Combinatorial Biochemistry of Triterpene Saponins in Plants

Jacob Pollier

Ghent University - Faculty of Sciences
Department of Plant Biotechnology and Bioinformatics
VIB - Department of Plant Systems Biology

Combinatorial Biochemistry of Triterpene Saponins in Plants

Jacob Pollier

Thesis submitted in partial fulfillment of the requirements
for the degree of Doctor (PhD) in Sciences: Biotechnology

Academic year: 2010-2011

Promotor: Prof. Dr. Alain Goossens

Dit onderzoek werd uitgevoerd in het departement Planten Systeembiologie van het Vlaams Instituut voor Biotechnologie (VIB) en de Universiteit Gent.

This work was conducted in the department of Plant Systems Biology of the Flanders Institute for Biotechnology (VIB) and the Ghent University.

Dit onderzoek werd gefinancierd door het agentschap voor Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen, Strategisch Basisonderzoek project SBO040093).

This work was supported by the Agency for Innovation by Science and Technology in Flanders (IWT-Vlaanderen, Strategisch Basisonderzoek project SBO040093).

Board of Examiners

Promotor

Prof. Dr. Alain Goossens*	VIB Department of Plant Systems Biology Department of Plant Biotechnology and Bioinformatics Faculty of Sciences Ghent University

Examination Committee

Prof. Dr. Ann Depicker (Chair)	VIB Department of Plant Systems Biology Department of Plant Biotechnology and Bioinformatics Faculty of Sciences Ghent University
Prof. Dr. Danny Geelen*	Department of Plant Production Faculty of Bioscience Engineering Ghent University
Prof. Dr. Jan Van Bocxlaer*	Department of Bio-analysis Faculty of Pharmaceutical Sciences Ghent University
Prof. Dr. Luc Pieters*	Department of Pharmaceutical Sciences Faculty of Pharmaceutical, Biomedical and Veterinary Sciences University of Antwerp
Dr. Kris Morreel*	VIB Department of Plant Systems Biology Department of Plant Biotechnology and Bioinformatics Faculty of Sciences Ghent University

Prof. Dr. Dirk Inzé — VIB Department of Plant Systems Biology
Department of Plant Biotechnology and Bioinformatics
Faculty of Sciences
Ghent University

Prof. Dr. Sofie Goormachtig — VIB Department of Plant Systems Biology
Department of Plant Biotechnology and Bioinformatics
Faculty of Sciences
Ghent University

* Members of the Reading Committee

Table of Contents

List of Abbreviations

ACN	Acetonitrile
AFLP	Amplified Fragment Length Polymorphism
BAS	β-Amyrin Synthase
BEH	Bridged Ethyl Hybrid
BLAST	Basic Local Alignment Search Tool
bp	base pair
BY-2	Bright Yellow 2
CAS	Cycloartenol Synthase
CHX	Cycloheximide
CID	Collision-induced Dissociation
CKD	Chronic Kidney Disease
CV	Coefficient of Variation
CytP450	Cytochrome P450
DDS	Dammarenediol Synthase
E-DMNT	4,8-dimethyl-1,3(E),7-nonatriene
ERAD	Endoplasmic Reticulum-associated Degradation
ESI	Electrospray Ionization
EST	Expressed Sequence Tag
FL	Full-length
FPS	Farnesyl Pyrophosphate Synthase
FT-ICR MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
HMGR	3-hydroxy-3-methylglutaryl-CoA reductase
HPLC	High-performance Liquid Chromatography
hpRNAi	Hairpin RNA-mediated Interference
HRD1	HMGR DEGRADATION 1

HTS	High-throughput Screening
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IgG	Immunoglobulin G
IPP	Isopentenyl Pyrophosphate
ISCOM	Immunostimulating Complex
IT	Ion Trap
JA	Jasmonic Acid
JA-Ile	Jasmonoyl-isoleucine
JAZ	Jasmonate ZIM-domain
LC	Liquid Chromatography
LOF	Lack of Fit
LUS	Lupeol Synthase
MeJA	Methyl Jasmonate
MEP	2-*C*-methyl-D-erythritol 4-phosphate
MKB1	MAKIBISHI1
MS	Mass Spectrometry
MTGI	*Medicago truncatula* Gene Index
MVA	Mevalonate
NF-κB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NMR	Nuclear Magnetic Resonance
ORF	Open Reading Frame
OSC	Oxidosqualene Cyclase
PCS	Pentaketide Chromone Synthase
PDB	Precursor-Directed Biosynthesis
PPD	Protopanaxadiol
PPT	Protopanaxatriol

qRT-PCR	Quantitative Real-time Polymerase Chain Reaction
RACE	Rapid Amplification of cDNA Ends
RSD	Relative Standard Deviation
QTL	Quantitative Trait Locus
RIL	Recombinant Inbred Line
RING	Really Interesting New Gene
RMA1	RING-FINGER PROTEIN WITH MEMBRANE ANCHOR 1
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SD	Standard Deviation
SEM	Scanning Electron Microscopy
SLIP	Self-Ligation of Inverse PCR Products
SPE	Solid-phase Extraction
SQE	Squalene Epoxidase
SQS	Squalene Synthase
STL	Sesquiterpene Lactone
TC	Tentative Consensus
TEM	Transmission Electron Microscopy
THAS	Thalianol Synthase
TIA	Terpenoid Indole Alkaloid
TIC	Total Ion Current
TIGR	The Institute for Genomic Research
UPLC	Ultra Performance Liquid Chromatography
UTR	untranslated region

Abstract

Plants are capable of synthesizing an overwhelming variety of secondary metabolites, many of which possess biological activities relevant for the pharmaceutical and chemical industries. Furthermore, there is an ever increasing demand for novel compounds, due to, among others, the growing drug tolerance and resistance in microorganisms and newly emerging diseases. In microorganisms, combinatorial biochemistry is a widely used tool to increase structural variation in several classes of (microbial) natural products. Despite the potential importance of plant secondary metabolites, only a limited fraction of these molecules is currently used, mostly due to their complex structure and the low production levels *in planta*. Metabolic engineering of plants has offered limited help because the molecular mechanisms steering plant secondary metabolism remain poorly characterized. Here, we used a functional genomics approach to identify candidate genes involved in the saponin biosynthesis of five different plants. After targeted metabolite profiling confirmed the induction of triterpene saponin biosynthesis by methyl jasmonate treatment, a genome-wide cDNA-AFLP transcript profiling was carried out for the five plants. Taking into account the putative functional annotation and the expression pattern of the visualized transcript tags, a set of 259 candidate genes potentially involved in saponin biosynthesis and its regulation were identified. The generated gene list provided the basis for a combinatorial biochemistry platform that targets triterpene saponins in plants. Proof of concept of combinatorial biochemistry was achieved by heterologous expression of the candidate saponin biosynthesis genes in *M. truncatula* hairy roots. Three of the generated transgenic hairy root lines were found to accumulate novel molecules, two of which were shown to be novel triterpene saponins, whereas the third line produced a set of novel, non-saponin compounds. Furthermore, the identified transcription factors and other regulators were lead candidates for studies investigating the control of the saponin biosynthesis *in planta*. This led to the identification of a RING membrane-anchor E3 ubiquitin ligase, MAKIBISHI1 (MKB1), that targets 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), the enzyme catalyzing the rate-limiting step in the mevalonate pathway, for ubiquitin-mediated proteasomal degradation, thereby controlling saponin biosynthesis.

Chapter 1

Combinatorial biochemistry in plants:

A (p)review on its potential and future exploitation

Jacob Pollier*, Tessa Moses, and Alain Goossens

* Author contribution: writing of the (review) manuscript.

Publication status:

A synopsis was accepted by the editorial board of Natural Product Reports (IF: 9.202).
The full manuscript is under review.

Abstract

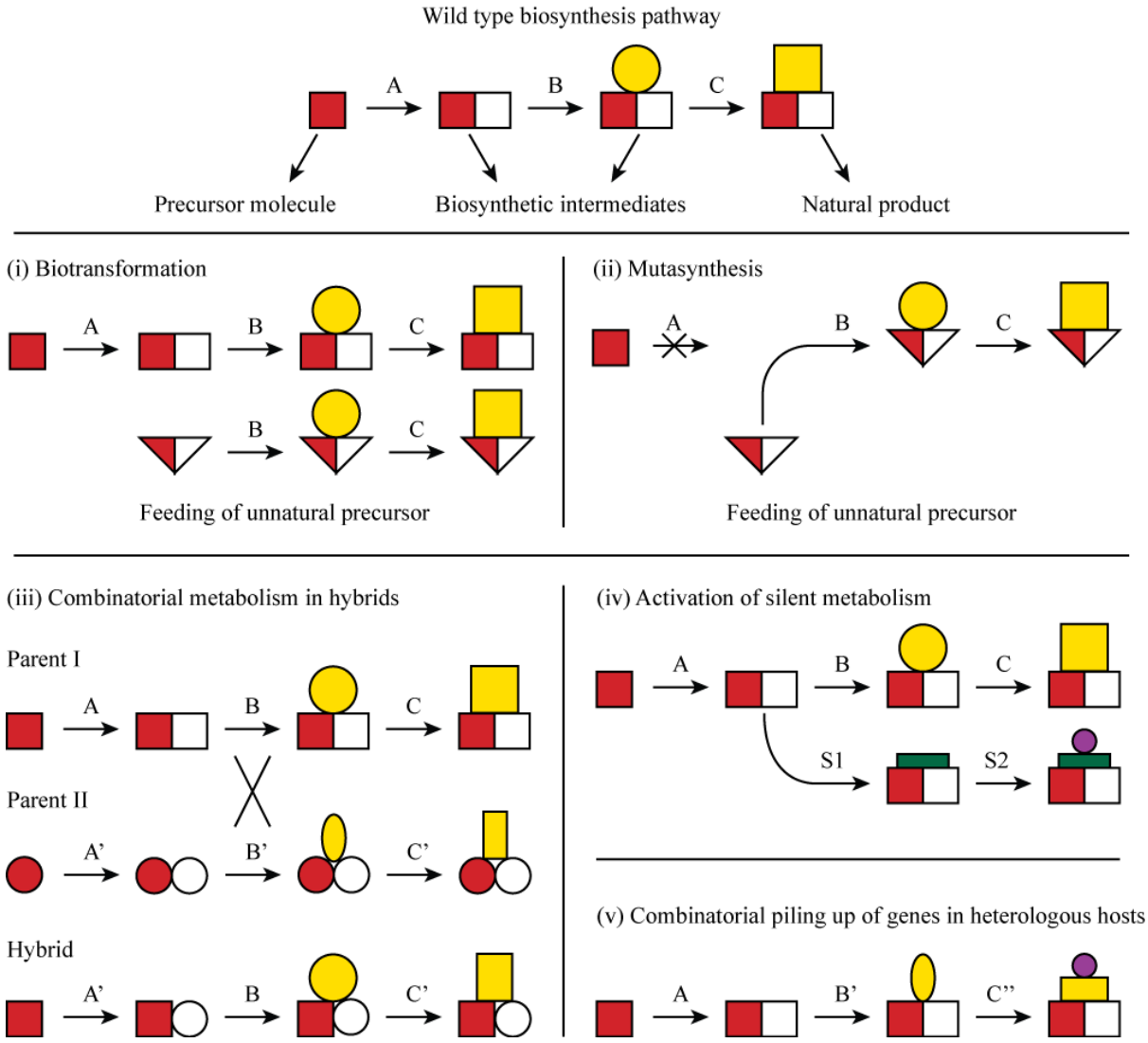
Combinatorial biochemistry, also called combinatorial biosynthesis, comprises a series of methods that establish novel enzyme-substrate combinations *in vivo*, which in turn lead to the biosynthesis of new, natural product-derived compounds that can be used in drug discovery programs. Plants are an extremely rich source of bioactive natural products and continue to possess a huge potential for drug discovery. In this review, we discuss the state-of-the-art in combinatorial biosynthesis methods to generate novel molecules from plants. We debate on the progress and potential in biotransformation, mutasynthesis, combinatorial metabolism in hybrids, activation of silent plant metabolism and synthetic biology in plants and microbes to create opportunities for combinatorial biosynthesis of plant-derived natural products, and, ultimately, for drug discovery. The therapeutic value of two classes of natural products, the terpenoid indole alkaloids and the triterpene saponins is particularly highlighted.

1. Introduction

For already thousands of years, mankind has made use of plant-derived natural products to treat diseases and injuries (Ji et al., 2009). Today, these small molecules are still used in a variety of therapeutics and play an important role in drug discovery programs (Newman and Cragg, 2007; Butler, 2008; Cragg et al., 2009; Li and Vederas, 2009). Nonetheless, there is a continuous demand for novel molecules with novel or superior pharmaceutical activities, due to, among others, newly emerging diseases and the growing drug tolerance and resistance in microorganisms. For instance, the methicillin-resistant *Staphylococcus aureus* (MRSA) causes an estimated 19,000 deaths per year in the United States (Klevens et al., 2007; Fischbach and Walsh, 2009).

Recent advancements in functional genomics and the application of multidisciplinary approaches allow picking the inherent bioactive molecules from medicinal plants, resulting in the discovery and development of new and important drugs, such as the recently approved plant-derived drugs elliptinium, galantamine and huperzine (Butler, 2008; Harvey, 2008). Chemical synthesis of natural compounds generates a better understanding of how the molecules function, and allows targeted design of novel compounds with improved or even new functions (Wender and Miller, 2009). Unfortunately, due to their structural complexity, it is often commercially unfeasible to chemically synthesize the bioactive plant-derived molecules (Chemler and Koffas, 2008). Therefore, development of novel drugs from natural plant sources by using the present-day experimental tools remains a rational approach.

Traditionally, medicinal plant extracts were the only source of plant-derived natural products used in pharmacognosy and their bioactive compounds were discovered by traditional screening (Kong et al., 2003). For instance, a large-scale screening for antitumor agents, conducted by the National Cancer Institute, led to the discovery of paclitaxel from the bark of the Pacific Yew tree, *Taxus brevifolia* (Wani et al., 1971). Yet, such traditional pharmacological screenings of medicinal plants are time consuming and expensive, drawbacks that may be partially overcome by high-throughput pharmacological screening of crude plant extracts (Littleton et al., 2005). Despite the potential of plant natural products, the pharmaceutical industry has shifted focus to the high-throughput screening of combinatorial chemistry libraries, but, to date, the output in terms of new drugs is far below the initial



Scheme 1. Various approaches for combinatorial biosynthesis of plant-derived natural products. A, B, C, S1, and S2, enzymes. Coloured symbols represent small molecule scaffolds and modifications thereof.

expectations (Ortholand and Ganesan, 2004; Koehn and Carter, 2005; Walsh and Fischbach, 2010; Welsch et al., 2010).

An alternative approach to generate new molecules or scaffolds is combinatorial biosynthesis, that can be defined as an approach in which gene products from different organisms are combined to produce new bioactive compounds. In a broad sense, combinatorial biosynthesis can be practically achieved using various approaches, such as (i) biotransformation, where the chemical modifications of compounds are carried out by biocatalysts; (ii) mutasynthesis, or feeding of compounds to mutant biocatalysts; (iii) combinatorial metabolism in hybrids;

(iv) activation of silent metabolism; and (v) synthetic biology or the combinatorial piling up of genes in heterologous hosts (Scheme 1).

In this review, we focus on the potential and the progress made in combinatorial biosynthesis in plants. The practical mechanisms available to date are discussed and a possible future outline to use plant-based combinatorial biosynthesis platforms for novel drug discovery is given.

