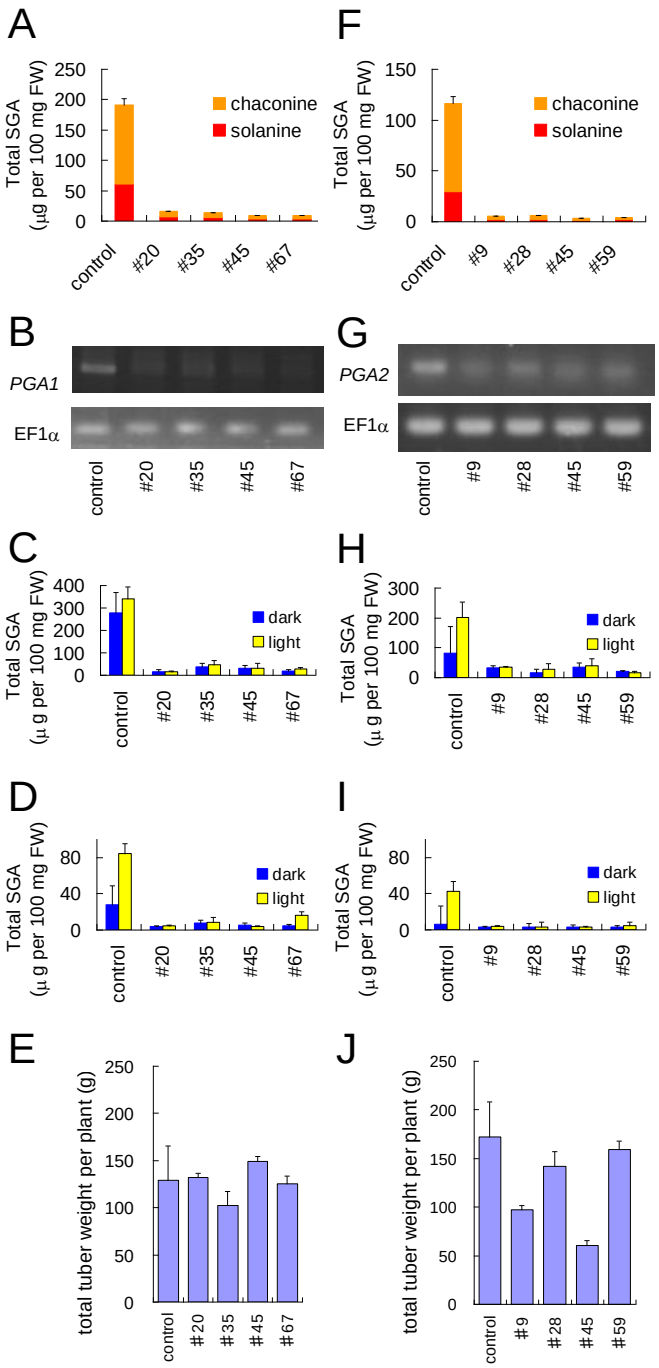


Figure 2-4. SGAs Content and yield of the *PGA1*- and *PGA2*-silenced transgenic potato plants. (A and F) LC-MS analysis of SGAs content α-solanine and α-chaconine) levels in the stems of in vitro-grown shoots *PGA1*- (A) and *PGA2*- (F) silenced plants; (B) Semi-quantification RT-PCR analysis of *PGA1* in the in vitro-grown shoots of independent *PGA1*-silenced lines; (C and D) LC-MS analysis of the SGAs levels of the *PGA1*-silenced plants in the peel (C) and cortex (D) of harvested tubers with/without light exposure; (G) Semi-quantification RT-PCR analysis of *PGA2* in the in vitro-grown shoots of independent *PGA2*-silenced lines; (H and I) LC-MS analysis of the SGAs levels of the *PGA2*-silenced plants in the peel (H) and cortex (I) of harvested tubers with/without light exposure; (E and J) Yields of the tubers from *PGA1*(E)- and *PGA2*(J)-knockdown plants. Bars indicate standard deviation from the mean (n=3). FW, fresh weight.

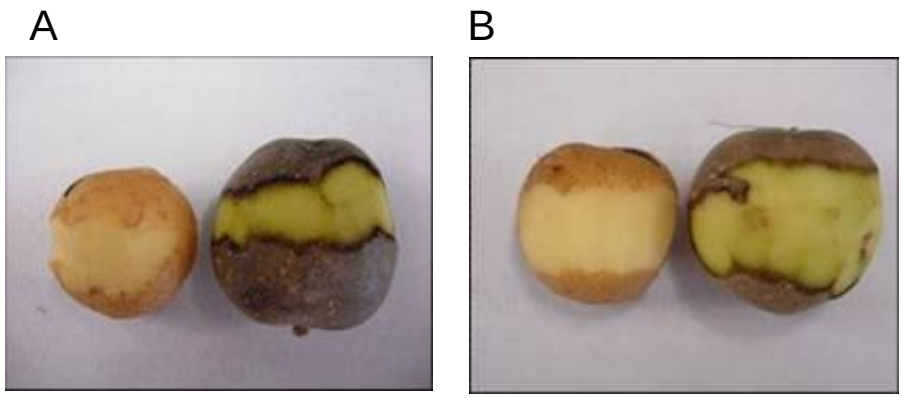


Figure 2-5. Tubers of *PGA1*-silenced transgenic potato plants with light exposure. (A and B) Partly peeled tubers of control (A) and the *PGA1*-silenced plant pKT226-#67 (B) after dark exposure (left) and after light exposure (right). Tubers of both the control and *PGA1*-silenced plants accumulated chlorophyll and anthocyanin after light exposure and their peels darkened.

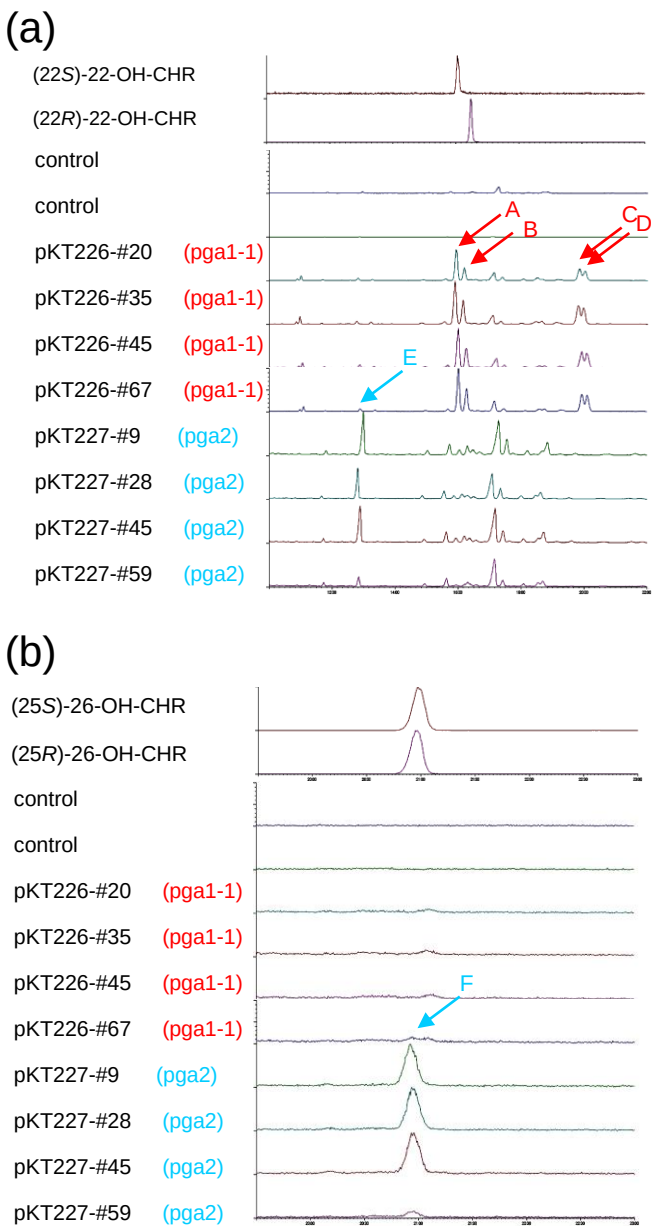


Figure 2-6A. EIC obtained by GC-MS analysis of the authentic compounds and accumulated compounds in the non-transgenic plants, *PGA1*- and *PGA2*-silenced transgenic potato plants. (a) EIC (*m/z* 173) of the authentic compounds of 22-hydroxycholesterols and accumulated compounds in the independent lines; (b) EIC (*m/z* 456) of the authentic compounds of 26-hydroxycholesterols and accumulated compounds in the independent lines. Peaks A-D present in accumulated compounds in *PGA1*-silenced plants were indicated with solid red arrows. Peaks E and F present in accumulated compounds in *PGA2*-silenced plants were indicated with solid light blue arrows. Mass spectra of the peaks are shown in Figure 2-6B.

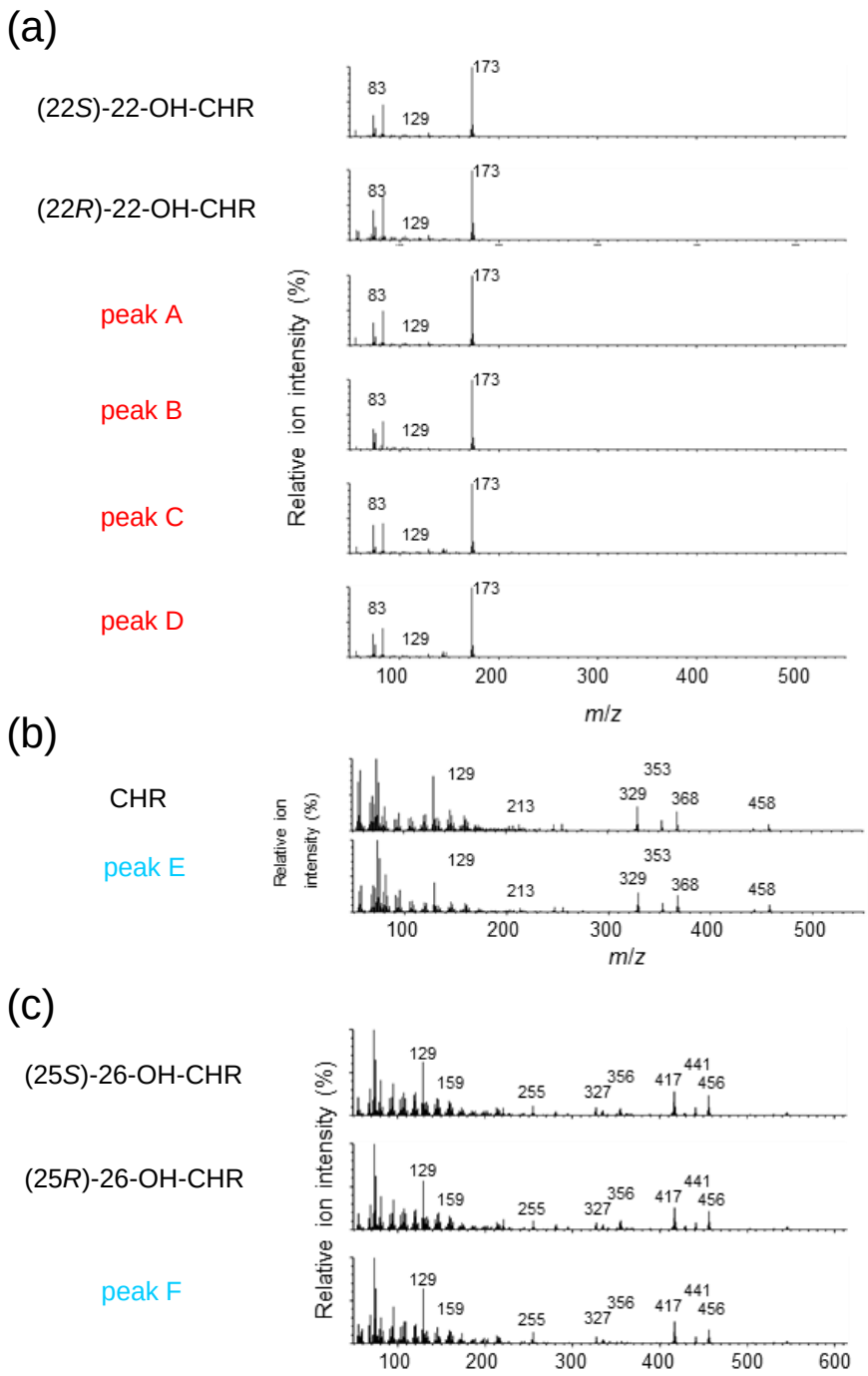


Figure 2-6B. Mass spectra of accumulated compounds in *PGA1*- and *PGA2*-silenced transgenic potato plants. (a) The mass spectra of peaks A-D in Figure 2-6A and authentic compounds of (22*S*)-22-hydroxycholesterol (22*S*)-22-OH-CHR) and (22*R*)-22-hydroxycholesterol (22*R*)-22-OH-CHR). (B) The mass spectra of peak E in Figure 2-6A and authentic compound of cholesterol (CHR). (C) The mass spectra of peak F in Figure 2-6A and authentic compounds of (25*S*)-26-hydroxycholesterol ((25*S*)-26-OH-CHR) and (25*R*),26-hydroxycholesterol ((25*R*)-26-OH-CHR).

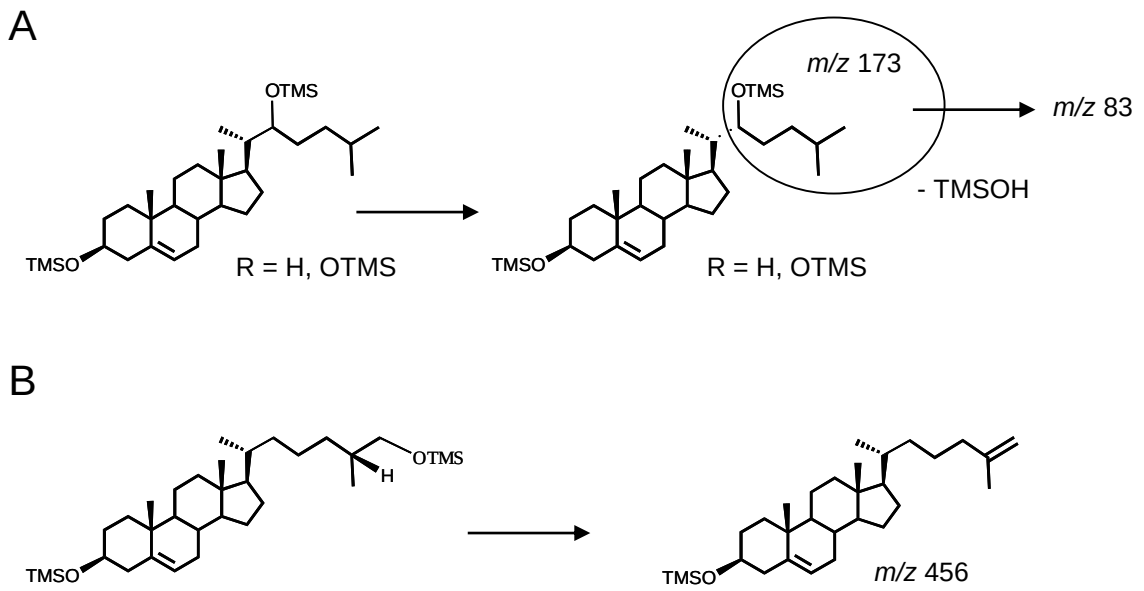


Figure 2-7. Proposed assignment of fragment ions of hydroxycholesterol derivatives. (A) The mass spectrum of trimethylsilylated 22-hydroxycholesterol and 22-hydroxycholesterol derivatives such as 16, 22-dihydroxycholesterols show characteristic fragment ions at *m/z* 173 and 83, although their authentic standards were not available. (B) The mass spectrum of trimethylsilylated 26-hydroxycholesterols showed a characteristic fragment ion at *m/z* 456.

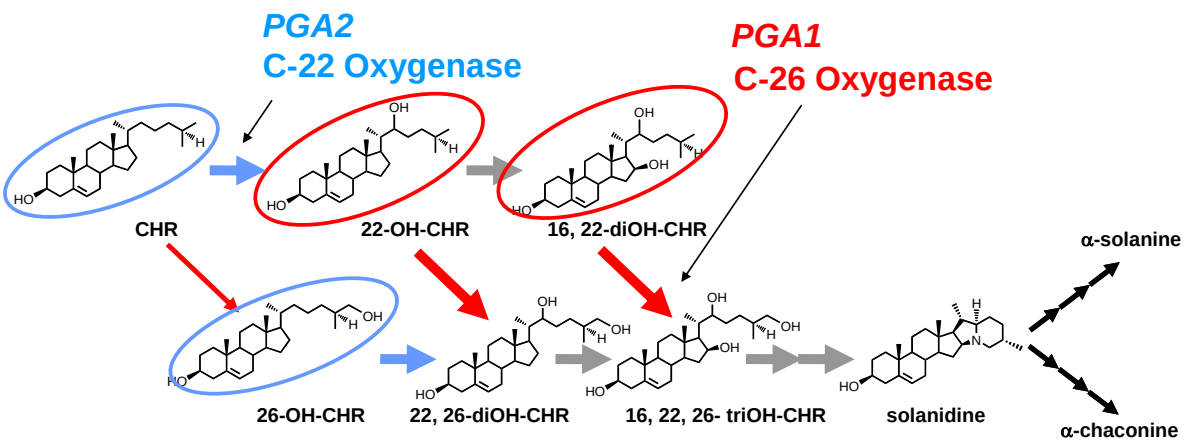


Figure 2-8. Proposed SGAs biosynthetic pathway catalyzed by PGA1 and PGA2. A summary of the results obtained in this chapter. Compounds that accumulated in the *PGA1*- and *PGA2*-silenced transgenic plants are indicated with red and light blue circles, respectively. Metabolic flow of PGA1, PGA2 and other metabolites are indicated with red, light blue and black arrows, respectively; cholesterol (CHR); hydroxycholesterol (OH-CHR).