2. Combinatorial biosynthesis in microorganisms: setting the exemplar

When Hopwood and coworkers introduced genes involved in the biosynthesis of the antibiotic actinorhodin into a medermycin-producing *Streptomyces* strain, the transformed strain produced a novel antibiotic, mederrhodin A. This compound had a structure similar to that of medermycin, but with an additional hydroxyl group typical of actinorhodin (Figure 1) (Hopwood et al., 1985). This was the first report that established the feasibility and potential of interchanging and combining genes from different organisms as a way to generate hybrid pathways that lead to the synthesis of 'hybrid' or 'new-to-nature' compounds. Since then, combinatorial biosynthesis has become a powerful platform for generating structural diversity in microbial natural products (Floss, 2006; Lam, 2007; Menzella and Reeves, 2007; Zhou et al., 2008).

In particular, combinatorial biosynthesis proved to be an excellent approach for the development of novel polyketides, a large family of natural products that include many pharmaceuticals (Cane et al., 1998; Hutchinson, 1998; Staunton and Weissman, 2001; Rix et al., 2002; Weissman and Leadlay, 2005; Floss, 2006; Lam, 2007; Menzella and Reeves, 2007; Zhou et al., 2008). Polyketides are biosynthesized from acyl-CoA monomers of simple carboxylic acids, by large, multifunctional proteins (megasynthases) that comprise several active catalytic sites. Each active site or module catalyzes one specific cycle of the polyketide chain elongation and the associated modification of the polyketide backbone (Staunton and Weissman, 2001; Floss, 2006). The large natural diversity of polyketide structures is obtained by combinations of different catalytic modules (incorporation of different acyl-CoA residues), in a different sequence and a different number of modules. After the synthesis of the polyketide backbone, various cyclization and tailoring enzymes increase the structural variability. The tremendous combinatorial biosynthesis potential of polyketides was underscored by McDaniel and coworkers. By engineering the multiple biosynthetic steps in a

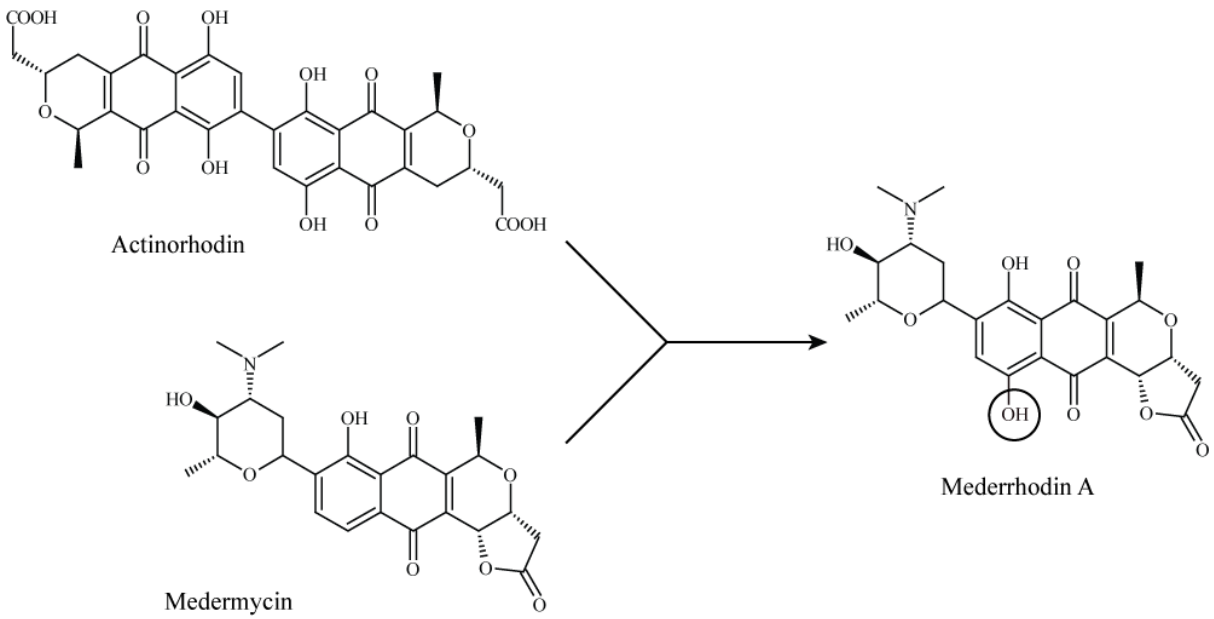


Figure 1. Mederrhodin A, produced by cloning of actinorhodin biosynthesis genes into a medermycin producing *Streptomyces* strain (Floss, 2006).

polyketide synthase that produced the macrolide ring of the widely used antibiotic erythromycin, a small combinatorial library of more than 50 novel polyketides was generated (McDaniel et al., 1999). Later, the molecular nature of the polyketide synthesis allowed the establishment of a modular method, referred to as 'Lego-ization of polyketide biosynthesis', to assemble synthetic polyketide synthases that catalyzed the production of new polyketides (Sherman, 2005; Kittendorf and Sherman, 2006).

Combinatorial biosynthesis in microorganisms has been successfully applied to other classes of microbial natural products as well, such as the indolocarbazoles, a class of compounds with potential applications in cancer therapy (Sánchez et al., 2006). Characterization of the rebeccamycin biosynthetic gene cluster paved the way for its combinatorial stacking with selected genes from other indolocarbazole biosynthetic pathways, resulting in the production of over 24 novel indolocarbazole derivatives in the microbial host (Sánchez et al., 2005; Salas and Méndez, 2009).

Recent, groundbreaking, technological advances in molecular biology, functional genomics and biotechnology will substantially increase the capacity of combinatorial biosynthesis (Zhou et al., 2008). For instance, with the emergence of next-generation sequencing techniques entire microbial genomes can be sequenced, facilitating the discovery of biosynthetic gene clusters that can be applied in combinatorial biosynthesis platforms.

Obviously, because these generic technologies are also applicable to plants, they undoubtedly will allow the exploitation of the unrivalled complex metabolic arsenal of plants (see below) and open a new era for plant-based drug discovery in the near future.

3. The metabolic power of plants

3.1.Broad substrate specificity

Although plants produce a vast array of metabolites with very diverse structures and biological activities, a large majority of the plant metabolites are derived from only a limited number of chemical scaffolds. For instance, in *Catharanthus roseus* (Madagascar periwinkle), over 120 different terpenoid indole alkaloids (TIAs) are biosynthesized from one precursor molecule, i.e. strictosidine. Starting from this precursor, the structural variation is obtained via a highly branched pathway, containing tens of specific enzymatic conversions and spontaneous chemical reactions (van der Heijden et al., 2004). Similar cases have been reported in multiple plant species and various metabolite classes. For example, over 60 different ginsenosides derived from the ubiquitous precursor 2,3-oxidosqualene have been isolated from *Panax quinquefolius* (American ginseng), and novel structures continue to be discovered (Qi et al., 2010, 2011).

Several factors collectively contribute to the enormous structural diversity of plant metabolites. Plants contain a large number of proteins capable of catalyzing regio- and stereospecific reactions. Cells from any plant species or organ express over 10,000 genes, of which a substantial fraction encode enzymes with a role in metabolism. For instance, in the model plant *Arabidopsis thaliana*, approximately 20% (5,506 out of 27,416) of the annotated protein-encoding genes in the genome are currently predicted to catalyze enzymatic reactions (TAIR10 Genome release, http://www.arabidopsis.org/ and AraCyc 7.0 Release, http://www.plantcyc.org/). Furthermore, alternative reading frames, gene fusions and alternative splicing of pre-mRNA can lead to structurally and functionally different proteins from a single gene (Reddy, 2001; Schwab, 2003). Also at the protein level, structural variation is further enlarged. Multifunctional enzymes can accept a range of substrates, catalyze distinct reactions, or form different products (Schwab, 2003). The multi-amyrin synthase from *Pisum sativum* (pea), produces eight different triterpene scaffolds from the 2,3-oxidosqualene precursor. Consequently, only one enzyme can already account for nearly 10% of the natural variation in triterpene skeletons (Morita et al., 2000). Perhaps the most

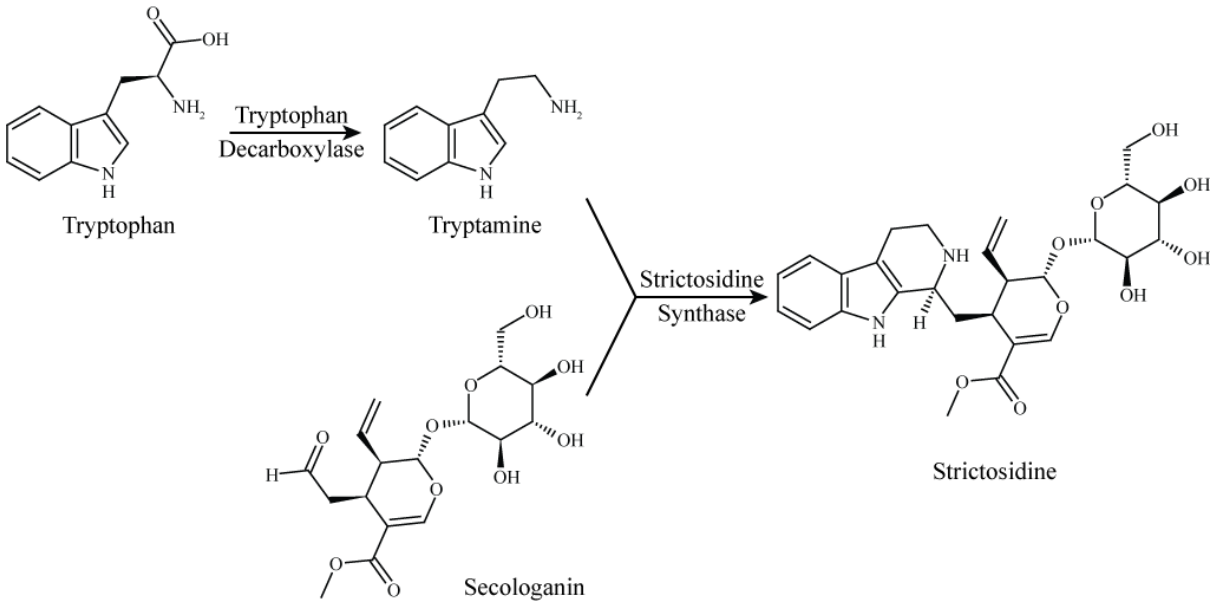


Figure 2. The condensation of tryptophan-derived tryptamine and secologanin to form strictosidine by the enzyme strictosidine synthase.

impressive cases reported to date are represented by the sesquiterpene synthases, δ-selinene synthase and γ-humulene synthase, from *Abies grandis* (Grand fir), that are capable of producing 34 and 52 different sesquiterpene scaffolds, respectively (Steele et al., 1998). Interestingly, directed evolution strategies based on the knowledge of the active sites of these promiscuous enzymes allowed the creation of enzymes with altered (new or narrower) specificity and activity (Yoshikuni et al., 2006), illustrating the enormous potential of plant enzymes for metabolic engineering towards the production of new-to-nature plant-derived products.

Ultimately, the successful development of a combinatorial biosynthesis platform in plants will depend on the availability of an arsenal of biosynthetic enzymes to ensure a broad substrate tolerance and of a wide range of possible substrates for the substrate-flexible enzymes. Then, also unnatural substrates may be incorporated into biosynthetic pathways, as illustrated by strictosidine synthase, the enzyme that catalyzes the first committed step in the TIA biosynthesis, the condensation of tryptamine and secologanin to form strictosidine (Figure 2). Strictosidine synthase tolerates various modifications on the indole ring of tryptamine and can incorporate analogs containing benzofuran and benzothiophene heterocycles. Furthermore, the resulting unnatural condensation products are accepted by the downstream glucosidase, which leads to the production of unnatural analogs of alkaloid intermediates (McCoy and O'Connor, 2006; O'Connor and Maresh, 2006). Similarly, by

introducing prokaryotic halogenases that chlorinate the indole ring of tryptophan into *C. roseus* hairy roots, chlorinated tryptophan was produced and used by the downstream TIA biosynthesis enzymes to form halogenated alkaloids *in planta* (Runguphan et al., 2010). This first example of combinatorial biosynthesis in plants underscores its huge potential.

In future combinatorial biosynthesis programs, the availability of substrates for the substrate-flexible enzymes will become key. For instance, the recombinant alcohol acyltransferase of *Fragaria* x *ananassa* (strawberry) has been shown to accept a broad range of acyl-CoAs as precursor to form the volatile esters of strawberry fruit. However, the observed substrate tolerance of the recombinant enzyme and the observed volatile esters occurring in strawberries did not correspond, indicating that the availability of the precursors rather than the substrate specificity of the enzymes governs the volatile ester formation in strawberry (Aharoni et al., 2000).

3.2.Biotransformation

Biotransformation is defined as the process in which one chemical is transformed into another, by means of a biocatalyst. As biocatalyst, whole cells (either from microorganisms, animals or plants), cell extracts, or (partially) purified enzymes can be used (Julsing et al., 2006). As purified enzymes are expensive, difficult to isolate, and need to be externally supplied with cofactors during the reaction, whole cells are considered as the preferred biocatalysts, because they are relatively cheap, easy to obtain, provide a protective environment for the enzymes from shear forces, and provide essential cofactors for the biotransformation. However, in whole cells it might be more difficult to avoid alongside side reactions and to control or reproduce the cells (de Carvalho and da Fonseca, 2006).

Because of the dazzling high number of enzymes present in whole plant cells (see above), biotransformation can serve as an excellent tool for combinatorial biosynthesis to obtain novel compounds. Typically, the biotransformation substrates are cheap and highly available products that are transformed into more rare, and, thus, more expensive compounds (Ishihara et al., 2003; Zhou and Wu, 2006). Over the last few decades, numerous studies have reported on the biotransformation of exogenous substrates by plant cells (for an excellent review from 1978 to 2003, see Ishihara et al., 2003). The advances in this field after 2003 are summarized in Table 1.

Table 1. An overview of biotransformations by plant biocatalysts from 2003 onwards. For earlier literature on plant-based biotransformation see (Ishihara et al., 2003).

Substrate	Species	Reference
Hydroxylation		
Ingenol-3-angelate	*Hordeum vulgare*, *Oryza sativa*, *Panax quinquefolium*, *Nicotiana tabacum*	(Teng et al., 2009)
Tyrosine	*Portulaca grandiflora*	(Rani et al., 2007)
Benzofuran-2-yl methyl ketone	*Daucus carota*	(Ravía et al., 2006)
Acyclic monoterpene	*Catharanthus roseus*	(Matsuki et al., 2006)
Oxidation of hydroxyl groups		
Taxadienes	*Ginkgo biloba*	(Zou et al., 2008)
1-arylethanols	*Ocimum basilicum cv. Purpurascens*	(Itoh et al., 2008)
α-pinene	*Psychotria brachyceras*, *Rauvolfia sellowii*, *Picea abies*	(Vaněk et al., 2005; Limberger et al., 2007)
Prenyl alcohols	*Cucurbita maxima*	(Nagaki et al., 2007)
Valencene	*Gynostemma pentaphyllum*	(Sakamaki et al., 2005)
Enol acetates	*Marchantia polymorpha*	(Shimoda et al., 2004)
α-ionone and β-ionone	*Caragana chamlagu*	(Sakamaki et al., 2004)
Thujopsene	*Nicotiana tabacum*, *Catharanthus roseus*	(Chai et al., 2004)
Reduction of carbonyl groups		
Aliphatic and aromatic aldehydes and ketones	*Cocos nucifera*	(Fonseca et al., 2009)
Cyclic 3-oxo-amines	*Daucus carota*	(Lacheretz et al., 2009)
Substituted fluorenones	*Malus pumila Mill.*, *Musa balbisiana Colla*, *Citrus reticulata Blanco*, *Solanum tuberosum*, *Fragaria ananassa Duch.*, *Allium fistulosum*, *Prunus spp.*, *Allium saticum*, *Allium cepa*, *Prunus pseudocerasus Lindl.*, *Zizyphus jujuba Mill.*, *Jerusalem artichoko*, *Vitis spp.*	(Xie et al., 2009a)
Aromatic ketones	*Phaseolus angularis*	(Xie et al., 2009b)
Ketones	*Raphanus sativus*, *Baccharis crispa*, *Tessaria absinthioides*, *Malus pumila*, *Daucus carota*, *Cucumis sativus*, *Allium cepa*, *Solanum tuberosum*, *Raphanus sativus*, *Ipomoea batatas*, *Arctium lappa*, *Cyanidioschyzon merolae*, *Cyanidium caldarium*	(Utsukihara et al., 2006b; Utsukihara et al., 2006a; Matsuo et al., 2008; Orden et al., 2008; Yang et al., 2008)
Substituted acenaphthenequinones	*Prunus persica*, *Daucus carota*	(Tong et al., 2008)
Aliphatic and aromatic aldehydes, ketone, ester, nitrile and amide	*Saccharum officinarum*	(Assunção et al., 2007)
Dydrogesterone	*Azadirachta indica*	(Azizuddin et al., 2008)
Substituted α-tetralones	*Daucus carota*	(Ferraz et al., 2008)
Aldehydes, ketones	*Passiflora edulis*, *Manihot esculenta*, *Manihot dulcis*	(Machado et al., 2006; Machado et al., 2008a; Machado et al., 2008b)
Carbonylic and carboxylic groups	*Actinidia chinensis*, *Allium porro*, *Convolvulus sepium*, *Daucus carota*, *Glycine soja*, *Helianthus annuus*, *Nicotiana tabacum*, *Philadelphus virginalis*, *Phytolacca decandra*, *Polygonum persicaria*, *Rauvolfia manii*, *Scorzonera hispanica*, *Solanum melanogena*, *Tagetes patula*, *Vitis vinifera*	(Villa and Molinari, 2008)
2-substituted tetrahydropyran-4-ones	*Daucus carota*	(Yadav et al., 2008)
Prochiral ketones	*Daucus carota*	(Caron et al., 2005; Blanchard and de Weghe, 2006; Yadav et al., 2007)
Phenolic 1-Benzyl-N-methyltetrahydroisoquinolines	*Corydalis spp.*, *Macleaya spp.*	(Cui et al., 2007)
Nitro-aromatic compounds	*Arracacia xanthorrhiza*, *Beta vulgaris*	(Pacheco et al., 2007)
Homochiral secondary alcohols	*Brassica oleracea botrytis*, *Cucurbita maxima*, *Cucurbita pepo*, *Cynara scolimus*, *Daucus carota*, *Foeniculum vulgare*, *Musa sapientum*	(Bruni et al., 2006)
Xenobiotic diketones	*Brassica napus*	(Orden et al., 2006)
Aromatic γ-nitroketones	*Daucus carota*	(Scarpi et al., 2005)
Aromatic nitro compounds	*Vitis vinifera*	(Li et al., 2005a)
Chinensiolide B	*Catharanthus roseus*	(Dai et al., 2005a)
3-acetylisoxazoles	*Ocimum basilicum cv. Purpurascens*	(Itoh et al., 2005)

Bromo- and methoxy-acetophenone derivatives	*Daucus carota*, *Apium graveolens*	(Mączka and Mironowicz, 2004a)
Hydrogenation of double bonds		
Progesterone	*Marchantia polymorpha*	(Haridy et al.)
Dihydroartemisinic acid	*Catharanthus roseus*, *Panax quinquefolium*	(Zhu et al.)
N-substituted maleimides	*Nicotiana tabacum*, *Catharanthus roseus*, *Marchantia polymorpha*, *Parthenocissus tricuspidata*, *Gossypium hirsutum*	(Hirata et al., 2005)
Glycosylation		
Capsaicin	*Panax ginseng*	(Katsuragi et al.)
Cinnamic acid, p-coumaric acid, caffeic acid and ferulic acid	*Eucalyptus perriniana*	(Katsuragi et al., 2010)
Naringin and naringenin	*Eucalyptus perriniana*	(Shimoda et al., 2010)
Sesamol	*Nicotiana tabacum*, *Eucalyptus perriniana*	(Shimoda et al., 2009a)
Fluorophenols	*Nicotiana tabacum*	(Shimoda et al., 2009b)
Farnesol, geraniol and (S)-perillyl alcohol	*Eucalyptus perriniana*, *Strophanthus gratus*, *Phytolacca americana*	(Shimoda et al., 2008b)
Genistein and quercetin	*Ipomoea batatas*, *Marchantia polymorpha*	(Shimoda et al., 2008a)
Hesperetin	*Ipomoea batatas*, *Eucalyptus perriniana*	(Shimoda and Hamada, 2008)
p-hydroxybenzyl alcohol	*Nicotiana tabacum*	(Peng et al., 2008)
Daidzein	*Eucalyptus spp.*, Various plants	(Sato et al., 2007; Shimoda et al., 2008c)
β-thujaplicin	*Nicotiana tabacum*	(Kwon et al., 2008)
Cholesterol	*Dioscorea bulbifera*	(Jyothishwaran and Seetharam, 2008)
Mono-, sesqui- and diterpenes	*Arabidopsis thaliana*	(Caputi et al., 2008)
p-, m-, and o-hydroxybenzoic acids	*Panax ginseng*	(Chen et al., 2008)
Taxifolin and Quercetin	*Eucalyptus perriniana*	(Kasai et al., 2007)
Monoterpene alcohols	*Eucalyptus perriniana*	(Sakamoto et al., 2007)
Capsaicin and 8-nordihydrocapsaicin	*Catharanthus roseus*	(Shimoda et al., 2007c)
Ketone and zingerone	*Phytolacca americana*	(Shimoda et al., 2007a)
Tocopherol	*Eucalyptus perriniana*	(Shimoda et al., 2007b)
Podophyllotoxin	*Hordeum vulgare*	(Teng et al., 2007)
4-hydroxybenzene derivatives	*Polygonum multiflorum*	(Yan et al., 2007)
p-coumaric acid	*Eucalyptus perriniana*	(Hirabayasi et al., 2006)
Thymol, carvacrol, and eugenol	*Eucalyptus perriniana*	(Shimoda et al., 2006b)
Vitamin E, its homologues, and vitamin A	*Phytolacca americana*, *Catharanthus roseus*	(Shimoda et al., 2006a)
Nitrosobenzene	*Spinacia oleracea*	(Tatsunami and Yoshioka, 2006)
Vitamin E	*Phytolacca americana*, *Catharanthus roseus*, *Eucalyptus perriniana*	(Kondo et al., 2006)
Multiple transformations		
Paeonol	*Panax ginseng*	(Li et al., 2005b)
Curcumin	*Catharanthus roseus*	(Kaminaga et al., 2003)
Deoxypodophyllotoxin	*Linum flavum*	(Koulman et al., 2003)
Paeonol	*Panax quinquefolium*	(Ma et al., 2010)
Carbonyl compounds	*Taxus spp*	(Sandoval et al., 2009)
Tryptophan	*Achillea millefolium*	(Marioni et al., 2008)
Valdecoxib	*Catharanthus roseus*, *Azadirachta indica*, *Capsicum annuum*, *Ervatamia heyneana*, *Nicotiana tabacum*	(Molmoori et al., 2008)
Aldehydes, ketones, α-keto esters, β-keto esters	Various plants	(Liu et al., 2008)
Zerumbone	*Caragana chamlagu*	(Sakamaki et al., 2008)
4(20),11-taxadienes	*Asparagus officinalis*, *Platycodon grandiflorum*, *Catharanthus roseus*, *Ginkgo biloba*	(Dai et al., 2005b; Dai et al., 2008)
Bromo- and methoxy-substituted 1-phenylethanol acetates	*Daucus carota*, *Apium graveolens*	(Mączka and Mironowicz, 2007)
Acetophenone derivatives, (RS)-1-phenylethanol derivatives	*Allium schoenoprasum*, *Arctium lappa*, *Arracacia xanthorrhiza*, *Beta vulgaris*, *Brassica rapa*, *Colocasia esculenta*, *Coriandrum sativum*, *Dioscorea alata*, *Ipomoea batatas* (white), *Ipomoea batatas* (red), *Manihot esculenta*, *Polymnia sonchifolia*, *Raphanus*	(Andrade et al., 2006)

	sativus, Solanum tuberosum, Zingiber officinale	
Sesquiterpenoids	*Marchantia polymorpha*	(Hegazy et al., 2005)
α-bromo and α,α'-dibromo alkanone	Various plants	(Utsukihara et al., 2007a)
α- and 6 β-santonin	*Asparagus officinalis, Catharanthus roseus, Ginkgo biloba, Platycodon grandiflorum, Taxus cuspidata, Phytolacca asinosa*	(Yang et al., 2005; Yang et al., 2006)
Terpenes	Various plants	(de Carvalho and da Fonseca, 2006)
Taxoids, cinobufagin, chinensiolide B, α-santonin, artemisinin, hydroquinone, podophyllotoxin, cantharidin	Various plants	(Zhou and Wu, 2006)
Hesperedin	*Daucus carota, Nicotiana tabacum*	(Frydman et al., 2005)
Terpenoids	Various plants	(Ishida, 2005)
14-deacetoxyl sinenxan A	*Ginkgo spp*	(Zhan et al., 2005)
Steroids, terpenes, shikimic acid derivatives	*Apium graveolens var. rapaceum, Daucus carota, Armoracia lapathifolia Gilib.*	(Mączka and Mironowicz, 2004b)
Others		
2',4'-dihydroxychalcone	*Morus nigra*	(Bolasco et al.)
Racemic aromatic alcohols	*Manihot esculenta*	(Machado et al., 2008a)
1-arylethanols	*Cyanidioschyzon merolae*	(Utsukihara et al., 2008)
Alkylcycloalkanediones; Aromatic heterocyclic compounds	*Caragana chamlagu, Wasabia japonica*	(Chai et al., 2003; Utsukihara et al., 2007b)

The applications of biotransformation in pharmacognosy are very broad and are illustrated below with some of the most representative examples. First, biotransformation can generate novel potentially bioactive molecules that may serve as candidate drugs. For instance, paeonol supplied to hairy roots of *Panax ginseng* was converted into glycosylated derivatives. Paeonol glycosides have radical scavenging effects (Matsuda et al., 2001) and after biotransformation, two of the five detected paeonol glycosides were novel (Li et al., 2005b). Second, biotransformation can alter the physical properties of a compound. For instance, the natural antioxidants genistein and quercetin that show antiallergic and anti-inflammatory activities (Wang et al., 2006) cannot be applied as drugs because of their water-insolubility and low absorbability after oral administration (Shimoda et al., 2008a). Glycosylation of flavonoids increases their water solubility (Nakajima et al., 2001), which, in turn, can possibly lead to a higher bioavailability of the compound. Indeed, feeding genistein and quercetin to cultured cells of *Ipomoea batatas* (sweet potato) and *Marchantia polymorpha* (liverwort) resulted in the glycosylation of these compounds and consequently enhanced their pharmaceutical applicability (Shimoda et al., 2008a). Third, the production of enantiopure compounds or molecules of one chirality is essential in the synthesis of organic molecules of biological interest (Rentsch, 2002). As enzymes are stereospecific biocatalysts, biotransformation can be applied to produce enantiopure compounds. For instance, hairy root cultures of *Daucus carota* (carrot) reduced the supplemented prochiral ketone acetophenone to 1-phenylethanol that consisted for more than 98% of the *S*-isomer (Caron et al., 2005).

3.3.Mutasynthesis

A major limitation of biotransformation is its inability to generate targeted modifications on the applied substrates. Knowledge of metabolic pathways generates the opportunity to introduce particular, desired modifications on the supplied precursor molecules. This is the principle behind Precursor-Directed Biosynthesis (PDB), in which a compound analogous to the precursor is added to a secondary metabolite-producing organism. This precursor analogue is used by the wild-type organism in competition with the naturally occurring precursor, finally resulting in a mixture of the desired and the wild-type compounds (Birch, 1963; Weist and Süssmuth, 2005; Kirschning et al., 2007; Weissman, 2007). PDB was originally developed for generating structural diversity in secondary metabolites of microorganisms (Kirschning et al., 2007). Nonetheless, it already proved to be applicable in plant secondary metabolism as well. Feeding of unnatural tryptamine analogues to *C. roseus* seedlings or hairy roots, led to the production of a whole range of novel alkaloids (McCoy and O'Connor, 2006). A major disadvantage of PDB, however, is the substrate specificity of the involved biosynthetic enzymes, that limits the range of unnatural precursors that can be applied. Furthermore, the internal competition of the analogous precursor with the naturally occurring endogenous precursor(s) may lead to low yields of the novel compound or to complex product mixtures from which the desired molecule is difficult to isolate.

To overcome these limitations, mutational biosynthesis or mutasynthesis can be used. Mutasynthesis is a technique in which the analogous compound or mutasynthon is supplemented to a 'biosynthetic block mutant'. In this mutant, the biosynthetic pathway of the natural compound is blocked, so that only the analogous compound is produced (Rinehart, 1977; Weist and Süssmuth, 2005; Kirschning et al., 2007). In the first study published on mutasynthesis in plants, RNA silencing of tryptophan decarboxylase (Figure 2) effectively blocked the biosynthesis of tryptamine in *C. roseus* hairy roots. As a consequence, accumulation of TIAs was eliminated in these cultures. Subsequent feeding of these 'mutant' hairy root cultures with an unnatural analogue of tryptamine yielded several novel TIAs. Importantly, the new compounds were not 'contaminated' by the naturally present TIAs in the 'wild-type' *C. roseus* (Runguphan et al., 2009). Another way to generate novel molecules via mutasynthesis is to reorient the plant biosynthetic pathways via mutated enzymes. Engineering enzymes via deliberate mutations can alter their substrate selectivity. For instance, by engineering the binding pocket of strictosidine synthase of *C. roseus*, mutant

enzymes were created that accepted strictosidine analogues that were not accepted by the native enzyme. Ultimately, novel, unnatural alkaloids were produced both in *in vitro* assays (Chen et al., 2006; Bernhardt et al., 2007; Loris et al., 2007) and in transgenic plant cell cultures (Bernhardt and O'Connor, 2009; Runguphan and O'Connor, 2009). Mutasynthesis has also been applied to generate novel plant polyketides. By mutation of three amino acids in the sequence of the pentaketide chromone synthase (PCS) from the medicinal plant *Aloe arborescens* (Krantz aloe), an enzyme was created that could condensate nine molecules of malonyl-CoA to form a novel nonaketide naphthopyrone (Abe, 2008).

3.4.Combinatorial metabolism in hybrids

Hybridization is a frequently occurring phenomenon in plants that can alter their resistance against biotic or abiotic stresses, facilitate introgression of plant traits, or ultimately lead to the formation of new plant species. Often, quantitative and qualitative variations are found in the secondary chemistry of hybrid plants (Orians, 2000). Hybrids may contain all of the secondary chemicals of both parental taxa, or may not contain certain parental chemicals, or may contain novel chemicals that are absent in either of the parental taxa (Orians, 2000; Lewinsohn and Gijzen, 2009). The latter form of hybrid metabolism is a "naturally" occurring form of combinatorial biochemistry in plants.

A classical example of natural hybridization leading to novel compounds occurs in *Helianthus* (sunflower) species that produce sesquiterpene lactones (STLs) in their flowers. *H. annuus*, a species that synthesizes niveusin- and argophyllin-type STLs in its foliar glandular trichomes, hybridizes naturally with *H. petiolaris*, a species without foliar glandular trichomes that produces budlein-type STLs. Natural hybridization between these two species resulted in three hybrid-derived species in which the major STLs differ from those produced by either of the parental taxa (Orians, 2000). Furthermore, the F_1 progeny of crosses between *H. annuus* and *H. debilis* contained novel STLs that were biosynthetic chimeras bearing the basic skeleton of one parent and the side chain of the other parent (Buschmann and Spring, 1995).