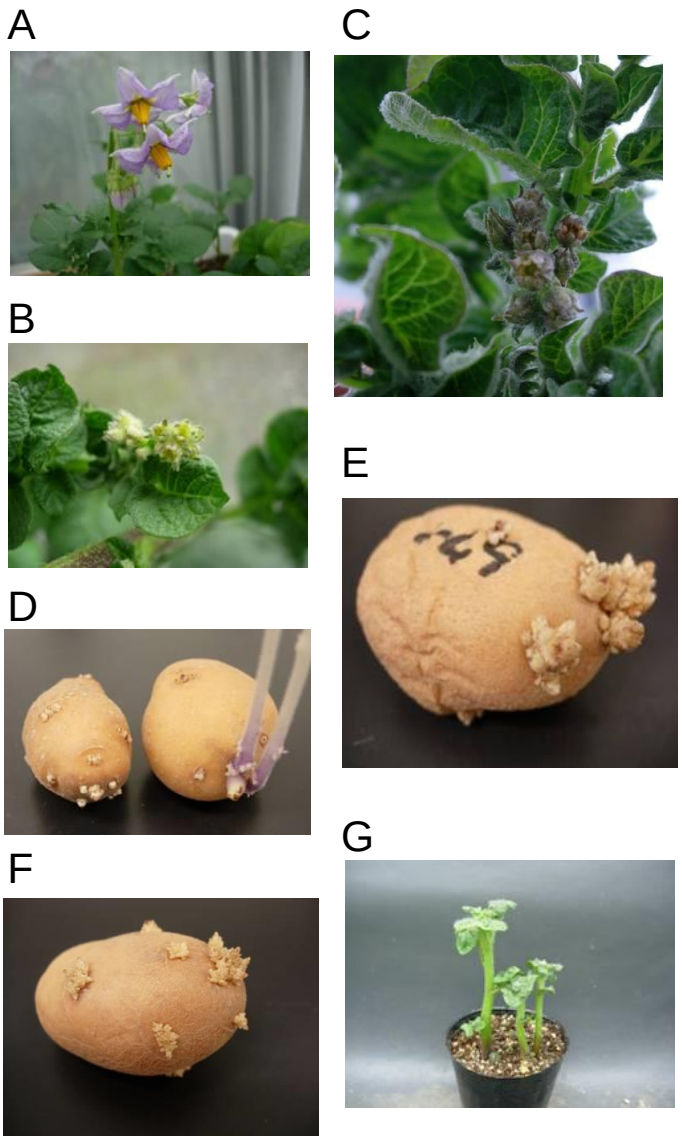


Figure 2-9. Phenotypes of the *PGA1*- and *PGA2*-silenced transgenic potato plants. (A) Flowers of the control plant; (B) Flowers of the *PGA1*-silenced plant (pKT226-#35); (C) Flowers of the *PGA2*-silenced plant (pKT227-#45); (D) Sprouted potatoes of the *PGA1*-silenced plant (pKT226-#67, left) and control (right) two weeks after the cessation of plant dormancy in the control; (E and F) Sprouted potatoes of the *PGA1*-silenced plant (pKT226-#67) and the *PGA2*-silenced plant (pKT227-#45) three months and two months after the cessation of plant dormancy in the control, respectively; (G) Sprouted potatoes of the *PGA1*-knockdown plant (pKT226-#67) three weeks after being planted into soil.

Figure 2-10. In vitro growth of sprout tips on tissue culture media. Sprout tips cut from tubers of the *PGA1*-silenced plants pKT226-#67 were placed on tissue culture media without plant hormones. Plantlets one week (A), two weeks (B) and six weeks (C) after transfer to media.

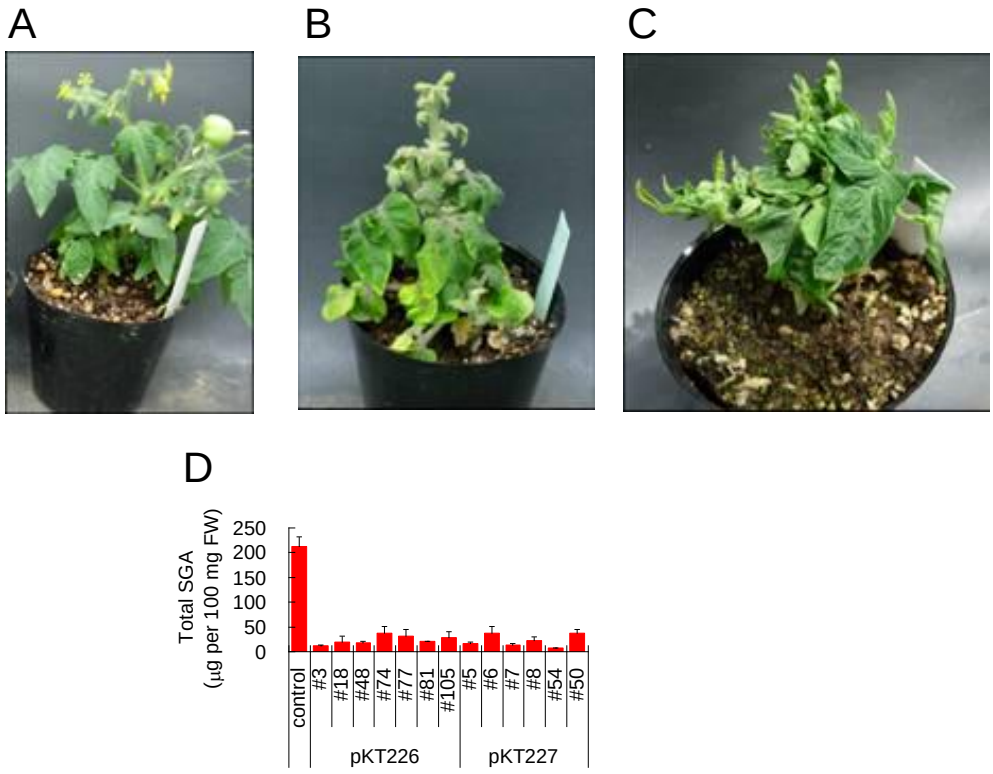


Figure 2-11. Phenotype and SGAs content of the *PGA1*- and *PGA2*- homolog (*PGA1H* and *PGA2H*)-silenced tomato plants. (A) Control; (B) The *PGA1H*-silenced tomato plant (pKT226-#3); (C) The *PGA2H*-silenced tomato plant (pKT227-#6); (D) LC-MS analysis of α-tomatine content levels in the leaves of the *PGA1H* (pKT226)- and *PGA2H* (pKT227)-silenced tomato plants grown in the greenhouse. FW, fresh weight.

Chapter 3

Identification of a 2-oxoglutarate-dependent dioxygenase catalyzing steroid 16α-hydroxylation for steroidal glycoalkaloid biosynthesis in potato

Introduction

In chapter 2, potato two CYPs (PGA1 and PGA2) were identified as cholesterol C-26 and C-22 hydroxylase involved in the SGAs biosynthesis, respectively. However, C-16 hydroxylase could not be found from CYPs as the candidate oxidases. Then, to further investigate SGAs biosynthesis, this chapter was focused on 2–oxoglutarate-dependent dioxygenases (2OGDs) superfamily as the additional candidate genes.

2OGDs are also involved in various oxidation reactions as well as CYPs. 2OGDs are non-heme iron-containing proteins that localize in the cytosol as soluble proteins while CYPs which are heme-thiolate membrane proteins, generally bound to the cytoplasmic surface of the endoplasmic reticulum. However, no 2OGDs involved in the biosynthesis and catabolism of triterpenoids and steroids, have been identified.

In this chapter, a 2OGD gene, named 16DOX, highly expressing in the SGAs-rich organs of potato and tomato by using the public EST databases, was selected. Previously, this gene was reported as pistil expressed dioxygenase from tomato named TPP1 (Milligan and Gasser, 1995), as a gibberellin down-regulated partial cDNA clone named GAD2 isolated from a tomato leaf cDNA library (Jacobsen and Olszewski, 1996), and as a developmentally regulated pistil dioxygenase, induced by pollination, wounding and jasmonate treatments, from *S. chacoense* named SPP2 (Sylviane, 1999). Would-induced expression of this gene was suppressed by hyphal wall components elicitor prepared from the mycelia of *Phytophthora*

infestans, the late blight pathogen (Nakane, 2003).

Here, 16DOX gene was co-expressed with the previously identified SGA biosynthetic genes in potato and tomato. This gene encodes enzyme catalyzing hydroxylation step of cholesterol at C-16 position. 16DOX-silencing transgenic plants showed dramatic reduction in SGAs composition and the suppression of potato tuber sprouting similar to *PGA1*- and *PGA2*-silencing plants. *16DOX* is first identified as the 2OGD gene involved in the oxidation of triterpenoids and steroids.

Materials and Methods

Chemicals

Authentic samples of α-solanine, α-chaconine, α-tomatine and the two 22-hydroxycholesterols were purchased from Sigma-Aldrich. Cholesterol was purchased from Tama Biochemical Co. Authentic compounds of the two 26-hydroxycholesterols, the two (16, 22*S*)-16, 22-dihydroxycholesterol, (22*S*, 25*S*)-22, 26-dihydroxycholesterol, (22*R*, 25*S*)-22-hydroxy, 26-oxocholesterol, and (22*S*, 25*R*)-22-hydroxy, 26-oxocholesterol were kindly provided by Bunta Watanabe. (15, 15, 16, 17,-$^{2}H_{4}$)cholesterols and (15, 15, 17,-$^{2}H_{3}$)cholesterol were synthesized by Kiyoshi Ohyama.

RNA extraction and reverse transcription

Total RNA of potato and tomato was prepared from leaves, flowers, tuber skins, stems, roots, stolons and sprouts of *Solanum tuberosum* cv Sassy and leaves, flowers, mature green fruits, yellow fruits, orange fruits and red fruits of *S. lycopersicum* cv Micro-Tom, respectively, using the RNAeasy plant mini-kit (QIAGEN, Hilden, Germany) and RNase-Free DNase Set (QIAGEN). The total RNA was used to synthesize the first strand cDNA of potato and tomato

using SuperScript First-Strand Synthesis System for RT-PCR (Life Technologies) and Transcriptor First Strand cDNA Synthesis Kit (TOYOBO), respectively.

Cloning of *16DOX* cDNAs.

The full open reading frames of *16DOX* genes were PCR amplified with primer sets designed from the potato and tomato unigene sequences (Sotub07g016570 in the Potato ITAG protein database and Solyc07g043420 in the Tomato ITAG protein database, respectively); U997: 5'-CACCATGGCGGAGCTTCTTTCAAAC-3' / U998: 5'-TTAAGCATCGATTTTGAAGGGC-3' and Sl16DOX_Fw: 5'- CACCATGGCGGATCTTCTCTCGAAC-3' / Sl16DOX_Rv: 5'-TTAATTAGCTTCTGTTTTGAAG-3', respectively. The PCR products of *St16DOX* and *Sl16DOX* were cloned into the pENTR/D-TOPO plasmid (Life technologies).

Expression of recombinant St16DOX in *E. coli*

The coding sequences for the *16DOX* genes were amplified from the pENTR/D-TOPO plasmid using primer set containing restriction sites (underlined); BamHI-NdeI-St16DOX Fw: 5'-GGATCCCATATGGCGGAGCTTCTTTCAAACTGG-3' / SalI-St16DOX Rv: 5'-GTCGACTTAAGCATCGATTTTGAAGGGCTC-3' and BamHI-NdeI-Sl16DOX Fw: 5'-GGATCCCATATGGCGGATCTTCTCTCGAACTGG-3' / SalI-Sl16DOX Rv: 5'-GTCGACTTAATTAGCTTCTGTTTTGAAGGGCTC-3', respectively. The amplified DNA fragments were ligated into the pMD19 vector (TaKaRa) and digested with BamHI and SalI. The DNA fragments were ligated into BamHI-SalI sites of the pGEX4T-1. *E. coli* strain BLR (DE3) transformed with constructed plasmid was grown at 37°C in LB medium with 50 μg ml^{-1} ampicillin until its OD_{600} reached appropriate 0.5. The recombinant protein expression was induced by adding IPTG to 0.1 mM and continued for 20 h at 18°C. The culture was then centrifuged at 3500 rpm for 30 min at 4°C. The cell pellets were resuspended in 5 ml of cold sonication buffer containing 50 mM sodium phosphate (pH 7.4), 300 mM NaCl and 20% (v/v)

glycerol, sonicated using a Bandelin Sonopuls HD 2070 ultrasonic homogenizer (Sigma) typeMS73 at a sound intensity of 200 W cm^{-2}, three times for 30 sec each on ice, and centrifuged at 15,000 rpm for 10 min at 4°C. The GST-tagged proteins present in the supernatant were purified using GST Spin Trap columns (GE Healthcare) according to the manufacturer's instructions. After two column washes, the adsorbed proteins were eluted twice in 200 µl of a solution, containing 50 mM Tris-HCl (pH 8.0), 20 mM reduced glutathione and 20% (v/v) glycerol and mixed. The concentration of the purified proteins were determined by Bradford system. The purified recombinant proteins were visualized by SDS–PAGE. The proteins were revealed by staining the gel using Coomassie brilliant blue R-250. The proteins were used for further analyses.

In vitro enzyme activity assay

The reaction mixture (100 µl) consisted of 100 mM Bistris-HCl (pH 7.2), 5 mM disodium 2-oxoglutarate, 10 mM sodium ascorbate, 0.2 mM $FeSO_4$, 25 µM cholesterol or hydroxycholesterol as substrate, and the purified recombinant 16DOX protein. The reaction was initiated by addition of 16DOX and was carried out at 30˚C for 3 h. The reaction was stopped after by the addition to 100 µl of ethyl acetate, followed by the addition of 0.2 µg of 25-hydroxycholesterol in 100% ethanol as an internal standard. The reaction products were extracted three times with an equal volume of ethyl acetate. The organic phase was collected and evaporated. The residue was trimethylsilylated with TMS-HT (=HMDS and TMCS in Anhydrous Pyridine) [for Gas Chromatography] (Tokyo Chemical Industry Co., Ltd.) at 80 ºC for 45 min. GC-MS analyses of the 16DOX reaction products were performed with the similar method as previously described (Seki et al., 2008). GC-MS was conducted using a GC-MS-QP2010 Ultra (Shimadzu) with a DB-5MS (30 m × 0.25 mm, 0.25 µm film thickness; J&W Scientific) capillary column to analyze the 16DOX reaction product. The injection temperature was 250 ºC. The column temperature program for analysis of the 16DOX reaction products was

as follows: 80 ºC for 1 min, followed by a rise to 300 ºC at a rate of 20 ºC min^{-1}, and a hold at 300 ºC for 20 min. The carrier gas was He, and the flow rate was 1.0 ml min^{-1}; the interface temperature was 300 ºC, with a splitless injection.

Biochemical analysis of recombinant St16DOX

I determined the kinetics parameters of recombinant St16DOX in triplicated assays. The activity was assayed using (22*S*, 25*S*)-22, 26-dihydroxycholesterol at concentration ranging from 1 to 50 μM. Reaction The reaction was carried out at 30°C for 30 minutes. Extraction and GC-MS analysis of the reaction product were performed as described above. Kinetic parameters were determined by non-linear regression with ANEMONA (Hernandez and Ruiz, 1998). The substrate specificity of St16DOX toward cholesterol and several oxygenated cholesterols was determined. The activity for each substrate was assayed using cholesterol, the two 22-hydroxycholesterols, 22-oxocholesterol, the two 26-hydroxycholesterols, (22*S*, 25*S*)-22, 26-dihydroxycholesterol, (22*R*, 25*S*)-22-hydroxy, 26-oxocholesterol and (22*S*, 25*R*)-22-hydroxy, 26-oxocholesterol at 25 μM. The reaction was carried out at 30°C for 3 h. The reaction product for (22*S*, 25*S*)-22, 26-dihydroxycholesterol was also analyzed by LC-MS. The product was extracted and evaporated as described above, and the residue was dissolved in 200 μl of ethanol. LC-MS analysis was performed using a system consisting of an ACQUITY UPLC H-Class (Waters, Milford, MA) and an SQ Detector 2 (Waters), and data acquisition and analysis were performed using MassLynx 4.1 software (Waters). Chromatographic column was an ACQUITY UPLC BEH C-18 column (50×2.1 mm, 1.7 μm, Waters). The column temperature was set at 40°C. For each sample, 5 μL was injected. The flow rate was set at 0.2 ml min^{-1}.The mobile phases were water with 0.1% (v/v) formic acid (A) and acetonitrile (B), using a gradient elution of 10-90% B at 0-30 min, 90-100% B at 30-40 min and 100% B at 40-46 min (0-30 min and 30-40 min, linear gradient). The mass spectra were obtained in positive ESI mode. In ESI conditions, the capillary voltage at 3 kV and sample cone voltage at 60 V were applied. MS

scan mode with a mass range of *m/z* 250-1400 was used.