In an extensive literature review by Orians (2000) that underscored the potential to achieve combinatorial biochemistry in plants via hybridization, 32 hybrid crosses were cited that led to the accumulation of novel compounds (Orians, 2000). Several recent reports confirm this remarkable potential, a few of which are highlighted here. F_1 poplar hybrids (*Populus*

deltoides cv ‘S9-2’ x *P. trichocarpa* cv ‘V24’) have been shown to accumulate an unknown flavanone compound that could not be detected in either of the parents. Furthermore, a metabolite quantitative trait locus had been identified that was postulated to control the accumulation level of the unknown flavanone compound (Morreel et al., 2006). Novel flavonoid compounds have also been detected in hybrids of *Ziziphus mauritiana* (jujube) (Bhargava et al., 2005) and *Zea mays* (maize) (Lee et al., 2009). In addition, three novel terpenoids were detected in hybrids of *Artemisia tridentata* (sagebrush) (Byrd et al., 1999), and two new odour compounds were found in *Ophrys* (orchid) hybrids (Vereecken et al., 2010).

Even crossings within different accessions, cultivars, or genotypes of a single plant species can result in combinatorial biochemistry, as demonstrated by a striking metabolite study of a collection of *A. thaliana* L*er*/Cvi Recombinant Inbred Lines (RILs). Using untargeted liquid chromatography-time of flight mass spectrometry-based metabolomics, it was shown that more than one third of the masses detected in the RILs were not detectable in either of the parental accessions. This qualitative variety was attributed to the recombination of the parental genomes (Keurentjes et al., 2006).

Nonetheless, in the above ‘classical’ breeding cases, interspecific crossing barriers represent a major pitfall, that in laboratory practice can partially be overcome by somatic hybridization methods. In symmetric somatic hybridization, hybrids with the complete genomes of both parents are obtained through protoplast fusions, whereas in asymmetric somatic hybridizations, the donor genome is fragmented prior to protoplast fusion, yielding hybrids with the complete recipient genome, but a partial donor genome. An elegant example of a somatic hybridization experiment aimed at overcoming crossing barriers between wild and cultivated species of potatoes, more specifically of the wild-type, non-tuberous *Solanum brevidens* and the commonly cultivated, tuberous species, *S. tuberosum* (Laurila et al., 1996). The major glycoalkaloids normally found in the parental strains are tomatine (aglycone tomatidine) in *S. brevidens* and solanidine in *S. tuberosum*. The somatic hybrids, however, possessed large quantities of a novel glycoalkaloid, demissidine, not found in any of the parents, and four other unidentified compounds, along with all of the parental glycoalkaloids (Figure 3) (Laurila et al., 1996). In another example, the interspecific crossing barrier of *Dendranthema morifolium* (chrysanthemum) and *Artemisia vulgaris* (mugwort) was overcome via embryo rescue. The resulting intergeneric hybrids produced a small number of

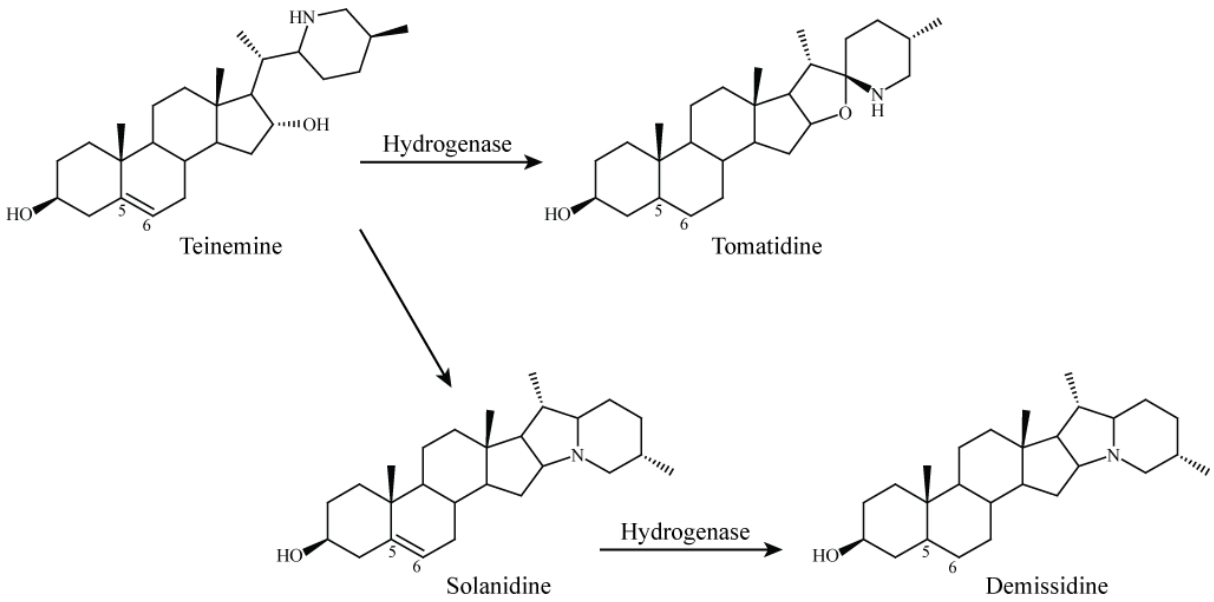


Figure 3. Biosynthesis of a novel alkaloid, demissidine, in somatic hybrid plants of *Solanum brevidens* and *Solanum tuberosum*. Both parental plants produce teinemine, whereas solanidine and tomatidine are produced only in *S. tuberosum* and *S. brevidens*, respectively (Laurila et al., 1996; Oksman-Caldentey and Inzé, 2004).

hybrid-specific terpenes besides the terpenes present in one or both of the parental plants (Deng et al., 2010).

3.5.Activation of silent metabolism

Recent advances in genomics indicate that the full metabolic capacity of organisms is far from comprehensively exploited. Bioinformatics has been used as a tool in microorganisms to identify a number of 'orphan' biosynthetic pathways, i.e. pathways of which the encoded natural products are unknown (Gross, 2007; Chiang et al., 2011). For example, in the actinomycete *Streptomyces coelicolor*, only three antibiotics and a spore pigment had been described before its complete genome had been sequenced. Predictions based on the genome sequence revealed 18 more gene clusters potentially encoding the biosynthesis of various secondary metabolites (Bentley et al., 2002). Such observations illustrate that the currently used natural product compendia do not represent the full potential of the producing organisms (Cichewicz, 2010; Chiang et al., 2011).

More than microorganisms, plant cells are capable of accumulating several thousands of secondary metabolites simultaneously, of which the exact biochemical role remains mostly unknown to date. Despite the already dazzling number of accumulating compounds, it is becoming clear that in plants, like in microorganisms, many biosynthetic pathways are not

fully active, or at least not detectably at the resolution of the present metabolite profiling technologies. Accordingly, many enzymes that have been annotated exist without any apparent endogenous substrate, pointing towards the occurrence of 'occult' metabolic capabilities in plants. This phenomenon is also defined as silent metabolism (Lewinsohn and Gijzen, 2009) and can be 'awakened' by either providing the appropriate substrate to the existing enzymes, or, alternatively, by triggering the inactive parts of the biosynthetic pathways. As such, ectopic onset of silent metabolism might generate unprecedented opportunities to obtain new enzyme-substrate combinations and serve as a tool to obtain novel plant-derived compounds.

In the literature, several silent metabolism cases in plants have been described. For instance, only recently *Arabidopsis* has been found capable of producing volatile terpenes in its flowers (Chen et al., 2003). Before this discovery, the terpene-producing capacity of the plant was only suspected, based on the 32 predicted terpene synthase genes in its genome. In another study, scanning of the genomic regions of predicted oxidosqualene cyclases revealed a gene operon that was exclusively expressed in *Arabidopsis* roots and potentially involved in plant secondary metabolism. Functional analysis of the oxidosqualene cyclase gene in yeast showed that it catalyzed the formation of thalianol, a compound that had never been reported in plants before. Subsequently, the occurrence of low amounts of the triterpene thalianol was demonstrated in *Arabidopsis* roots (Field and Osbourn, 2008). Similarly, although in *Arabidopsis* no complex alkaloids have been detected yet, genome scanning revealed the presence of numerous homologues of enzymes involved in alkaloid biosynthesis in medicinal plants, including of the tropane, pyrrolidine and indole alkaloid types (Facchini et al., 2004). Thus, a well-studied plant such as *Arabidopsis* might also be capable of producing still undiscovered alkaloids, a hypothesis that nonetheless still needs to be substantiated (D'Auria and Gershenzon, 2005). More examples of silent metabolism have been recently described elsewhere (Lewinsohn and Gijzen, 2009).

In a broader sense, tissue-, cell-, or even organelle-specific expression of genes in plants can be seen as silent metabolism as well. Activation of this alternative form of silent metabolism via ectopic expression can give rise to novel enzyme-substrate combinations, resulting in the generation of novel compounds. For instance, ectopic overexpression of the *Vitis vinifera* (grapevine) fruit-specific VvMybPA1 and VvMybPA2 MYB transcription factors activated protoanthocyanin biosynthesis in hairy roots. The MYB-overexpressing grapevine hairy roots

accumulated one specific class of protoanthocyanins, B-ring trihydroxylated units, that are absent in control hairy roots (Terrier et al., 2009).

Finally, even (re)directing a metabolic pathway to a specific cellular compartment may activate silent metabolism and subsequent production of novel compounds. The targeting of a strawberry linalool/nerolidol synthase to the mitochondria in transgenic *Arabidopsis*, instead of to its natural site, the cytosol, produced small amounts of 4,8-dimethyl-1,3(*E*),7-nonatriene (*E*-DMNT), a new compound for *Arabidopsis* (Kappers et al., 2005). Hence, *Arabidopsis* contains the necessary enzymes to recognize and convert the product of the linalool/nerolidol synthase, (3*S*)-(*E*)-nerolidol, to *E*-DMNT, a signal molecule in plant defence that attracts enemies of herbivores (Bouwmeester et al., 1999; Kappers et al., 2005). Similarly, reallocating patchoulol synthase from its natural cytosolic location to the chloroplast of *Nicotiana tabacum* (tobacco), resulted in the production of high amounts of patchoulol, and an entire suite of novel sesquiterpenes (Wu et al., 2006).

4. Synthetic biology - heterologous biosynthesis of plant products in microorganisms

One of the major challenges in the development of plant-derived drugs is obtaining sufficient amounts of the compounds. In a lot of cases, the complex bioactive molecules are available at very low concentrations *in planta*. Furthermore, plant growth can be very slow, or the molecules are found in plant species that are rare, threatened with extinction, or that are recalcitrant to genetic transformation. As such, the establishment of combinatorial biosynthesis programs *in planta* may be compromised, at least in some plant species.

As an alternative approach, sets of plant biosynthesis enzymes can be expressed in microorganisms, such as *Escherichia coli* or *Saccharomyces cerevisiae* (yeast), in combination, if necessary, with modified genes to optimize production in the microbial host. This approach is called synthetic biology and has become fashion thanks to the latest technological and strategical advances in this field (Li and Vederas, 2009; Roberts and Kolewe, 2010). Heterologous expression of plant enzymes in microorganisms, may generate unlimited opportunities for combinatorial biochemistry, provided the necessary genetic information is available. Genes from different plant species can be merged, thus generating new-to-nature enzyme combinations, that finally lead to production of novel compounds and

their structural variation can be increased by addition of unnatural substrates to the recombinant microorganisms.

To obtain combinatorial biosynthesis it is essential to engineer the recombinant microorganisms to produce the desired plant compounds. To date, several classes of medicinal plant natural products can be produced in microorganisms. For instance, the simultaneous expression of phenylalanine ammonia lyase from the yeast *Rhodotorula rubra*, 4-coumarate CoA ligase from the actinomycete *Streptomyces coelicolor*, and chalcone synthase from the plant *Glycyrrizha echinata* (licorice) in *E. coli* led to the production of plant-specific flavanones in the recombinant bacteria (Hwang et al., 2003; Kaneko et al., 2003). Later metabolic engineering efforts in *E. coli* led to the fermentative production of several other types of plant secondary metabolites with increased amounts of the desired compounds including flavonoids compounds (Watts et al., 2004; Fowler and Koffas, 2009; Horinouchi, 2009; Du et al., 2010), terpenoids (Chemler and Koffas, 2008), and alkaloids (Minami et al., 2008). For instance, amorphadiene, the sesquiterpene precursor molecule of the powerful anti-malarial drug artemisinin, is produced in *E. coli* by simultaneous expression of the genes of the yeast mevalonate pathway and a synthetic amorphadiene synthase gene in the bacteria (Martin et al., 2003). As more enzymes in the artemisinin biosynthesis pathway were characterized, the production of up to 100 mg/L artemisinic acid was established in *E. coli* (Chang et al., 2007). Similarly, the production of the diterpene taxadiene, an intermediate of the biosynthesis of the anticancer drug paclitaxel was obtained in *E. coli* via the ectopic expression of genes encoding isopentenyl diphosphate isomerase from the yeast *Schizosaccharomyces pombe*, geranylgeranyl diphosphate from the bacteria *Erwinia herbicola* and taxadiene synthase from the plant *Taxus brevifolia* (Huang et al., 2001). Subsequent metabolic engineering of taxadiene producing *E. coli* increased production levels of taxadiene and taxadien-5α-ol, the next step in the biosynthesis of paclitaxel (Ajikumar et al., 2010; Liu and Khosla, 2010).

Besides *E. coli*, *S. cerevisiae* cells can be used for the biosynthesis of plant natural products. The production of up to 100 mg/L of artemisinic acid is achieved in yeast by engineering its mevalonate pathway and simultaneous overexpression of amorphadiene synthase and CYP71AV1, a cytochrome P450 monooxygenase, from *A. annua* (Ro et al., 2006). Furthermore, the biosynthesis of various plant metabolites, such as taxol (DeJong et al., 2006; Engels et al., 2008), several types of flavonoids (Jiang et al., 2005; Chemler and Koffas,

2008; Fowler and Koffas, 2009; Horinouchi, 2009), alkaloids (Chemler and Koffas, 2008; Hawkins and Smolke, 2008), and saponins (Shibuya et al., 2006; Seki et al., 2008) has been (partially) reconstructed in yeast.

Once a (partial) biosynthesis pathway is established in the target microbe, novel compounds can be produced via the combinatorial biosynthesis techniques discussed above. For instance, when unnatural carboxylic acids were fed to recombinant *E. coli* cells harboring artificial flavanone and stilbene biosynthesis pathways, more than 30 unnatural, novel polyketides were produced (Katsuyama et al., 2007). Moreover, the produced unnatural scaffolds could be further modified by the incorporation of downstream modification enzymes within the artificial biosynthetic pathways (Horinouchi, 2008, 2009; Katsuyama et al., 2010; Werner et al., 2010). Combinatorial biosynthesis of plant products in microorganisms can also be achieved by combinatorial piling up of plant biosynthetic genes in the microbes. Expressing genes from different plant species may lead to new-to-nature enzyme combinations, finally resulting in the biosynthesis of new-to-nature compounds in the recombinant microorganism.

5. Bioactive terpenoid indole alkaloids

To highlight the relevance of the structural variety for bioactivity, in terms of drug potential, two classes of plant natural products, the terpenoid indole alkaloids (TIAs) and the saponins will be discussed.

Alkaloids are nitrogen-containing molecules with a low molecular weight and are mostly derived from the amino acids Phe, Tyr, Trp, Lys, and Arg. Approximately 20% of all plant species produce alkaloids. Traditionally, humans make use of the biological potency of alkaloids for hunting, or treatment of diseases, but also for less noble purposes, such as executions and warfare (De Luca and St Pierre, 2000).

An important set of alkaloids are the TIAs, a class of structurally diverse molecules of which many show interesting pharmaceutical activities. TIAs are mainly found in plants belonging to the Apocynaceae, Icacinaceae, Loganiaceae, Nyssaceae, and Rubiaceae families and are derived from the common precursor molecule strictosidine, a condensation product of the terpenoid moiety secologanin and the tryptophan-derived indole moiety tryptamine (Figure 2). The structural variability of TIAs in the different TIA-producing plant species is due to species-specific conversions of the common strictosidine precursor molecule that are

either enzymatically catalyzed or occur through spontaneous chemical reactions (Memelink et al., 2001; O'Connor and Maresh, 2006). These species-specific decorations have profound effects, not only for the final complexity of the structure itself, but above all on its bioactivity, which sets a splendid example of the enormous resources for combinatorial biochemistry available within the plant kingdom, but still remains hardly explored. The structure and bioactivity of the most potent TIA drugs currently known are briefly described below.

The predominantly investigated TIA-producing plant species is *C. roseus* that is capable of producing over 120 different TIAs in its various organs (Memelink et al., 2001; van der Heijden et al., 2004), some of which have therapeutic activities and are used as pharmaceuticals. Vinblastine and its oxidized form vincristine (Figure 4), two minor constituents of the complex *C. roseus* TIA mixture, are commercially used as anti-cancer drugs. They are antineoplastic drugs that bind tubulin, thereby inhibiting the assembly of microtubule structures, that, in turn, disrupts the assembly of the mitotic spindle and the kinetochore, and ultimately leads to an arrest in the metaphase of rapidly dividing cells, including cancer cells (Jordan et al., 1991; van der Heijden et al., 2004). Vinblastine (brand names, e.g. Velban® and Velbe®) is mainly administered to treat lymphoma, but is also applied to treat bladder cancer, testicular cancer or Kaposi's sarcoma, whereas vincristine (brand names, e.g. Oncovin® and Vincasar Pfs®) is used to treat lymphoma, leukaemia, and other types of cancer. *C. roseus* also accumulates the antihypertensive and antiarrhythmic TIA compounds ajmalicine (or raubasine; brand names, e.g. Circolene®, Hydrosarpan®, Isoarteril® and Lamuran®) and serpentine (Figure 4) (van der Heijden et al., 2004; Dutta et al., 2005; O'Connor and Maresh, 2006).

Another important genus of TIA producing plants is *Rauwolfia*, in which the most pharmacologically important TIAs are reserpine, ajmaline, yohimbine, and ajmalicine (Figure 4). Reserpine (brand name Harmonyl®) is administered as an antihypertensive drug to control high blood pressure and as an antipsychotic to treat, for instance, schizophrenia, but is nowadays rarely used because of its numerous side-effects and the development of drugs with better efficacy. Ajmaline is an antiarrhythmic drug that is, due to its short half-life, very useful for acute intravenous treatments of abnormal heart rhythms, such as ventricular tachycardia and atrial fibrillations in patients with the Wolff–Parkinson–White syndrome (Wellens et al., 1980). Furthermore, it is also the drug of choice to search for ST elevations in

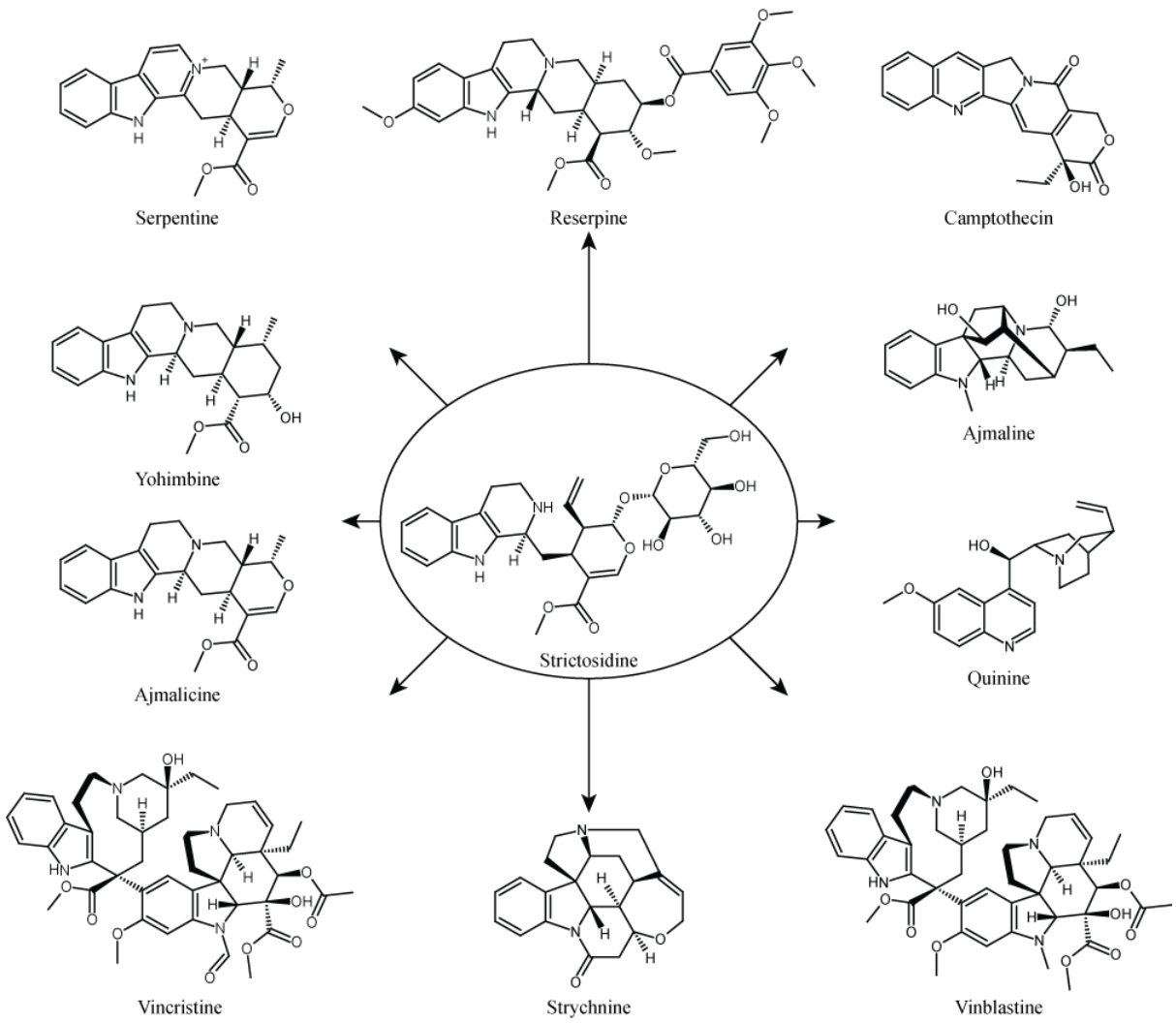


Figure 4. Overview of the molecular structures of pharmacologically important TIAs derived from strictosidine.

electrocardiograms of patients suspected of suffering from the Brugada syndrome (Miyazaki et al., 1996), a genetic disease causing unanticipated ventricular fibrillation, which can lead to cardiac arrest and death (Brugada and Brugada, 1992). Yohimbine (brand name Procomil®) is a psychoactive drug with stimulant and aphrodisiac effects. Yohimbine hydrochloride has potential clinical applications for the treatment of erectile dysfunction (Ernst and Pittler, 1998; Drewes et al., 2003), and can also be used as a mydriatic drug to aid eye examinations (Phillips et al., 2000). *Rauwolfia* species produce yohimbine only in limited amounts, hence the main natural source of the molecule is the African tree *Pausinystalia yohimbe* (yohimbe) (Drewes et al., 2003).

The Chinese ‘happy tree’ *Camptotheca acuminata* was the first source of another important TIA, the antineoplastic compound camptothecin (Wall et al., 1966), which was later found in

a few other plant species as well (Govindachari and Viswanathan, 1972; Tafur et al., 1976; Gunasekera et al., 1979; Arisawa et al., 1981; Dai et al., 1999; Zhou et al., 2000; Saito et al., 2001; Klausmeyer et al., 2007). Camptothecin (Figure 4) binds to the mammalian DNA topoisomerase I and, thereby, interferes with the breakage-reunion reaction of the topoisomerase I via the stabilization of a reaction intermediate, namely the cleavable complex. The S-phase specific toxicity of camptothecin is caused by the interaction between the stabilized topoisomerase I cleavable complexes and the moving replication forks, that provokes an irreversible fork arrest and ultimately highly toxic double-strand DNA breaks, that finally cause cell death (Hsiang et al., 1985; Hsiang et al., 1989; Liu, 1989; D'Arpa et al., 1990; Ryan et al., 1991; Thomas et al., 2004). Nowadays, camptothecin and its more water-soluble derivatives topotecan and irinotecan are common anticancer agents (Thomas et al., 2004). Topotecan hydrochloride (Hycamtin®) is mainly used to treat ovarian and lung cancer. CPT-11 or irinotecan hydrochloride (Camptosar® and Campto®) is used to treat colon cancer, frequently in combination with other types of chemotherapeutical drugs.

Other important well-known TIAs include strychnine and quinine (Figure 4). Strychnine is a poison produced by *Strychnos nux-vomica* (poison nut) that provokes muscular convulsions and eventually leads to death through asphyxia or sheer exhaustion and is used as a pesticide against small vertebrates, such as birds and rodents. Quinine is a bitter tasting TIA with antipyretic, antimalarial, analgesic, and anti-inflammatory activities. It is extracted from the bark of the Cinchona tree and was for a long time the drug-of-choice to treat malaria. The distinctive bitter taste of tonic water is due to the added quinine.

To date, most advances in molecular characterization of the complex and strictly regulated TIA biosynthesis pathway have been made in *C. roseus*. The first step in the biosynthesis of all TIAs is the condensation of the precursor molecules tryptamine and secologanin by the enzyme strictosidine synthase. The indole moiety tryptamine is derived from the primary metabolite tryptophan via a single enzymatic step, the decarboxylation of the amino acid by the enzyme tryptophan decarboxylase (Figure 2). The biosynthesis of the terpenoid (iridoid) moiety secologanin consists of several enzymatic steps, of which only a few are characterized, and starts with the hydroxylation of the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway derived monoterpenoid geraniol (Facchini and De Luca, 2008; Murata et al., 2008).

After the condensation of secologanin and tryptamine to strictosidine, the various TIA structures are generated via species-specific enzymatic reactions and spontaneous chemical modifications (Memelink et al., 2001). Because of the complexity of the pathway, most of the enzymes involved in the biosynthesis of the various TIA backbones are still unknown. Nonetheless, several subpathways, such as the six-step enzymatic conversion of tabersonine to vindoline in *C. roseus* (Loyola-Vargas et al., 2007) or the 10-step ajmaline biosynthesis in *R. serpentina* (Ruppert et al., 2005) have already been elucidated.

6. Bioactive triterpene saponins

Saponins compose a structurally diverse group of amphipathic glycosides, with a steroid, steroidal alkaloid, or triterpenoid aglycone backbone. The aglycone or sapogenin can be covalently linked to one (monodesmosidic) or more (di- or tridesmosidic) sugar chains *via* a glycosidic bond. Saponins are surface-active agents that, when dissolved in water, form a colloidal solution that foams upon shaking. Accordingly, they are used as emulsifier and foaming agents in the food and beverage industries. Importantly, many saponins also possess pharmacological properties and are, therefore, commonly used in phytotherapy and the cosmetic industry.

Here, we will focus on the structure-activity relationships of only one type of saponins, the triterpene saponins. All terpenoids are derived by repetitive fusions of a branched five-carbon unit, isopentenyl diphosphate (IPP). The number of fused IPP units determines the classification of the terpenoids. Triterpenoids, to which also the steroids belong, contain 30 carbon atoms. The precursor C_{30} unit, called squalene, is generated by the joining of two C_{15} units, or farnesyl diphosphate, each of which constitutes of three IPP units (Croteau et al., 2000) and is usually oxidized to its 2,3-epoxide, 2,3-oxidosqualene. The first committed step in the biosynthesis of triterpene saponins is the cyclization of 2,3-oxidosqualene. This step forms the branch point between the primary sterol and the secondary saponin metabolism (Figure 5). Predating the emergence of plants, cycloartenol synthase (CAS), involved in primary steroid metabolism, is the ancestor enzyme of all plant oxidosqualene cyclases (OSCs) involved in secondary metabolism (Figure 5) (Phillips et al., 2006). Most plants contain several OSCs that catalyze the formation of different triterpene skeletons. For instance, the genome of *Arabidopsis* contains up to 13 predicted OSC-encoding genes, producing various triterpenes of which the function *in planta* remains mostly unknown (Field

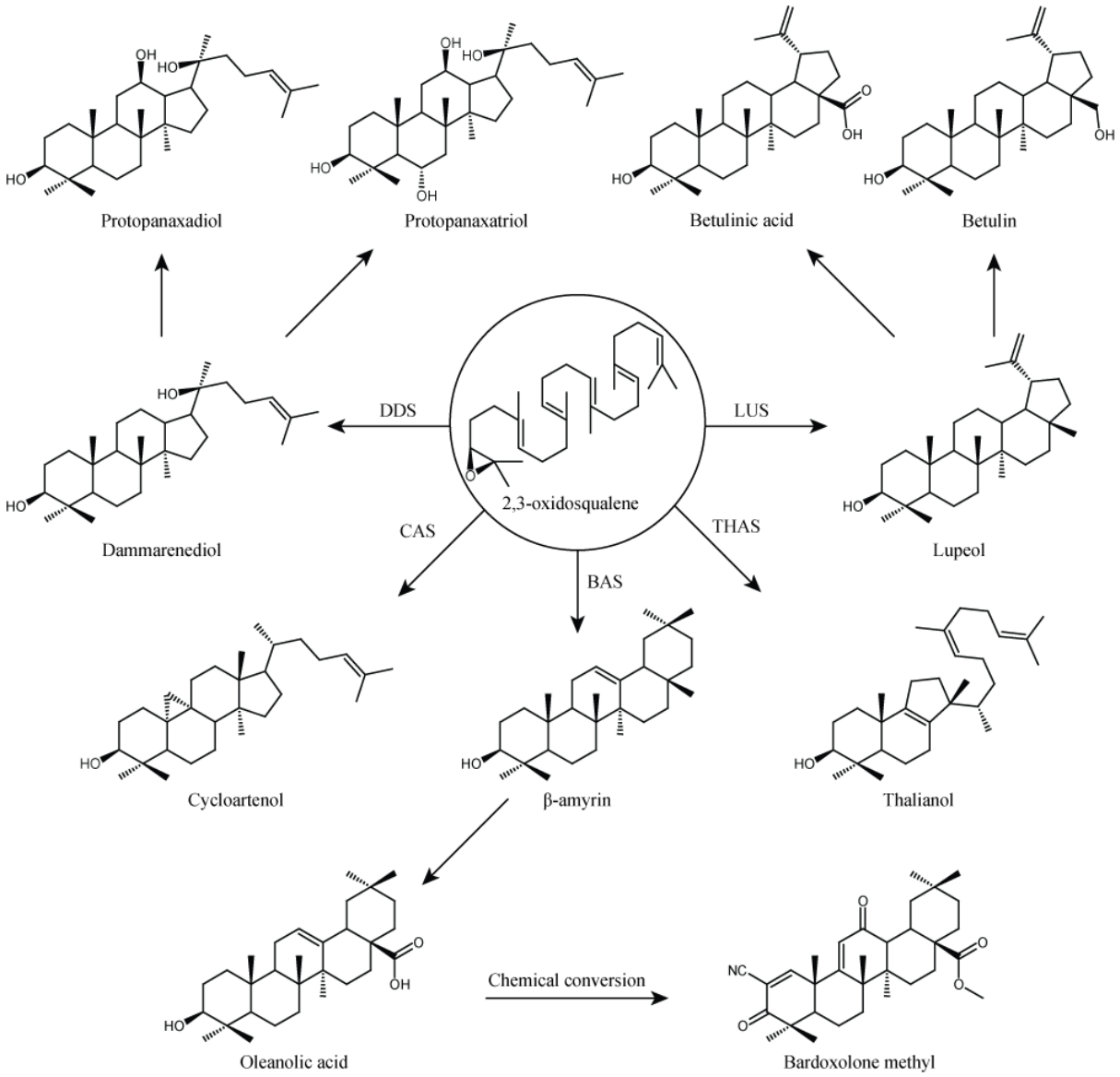


Figure 5. Biosynthesis of triterpene saponin aglycones from the common precursor 2,3-oxidosqualene. Following cyclization of 2,3-oxidosqualene by specific oxidosqualene cyclases, the cyclization products are oxidized at various positions. BAS, β-amyrin synthase; CAS, cycloartenol synthase; DDS, dammarenediol synthase; LUS, lupeol synthase; THAS, thalianol synthase.