Generation of transformation vectors, plant transformation and growth conditions

A 370 bp fragment of *St16DOX* cDNA was PCR amplified using primer set containing restriction sites (underlined); SacI-U1003: 5'-GAGCTCTAGATTTGGGAAAAGCTAATGGT-3' and BamHI-U1004: 5'-GGATCCATATGCACCAATAACCTCC-3'. An RNAi binary vector pKT258 targeting the *St16DOX* gene was constructed from the binary vector pKT11 (Umemoto et al., 2001) by locating two *St16DOX* fragments in opposite interposing the third intron of the Arabidopsis thaliana *At4g14210* gene under the control of the cauliflower mosaic virus 35S (CaMV35S) promoter in the T-DNA region (Figure 3-1). The RNAi binary vector pKT258 was transformed into *Agrobacterium tumefaciens* GV3110. Potatoes (*S. tuberosum* cv Sassy) were transformed using *Agrobacterium* GV3110 cells with pKT258 as previously reported (Momma, 1990). In vitro-grown plants were cultured at 20°C under a 16-hlight/ 8-h-dark cycle. 41 transformants were individually selected by genomic PCR of the shoots with the primer set; NP2: 5'-TAAAGCACGAGGAAGCGGT-3' and NP3: 5'-GCACAACAGACAATCGGCT-3' targeting the kanamycin resistance gene on the T-DNA region integrated into the potato genome. RT-PCR analysis of *St16DOX* was performed using the primer sets; U1060: 5'-TGGTTTACCAATTACGGTAGTTAGCA-3' and U998: 5'-TTAAGCATCGATTTTGAAGGGC-3' and total RNA prepared from stems of five independent lines of in vitro-cultured plants, #5, #16, #28, #39 and #41. As a control, the potato *elongation factor 1α* gene (*EF1α*) and the primer set; 5'-ATTGGAAACGGATATGCTCCA-3' and 5'-TCCTTACCTGAACGCCTGTCA-3', were used (Nicot et al., 2005).

LC-MS analysis of SGA in *St16DOX*-silenced transgenic potato plants

Extractions and LC-MS analyses of the plant materials were performed with the similar method

as previously described (Ohyama et al., 2013). Fresh plant materials (100 mg) were homogenized with a mixer mill at 4°C in a 1 mL solution containing 80% (v/v) methanol and 0.1% (v/v) formic acid. For analyses of the levels of α-chaconine and α-solanine in the stems of in vitro-grown *St16DOX*-silenced plants, 10 mg brassinolide was added as an internal standard. After centrifugation, 25 mL of supernatant was diluted with 475 mL 0.1% (v/v) formic acid solution and filtered with a MultiScreen Solvinert (Millipore). An aliquot (10 mL) was analyzed by liquid chromatography-mass spectrometry (LC-MS) using 10 mM ammonium hydrogen carbonate in water (pH 10): acetonitrile (2:3, v/v) as eluent at a flow rate of 0.2 mL min^{-1} at 40°C. LC-MS was performed with a Shimadzu LCMS-2010EV apparatus operating in ESI mode attached to an XBridge Shield RP18-5 column (150 mm × 2.1 mm i.d.; Waters). Quantifications of α-solanine and α-chaconine were calculated from the ratio of peak area at *m/z* 868 and 852 from positive ion scans using a calibration curve of authentic samples (with both coefficients of determination: r2 > 0.999), respectively.

GC-MS analysis of metabolites in *St16DOX*-silenced transgenic potato plants

GC-MS analyses were conducted with the same conditions as described before (Seki et al., 2008). The steroidal compounds that were fluctuated in the *St16DOX*-silenced plants were analyzed. Peaks were identified by comparing the retention times and mass spectra with those of the authentic standards.

Real-Time quantitative RT-PCR analysis

Quantitative RT-PCR was performed with LightCycler®Nano (Roche) using THUNDERBIRDTM SYBR® qPCR Mix (TOYOBO) with the following primer sets; 16DOX qPCR Fw: 5’-GGGGTGCAGCTACACTGTACAGTAG-3’/ 16DOX qPCR Rv: 5’-CACCAATAACCTCCCTATATCTTGGAGGGT-3’, PGA1 qPCR Fw: 5’-CATTATGGTGGTGGCCAAAGATGATCG-3’/ PGA1 qPCR Rv: 5’-

CCCAAGGAACGATATCATGATTTACTAACGGTG-3', PGA2 qPCR Fw: 5'-TGTATGGTGGCGTCCCAAAACAGTAG-3'/ PGA2 qPCR Rv: 5'-TGTGAAATCGTGGTGCAATGGCATG-3', SGT1 qPCR Fw: 5'-TTCTTTTCCTTCCCTTCTTATCCGCTGG-3'/ SGT1 qPCR Rv: 5'-GGGAAATCCGGAAATTCGAACATCATCG-3', SGT3 qPCR Fw: 5'-TGTGTTCATACCATACGCCATGACGAG-3'/ SGT3 qPCR Rv: 5'-CTGCCCGAAAAGAGACGGTCTCT-3' and EF1α primer set as described above using cDNA of potato various tissues as templates, and 16DOX_qPCR primer set, PGA1H qPCR Fw: 5'-GGAAGGTATTCATGGTCAGCCGTACC-3'/ PGA1H qPCR Rv: 5'-TCCAGCCCACATCACAAATAATCTCTCG-3', PGA2H qPCR Fw: 5'-CATGCCATTGCACCACGATTTCACA-3'/ PGA2H qPCR Rv: 5'-GCACTGATTTTTGGTTTTCTGAACTCACCAG-3', GAME1 qPCR Fw: 5'-GGTGTTAAAGCCACAATCCTCACTACC-3'/ GAME1 qPCR Rv: 5'-GGCATTTCAGGTGAAGTGGCAGAG-3' and Ubiquitin qPCR Fw: 5'-CACCAAGCCAAAGAAGATCAAGC-3'/ Ubiquitin qPCR Rv: 5'-TCAGCATTAGGGCAVTCCTTACG-3' using cDNA of tomato various tissues as templates. Cycling was undertaken for 10 min at 95˚C, 45 cycles of 10 sec at 95°C, 10 sec at 60°C, and 15 sec at 72°C for amplification, followed by holding for 30 sec at 95°C and ramping up from 60°C to 95°C at 0.1°C sec^{-1} for melting curve analysis. Three biological repeats were analyzed in duplicate. The genes expression levels were normalized against the values obtained for the *EF1α* and *Ubiquitin* genes, which were used as an internal reference in potato and tomato, respectively. Data acquisition and analysis were performed using LightCycler®Nano software (Roche).

Tracer experiments

Tracer experiments were performed according to a reported method (Ohyama et al., 2013).

Seedlings of tomato, *S. lycopersicum* cultivar Micro-Tom, and shoots of potato, *S. tuberosum* cultivar Sassy, for tracer experiments were prepared and fed with stable isotope labeled compounds, such as (15, 15, 16, 17,-$^{2}H_{4}$)cholesterols and (15, 15, 17,-$^{2}H_{3}$)cholesterol. The accumulated SGAs in the harvested tomato seedlings and potato shoots were extracted and analyzed by LC-MS.

Results

Identification of candidate 16DOXs of potato and tomato

In order to identify the candidate genes involved in the oxidative modification of cholesterol for SGA biosynthesis, several enzyme superfamilies were surveyed, which are present in the EST databases of potato from the DFCI Plant Gene Indices (http://compbio.dfci.harvard.edu/tgi/plant.html) and the tomato databases from MiBASE (http://www.pgb.kazusa.or.jp/mibase/) and Sol Genomics (http://solgenomics.net). In Chapter 2, several *CYP*s were first selected based on the correlation between the read numbers of the EST contigs and the tissue specific accumulation of SGAs in potato and tomato. Then two CYPs of potato, PGA1 (Sotub06g021140, CYP72A208) and PGA2 (Sotub07g016580, CYP72A188), have been identified as steroid 26-hyroxylase and steroid 22-hydroxylase, respectively, for potato SGA biosynthesis. However, a 16-hydroxylase could not be identified among the CYP superfamily. The 2OGD superfamily is also associated with the oxygenation reactions of various secondary metabolites (Kawai et al., 2014), and the unigenes encoding 2OGDs were next surveyed from the potato database as the second focus. A 2OGD candidate, designated as St16DOX of potato, was identified among 150 2OGD genes in potato genome. qRT-PCR analysis showed that potato *St16DOX* was highly expressed in the tuber sprouts (Figure 3-2A), where the high levels of SGAs are accumulated. BLAST searches against the

genome databases of *Solanum* species (Sol Genomics Network: http://solgenomics.net) showed that St16DOX is identical to Sotub07g016570 in the Potato ITAG protein database and shows 93% identity to tomato Solyc07g043420 (designated as Sl16DOX) in the Tomato ITAG protein database. Tomato *Sl16DOX* was highly expressed in the flowers, which accumulate the high amount of α-tomatine (Figure 3-2B). These results suggested 16DOX genes as a candidate for 16-hydroxylase in SGA biosynthesis. Based on a recent phylogenetic classification of plant 2OGD superfamily (Kawai et al., 2014), 16DOXs were found to belong to clade DOXC41. The DOXC41 clade includes hyoscyamine 6β-hydroxylase (H6H) in the biosynthesis of a tropane alkaloid (Matsuda et al., 1991), scopolamine, and the 16DOXs share around 44% amino acid sequence identity to H6H in *Hyoscyamus niger*. The deduced 16DOX contained several sequence motifs that are highly conserved among 2OGDs (Figure 3-3). The St16DOX and Sl16DOX proteins contain a Fe(II)-binding motif His–X–Asp–X_n–His that is conserved in the 2OGD superfamily (His-217, Asp-219 and His-272) (Bugg, 2003; Lukacˇin and Britsch, 1997; Wilmouth et al., 2002). The Arg–X–Ser motif binding to the C-5 carboxy group of 2-oxoglutarate is also conserved in the proteins (Arg-282 and Ser-284) (Lukacˇin et al., 2000; Wilmouth et al., 2002).

In vitro functional analysis of the recombinant 16DOX proteins.

To investigate the catalytic functions of 16DOXs, the recombinant St16DOX protein was prepared with a bacterial expression system in *E.coli*, and an in vitro enzyme assay was performed with 22*S*-hydroxycholesterol as a substrate. The St16DOX protein metabolized 22*S*-hydroxycholesterol to a product with a retention time at 20.0 min (Figure 3-4). The product was identical to the authentic compound (16α, 22*S*)-16, 22-dihydroxycholesterol in terms of the retention time and the mass spectra but different from the authentic compound (16β, 22*S*)-16, 22-dihydroxycholesterol, indicating that St16DOX catalyzes the 16-hydroxylation of 22*S*-hydroxycholesterol specifically at 16α-configuration. The substrate specificity of St16DOX

toward cholesterol and several oxygenated cholesterols was next determined (Figure 3-5). St16DOX showed the highest activity for (22*S*, 25*S*)-22, 26-dihydroxycholesterol, approximately 50 times higher than that for 22*S*-hydroxycholesterol (Figure 3-6). The product from (22*S*, 25*S*)-22, 26-dihydroxycholesterol was not identical to any authentic compounds tested, however LC-MS analysis showed a fragment ion *m/z* 435 of the product was corresponding to the deduced parent ion of 16, 22, 26-trihydroxycholesterol (Figure 3-7). St16DOX could also weakly metabolize (22*R*)-22-hydroxycholesterol to a new product, but the structure of that was unknown. In contrast, the assays with the other substrates did not give any product peaks. The K_m value for 16α-hydroxylation activity of 16DOX toward (22*S*, 25*S*)-22, 26-dihydroxycholesterol was determined to be 4.19 ± 0.17 μM. These results strongly suggest that St16DOX is a 16α-hydroxylase of (22*S*, 25*S*)-22, 26-dihydroxycholesterol in the SGA biosynthesis (Figure 3-8). Sl16DOX of tomato also showed the catalytic activity and specificity consistent with St16DOX (Figure 3-4).

SGA analysis of *St16DOX*-silenced transgenic potato plants.

To confirm the contribution of *St16DOX* to SGA biosynthesis in potato plants, potato plants were transformed with an RNA interference vector to create *St16DOX*-silenced transgenic potato plants (Figure 3-1, 16dox-1). Of the 41 16DOX-silenced transgenic lines, the in vitro shoots of five independent lines (#15, #16, #28, #39 and #41 in Figure 3-9A-E) had the significantly reduced *St16DOX* transcript levels than the control (Figure 3-9A), and consistently had a much lower SGA content (Figure 3-9B). All the silenced lines grew normally in a greenhouse, and the potato tubers were harvested. The SGA contents in tuber peel and tuber cortex of the five silenced lines were significantly lower than the control (Figure 3-9C and D). Similarly, SGA contents in *Sl16DOX*-silenced transgenic tomato plants also were severely reduced (Figure 3-10). Most of the *St16DOX*-silenced plants had similar tuber yields as compared to the control (Figure 3-9E), although the #28 line had lower tuber yields than the

control. These results indicate the involvement of *St16DOX* in SGA biosynthesis in potato plants.

Metabolite analyses of St16DOX-silenced transgenic potato plants.

To examine the effects of silencing of the *St16DOX* expression on the endogenous metabolites, the steroidal compounds which were fluctuated in the *St16DOX*-silenced plants were analyzed by gas chromatography-mass spectrometry (GC-MS). The *St16DOX*-silenced plants had two new peaks in the total ion chromatogram (TIC) (Figure 3-11A, peaks A and B), as compared to non-transgenic plants and vector control plants. The retention time and mass spectrum of peak B was identical to those of authentic (22*S, 25S*)-22, 26-dihydroxycholesterol (Figure 3-11), although the retention time and the mass spectra of peaks A were not identical to the authentic compounds tested. Therefore, It was hypothesized that 16DOX, a C-16 hydroxylase, oxidizes 22, 26-diOH-CHR to 16, 22, 26-triOH-CHR.

SGA analysis of potato shoots and tomato seedlings fed with stable isotope labeled compounds

To investigate the fates of the C-16 hydrogens during SGAs biosynthesis, stable isotope labeled compounds, (15, 15, 16α, 17α-$^{2}H_4$)cholesterol and (15, 15, 16β, 17α-$^{2}H_4$)cholesterol, were fed with potato shoots and tomato seedlings. The accumulated SGAs in the harvested potato shoots and tomato seedlings were extracted and analyzed by LC-MS (Figure 3-12C-F and 3-13B-C). Non-labeled solanine and chaconine contained in potato shoots were detected by monitoring the *m/z* 868 and 852 ions $[M+H]^+$, respectively, while labeled solanine and chaconine were detected by monitoring the *m/z* 871 and 855 ions $[M+H]^+$, respectively, when both labeled compounds were administered (Figure 3-12C-F). On the other hand, non-labeled tomatine contained in tomato seedlings was detected by monitoring the *m/z* 1034 ions $[M+H]^+$, while labeled tomatine was detected by monitoring the *m/z* 1037 ions $[M+H]^+$ when (15, 15, 16α,

17α-$^{2}H_4$)cholesterol was administered (Figure 3-13B). Whereas, labeled tomatine was detected by monitoring the *m/z* 1038 ions $[M+H]^+$ when (15, 15, 16β, 17α-$^{2}H_4$)cholesterol was administered (Figure 3-13C). Next, to confirm the retention of the C-15 and C-17 hydrogens during SGAs biosynthesis, (15, 15, 17α-$^{2}H_3$)cholesterol was fed with potato shoots and tomato seedlings (Figure 3-12A-B and Figure 3-13A). Labeled solanine and chaconine in potato shoots were detected by monitoring the *m/z* 871 and 855 ions $[M+H]^+$, respectively (Figure 3-12A-B), and labeled tomatine in tomato seedlings was detected by monitoring the *m/z* 1037 ions $[M+H]^+$, when both labeled compounds were administered (Figure 3-13A). All labeled SGAs were a little shorter retention time than the non-labeled SGAs (Figure 3-12 and Figure 3-13). These results indicate that both hydrogens at C-16 are eliminated during SGA biosynthesis in potato, whereas only 16β hydrogen is removed in tomato, although all hydrogens at C-15 and C-17 are retained in both potato and tomato.

Phenotypes of *16DOX*-silenced transgenic potato plants

The *16DOX*-silenced transgenic plants demonstrated that the tubers of these plants could not sprout without treatment similar to that of *PGA1*- and *PGA2*-silenced plants as described in Chapter 2. Sprouting initiates normally, but the sprout clusters did not grow even after more than three months at 20 ºC or for a one year at 4 ºC after the cessation of usual dormancy in the control plants (Figure 3-14A). However, the sprout clusters could grow after the tubers were planted in soil (Figure 3-14B). After placing excised sprout tips on tissue culture media, the sprouts could also grow (Figure 3-14C). Whereas, the *16DOX*-silenced plants produced normal flowers different from that of *PGA1*- and *PGA2*-silenced plants as described in Chapter 2.

Discussion

The SGAs biosynthesis is supposed to early require the oxidation steps of cholestrol at C-16, C-22 and C-26 position (General introduction; Chapter 2). In Chapter2, two CYPs (PGA1 and PGA2) was identified as the enzymes catalyzing the C-26 hydroxylation and C-22 hydroxylation of cholesterol, respectively.

In this chapter, in vitro functional analysis of the recombinant 16DOX protein demonstrated that 16DOX catalyzes the stereospecific hydroxylation at the C-16α position (Figure 3-4). This result was supported by tracer experiments using tomato seedling (Figure 3-13). Additionally, the best substrate for recombinant 16DOX is (22*S*, 25*S*)-22, 26-dihydroxycholesterol among 9 compounds tested (Figure 3-5). The results were also clearly supported by analysis of the metabolites accumulated in *16DOX*-silenced transgenic plants (Figure 3-11). Furthermore, the transgenic plants contained significantly low amounts of SGAs (Figure 3-9B-D) and grew normally in the produced yields of tubers in the greenhouse similar to control plants (Figure 3-9E). Then, 16DOX catalyze the later hydroxylation step, following the hydroxylation reactions by PGA1 and PGA2, in the SGAs biosynthesis.

Based on the phylogenetic analysis, 16DOX was found to belong to clade DOXC41. DOXC41 includes hyoscyamine 6β–hydroxylase (H6H), which is involved in the biosynthesis of scopolamine, a tropane alkaloid, and *Hordeum vulgare* (barley) iron deficiency-specific clones 2 and 3 (HvIDS2 and HvIDS3), which are involved in phytosiderophore mugineic acid biosynthesis (Matsuda et al., 1991; Nakanishi et al., 2000; Kobayashi et al., 2001; Kawai et al., 2014). 16DOX is also first identified as a DOXC41 gene involved in terpenoid biosynthesis. It has been indicated that DOXC41 contains functionally diverse 2OGDs involved in various specialized metabolisms (Kawai et al., 2014).

Quantitative RT-PCR analysis in various tissues of potato and tomato showed that *16DOX* gene was co-expressed with *PGA1*, *PGA2*, and the previously identified UGTs responsible for the glycosylation steps in the SGAs biosynthesis (Figure 3-2). Recently, it has

been reported that six in tomato or five in potato of SGAs biosynthetic genes, including *16DOX* and the *UGTs*, exist as a cluster on chromosome 7, while an additional two are adjacent on chromosome 12 (Itkin et al., 2013). Physical clustering of functionally related genes may enable coordinate regulation of gene expression at the level of chromatin (Field et al., 2011).

As for subcellular localization, SSR2 proteins, the cholesterol synthase involved in the SGAs biosynthesis, have been predicted to contain transmembrane regions (Sawai et al., 2014). CYPs are also membrane proteins and generally bound to the endoplasmic reticulum membrane. On the other hand, 2OGDs and UGTs are soluble proteins and generally localized in the cytosol. Based on the observation of the accumulated metabolites in *SSR2*-, *PGA1*-, *PGA2*- and *16DOX*-silenced transgenic plants and the subcellular localization of the proteins encoded by these genes, It is suggested that cholesterol synthesis and the C-22 and C-26 hydroxylation steps are associated with the endoplasmic reticulum membrane, and the reaction product more hydrophilic than cholesterol are likely transported to the cytosol and modified by 16DOX, several UGTs and the other unidentified enzymes during the SGAs biosynthesis.

Tracer experiments of stable isotope labeled compounds to potato shoot and tomato seedling showed that two hydrogen atom were lost from C-16 during solanidine biosynthesis (Figure 3-12), while the 16β hydrogen was retained, but is inverted to 16α position during tomatidine biosynthesis (Figure 3-13), corresponding to the earlier study (Canonica et al., 1977). The results in potato suggest that an unknown enzyme further oxidizes at C-16 and forms the carbonyl group after 16α hydroxylation by 16DOX, and subsequently solanidine is biosynthesized via the event such as the nucleophilic attack to the C-16 by the nitrogen atom in the piperidine ring of a putative biosynthetic intermediate, which has the structure similar to verazine, a steroidal alkalid (Figure 3-15). Recently, verazine biosynthetic genes have been identified in *Veratrum californicum* (Augustin et al., 2015). Verazine are proposed intermediates in solanidine biosynthesis (Kaneko et al., 1976). On the other hand, considering that the configuration of the oxygen atom in tomatidine E-ring is 16β position, the results in

tomato propose that tomatidine is biosynthesized via the event that the E-ring is formed, accompanied by the desorption of 16α hydroxy group, the inversion of 16β hydrogen to the 16α position, and the nucleophilic attack of C-22 hydroxy group bound to the piperidine ring of a putative biosynthetic intermediate distinct from that of solanidine (Figure 3-15).

The tuber sprouts of *16DOX*-silenced transgenic potato plants did not grow (Figure 3-14A), however the sprouts could grow by planting the tubers in soil or placing excised sprout tips on tissue culture media, similar to those of *PGA1*- and *PGA2*-silenced plants. As described in Chapter 2, the results support the hypothesis that SGAs biosynthesis are associated with the sprouting of the tuber by an unknown mechanism and that the silenced plants might accumulate sprouting inhibitors which can be disabled by planting tubers in soil in the same way as that of *PGA1*- and *PGA2*-silenced plants (Figure 3-14).

Finally, 16DOX was first identified as a 2OGD gene involved in the oxidation of steroids, although several 2OGD genes have been reported as the biosynthetic and catabolic genes of gibberellin (GA), tetracyclic diterpenoid plant hormone (Yamaguchi, 2008; Xu et al., 1995; Lange, 1997; Chiang et al., 1995), and as the biosynthetic gene of monoterpenoid indole alkaloids, which is pharmaceutically important drugs (Vazquez-Flota et al., 1997). The accomplishment offers the possibility that 2OGDs also are associated with the SGAs and steroidal saponins biosynthesis in various plants.

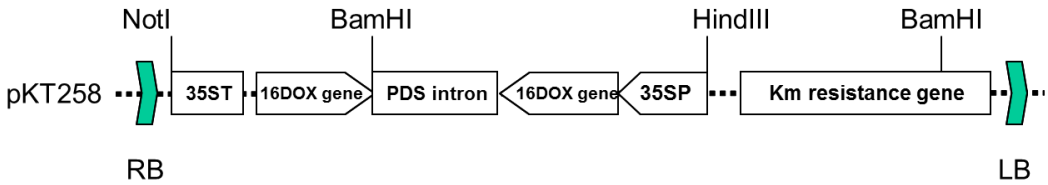


Figure 3-1. *16DOX*-knockdown vectors. Structure of control (pKT19) and knockdown vectors and the number of transgenic lines obtained. 35SP, Cauliflower mosaic virus 35S promoter; PDS intron, Third intron of the Arabidopsis *phytoene desaturase* gene (*PDS*, At4g14210); 35ST, terminator.

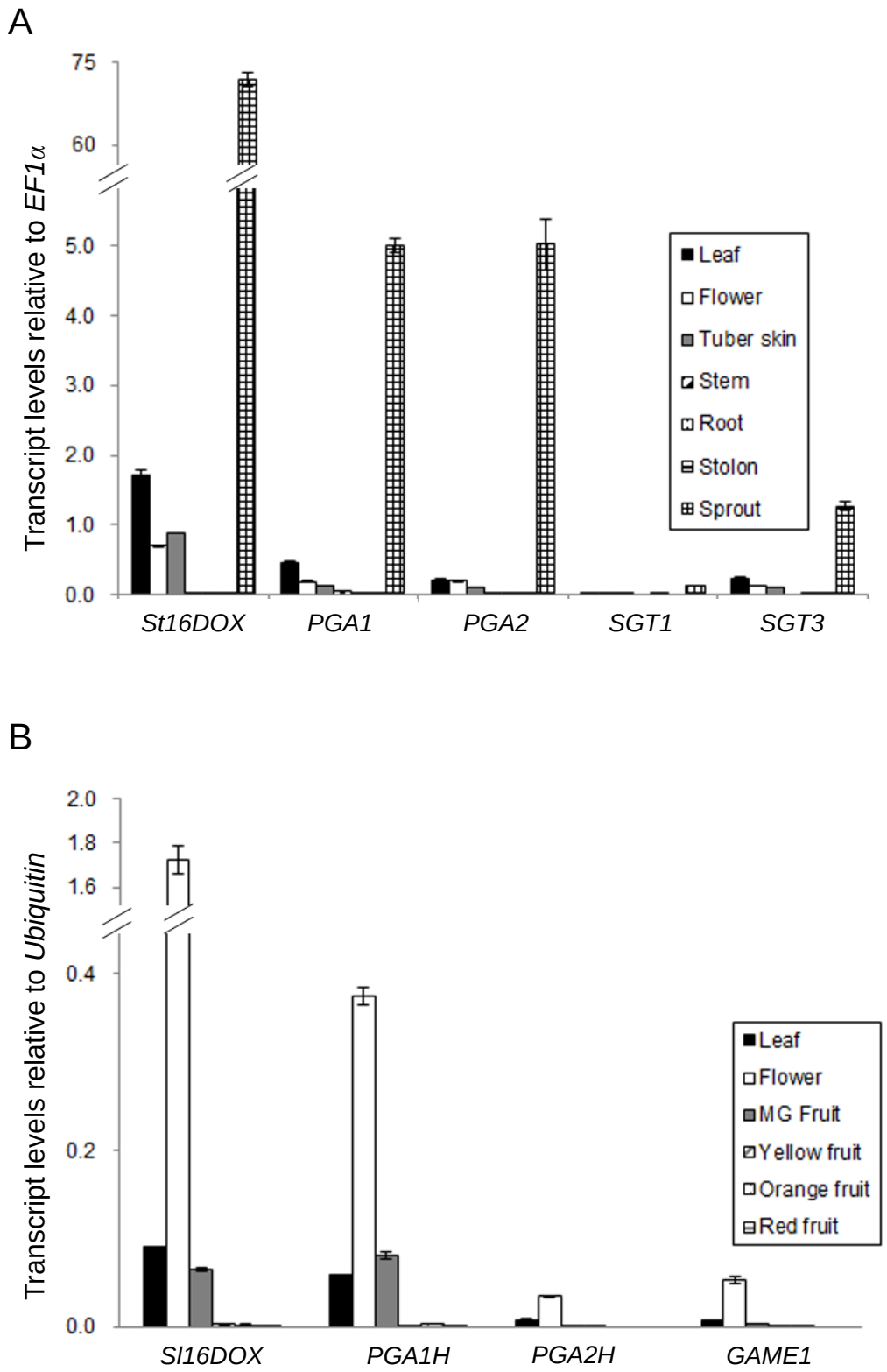


Figure 3-2. Quantitative RT-PCR analysis of SGAs biosynthetic gene in various organs of (A) potato and (B) tomato. Values represent the ration between each the biosynthetic genes and the corresponding *EF1α* and *Ubiquitin* levels of potato and tomato, respectively. Bars indicate standard deviation from the mean (*n*=3).

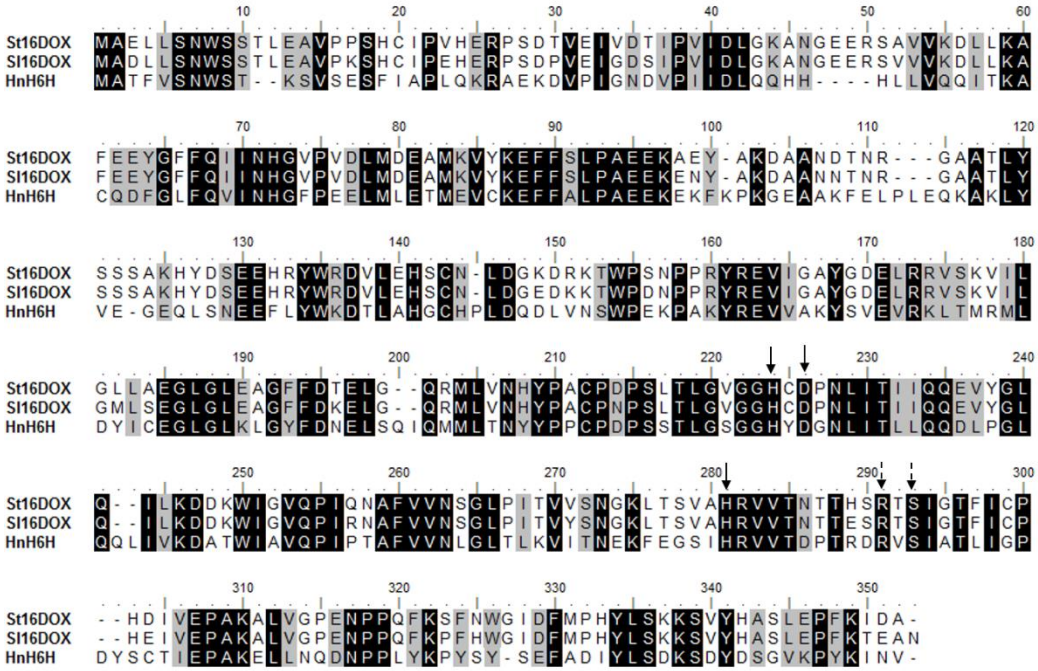


Figure 3-3. Sequence alignment of amino acid of St16DOX, Sl16DOX and HnH6H. Multiple sequence alignment was performed using the ClustralW multiple alignment analysis tool of BioEdit. Identical and similar amino acid residues are shaded in black and gray respectively. Conserved Fe(II)-binding motif among 2OGD super family are indicated with solid arrows. Conserved motif binding to the C-5 carboxy group of 2-oxoglutarate among 2OGD super family are indicated with dashed arrows.

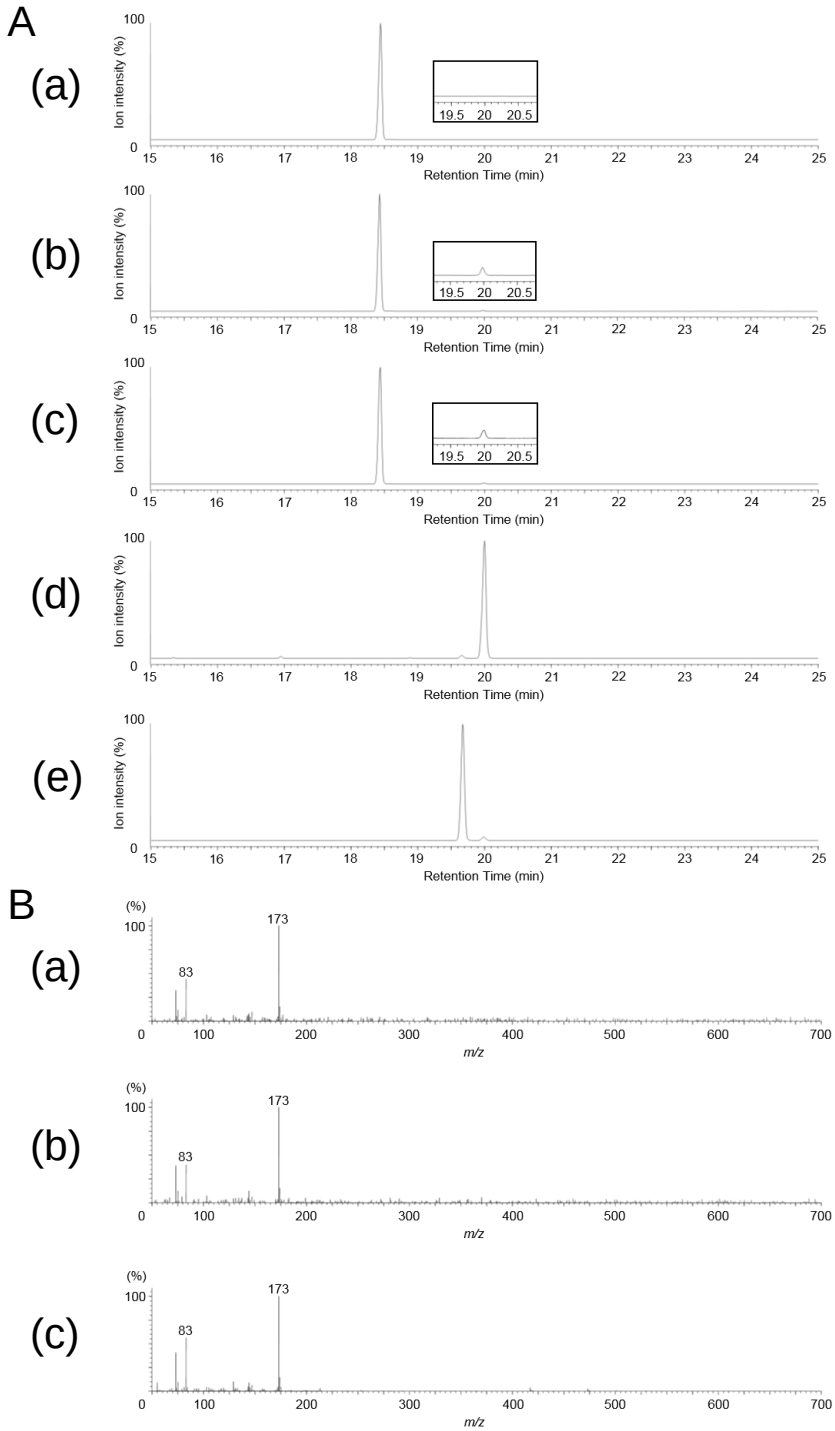


Figure 3-4. GC-MS analysis of the reaction products from the recombinant GST fusion proteins with (22*S*)-22-hydroxycholesterol as substrate. (A) The extracted ion chromatogram (EIC) of *m/z* 173 of the authentic compounds and the reaction products from the recombinants. (a) the reaction products with the empty vector; (b) the reaction products with the St16DOX; (c) the reaction products with the Sl16DOX. (d) the authentic compound of (16α, 22*S*)-16, 22-dihydroxycholesterol; (e) the authentic compound of (16β, 22*S*)-16, 22-dihydroxycholesterol. (B) (a) Mass spectra of the product peak with a retention time at 20.0 min from the St16DOX; (b) mass spectra of the product peak with a retention time at 20.0 min from the Sl16DOX; (c) mass spectra of the peak with a retention time at 20.0 min from the authentic compound of (16α, 22*S*)-16, 22-dihydroxycholesterol.