and Osbourn, 2008). As such, nearly 100 distinct non-steroidal triterpene skeletons have been described in plants (Phillips et al., 2006). Usually, two distinct types of additional modifications of the cyclization products complete the saponin biosynthesis and dramatically increase the structural diversity. Firstly, cytochrome P450 mediated oxidations of the cyclization products give rise to different sapogenins. For instance, in the model plant *Medicago truncatula* (barrel medic), oxidation of the oleanane-type cyclization product β-amyrin leads to at least 10 different sapogenins (Pollier et al., 2011) and in the medicinal plant *Panax ginseng* (Asian ginseng), two different saponin backbones, protopanaxadiol

(PPD) and protopanaxatriol (PPT) are formed by oxidation of the dammarane-type cyclization product dammarenediol (Lü et al., 2009) (Figure 5). Subsequently (and/or concomitantly), the different sapogenins can be glycosylated at different positions. The attached sugar chains can be linear or branched and can contain different monosaccharides, such as glucose, galactose, glucuronic acid, xylose, or rhamnose. Other, less common, modifications, such as acylations, can further increase the structural variability of the saponins. Even within a single plant species, an enormously rich mixture of saponins can occur. For example, in *P. ginseng* alone already over 40 ginsenosides have been reported, some of which have only been discovered recently (Shin et al., 2000; Dou et al., 2001; Zou et al., 2002; Yoshikawa et al., 2007; Tung et al., 2009; Tung et al., 2010).

One of the most promising pharmacological activities of triterpenoid saponins is their ability to activate the mammalian immune system, making them candidate vaccine adjuvants (Skene and Sutton, 2006; Sun et al., 2009a). As candidate adjuvant, QS-21, a saponin isolated from the bark of *Quillaja saponaria* (soapbark), has been evaluated in several clinical trials (Sun et al., 2009a). QS-21 is far more effective than the currently accepted adjuvants for human vaccines, calcium and aluminium salts (Cipolla et al., 2008), but toxicity, haemolytic side effects and instability in the aqueous phase limit its current use in human vaccines (Liu et al., 2002; Skene and Sutton, 2006; Yang et al., 2007; Sun et al., 2009a). Insights into the structure-activity relationship of saponins, in regard to their adjuvant and haemolytic effects, might enable to overcome some of these undesired effects. A lot of adjuvant compounds have an amphipathic structure. Saponins, with their lipophilic aglycone backbone and hydrophilic sugar chains have amphipathic properties as well. A study on the correlation between adjuvant activity and amphipathic structure of saponins has clearly shown an increasing adjuvanticity with an increasing amphipathicity of the molecules (Oda et al., 2003), whereas haemolytic effects are considered to be influenced by the affinity of the lipophilic aglycone to cholesterol in the cell membranes. The degree of affinity depends on the type of aglycone and side chains occurring on the aglycones (Sun et al., 2009a).

Also the immunostimulating complex (ISCOM®)-based vaccines make use of the adjuvant properties of saponins, while overcoming their haemolytic effects (Morein et al., 1984; Pearse and Drane, 2005; Sun et al., 2009b). Iscoms are spherical, cage-like structures, typically 40 nm in diameter, that form spontaneously when saponin, cholesterol and phospholipid are mixed under controlled conditions. Traditionally, the ISCOM®-based

vaccines also make use of 'Quil A' saponins, which is a crude saponin extract from the bark of *Q. saponaria* (Morein et al., 1984). However, because the Quil A extract is a heterogeneous mixture of dozens of closely related saponins that vary in their chemical or biological activities, variations in the composition of the mixture may lead to unpredictable *in vivo* effects. Hence, the current tendency is to use more purified fractions of the Quil A saponin extract, such as QS-21 (Sjölander et al., 1998; Pearse and Drane, 2005; Skene and Sutton, 2006). Several other medicinal plants contain structurally different saponins with adjuvant activity (Song and Hu, 2009; Sun et al., 2009a) and novel immunostimulatory nanoparticles could be developed by incorporation of these saponins in iscom-like nanostructures. Ginsomes, for instance, are nanoparticles containing saponins from the medicinal plant *P. ginseng*, and enhance the IgG responses in mice, when they are administered together with the antigen, instead of the antigen alone (Song et al., 2009).

Triterpene saponins can also serve in cancer treatments. One of the most studied triterpene molecules in this context, is a compound isolated from the bark of *Betula alba* (white birch) trees, betulinic acid (Figure 5), which was shown to be a melanoma-specific cytotoxic agent able to inhibit tumour growth *in vitro* and *in vivo*, without systemic toxicity (Pisha et al., 1995; Fulda, 2009; Laszczyk, 2009). Betulinic acid affects the mitochondrial membrane potential of cancer cells, resulting in the release of mitochondrial proteins, such as cytochrome c, in the cytosol that trigger apoptosis in the cancer cells (Fulda et al., 1998; Laszczyk, 2009). In contrast to other antineoplastic drugs, betulinic acid has no cytotoxic effects on nonmalignant cells, but the exact mechanism for the selectivity of betulinic acid towards malignant cells remains to be discovered (Fulda, 2009). Betulin (Figure 5), which differs from betulinic acid only on position 28, can also induce apoptosis in cancer cells (Pyo et al., 2009; Rzeski et al., 2009; Li et al., 2010), but, despite its almost identical structure and similar effects on cancer cells, its mode of action is different. Unlike betulinic acid, betulin fails to trigger the release of cytochrome c from isolated mitochondria, suggesting a different mechanism of triggering the intrinsic apoptotic pathway (Li et al., 2010).

In addition to adjuvant and anti-cancer properties, a plethora of other bioactivities have been attributed to triterpenoid saponins (Francis et al., 2002; Sparg et al., 2004; Güçlü-Üstündağ and Mazza, 2007; Sun et al., 2009a; Augustin et al., 2011). For instance, ginsenosides from *P. ginseng* alone are pharmacologically significant as anti-cancer, anti-diabetic,

neuroprotective, cardioprotective, radioprotective, hypolipidemic, anti-amnesic and anti-aging agents (Ji and Gong, 2007; Liang and Zhao, 2008; Ng et al., 2008; Lu et al., 2009).

All of the pharmacologically potent triterpenes described above are naturally occurring compounds. Chemical modification of the triterpene scaffolds has led to the synthesis of semisynthetic triterpenoids with improved pharmacological properties, in comparison to the parent compounds. The most promising candidate drug of this sort is bardoxolone methyl, a chemical derivative of oleanolic acid (Figure 5). Recently, bardoxolone methyl has been positively evaluated in a Phase-IIB clinical trial, in which its oral administration reduced the stage of disease and improved measures of kidney function in patients with chronic kidney disease (CKD) and Type 2 diabetes (see http://www.abbott.com/global/url/pressReleases). Bardoxolone methyl acts by activating the transcription factor Nrf2, inducing the antioxidant and detoxification enzymes of the phase 2 response (Dinkova-Kostova et al., 2005; Yates et al., 2007). This increased cellular antioxidant content prevents activation of inflammatory signalling pathways, such as NF-κB and, thereby, prevents chronic "metabolic inflammation" that promotes Type 2 diabetes and CKD. Besides bardoxolone methyl, several other synthetic triterpenoids activate Nrf2. Essential features required for the activation of Nrf2 are the presence of Michael acceptor groups (enones) on both the A and C rings of the molecules, and a nitrile group at the C_2 position (Yates et al., 2007). Nrf2 activation may also be useful for treatment of several types of cancer. Tumour-associated immune suppression is mediated by reactive oxygen species and contributes to tumour growth and limits the activity of cancer immunotherapy. Accordingly, blocking the tumour-associated immune suppression by treatment with bardoxolone methyl improved the immune response in tumour-bearing mice and cancer patients (Nagaraj et al., 2010).

7. Outlook: Gene discovery for combinatorial biosynthesis of plant natural products

Plant-derived natural products are and will undoubtedly remain very important for various drug discovery programs. The structural diversity of these complex molecules may be further enhanced with various combinatorial biosynthesis techniques, and the low production amounts *in planta* may be tackled with the heterologous expression of plant genes in microorganisms, such as *E. coli* or yeasts, or ultimately, plants themselves might be metabolically engineered. To exploit the full potential of plant natural products, knowledge of

the genes involved in the biosynthesis of the metabolites is essential, but, until now, most, if not all, of the biosynthesis pathways remain incompletely understood.

Over the past five years, the commercialization of several next-generation sequencing techniques, such as pyrosequencing in the 454 system (Margulies et al., 2005), bridge PCR in the Solexa technology (Bentley, 2006), sequencing by ligation in the AB SOLiD and Polonator platforms (Shendure et al., 2005; Shendure and Ji, 2008), single molecule DNA sequencing in the HeliScope platform (Harris et al., 2008), or, more recently, the application of single-molecule real-time DNA sequencing in the PacBio RS system (Eid et al., 2009; Metzker, 2010), has revolutionized functional genomics. This tremendous progress now brings the whole genome or transcriptome of any plant species within reach and will undoubtedly greatly accelerate gene discovery for secondary metabolite pathways. Simultaneously, the enormous amount of data generated by such next-generation sequencing platforms demands for efficient strategies to identify and select the relevant biosynthetic genes from the pools of thousands of sequenced genes, or at least for efficient ways to exploit them. One way to solve this problem would be the use of elicitors, such as the phytohormone jasmonate. Elicitors induce the production of a compound of interest in plants or plant cell cultures and most often this effect is caused by the transcriptional activation of the genes involved in the biosynthesis of these compounds. Hence, comparison of transcriptomes at various stages before and after elicitor treatment will allow pinpointing the desired subsets of the total gene pools as candidate genes responsible for the biosynthesis of the secondary metabolites (Oksman-Caldentey and Inzé, 2004; Pauwels et al., 2009). Several studies have already demonstrated the feasibility of this approach. For instance, treatment of tobacco Bright Yellow 2 (BY-2) cell cultures with jasmonate induces nicotine biosynthesis in the cells. Genome-wide transcript profiling of the jasmonate-elicited cell cultures resulted in the establishment of a tobacco gene compendium (Goossens et al., 2003) from which several new proteins, including enzymes, regulators and transporters involved in nicotine biosynthesis were identified in subsequent functional screens (De Sutter et al., 2005; Häkkinen et al., 2007; Morita et al., 2009; De Boer et al., 2011; Lackman et al., 2011). In a similar series of studies in the triterpene saponin-producing model legume, *M. truncatula*, the response of the metabolome and transcriptome to different elicitors was investigated (Suzuki et al., 2005). Analysis of unknown genes co-regulated with known triterpene saponin biosynthesis genes led to the identification and characterization of glucosyltransferases involved in triterpene saponin biosynthesis (Achnine et al., 2005; Naoumkina et al., 2010). The identification of

such sets of candidate genes involved in the biosynthesis of plant natural products will allow a high-throughput approach for combinatorial biosynthesis in plants or heterologous microbial systems. Alternatively, 'metagenomics' approaches may be employed to screen for plant enzymes that can catalyze the synthesis of new-to-nature compounds. Indeed, capturing the genetic resources of complex microbial communities in metagenome libraries already allowed the discovery of a wealth of new enzymatic diversity, of entirely new sequence classes and novel functionalities (Ferrer et al., 2009).

Hence, we can conclude that, aided by the recent technical advances in gene discovery and functional genomics, the large-scale implementation of different combinatorial biosynthesis techniques in plants will offer a bright green future for drug discovery in plant-derived natural products.

Acknowledgements

The authors thank Martine De Cock for help in preparing the manuscript. This work was supported by the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" project SBO040093) and the VIB International PhD Fellowship Program for predoctoral fellowships (to T.M.).

References

Abe, I. (2008). Engineering of plant polyketide biosynthesis. Chem Pharm Bull **56**, 1505-1514.

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

Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A.M.M.L., De Vos, R.C.H., van der Voet, H., Jansen, R. C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J., and O'Connell, A.P. (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell **12**, 647-661.

Ajikumar, P.K., Xiao, W.H., Tyo, K.E.J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., and Stephanopoulos, G. (2010). Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science **330**, 70-74.

Andrade, L.H., Utsunomiya, R.S., Omori, A.T., Porto, A.L.M., and Comasseto, J.V. (2006). Edible catalysts for clean chemical reactions: Bioreduction of aromatic ketones and biooxidation of secondary alcohols using plants. J Mol Catal, B Enzym **38**, 84-90.

Arisawa, M., Gunasekera, S.P., Cordell, G.A., and Farnsworth, N.R. (1981). Plant anticancer agents XXI. Constutuents of *Merrilliodendron megacarpum*. Planta Med **43**, 404-407.

Assunção, J.C.C., Machado, L.L., Lemos, T.L.G., Cordell, G.A., and Monte, F.J.Q. (2007). Sugar cane juice for the bioreduction of carbonyl compounds. J Mol Catal, B Enzym **52-53**, 194-198.

Augustin, J.M., Kuzina, V., Andersen, S.B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry **72**, 435-457.

Azizuddin, Saifullah, Khan, S., Choudhary, M.I., and Atta-Ur-Rahman. (2008). Biotransformation of dydrogesterone by cell suspension cultures of *Azadirachta indica*. Turk J Chem **32**, 141-146.

Bentley, D.R. (2006). Whole-genome re-sequencing. Curr Opin Genet Dev **16**, 545-552.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., and Hopwood, D.A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417**, 141-147.

Bernhardt, P., and O'Connor, S.E. (2009). Opportunities for enzyme engineering in natural product biosynthesis. Curr Opin Chem Bio **13**, 35-42.

Bernhardt, P., McCoy, E., and O'Connor, S.E. (2007). Rapid identification of enzyme variants for reengineered alkaloid biosynthesis in periwinkle. Chem Biol **14**, 888-897.

Bhargava, R., Shukla, A.K., Chauhan, N., Vashishtha, B.B., and Dhandar, D.G. (2005). Impact of hybridity on flavonoid spectrum of ber (*Ziziphus mauritiana* Lamk.). Environ Expl Bot **53**, 135-138.

Birch, A.J. (1963). The biosynthesis of antibiotics. Pure Appl Chem **7**, 527-537.

Blanchard, N., and van de Weghe, P. (2006). *Daucus carota* L. mediated bioreduction of prochiral ketones. Org Biomol Chem **4**, 2348-2353.

Bolasco, A., Fioravanti, R., Rossi, F., Rossi, P., and Vitali, A. (2010). Use of cyclodextrins in biotransformation reactions with cell cultures of *Morus nigra*: biosynthesis of prenylated chalcone isocordoin. Biotechnol Appl Biochem **56**, 77-84.