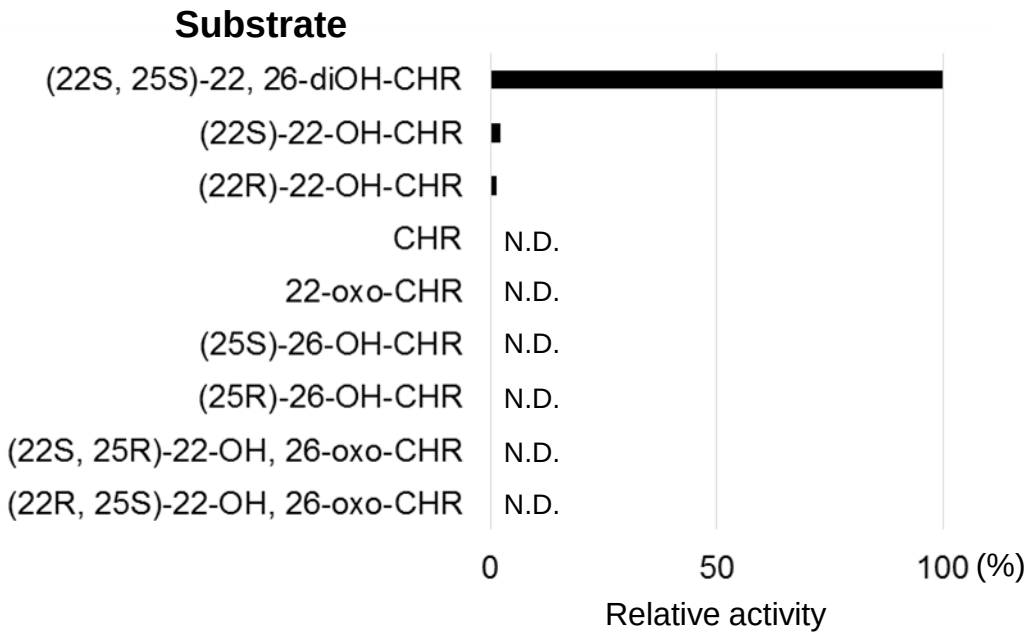


Figure 3-5. Relative activities of recombinant St16DOX toward cholesterol and several oxygenated cholesterols. N.D. indicates not detected.

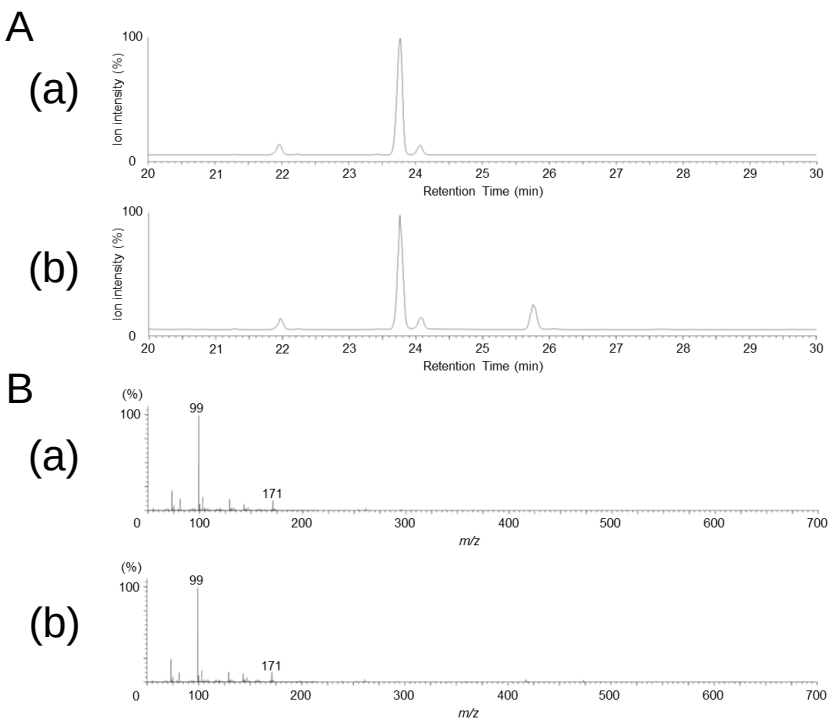


Figure 3-6. GC-MS analysis of the reaction products from the recombinant GST fusion proteins with (22*S*, 25*S*)-22, 26-dihydroxycholesterol as substrate. (A) The extracted ion chromatogram (EIC) of *m/z* 99 of the reaction products from the recombinants. (a) the reaction products with the empty vector; (b) the reaction products with the St16DOX. (B) (a) Mass spectra of the substrate peak with a retention time at 23.76 min from the St16DOX; (b) mass spectra of the product peak with a retention time at 25.77 min from the St16DOX

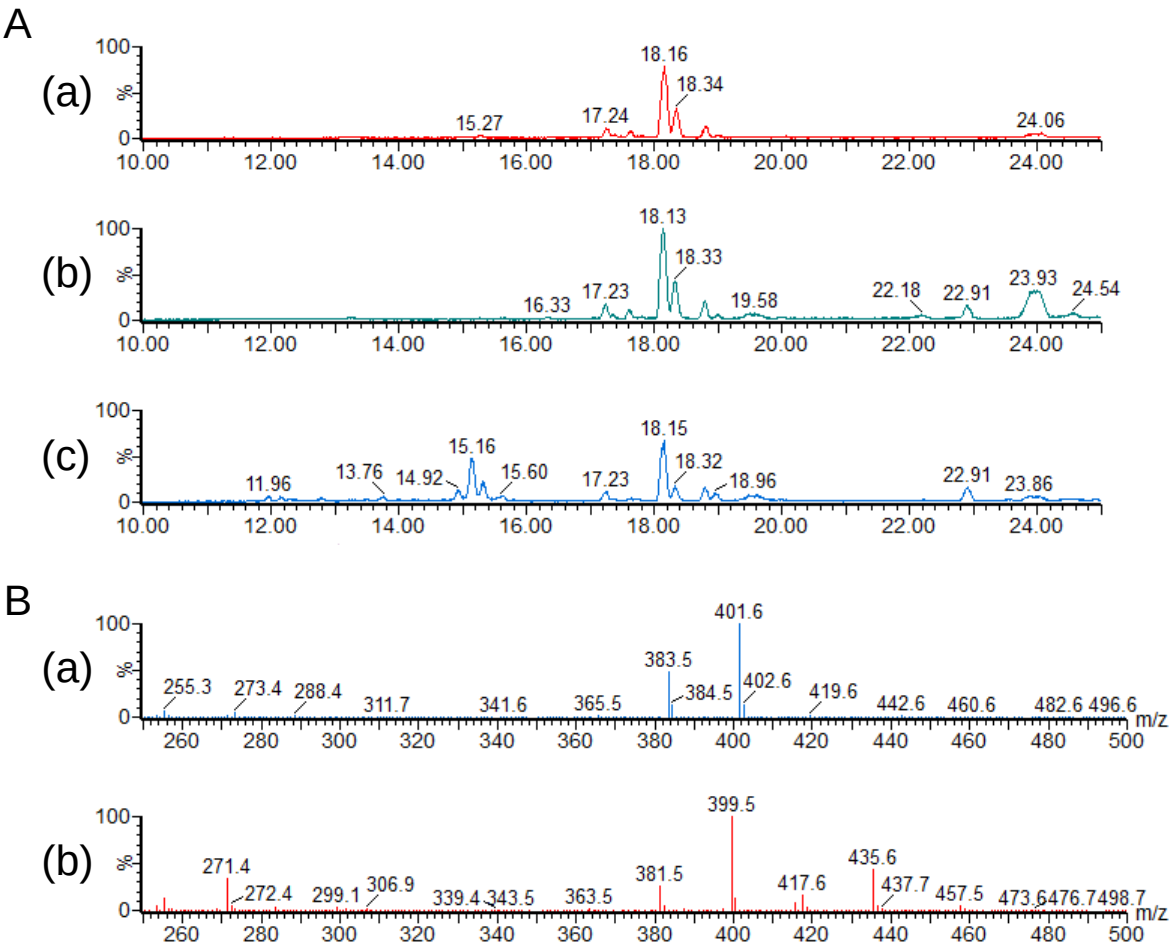


Figure 3-7. LC-MS analysis of the reaction products from the recombinant GST fusion proteins with (22*S*, 25*S*)-22, 26-dihydroxycholesterol as substrate. (A) Total ion chromatogram (TIC) of the authentic compound and the reaction products from the recombinants. (a) the authentic compound of (22*S*, 25*S*)-22, 26-dihydroxycholesterol; (b) the reaction products with the empty vector; (c) the reaction products with the St16DOX. (B) Mass spectra of the authentic compound and the reaction products from the recombinants. (a) Mass spectra of the substrate peak with a retention time at 18.16 min from the authentic compound of (22*S*, 25*S*)-22, 26-dihydroxycholesterol; (b) mass spectra of the product peak with a retention time at 15.16 min from the St16DOX

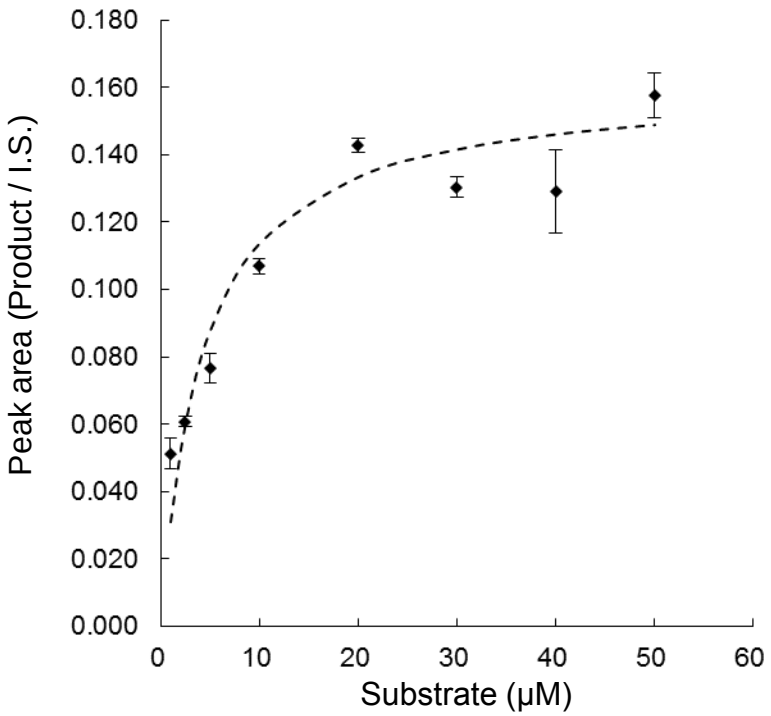


Figure 3-8. St16DOX recombinant enzyme activity curves. Enzyme activities were measured with substrate concentrations up to 50 μM (22*S*, 25*S*)-22, 26-dihydroxycholesterol. Michaelis–Menten curve (featuring *K*m of 4.19 μM) was fitted to the values obtained. Kinetic parameters were determined by non-linear regression with ANEMONA (Hernandez and Ruiz, 1998).

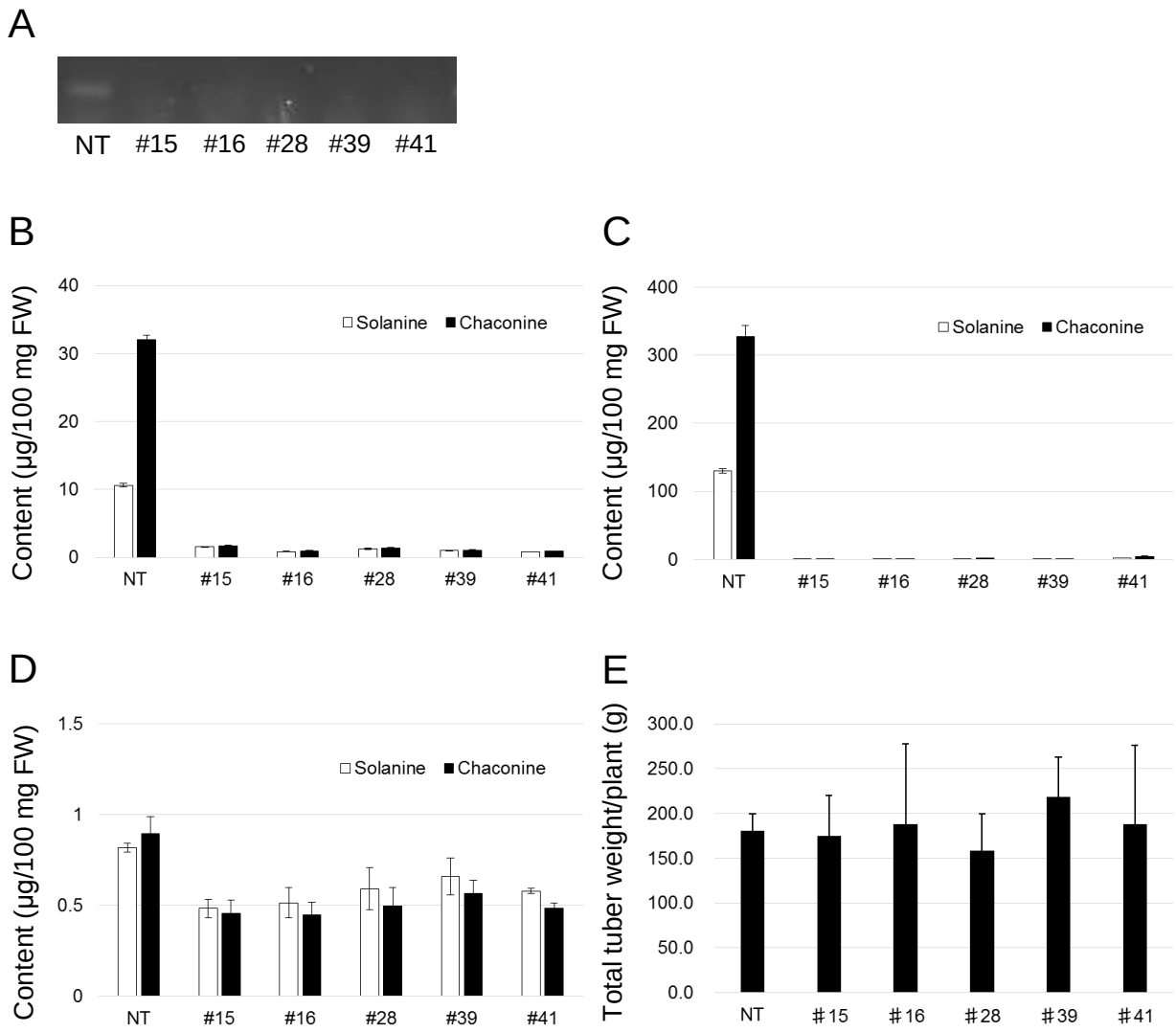


Figure 3-9. SGAs Content and yield of the *16DOX*-silenced transgenic potato plants. (A) Semi-quantification RT-PCR analysis of *16DOX* in the in vitro-grown shoots of independent *16DOX*-silenced lines; (B) LC-MS analysis of SGAs content α-solanine and α-chaconine) levels in the stems of in vitro-grown shoots *16DOX*-silenced plants; (C and D) LC-MS analysis of the SGAs levels of the *16DOX*-silenced plants in the peel (C) and cortex (D) of harvested tubers with/without light exposure; (E) Yields of the tubers from *16DOX*-silenced plants. Bars indicate standard deviation from the mean (n=3). FW, fresh weight. NT, Non-transgenic potato plants as control

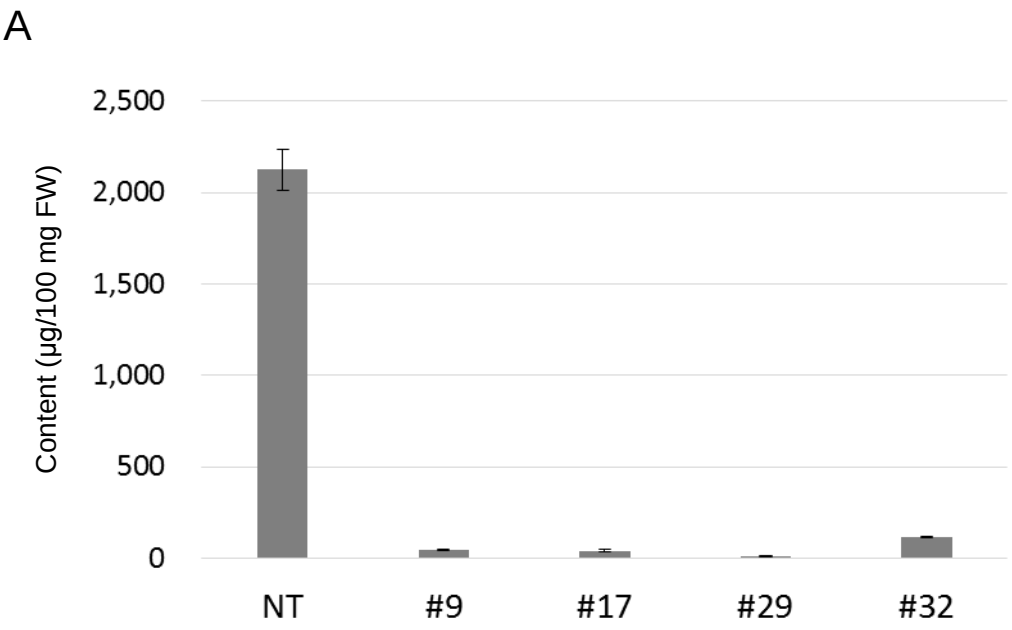


Figure 3-10. α-tomatine content and the accumulated metabolites in the leaves of the *16DOX*-silenced transgenic tomato plants, (NT) Non-transgenic line as control; (#9, #17, #29 and #32) Independent transgenic lines. (A) LC-MS analysis of α-tomatine content. Bars indicate standard deviation from the mean (n=3). FW, fresh weight; (B) GC-MS analysis of the accumulated metabolites.

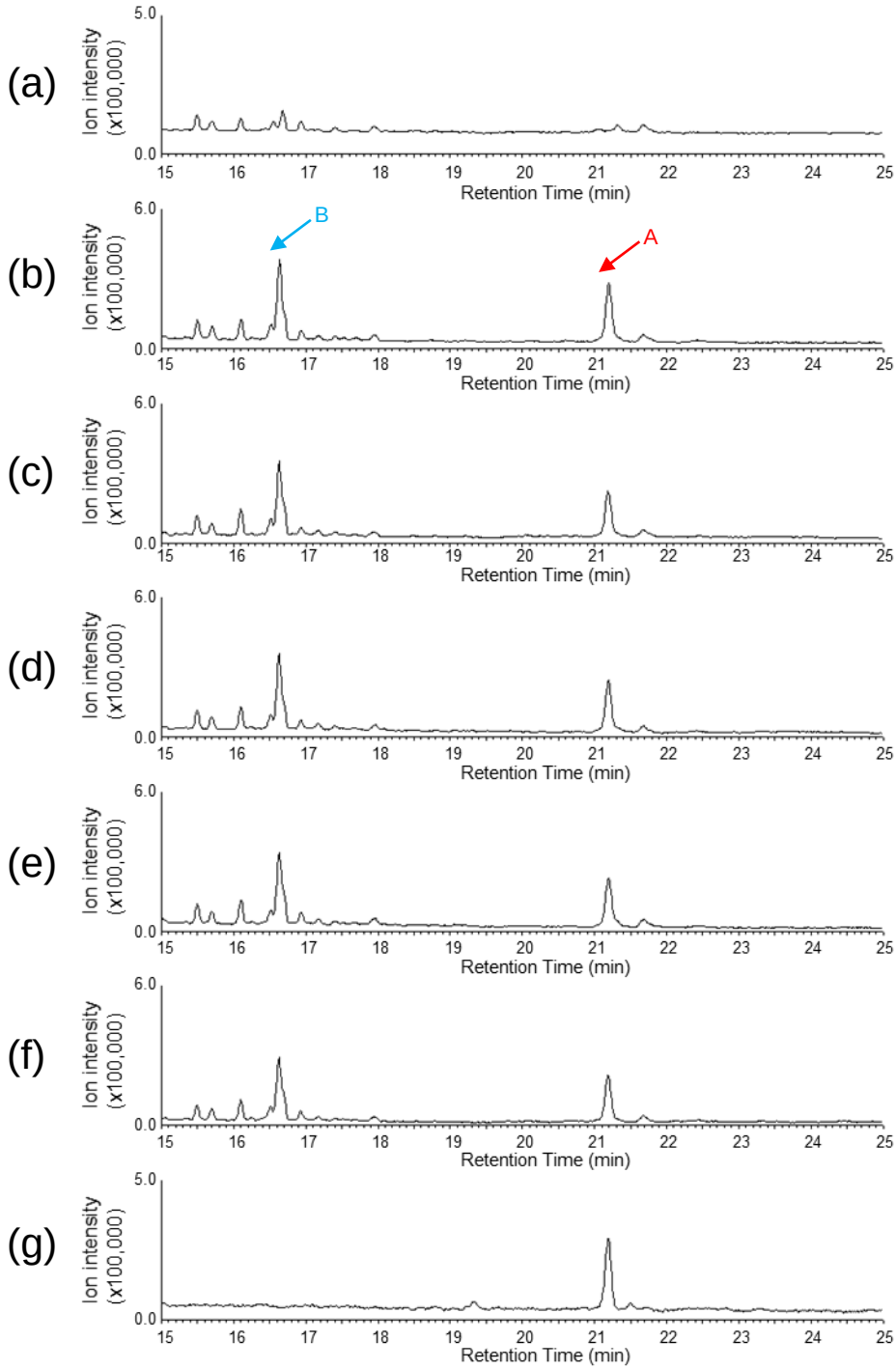


Figure 3-11A. TIC obtained by GC-MS analysis of the authentic compound and the accumulated compounds in the leaves of *16DOX*-silenced transgenic potato plants. (a) TIC of the accumulated compounds in non-transgenic plant as control; (b-f) TIC of the accumulated compounds in the independent transgenic plants (#15, #16, #28, #39 and #41); (g) TIC of the authentic compound of (22*S*, 25*S*)-22, 26-dihydroxycholesterol.

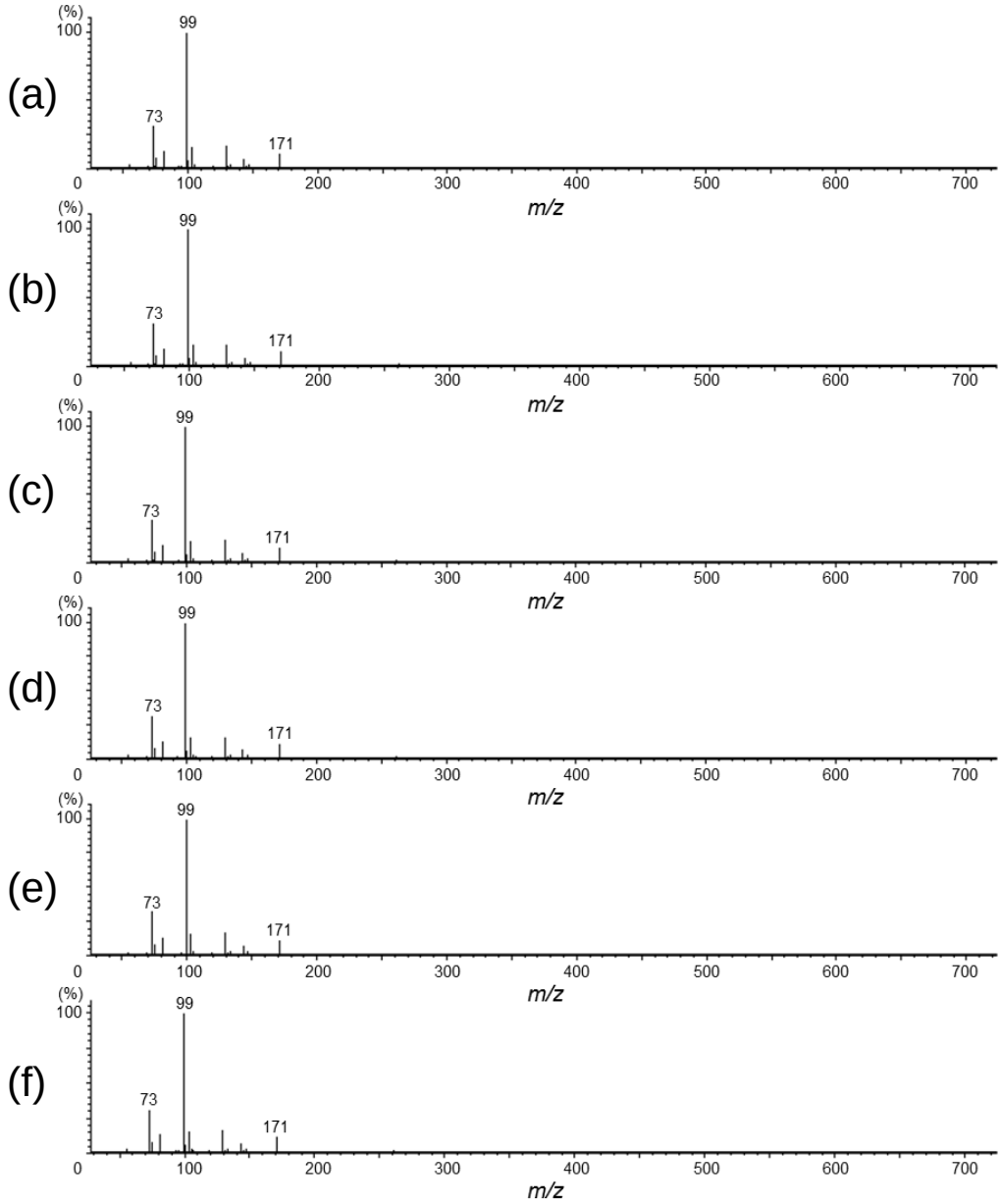


Figure 3-11B. Mass spectra of the peak A with retention time 21.2 min in Figure 3-11A. (a-f) are the mass spectra of the peaks in Figure 3-11A (b-g), respectively.

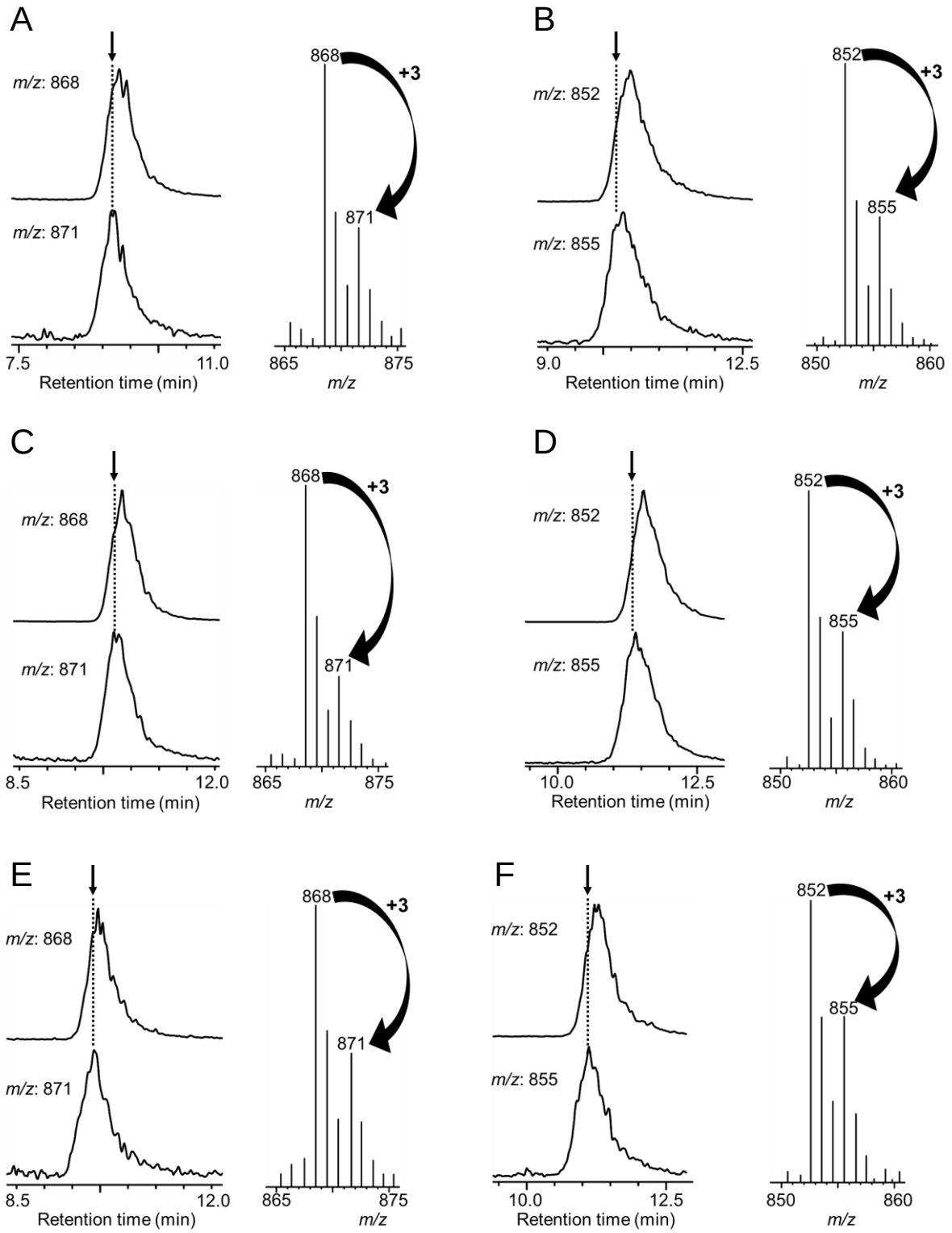


Figure 3-12. LC-MS analysis of SGAs upon feeding the stable isotope labeled compounds to the seedlings of potato. The left two traces were mass-chromatograms for the indicated *m/z* ions. The right mass spectra were obtained at the retention time indicated by arrow. (A, C and E) LC-MS analysis of α-solanine; (B, D and F) LC-MS analysis of α-chaconine; (A and B) feeding (15, 15 , 17a-$^{2}H_{3}$)cholesterol; (C and D) feeding (15, 15, 16α, 17α-$^{2}H_{4}$)cholesterol; (E and F) feeding (15, 15, 16β, 17α-$^{2}H_{4}$)cholesterol.

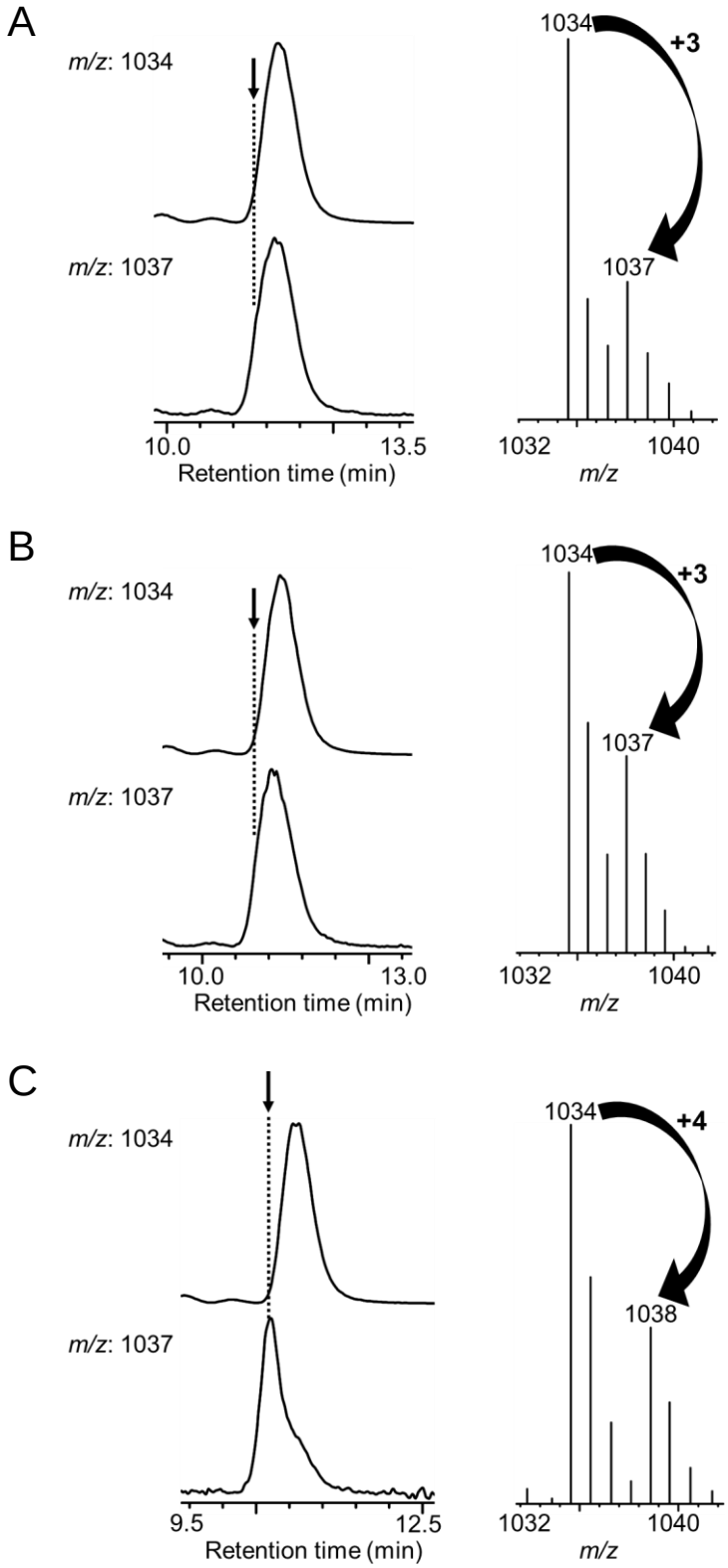


Figure 3-13. LC-MS analysis of α-tomatine upon feeding the stable isotope labeled compounds to the seedlings of tomato. The left two traces were mass-chromatograms for the indicated *m/z* ions. The right mass spectra were obtained at the retention time indicated by arrow. (A) feeding (15, 15 , 17a-$^{2}H_{3}$)cholesterol; (B) feeding (15, 15, 16α, 17α-$^{2}H_{4}$)cholesterol; (C) feeding (15, 15, 16β, 17α-$^{2}H_{4}$)cholesterol.