Bouwmeester, H.J., Verstappen, F.W.A., Posthumus, M.A., and Dicke, M. (1999). Spider mite-induced (3*S*)-(*E*)-nerolidol synthase activity in cucumber and Lima bean. The first dedicated step in acyclic C11-homoterpene biosynthesis. Plant Physiol **121**, 173-180.

Brugada, P., and Brugada, J. (1992). Right bundle branch block, persistent ST segment elevation and sudden cardiac death: A distinct clinical and electrocardiographic syndrome - A multicenter report. J Am Coll Cardiol **20**, 1391-1396.

Bruni, R., Fantin, G., Maietti, S., Medici, A., Pedrini, P., and Sacchetti, G. (2006). Plants-mediated reduction in the synthesis of homochiral secondary alcohols. Tetrahedron-Asymmetry **17**, 2287-2291.

Buschmann, H., and Spring, O. (1995). Sesquiterpene lactones as a result of interspecific hybridization in *Helianthus* species. Phytochemistry **39**, 367-371.

Butler, M.S. (2008). Natural products to drugs: natural product-derived compounds in clinical trials. Nat Prod Rep **25**, 475-516.

Byrd, D.W., McArthur, E.D., Wang, H., Graham, J.H., and Freeman, D.C. (1999). Narrow hybrid zone between two subspecies of big sagebrush, *Artemisia tridentata* (Asteraceae). VIII. Spatial and temporal pattern of terpenes. Biochem Syst Ecol **27**, 11-25.

Cane, D.E., Walsh, C.T., and Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. Science **282**, 63-68.

Caputi, L., Lim, E.K., and Bowles, D.J. (2008). Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. Chemistry **14**, 6656-6662.

Caron, D., Coughlan, A.P., Simard, M., Bernier, J., Piché, Y., and Chênevert, R. (2005). Stereoselective reduction of ketones by *Daucus carota* hairy root cultures. Biotechnol Lett **27**, 713-716.

Chai, W., Sakamaki, H., Kitanaka, S., and Horiuchi, C.A. (2003). Biotransformation of cycloalkanediones by *Caragana chamlagu*. Bull Chem Soc Jpn **76**, 177-182.

Chai, W., Hayashida, Y., Sakamaki, H., and Horiuchi, C.A. (2004). The biocatalytic oxidation of thujopsene by plant cultured-cells. J Mol Catal, B Enzym **27**, 55-60.

Chang, M.C.Y., Eachus, R.A., Trieu, W., Ro, D.K., and Keasling, J.D. (2007). Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. Nat Chem Biol **3**, 274-277.

Chemler, J.A., and Koffas, M.A.G. (2008). Metabolic engineering for plant natural product biosynthesis in microbes. Curr Opin Biotechnol **19**, 597-605.

Chen, F., Tholl, D., D'Auria, J.C., Farooq, A., Pichersky, E., and Gershenzon, J. (2003). Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. Plant Cell **15**, 481-494.

Chen, S., Galan, M.C., Coltharp, C., and O'Connor, S.E. (2006). Redesign of a central enzyme in alkaloid biosynthesis. Chem Biol **13**, 1137-1141.

Chen, X., Zhang, J., Liu, J.H., and Yu, B.Y. (2008). Biotransformation of *p*-, *m*-, and *o*-hydroxybenzoic acids by *Panax ginseng* hairy root cultures. J Mol Catal, B Enzym **54**, 72-75.

Chiang, Y.M., Chang, S.L., Oakley, B.R., and Wang, C.C.C. (2011). Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. Curr Opin Chem Biol **15**, 137-143.

Cichewicz, R.H. (2010). Epigenome manipulation as a pathway to new natural product scaffolds and their congeners. Nat Prod Rep **27**, 11-22.

Cipolla, L., Peri, F., and Airoldi, C. (2008). Glycoconjugates in cancer therapy. Anticancer Agents Med Chem **8**, 92-121.

Cragg, G.M., Grothaus, P.G., and Newman, D.J. (2009). Impact of natural products on developing new anti-cancer agents. Chem Rev **109**, 3012-3043.

Croteau, R., Kutchan, T.M., and Lewis, N.G. (2000). Natural products (secondary metabolites). In Biochemistry & Molecular Biology of Plants, B. Buchanan, W. Gruissem, and R. Jones, eds (Rockville, MD: American Society of Plant Biologists), pp. 1250-1268.

Cui, W., Iwasa, K., Sugiura, M., Takeuchi, A., Tode, C., Nishiyama, Y., Moriyasu, M., Tokuda, H., and Takeda, K. (2007). Biotransformation of phenolic 1-benzyl-*N*-methyltetrahydroisoquinolines in plant cell cultures followed by LC/NMR, LC/MS, and LC/CD. J Nat Prod **70**, 1771-1778.

D'Auria, J.C., and Gershenzon, J. (2005). The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. Curr Opin Plant Biol **8**, 308-316.

Dai, J.G., Yang, L., Sakai, J., and Ando, M. (2005a). Biotransformation of chinensiolide B and the cytotoxic activities of the transformed products. J Mol Catal, B Enzym **33**, 87-91.

Dai, J.G., Yang, L., Sakai, J., and Ando, M. (2005b). Combined biotransformations of 4(20),11-taxadienes. Tetrahedron **61**, 5507-5517.

Dai, J.G., Qu, R.J., Zou, J.H., and Chen, X.G. (2008). Structural diversification of taxanes by whole-cell biotransformation. Tetrahedron **64**, 8102-8116.

Dai, J.R., Hallock, Y.F., Cardellina, J.H., and Boyd, M.R. (1999). 20-*O*-β-glucopyranosyl camptothecin from *Mostuea brunonis*: A potential camptothecin pro-drug with improved solubility. J Nat Prod **62**, 1427-1429.

D'Arpa, P., Beardmore, C., and Liu, L.F. (1990). Involvement of nucleic acid synthesis in cell killing mechanisms of topoisomerase poisons. Cancer Res **50**, 6919-6924.

De Boer, K., Tilleman, S., Pauwels, L., Vanden Bossche, R., De Sutter, V., Vanderhaeghen, R., Hilson, P., Hamill, J.D., and Goossens, A. (2011). AP2/ERF and bHLH tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. Plant J, **in press** doi: 10.1111/j.1365-313X.2011.04566.x.

de Carvalho, C.C.C.R., and da Fonseca, M.M.R. (2006). Biotransformation of terpenes. Biotechnol Adv **24**, 134-142.

De Luca, V., and St Pierre, B. (2000). The cell and developmental biology of alkaloid biosynthesis. Trends Plant Sci **5**, 168-173.

De Sutter, V., Vanderhaeghen, R., Tilleman, S., Lammertyn, F., Vanhoutte, I., Karimi, M., Inzé, D., Goossens, A., and Hilson, P. (2005). Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J **44**, 1065-1076.

DeJong, J.M., Liu, Y.L., Bollon, A.P., Long, R.M., Jennewein, S., Williams, D., and Croteau, R.B. (2006). Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*. Biotechnol Bioeng **93**, 212-224.

Deng, Y.M., Chen, S.M., Lu, A.M., Chen, F.D., Tang, F.P., Guan, Z.Y., and Teng, N.J. (2010). Production and characterisation of the intergeneric hybrids between *Dendranthema morifolium* and *Artemisia vulgaris* exhibiting enhanced resistance to chrysanthemum aphid (*Macrosiphoniella sanbourni*). Planta **231**, 693-703.

Dinkova-Kostova, A.T., Liby, K.T., Stephenson, K.K., Holtzclaw, W.D., Gao, X.Q., Suh, N., Williarrli, C., Risingsong, R., Honda, T., Gribble, G.W., Sporn, M.B., and Talalay, P. (2005). Extremely potent triterpenoid inducers of the phase 2 response: Correlations of protection against oxidant and inflammatory stress. Proc Natl Acad Sci U S A **102**, 4584-4589.

Dou, D.Q., Chen, Y.J., Liang, L.H., Pang, F.G., Shimizu, N., and Takeda, T. (2001). Six new dammarane-type triterpene saponins from leaves of *Panax ginseng*. Chem Pharm Bull **49**, 442-446.

Drewes, S.E., George, J., and Khan, F. (2003). Recent findings on natural products with erectile-dysfunction activity. Phytochemistry **62**, 1019-1025.

Du, H., Huang, Y.B., and Tang, Y.X. (2010). Genetic and metabolic engineering of isoflavonoid biosynthesis. Appl Microbiol Biotechnol **86**, 1293-1312.

Dutta, A., Batra, J., Pandey-Rai, S., Singh, D., Kumar, S., and Sen, J. (2005). Expression of terpenoid indole alkaloid biosynthetic pathway genes corresponds to accumulation of related alkaloids in *Catharanthus roseus* (L.) G. Don. Planta **220**, 376-383.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X.X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C.C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., and Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. Science **323**, 133-138.

Engels, B., Dahm, P., and Jennewein, S. (2008). Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng **10**, 201-206.

Ernst, E., and Pittler, M.H. (1998). Yohimbine for erectile dysfunction: A systematic review and meta-analysis of randomized clinical trials. J Urol **159**, 433-436.

Facchini, P.J., and De Luca, V. (2008). Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. Plant J **54**, 763-784.

Facchini, P.J., Bird, D.A., and St-Pierre, B. (2004). Can *Arabidopsis* make complex alkaloids? Trends Plant Sci **9**, 116-122.

Ferraz, H.M.C., Bianco, G.G., Bombonato, F.I., Andrade, L.H., and Porto, A.L.M. (2008). Bioreduction of substituted α-tetralones promoted by *Daucus carota* root. Quim Nova **31**, 813-817.

Ferrer, M., Beloqui, A., Timmis, K.N., and Golyshin, P.N. (2009). Metagenomics for mining new genetic resources of microbial communities. J Mol Microbiol Biotechnol **16**, 109-123.

Field, B., and Osbourn, A.E. (2008). Metabolic diversification - independent assembly of operon-like gene clusters in different plants. Science **320**, 543-547.

Fischbach, M.A., and Walsh, C.T. (2009). Antibiotics for emerging pathogens. Science **325**, 1089-1093.

Floss, H.G. (2006). Combinatorial biosynthesis - Potential and problems. J Biotechnol **124**, 242-257.

Fonseca, A.M., Monte, F.J.Q., de Oliveira, M.D.F., de Mattos, M.C., Cordell, G.A., Braz-Filho, R., and Lemos, T.L.G. (2009). Coconut water (*Cocos nucifera* L.)-A new biocatalyst system for organic synthesis. J Mol Catal, B Enzym **57**, 78-82.

Fowler, Z.L., and Koffas, M.A.G. (2009). Biosynthesis and biotechnological production of flavanones: current state and perspectives. Appl Microbiol Biotechnol **83**, 799-808.

Francis, G., Kerem, Z., Makkar, H.P.S., and Becker, K. (2002). The biological action of saponins in animal systems: a review. Br J Nutr **88**, 587-605.

Frydman, A., Weisshaus, O., Huhman, D.V., Sumner, L.W., Bar-Peled, M., Lewinsohn, E., Fluhr, R., Gressel, J., and Eyal, Y. (2005). Metabolic engineering of plant cells for biotransformation of hesperedin into neohesperidin, a substrate for production of the low-calorie sweetener and flavor enhancer NHDC. J Agric Food Chem **53**, 9708-9712.

Fulda, S. (2009). Betulinic acid: A natural product with anticancer activity. Mol Nutr Food Res **53**, 140-146.

Fulda, S., Scaffidi, C., Susin, S.A., Krammer, P.H., Kroemer, G., Peter, M.E., and Debatin, K.M. (1998). Activation of mitochondria and release of mitochondrial apoptogenic factors by betulinic acid. J Biol Chem **273**, 33942-33948.

Goossens, A., Häkkinen, S.T., Laakso, I., Seppänen-Laakso, T., Biondi, S., De Sutter, V., Lammertyn, F., Nuutila, A.M., Söderlund, H., Zabeau, M., Inzé, D., and Oksman-Caldentey, K.M. (2003). A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proc Natl Acad Sci U S A **100**, 8595-8600.

Govindachari, T.R., and Viswanathan, N. (1972). Alkaloids of *Mappia foetida*. Phytochemistry **11**, 3529-3531.

Gross, H. (2007). Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. Appl Microbiol Biotechnol **75**, 267-277.

Güçlü-Üstündağ, O., and Mazza, G. (2007). Saponins: properties, applications and processing. Crit Rev Food Sci Nutr **47**, 231-258.

Gunasekera, S.P., Badawi, M.M., Cordell, G.A., Farnsworth, N.R., and Chitnis, M. (1979). Plant anticancer agents X. Isolation of camptothecin and 9-methoxycamptothecin from *Ervatamia heyneana*. J Nat Prod **42**, 475-477.

Häkkinen, S.T., Tilleman, S., Świątek, A., De Sutter, V., Rischer, H., Vanhoutte, I., Van Onckelen, H., Hilson, P., Inzé, D., Oksman-Caldentey, K.M., and Goossens, A. (2007). Functional characterisation of genes involved in pyridine alkaloid biosynthesis in tobacco. Phytochemistry **68**, 2773-2785.

Haridy, M.S.A., Hegazy, M.E.F., Mohamed, A.H., Paré, P.W., and Hirata, T. (2010). Biotransformation of progesterone by cultured cells of *Marchantia polymorpha*. Z Naturforsch, C J Biosci **65**, 599-602.

Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., Dimeo, J., Efcavitch, J.W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S.R., Steinmann, K., Thayer, E., Tyurina, A., Ward, R., Weiss, H., and Xie, Z. (2008). Single-molecule DNA sequencing of a viral genome. Science **320**, 106-109.

Harvey, A.L. (2008). Natural products in drug discovery. Drug Discov Today **13**, 894-901.

Hawkins, K.M., and Smolke, C.D. (2008). Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. Nat Chem Biol **4**, 564-573.

Hegazy, M.E.F., Kuwata, C., Matsushima, A., Ahmed, A.A., and Hirata, T. (2005). Biotransformation of sesquiterpenoids having α,β-unsaturated carbonyl groups with cultured plant cells of *Marchantia polymorpha*. J Mol Catal, B Enzym **39**, 13-17.

Hirabayasi, M., Miyata, A., Hamada, H., and Furuya, T. (2006). Biotransformation of cinnamic acid derivatives by plant cultured cells - Abstracts. J Mol Catal, B Enzym **42**, 124-124.

Hirata, T., Takarada, A., Hegazy, M.E.F., Sato, Y., Matsushima, A., Kondo, Y., Matsuki, A., and Hamada, H. (2005). Hydrogenation of the C-C double bond of maleimides with cultured plant cells. J Mol Catal, B Enzym **32**, 131-134.

Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A.M., Floss, H.G., and Ōmura, S. (1985). Production of hybrid antibiotics by genetic engineering. Nature **314**, 642-644.

Horinouchi, S. (2008). Combinatorial biosynthesis of non-bacterial and unnatural flavonoids, stilbenoids and curcuminoids by microorganisms. J Antibiot **61**, 709-728.

Horinouchi, S. (2009). Combinatorial biosynthesis of plant medicinal polyketides by microorganisms. Curr Opin Chem Biol **13**, 197-204.

Hsiang, Y.H., Lihou, M.G., and Liu, L.F. (1989). Arrest of replication forks by drug-stabilized topoisomerase I-DNA cleavable complexes as a mechanism of cell killing by camptothecin. Cancer Res **49**, 5077-5082.

Hsiang, Y.H., Hertzberg, R., Hecht, S., and Liu, L.F. (1985). Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. J Biol Chem **260**, 4873-4878.

Huang, Q.L., Roessner, C.A., Croteau, R., and Scott, A.I. (2001). Engineering *Escherichia coli* for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. Bioorg Med Chem **9**, 2237-2242.