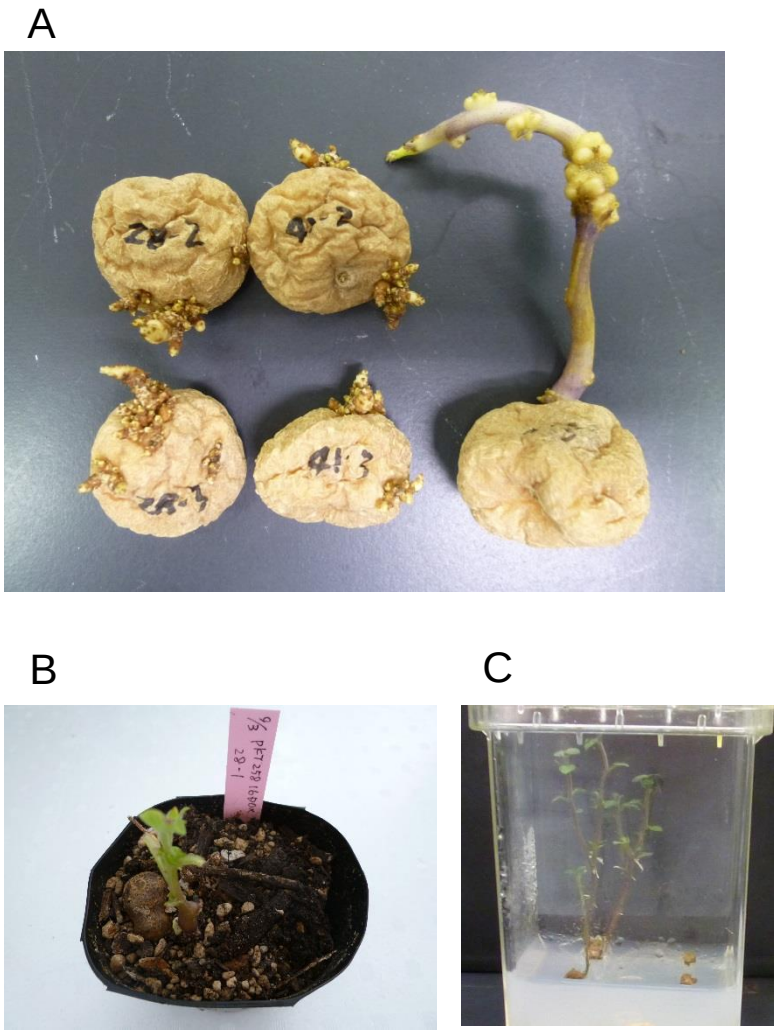


Figure 3-14. Phenotypes of the *16DOX*-silenced transgenic potato plants. (A) Sprouted potatoes of the *16DOX*-silenced plants (pKT258-#28 and #41, left and middle) and control (right) 18 weeks after the cessation of plant dormancy in the control; (B) Sprouted potatoes of the *16DOX*-sienced plant (pKT258-#28) planted into soil; (C) In vitro growth of sprout tips on tissue culture media. Sprout tips cut from tubers of the *16DOX*-silenced plant (pKT258-#28) were placed on tissue culture media without plant hormones.

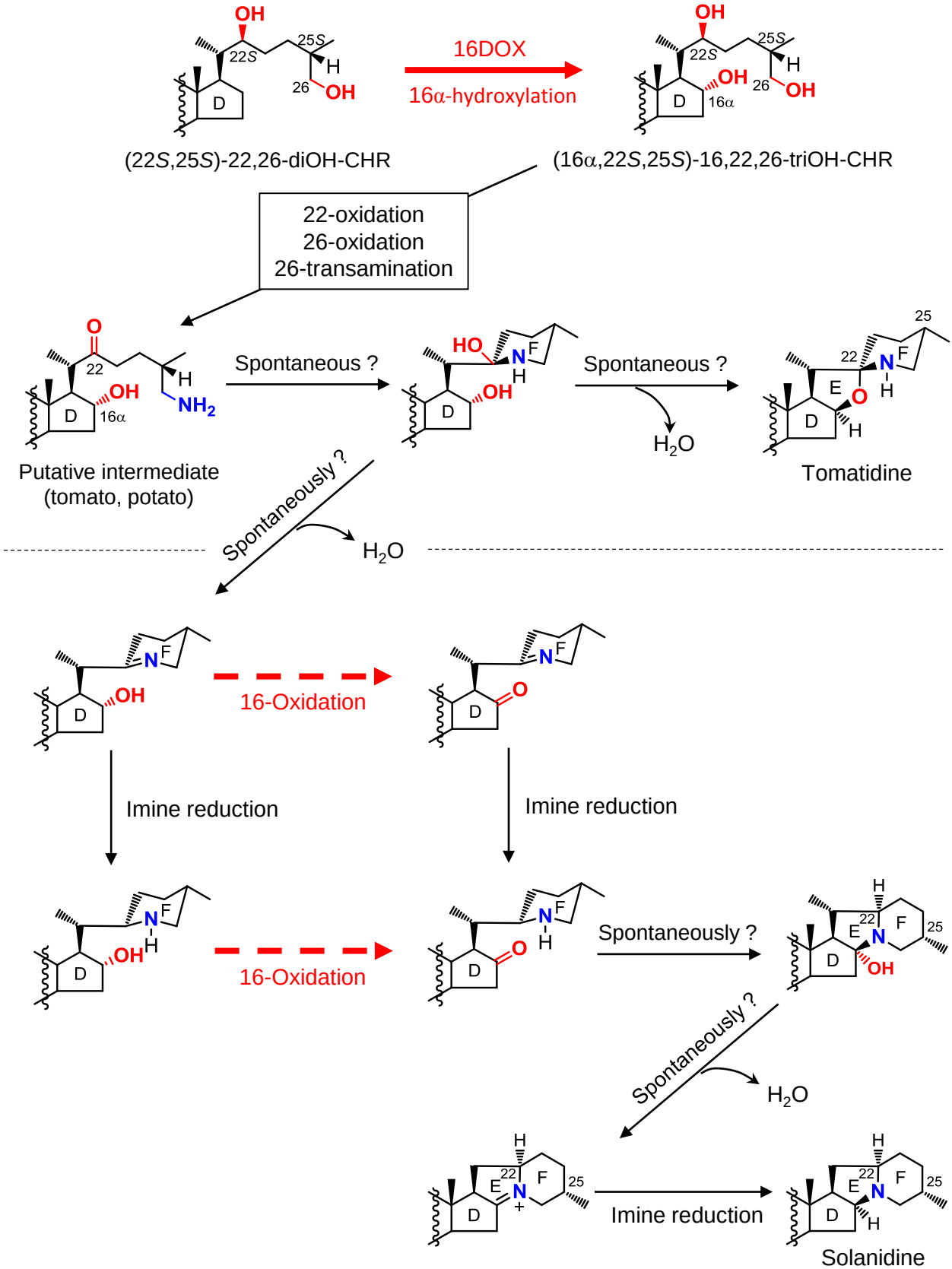


Figure 3-15. Putative biosynthetic pathways of solanidine and tomatidine in potato and tomato, respectively. Black solid arrows indicate unidentified reaction stages. Red solid arrow indicates the reaction stage identified in this chapter. Red dashed arrow indicates reaction stages suggested by tracer experiments in this chapter.

Concluding discussion

A variety of steroidal saponins and SGAs are found in many plant species. These are hypothesized to be biosynthesized from cholesterol via several sequential modification steps, but little has been known about the enzymes and genes involved in these biosynthesis. To elucidate these biosynthesis in plants, the functional analysis of the genes and enzymes for steroidal saponins biosynthesis in *Dioscorea* spp. and SGAs biosynthesis in potato and tomato, were performed.

Chapter 2 and Chapter 3 showed that PGA1, PGA2 and 16DOX catalyze the hydroxylation steps of cholesterol at C-22, C-26 and C-16, respectively, in both potato and tomato SGAs biosynthesis (Figure D). However, solanidine and tomatidine, the aglycones of potato and tomato SGAs, have a structural difference in the E- and F-rings. Tracer experiments in Chapter 3 suggested the further oxidation at C-16 after the 16α-hydroxylation occurring in only potato as the branching point in potato and tomato SGAs biosynthesis. On the other hand, Chapter 1 showed that DeF26G1 hydrolyzes protodioscin to dioscin in *D. esculenta*. This result supports that the E-ring of dioscin is formed followed by the closure of the F-ring by DeF26G1 in the biosynthesis. This order of the E- and F-rings formation is different from that of potato and tomato SGAs assumed in Chapter 3.

In Chapter 2, both PGA1 and PGA2 were found to belong to CYP72A subfamily. Whereas, C-26 hydroxylase/oxidase and C-22 hydroxylase and involved in verazine biosynthesis in *V. californicum,* belong to CYP94N and CYP90B subfamilies not CYP72A, respectively (Augustin et al., 2014). Additionally, in Chapte 3, 16DOX was found to belong to DOXC41 clade, although C-16 hydroxylase involved in solanidine biosynthesis in *V. californicum,* has been unidentified yet. Then, the genes resposible for the oxidation steps of dioscin biosynthesis in *D. esculenta* may belong to CYP72A, CYP90B, CYP94N or DOXC41.

Furthemore, the phylogenetic difference in these oxidases as described above imply the possibility of the convergent evolution of the steroidal saponins and SGAs biosynthetic genes among various plant species.

In Chapter 3, quantitative RT-PCR analysis showed that several SGAs biosynthetic genes were co-expressed in various tissues of plants, and these transcripts levels were the highest in the tissue accumulating the greatest level of SGAs content. Similarly, in Chapter 1, The distribution pattern of steroidal saponins in *Dioscorea* plants is well consistent with the *DeF26G1* gene expression. These results support that the expression of the steroidal saponins and SGAs biosynthetic genes is regulated coordinally. The coordinate regulation may be associated with the clustering of these genes in genome as described in Chapter 3. Then, to explore dioscin biosynthetic genes and the genes catalyzing the unclear modification steps such as the futher oxidaiton at C-16, C-22 and C-26 in SGAs biosynthesis, the candidate genes selected by co-expression analysis and the functionally unknown genes contained in the cluster should be investigated in future works.

Finally, Identification and biochemical characterization of the steroidal saponins and SGAs biosynthetic genes and enzymes will allow to modify the biosynthetic regulation in plants. Steroidal saponins such as protodioscin and dioscin accumulated in *Dioscorea* spp. are valuable compound used for semi-synthetic production of pharmaceutical steroidal drugs as described in General introduction. Since these steroidal saponins are likely biosynthesized similarly to the SGAs biosynthesis, the combination of the suppression of the SGAs biosynthetic genes as described in Chapter 2 and Chapter 3, and the introduction of the steroidal saponins biosynthetic genes will be able to construct the transgenic potato and tomato containing little toxic SGAs and accumulates useful steroidal saponins. Therefore, further research in these biosynthesis are required for the establishment of metabolic engineering techniques in plants.

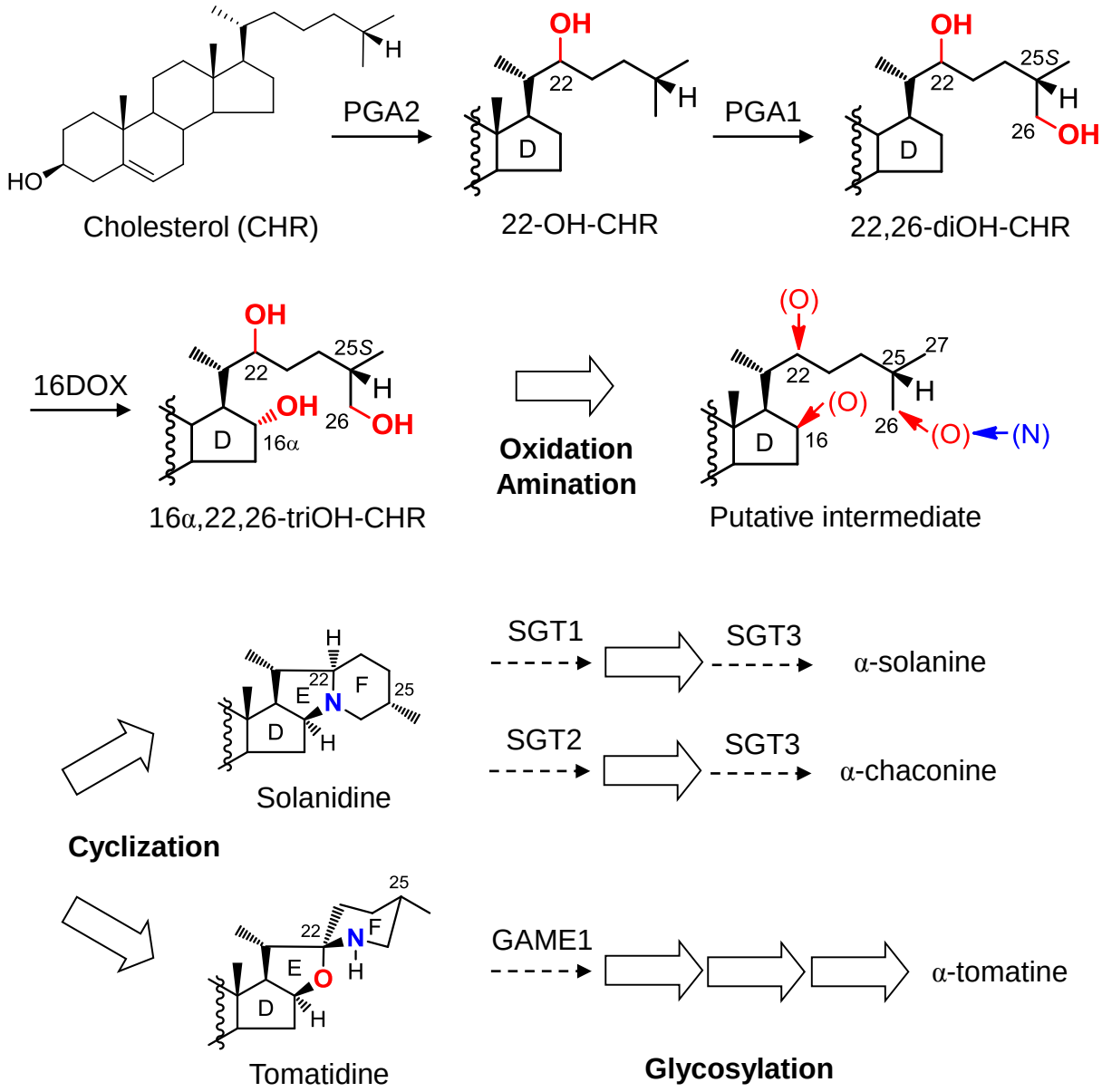


Figure D. The putative biosynthetic pathway of protodioscin and dioscin in *Dioscorea* spp. and SGAs in potato and tomato. Thick arrows indicate unidentified reaction stages. Dashed arrows indicated identified reaction stages in previous study. Solid arrows mean reaction steps characterized in this work.

Acknowledgements

I would like to express my cordial gratitude to Associate Professor Masaharu Mizutani (Lab. Funct. Phytochem., Grad. Sch. Agric. Sci., Kobe Univ.) for constant encouragement, excellent advice, extensive discussion and long-term support throughout my research and for reviewing the manuscript of this dissertation, to Professor Yukihiro Sugimoto (Lab. Funct. Phytochem., Grad. Sch. Agric. Sci., Kobe Univ.) for considerable encouragement, valuable advice and tremendous support throughout my research and for reviewing the manuscript of this dissertation, and to Assistant Professor Yasuo Yamauchi (Lab. Funct. Phytochem., Grad. Sch. Agric. Sci., Kobe Univ.) for sincere encouragement, useful comments and helpful discussion throughout my research. I also express my sincere appreciation to Professor Hiroshi Yamagata (Lab. Biochem., Grad. Sch. Agric. Sci., Kobe Univ.) for kindly reviewing the manuscript of this dissertation.

I express profound gratitude to Postdoctoral Fellow Lee Hyong Jae (Lab. Funct. Phytochem., Grad. Sch. Agric. Sci., Kobe Univ.) for the analysis of metabolites in plants and technical support; Senior Fellow Naoyuki Umemoto (CSRS, RIKEN) for the construction of transgenic plants and providing cDNA from various tissues of potato; Assistant Professor Kiyoshi Ohyama (CSRS, RIKEN and Grad. Sch. Chem. & Mater. Sci., Tokyo Tech. Univ.) for the analysis of accumulated compounds in transgenic plants, the tracer experiments and technical support; Professor Michio Onjo (Kagoshima Univ.) for providing plant materials of *Dioscorea* spp., Assistant Professor Bunta Watanabe (ICR, Kyoto University) for providing the authentic compounds of two 26-hydroxycholesterols, two (16, 22*S*)-16, 22-dihydroxycholesterol, (22*S*, 25*S*)-22, 26-dihydroxycholesterol, (22*R*, 25*S*)-22-hydroxy, 26-oxocholesterol, and (22*S*, 25*R*)-22-hydroxy, 26-oxocholesterol; Professor Toshiya Muranaka (Grad. Sch. Eng., Osaka Univ.) for the comprehensive support; Associate Professor Takashi

Kawasaki (Ritsumeikan Univ.) and Assistant Professor Kotomi Ueno (Tottori Univ.) for technical support.

I thank Mr. Shingo Urakawa, Mr. Daichi Kawaguchi, Ms. Rie Yamamura, Ms Mika Okada, Ms. Midori Kobayashi, Mr. Ryota Akiyama, Mr. Naoto Inoue, Ms. Haruka Miyachi, Mr. Junpei Kato, Ms. Tomoko Miyata, Technical Assistant Mina Yamamoto (Lab. Funct. Phytochem., Grad. Sch. Agric. Sci., Kobe Univ.) for focusing on the same research, and all the members in Lab. Funct. Phytochem., Grad. Sch. Agric. Sci., Kobe Univ. for encouragement and help during the research.

This study was supported by the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN), Japan; by Grant-in-Aid for JSPS Fellows; and by Cross-ministerial Strategic Innovation Promotion Program (SIP), Japan.

Finally, I would also like to express my gratitude to my parent Mr. Hidekazu Nakayasu and Ms. Kimiyo Nakayasu for their moral support and persistent encouragements, and my friends for warm encouragement.

References

Ahn YO, Mizutani M, Saino H, Sakata K (2004) Furcatin hydrolase from *Viburnum furcatum* Blume is a novel disaccharide-specific acuminosidase in glycosyl hydrolase family 1. *J Biol Chem* 279: 23405-23414

Ahn YO, Saino H, Mizutani M, Shimizu BI, Sakata K (2007) Vicianin hydrolase is a novel cyanogenic β-glycosidase specific to β-vicianoside (6-O-α-L-arabinopyranosyl-β-D-glucopyranoside) in seeds of *Vicia angustifolia*. *Plant Cell Physiol* 48: 938-947

Arnqvist L, Dutta PC, Jonsson L, Sitbon F (2003) Reduction of cholesterol and glycoalkaloid levels in transgenic potato plants by overexpression of a type 1 sterol methyltransferase cDNA. *Plant Physiol* 131:1792-1799

Arthan D, Kittakoop P, Esen A, Svasti J (2006) Furostanol glycoside 26-O-β-glucosidase from the leaves of *Solanum torvum*. *Phytochemistry* 67: 27-33

Augustin MM, Ruzicka DR, Shukla AK, Augustin JM, Starks CM, O'Neil-Johnson M, McKain MR, Evans BS, Barrett MD, Smithson A, Wong GK, Deyholos MK, Edger PP, Pires JC, Leebens-Mack JH, Mann DA, Kutchan TM (2015) Elucidating steroid alkaloid biosynthesis in Veratrum californicum: production of verazine in Sf9 cells. *Plant J* 82: 991-1003

Barrett T, Suresh CG, Tolley SP, Dodson EJ, Hughes MA (1995) The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* 3: 951-960

Bugg TDH (2003) Dioxygenase enzymes: catalytic mechanisms and chemical models. *Tetrahedron* 59: 7075–7101

Canonica L, Ronchetti F, Russo G (1977) Fate of the 16β-hydrogen atom of cholesterol in the biosynthesis of tomatidine and solanidine. *J Chem Soc Chem Commun*: 286–287

Chiang HH, Hwang I, Goodman HM (1995) Isolation of the Arabidopsis GA4 locus. *Plant Cell* 7: 195–201

Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci USA* 108: 16116-16121

Friedman M (2002) Tomato Glycoalkaloids: Role in the Plant and in the Diet. *J Agric Food Chem* 50: 5751-5780

Friedman M (2006) Potato glycoalkaloids and metabolites: roles in the plant and in the diet. *J Agric Food Chem* 54:8655-8681

Frisch DA, Harris-Haller LW, Yokubaitis NT, Thomas TL, Hardin SH, Hall TC (1995) Complete sequence of the binary vector Bin 19. *Plant Mol Biol* 27:405-409

Fukushima EO, Seki H, Sawai S, Suzuki M, Ohyama K, Saito K, Muranaka T (2013) Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol* 54:740-749

Ginzberg I, Tokuhisa J, Veilleux R (2009) Potato Steroidal Glycoalkaloids: Biosynthesis and Genetic Manipulation. *Potato Res* 52:1-15

Grünweller S, Kesselmeier J (1985) Characterization of a membrane bound β-glucosidase responsible for the activation of oat leaf saponins. *Phytochemistry* 24: 1941-1943

Gurielidze K, Gogoberidze M, Dadeshidze I, Vardosanidze M, Djaoshvili M, Lomkatsi N (2004) Tissue and subcellular localization of oligofurostanosides and their specific degrading β-glucosidase in *Dioscorea caucasica* Lipsky. *Phytochemistry* 65: 555-559

Gus-Mayer S, Brunner H, Schneider-Poetsch HA, Rüdiger W (1994) Avenacosidase from oat: purification, sequence analysis and biochemical characterization of a new member of the BGA family of β-glucosidases. *Plant Mol Biol* 26: 909-921

Harrison DM, (1990) Steroidal alkaloids. *Nat Prod Rep* 7: 139–147

Helmut R, (1998) Solanum steroid alkaloids – an update. Alkaloids: *Chem Biol Perspect* 12: 103–185

Hernandez A, Ruiz MT (1998) An EXCEL template for caluculation of enzyme kinetics parameters by non-linear regression. *Bioinformatics* 14: 227-228

Hoffmann D (2003) Medical herbalism: the science and practice of herbal medicine. Rochester (VT): *Inner Traditions, Bear & Co*

Hostettmann K, Marston A (2005) Saponins, chemistry and pharmacology of natural products. *Cambridge, UK: Cambridge University Press*

Iijima Y, Watanabe B, Sasaki R, Takenaka M, Ono H, Sakurai N, Umemoto N, Suzuki H, Shibata D, Aoki K (2013) Steroidal glycoalkaloid profiling and structures of glycoalkaloids in wild tomato fruit. *Phytochemistry* 95:145-157

Inoue K, Ebizuka Y (1996). Purification and characterization of furostanol glycoside 26-O-β-glucosidase from *Costus speciosus* rhizomes. *FEBS Lett* 378: 157-160

Inoue K, Shibuya M, Yamamoto K, Ebizuka Y (1996b) Molecular cloning and bacterial expression of a cDNA encoding furostanol glycoside 26-O-β-glucosidase of *Costus speciosus*. *FEBS Lett* 389: 273-277

Inoue K, Shimomura K, Kobayashi S, Sankawa U, Ebizuka Y (1996a) Conversion of furostanol glycoside to spirostanol glycoside by β-glucosidase in *Costus speciosus*. *Phytochemistry* 41: 725-727

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP, Aharoni A (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* 341:175-179

Itkin M, Rogachev I, Alkan N, Rosenberg T, Malitsky S, Masini L, Meir S, Iijima Y, Aoki K, de Vos R, Prusky D, Burdman S, Beekwilder J, Aharoni A (2011) GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell* 23:4507-4525

Jacobsen SE, Olszewski NE (1996) Gibberellins regulate the abundance of RNAs with sequence similarity to proteinase inhibitors, dioxygenases and dehydrogenases. *Planta* 198:78-86

Jenkins J, Leggio LL, Harris G, Pickersgill R (1995) β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes wit 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Lett* 362: 281-285

Joly RA, Bonner J, Bennett RD, Heftmann E (1969a) Conversion of an open-chain saponin to dioscin by a *Dioscorea floribunda* homogenate. *Phytochemistry* 8: 1445-1447

Joly RA, Bonner J, Bennett RD, Heftmann E (1969b) The biosynthesis of steroidal sapogenins in *Dioscorea floribunda* from doubly labelled cholesterol. *Phytochemistry* 8: 1709-1711

Kaneko K, Tanaka WM, Mitsuhashi H (1976) Origin of nitrogen in the biosynthesis of solanidine by Veratrum grandiflorum. *Phytochemistry* 15: 1391-1393

Kawai Y, Ono E, Mizutani M (2014) Evolution and diversity of the 2–oxoglutarate-dependent dioxygenase superfamily in plants. Plant J 78: 328–343

Keresztessy Z, Kiss L, & Hughes MA (1994) Investigation of the active site of the cyanogenic β-D-glucosidase (linamarase) from *Manihot esculenta* Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. *Arch Biochem Biophys* 315: 323-330

Kobayashi T, Nakanishi H, Takahashi M, Kawasaki S, Nishizawa N, Mori S (2001) In vivo evidence that Ids3 from Hordeum vulgare encodes a dioxygenase that converts 2 ‘-deoxymugineic acid to mugineic acid in transgenic rice. *Planta* 212: 864–871

Kozukue N, Mizuno S (1985) Studies on Glycoalkaloids of Potatoes (Part II) Analysis of Glycoalkaloid Content in Potato Tissues and Tubers. *J Jpn Soc Hortic Sci* 54 Suppl. 2:496-497

Kozukue N, Mizuno S (1989) Studies on Glycoalkaloids of Potatoes (Part IV) Changes of Glycoalkaloid Content in Four Parts of a Sprouted Potato Tuber and in Potato Tubers during Storage. *J Jpn Soc Hortic Sci* 58:231-235

Lange T (1997) Cloning gibberellin dioxygenase genes from pumpkin endosperm by heterologous expression of enzyme activities in *Escherichia coli*. *Proc Natl Acad Sci USA* 94: 6553–6558

Lantin S, O'Brien M, Matton DP (1999) Pollination, wounding and jasmonate treatments induce the expression of a developmentally regulated pistil dioxygenase at a distance, in the ovary, in the wild potato Solanum chacoense Bitt. *Plant Mol Biol* 41: 371-386

Lukacˇin R, Britsch L (1997) Identification of strictly conserved histidine and arginine residues as part of the active site in Petunia hybrida flavanone 3bhydroxylase. *Eur J Biochem* 249: 748–757

Lukacˇin R, Gröning I, Pieper U, Matern U (2000) Site-directed mutagenesis of the active

serine290 in flavanone 3b-hydroxylase from Petunia hybrida. *Eur J Biochem* 267: 853–860

Matsuda J, Okabe S, Hashimoto T, Yamada Y (1991) Molecular cloning of hyoscyamine 6β–hydroxylase, a 2-oxoglutarate-dependent dioxygenase, from cultured roots of Hyoscyamus niger. *J Biochem* 266: 9460–9464

McCue KF, Allen PV, Shepherd LV, Blake A, Maccree MM, Rockhold DR, Novy RG, Stewart D, Davies HV, Belknap WR (2007) Potato glycosterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. *Phytochemistry* 68:327-334

McCue KF, Allen PV, Shepherd LV, Blake A, Whitworth J, Maccree MM, Rockhold DR, Stewart D, Davies HV, Belknap WR (2006) The primary in vivo steroidal alkaloid glucosyltransferase from potato. *Phytochemistry* 67:1590-1597

McCue KF, Shepherd LVT, Allen PV, Maccree MM, Rockhold DR, Corsini DL, Davies HV, Belknap WR (2005) Metabolic compensation of steroidal glycoalkaloid biosynthesis in transgenic potato tubers: using reverse genetics to confirm the in vivo enzyme function of a steroidal alkaloid galactosyltransferase. *Plant Sci* 168:267-273

Milligan SB, Gasser CS (1995) Nature and regulation of pistil-expressed genes in tomato. *Plant Mol Biol* 28: 691-711

Mizutani M, Ohta D (1998) Two isoforms of NADPH:cytochrome P450 reductase in Arabidopsis thaliana. Gene structure, heterologous expression in insect cells, and differential regulation. *Plant Physiol* 116:357-367

Momma T (1990) Recent study for genetic engineering of soybean glycinin gene. *Plant tissue culture lett* 7:57-63

Moehs CP, Allen PV, Friedman M, Belknap WR (1997) Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. *Plant J* 11: 227–236

Morant AV, Jørgensen K, Jørgensen C, Paquette SM, Sánchez-Pérez R, Møller BL, Bak S (2008) β-Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69: 1795-1813

Moses T, Papadopoulou KK, Osbourn A (2014) Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit Rev Biochem Mol Biol* 49: 439-462

Moses T, Pollier J, Almagro L, Buyst D, Van Montagu M, Pedreño MA, Martins JC, Thevelein JM, Goossens A (2014) Combinatorial biosynthesis of sapogenins and saponins in Saccharomyces cerevisiae using a C-16α hydroxylase from Bupleurum falcatum. *Proc Natl Acad Sci USA* 111: 1634-1639

Nakane E, Kawakita K, Doke N, Yoshioka H (2003) Elicitation of primary and secondary metabolism during defense in the potato. *J Gen Plant Pathol* 69: 378-384

Nakanishi H, Yamaguchi H, Sasakuma T, Nishizawa NK, Mori S (2000) Two dioxygenase genes, Ids3 and Ids2, from Hordeum vulgare are involved in the biosynthesis of mugineic acid family phytosiderophores. *Plant Mol Biol* 44: 199–207

Nicot N, Hausman JF, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot* 56:2907-2914

Nisius A (1988) The stromacentre inAvena plastids: An aggregation of β-glucosidase responsible for the activation of oat-leaf saponins. *Planta* 173: 474-481

Ohnishi T, Szatmari AM, Watanabe B, Fujita S, Bancos S, Koncz C, Lafos M, Shibata K, Yokota T, Sakata K, Szekeres M, Mizutani M (2006a) C-23 hydroxylation by Arabidopsis CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis. *Plant Cell* 18:3275-3288

Ohnishi T, Watanabe B, Sakata K, Mizutani M (2006b) CYP724B2 and CYP90B3 function in the early C-22 hydroxylation steps of brassinosteroid biosynthetic pathway in tomato. *Biosci Biotechnol Biochem* 70:2071-2080

Ohnishi T, Yokota T, Mizutani M (2009) Insights into the function and evolution of P450s in plant steroid metabolism. *Phytochemistry* 70:1918-1929

Ohyama K, Okawa A, Moriuchi Y, Fujimoto Y (2013) Biosynthesis of steroidal alkaloids in Solanaceae plants: involvement of an aldehyde intermediate during C-26 amination. *Phytochemistry* 89:26-31

Petersen HW, Mølgaard P, Nyman U, Olesen CE (1993) Chemotaxonomy of the tuber-bearing Solanum species, subsection Potatoe (Solanaceae). *Biochem Syst Ecol* 21: 629–644

Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189-195

Roddick JG (1989) The acetylcholinesterase-inhibitory activity of steroidal glycoalkaloids and their aglycones. *Phytochemistry* 28: 2631–2634

Rye CS, Withers SG (2000) Glycosidase mechanisms. *Curr Opin Chem Biol* 4: 573-580

Saino H, Shimizu T, Hiratake J, Nakatsu T, Kato H, Sakata K, Mizutani M (2014) Crystal structures of β-primeverosidase in complex with disaccharide amidine inhibitors. *J Biol Chem* 289: 16826-16834

Saito S, Hirai N, Matsumoto C, Ohigashi H, Ohta D, Sakata K, Mizutani M (2004) Arabidopsis CYP707As encode (+)-abscisic acid 8'-hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid. *Plant Physiol* 134:1439-1449

Sasaki K (2011). *JP Patent Publicaton (Kokai)* 2011-027429 A

Sautour M, Mitaine-Offer AC, Lacaille-Dubois MA (2007) The *Dioscorea* genus: a review of bioactive steroid saponins. *J Nat Med* 61: 91-101

Sautour M, Mitaine-Offer AC, Miyamoto T, Dongmo A, Lacaille-Dubois MA (2004a) A new steroidal saponin from *Dioscorea cayenensis*. *Chem Pharm Bull (Tokyo)* 52: 1353-1355

Sautour M, Mitaine-Offer AC, Miyamoto T, Dongmo A, Lacaille-Dubois MA (2004b) Antifungal steroid saponins from *Dioscorea cayenensis*. *Planta Medica* 70: 90-92

Sawai S, Ohyama K, Yasumoto S, Seki H, Sakuma T, Yamamoto T, Takebayashi Y, Kojima M, Sakakibara H, Aoki T, Muranaka T, Saito K, Umemoto N (2014) Sterol Side Chain Reductase 2 Is a Key Enzyme in the Biosynthesis of Cholesterol, the Common Precursor of Toxic Steroidal Glycoalkaloids in Potato. *Plant Cell* 26:3763-3774

Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T (2008) Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proc Natl Acad Sci USA* 105: 14204–14209

Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, Fukushima EO, Akashi T, Aoki T, Saito K, Muranaka T (2011) Triterpene Functional Genomics in Licorice for Identification of CYP72A154 Involved in the Biosynthesis of Glycyrrhizin. *Plant Cell* 23:4112-4123

Sparg S, Light ME, Van Staden J (2004) Biological activities and distribution of plant saponins. *J Ethnopharmacol* 94: 219-243

Sue M, Yamazaki K, Yajima S, Nomura T, Matsukawa T, Iwamura H, Miyamoto T (2006) Molecular and structural characterization of hexameric β-D-glucosidases in wheat and rye. *Plant Physiol* 141: 1237-1247

Taylor MA, McDougall GJ, Stewart D (2007) Potato flavour and texture. *Potato Biol Biotechnol*: Advances and Perspectives: 525-540

Umemoto N, Sasaki K (2013) *US Patent application* 20130167271 A1

Umemoto N, Tsukahara M, Yoshioka M (2001) *JP Patent Publication (Kokai)* 2001-161373 A

Valkonen JPT, Keskitalo M, Vasara T, Pietilä L, Raman KV (1996) Potato Glycoalkaloids: A Burden or a Blessing? *Crit Rev Plant Sci* 15:1-20

Varma KR, Wickramasinghe JAF, Caspi E (1969) Biosynthesis of Biosynthesis of plant sterols. IX. The mode of oxygenation at carbon atom 26 in the formation of sapogenins from cholesterol. *J Biol Chem* 244: 3951-3957

Vazquez-Flota F, De Carolis E, Alarco AM, De Luca V (1997) Molecular cloning and characterization of desacetoxyvindoline-4–hydroxylase, a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vindoline in Catharanthus roseus (L.) G. Don. *Plant Mol. Biol* 34: 935–948

Wilmouth RC, Turnbull JJ, Welford RWD, Clifton IJ, Prescott AG, Schofield, CJ (2002) Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana. *Structure* 10: 93–103

Xu YL, Li L, Wu K, Peeters AJ, Gage DA, Zeevaart JA (1995) The GA5 locus of Arabidopsis thaliana encodes a multifunctional gibberellin 20–oxidase: molecular cloning and functional expression. *Proc Natl Acad Sci USA* 92: 6640–6644

Yamaguchi S (2008) Gibberellin metabolism and its regulation. *Annu Rev Plant Biol* 59: 225–251

Yang DJ, Lu TJ, Hwang LS (2009) Effect of endogenous glycosidase on stability of steroidal saponins in Taiwanese yam (*Dioscorea pseudojaponica* Yamamoto) during drying processes.

Food Chem 113: 155-159

Zechel DL, Withers SG (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res* 33: 11-18