Hutchinson, C.R. (1998). Combinatorial biosynthesis for new drug discovery. Curr Opin Microbiol **1**, 319-329.

Hwang, E.I., Kaneko, M., Ohnishi, Y., and Horinouchi, S. (2003). Production of plant-specific flavanones by *Escherichia coli* containing an artificial gene cluster. Appl Environ Microbiol **69**, 2699-2706.

Ishida, T. (2005). Biotransformation of terpenoids by mammals, microorganisms, and plant-cultured cells. Chem Biodivers **2**, 569-590.

Ishihara, K., Hamada, H., Hirata, T., and Nakajima, N. (2003). Biotransformation using plant cultured cells. J Mol Catal, B Enzym **23**, 145-170.

Itoh, K., Sakamaki, H., Nakamura, K., and Horiuchi, C.A. (2005). Biocatalytic asymmetric reduction of 3-acetylisoxazoles. Tetrahedron-Asymmetry **16**, 1403-1408.

Itoh, K.I., Nakamura, K., Utsukihara, T., Sakamaki, H., and Horiuchi, C.A. (2008). Stereoselective oxidation of racemic 1-arylethanols by basil cultured cells of *Ocimum basilicum* cv. *Purpurascens*. Biotechnol Lett **30**, 951-954.

Ji, H.F., Li, X.J., and Zhang, H.Y. (2009). Natural products and drug discovery: Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? EMBO Rep **10**, 194-200.

Ji, W., and Gong, B.Q. (2007). Hypolipidemic effects and mechanisms of *Panax notoginseng* on lipid profile in hyperlipidemic rats. J Ethnopharmacol **113**, 318-324.

Jiang, H.X., Wood, K.V., and Morgan, J.A. (2005). Metabolic engineering of the phenylpropanoid pathway in *Saccharomyces cerevisiae*. Appl Environ Microbiol **71**, 2962-2969.

Jordan, M.A., Thrower, D., and Wilson, L. (1991). Mechanism of inhibition of cell proliferation by *Vinca* alkaloids. Cancer Res **51**, 2212-2222.

Julsing, M.K., Koulman, A., Woerdenbag, H.J., Quax, W.J., and Kayser, O. (2006). Combinatorial biosynthesis of medicinal plant secondary metabolites. Biomol Eng **23**, 265-279.

Jyothishwaran, G., and Seetharam, Y.N. (2008). Biotransformation of cholesterol to diosgenin by freely suspended and immobilised cells of *Dioscorea bulbifera* L. J Asian Nat Prod Res **10**, 139-145.

Kaminaga, Y., Nagatsu, A., Akiyama, T., Sugimoto, N., Yamazaki, T., Maitani, T., and Mizukami, H. (2003). Production of unnatural glucosides of curcumin with drastically enhanced water solubility by cell suspension cultures of *Catharanthus roseus*. Febs Lett **555**, 311-316.

Kaneko, M., Hwang, E.I., Ohnishi, Y., and Horinouchi, S. (2003). Heterologous production of flavanones in *Escherichia coli*: potential for combinatorial biosynthesis of flavonoids in bacteria. J Ind Microbiol Biotechnol **30**, 456-461.

Kappers, I.F., Aharoni, A., van Herpen, T.W.J.M., Luckerhoff, L.L.P., Dicke, M., and Bouwmeester, H.J. (2005). Genetic engineering of terpenoid metabolism attracts bodyguards to *Arabidopsis*. Science **309**, 2070-2072.

Kasai, F., Ishihara, K., Shimoda, K., Nakajima, N., and Hamada, H. (2007). Biotransformation of taxifolin and quercetin by plant cultured cells of *Eucalyptus perfiniana*. J Mol Catal, B Enzym **48**, 116.

Katsuragi, H., Shimoda, K., Ohiro, A., and Hamada, H. (2010). Glycosylation of capsaicinoids with *Panax ginseng* stimulated by salicylic acid. Acta Biol Hung **61**, 449-456.

Katsuragi, H., Shimoda, K., Kubota, N., Nakajima, N., and Hamada, H. (2010). Biotransformation of cinnamic acid, *p*-coumaric acid, caffeic acid, and ferulic acid by plant cell cultures of *Eucalyptus perriniana*. Biosci Biotechnol Biochem **74**, 1920-1924.

Katsuyama, Y., Funa, N., Miyahisa, I., and Horinouchi, S. (2007). Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in *Escherichia coli*. Chem Biol **14**, 613-621.

Katsuyama, Y., Hirose, Y., Funa, N., Ohnishi, Y., and Horinouchi, S. (2010). Precursor-directed biosynthesis of curcumin analogs in *Escherichia coli*. Biosci Biotechnol Biochem **74**, 641-645.

Keurentjes, J.J.B., Fu, J.Y., de Vos, C.H.R., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. (2006). The genetics of plant metabolism. Nat Genet **38**, 842-849.

Kirschning, A., Taft, F., and Knobloch, T. (2007). Total synthesis approaches to natural product derivatives based on the combination of chemical synthesis and metabolic engineering. Org Biomol Chem **5**, 3245-3259.

Kittendorf, J.D., and Sherman, D.H. (2006). Developing tools for engineering hybrid polyketide synthetic pathways. Curr Opin Biotechnol **17**, 597-605.

Klausmeyer, P., McCloud, T.G., Melillo, G., Scudiero, D.A., Cardellina II, J.H., and Shoemaker, R.H. (2007). Identification of a new natural camptothecin analogue in targeted screening for HIF-1α inhibitors. Planta Med **73**, 49-52.

Klevens, R.M., Morrison, M.A., Nadle, J., Petit, S., Gershman, K., Ray, S., Harrison, L.H., Lynfield, R., Dumyati, G., Townes, J.M., Craig, A.S., Zell, E.R., Fosheim, G.E., McDougal, L.K., Carey, R.B., and Fridkin, S.K. (2007). Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. J Am Med Assoc **298**, 1763-1771.

Koehn, F.E., and Carter, G.T. (2005). The evolving role of natural products in drug discovery. Nat Rev Drug Discov **4**, 206-220.

Kondo, Y., Shimoda, K., Takimura, J., Hamada, H., and Hamada, H. (2006). Glycosylation of vitamin E homologue by cultured plant cells. Chem Lett **35**, 324-325.

Kong, J.M., Goh, N.K., Chia, L.S., and Chia, T.F. (2003). Recent advances in traditional plant drugs and orchids. Acta Pharmacol Sin **24**, 7-21.

Koulman, A., Beekman, A.C., Pras, N., and Quax, W.J. (2003). The bioconversion process of deoxypodophyllotoxin with *Linum flavum* cell cultures. Planta Med **69**, 739-744.

Kwon, S., Shimoda, K., Hamada, H., Ishihara, K., Masuoka, N., and Hamada H. (2008). High production of β-thujaplicin glycosides by immobilized plant cells of *Nicotiana tabacum*. Acta Biol Hung **59**, 347-355.

Lacheretz, R., Pardo, D.G., and Cossy, J. (2009). *Daucus carota* mediated-reduction of cyclic 3-oxo-amines. Org Lett **11**, 1245-1248.

Lackman, P., González-Guzmán, M., Tilleman, S., Carqueijeiro, I., Cuéllar Pérez, A., Moses, T., Seo, M., Kanno, Y., Häkkinen, S.T., Van Montagu, M.C.E., Thevelein, J.M., Maaheimo, H., Oksman-Caldentey, K.M., Rodriguez, P.L., Rischer, H., and Goossens, A. (2011). Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in Arabidopsis and tobacco. Proc Natl Acad Sci U S A **108**, 5891-5896.

Lam, K.S. (2007). New aspects of natural products in drug discovery. Trends Microbiol **15**, 279-289.

Laszczyk, M.N. (2009). Pentacyclic triterpenes of the lupane, oleanane and ursane group as tools in cancer therapy. Planta Med **75**, 1549-1560.

Laurila, J., Laakso, I., Valkonen, J.P.T., Hiltunen, R., and Pehu, E. (1996). Formation of parental-type and novel glycoalkaloids in somatic hybrids between *Solanum brevidens* and *S. tuberosum*. Plant Sci **118**, 145-155.

Lee, E.A., Staebler, J.M., Grainger, C., and Snook, M.E. (2009). Robustness of QTLs across germplasm pools using a model quantitative trait. Genome **52**, 39-48.

Lewinsohn, E., and Gijzen, M. (2009). Phytochemical diversity: The sounds of silent metabolism. Plant Sci **176**, 161-169.

Li, F., Cui, J.N., Qian, X.H., Rong, Z.A., and Xiao, Y. (2005a). Highly chemoselective reduction of aromatic nitro compounds to the corresponding hydroxylamines catalysed by plant cells from a grape (*Vitis vinifera* L.). Chem Comm **14**, 1901-1903.

Li, J.W.H., and Vederas, J.C. (2009). Drug discovery and natural products: end of an era or an endless frontier? Science **325**, 161-165.

Li, W., Koike, K., Asada, Y., Yoshikawa, T., and Nikaido, T. (2005b). Biotransformation of paeonol by *Panax ginseng* root and cell cultures. J Mol Catal, B Enzym **35**, 117-121.

Li, Y., He, K., Huang, Y.H., Zheng, D.X., Gao, C., Cui, L., and Jin, Y.H. (2010). Betulin induces mitochondrial cytochrome c release associated apoptosis in human cancer cells. Mol Carcinog **49**, 630-640.

Liang, Y., and Zhao, S. (2008). Progress in understanding of ginsenoside biosynthesis. Plant Biol **10**, 415-421.

Limberger, R.P., Aleixo, A.M., Fett-Neto, A.G., and Henriques, A.T. (2007). Bioconversion of (+)- and (-)-alpha-pinene to (+)- and (-)-verbenone by plant cell cultures of *Psychotria brachyceras* and *Rauvolfia sellowii*. Electron J Biotechnol **10**, 500-507.

Littleton, J., Rogers, T., and Falcone, D. (2005). Novel approaches to plant drug discovery based on high throughput pharmacological screening and genetic manipulation. Life Sci **78**, 467-475.

Liu, G., Anderson, C., Scaltreto, H., Barbon, J., and Kensil, C.R. (2002). QS-21 structure/function studies: effect of acylation on adjuvant activity. Vaccine **20**, 2808-2815.

Liu, L.F. (1989). DNA Topoisomerase poisons as antitumor drugs. Annu Rev Biochem **58**, 351-375.

Liu, T.G., and Khosla, C. (2010). A balancing act for taxol precursor pathways in *E. coli*. Science **330**, 44-45.

Liu, X., Zhang, B.L., Xia, Y.M., Fang, Z.J., and Xu, J.H. (2008). Asymmetric transformation of carbonyl compounds using plant cells. Prog Chem **20**, 1108-1114.

Loris, E.A., Panjikar, S., Ruppert, M., Barleben, L., Unger, M., Schübel, H., and Stockigt, J. (2007). Structure-based engineering of strictosidine synthase: Auxiliary for alkaloid libraries. Chem Biol **14**, 979-985.

Loyola-Vargas, V.M., Galaz-Ávalos, R.M., and Kú-Cauich, R. (2007). *Catharanthus* biosynthetic enzymes: the road ahead. Phytochemistry rev **6**, 307-339.

Lü, J.M., Yao, Q.Z., and Chen, C.Y. (2009). Ginseng compounds: an update on their molecular mechanisms and medical applications. Curr Vasc Pharmacol **7**, 293-302.

Ma, W.L., Yan, C.Y., Zhu, J.H., Duan, G.Y., and Yu, R.M. (2010). Biotransformation of paeonol and emodin by transgenic crown galls of *Panax quinquefolium*. Appl Biochem Biotechnol **160**, 1301-1308.

Machado, L.L., Souza, J.S.N., de Mattos, M.C., Sakata, S.K., Cordell, G.A., and Lemos, T.L.G. (2006). Bioreduction of aldehydes and ketones using *Manihot* species. Phytochemistry **67**, 1637-1643.

Machado, L.L., Lemos, T.L.G., de Mattos, M.C., de Oliveira, M.D.F., de Gonzalo, G., Gotor-Fernández, V., and Gotor, V. (2008a). Immobilized *Manihot esculenta* preparation as a novel biocatalyst in the enantioselective acetylation of racemic alcohols. Tetrahedron Asymmetry **19**, 1419-1424.

Machado, L.L., Monte, F.J.Q., de Oliveira, M.D.F., de Maltos, M.C., Lemos, T.L.G., Gotor-Fernández, V., de Gonzalo, G., and Gotor, V. (2008b). Bioreduction of aromatic aldehydes and ketones by fruits' barks of *Passiflora edulis*. J Mol Catal, B Enzym **54**, 130-133.

Mączka, W.K., and Mironowicz, A. (2004a). Enantioselective reduction of bromo- and methoxy-acetophenone derivatives using carrot and celeriac enzymatic system. Tetrahedron Asymmetry **15**, 1965-1967.

Mączka, W.K., and Mironowicz, A. (2004b). Biotransformation of isoprenoids and shikimic acid derivatives by a vegetable enzymatic system. Z Naturforsch, C, J Biosci **59**, 201-204.

Mączka, W.K., and Mironowicz, A. (2007). Enantioselective hydrolysis of bromo- and methoxy-substituted 1-phenylethanol acetates using carrot and celeriac enzymatic systems. Z Naturforsch, C, J Biosci **62**, 397-402.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z.T., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P.G., Begley, R.F., and Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**, 376-380.

Marioni, F., Bertoli, A., and Pistelli, L. (2008). A straightforward procedure to biosynthesise melatonin using freshly chopped *Achillea millefolium* L. as reagent. Phytochem Lett **1**, 107-110.

Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol **21**, 796-802.

Matsuda, H., Ohta, T., Kawaguchi, A., and Yoshikawa, M. (2001). Bioactive constituents of Chinese natural medicines. VI. Moutan cortex. (2): Structures and radical scavenging effects of suffruticosides A, B, C, D, and E and galloyl-oxypaeoniflorin. Chem Pharm Bull **49**, 69-72.

Matsuki, A., Takimura, J., Ishihara, K., Nakajima, N., Kasai, F., and Hamada, H. (2006). Biotransformation of acyclic monoterpenes by biocatalysts of plant cultured cells and *cyanobacterium* - Abstracts. J Mol Catal, B Enzym **42**, 125-125.

Matsuo, K., Kawabe, S., Tokuda, Y., Eguchi, T., Yamanaka, R., and Nakamura, K. (2008). Asymmetric reduction of ketones with a germinated plant. Tetrahedron Asymmetry **19**, 157-159.

McCoy, E., and O'Connor, S.E. (2006). Directed biosynthesis of alkaloid analogs in the medicinal plant *Catharanthus roseus*. J Am Chem Soc **128**, 14276-14277.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. Proc Natl Acad Sci U S A **96**, 1846-1851.

Memelink, J., Verpoorte, R., and Kijne, J.W. (2001). ORCAnization of jasmonate - responsive gene expression in alkaloid metabolism. Trends Plant Sci **6**, 212-219.

Menzella, H.G., and Reeves, C.D. (2007). Combinatorial biosynthesis for drug development. Curr Opin Microbiol **10**, 238-245.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat Rev Genet **11**, 31-46.

Minami, H., Kim, J.S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H., and Sato, F. (2008). Microbial production of plant benzylisoquinoline alkaloids. Proc Natl Acad Sci U S A **105**, 7393-7398.

Miyazaki, T., Mitamura, H., Miyoshi, S., Soejima, K., Aizawa, Y., and Ogawa, S. (1996). Autonomic and antiarrhythmic drug modulation of ST segment elevation in patients with Brugada syndrome. J Am Coll Cardiol **27**, 1061-1070.

Molmoori, V., Srisailam, K., and Ciddi, V. (2008). Biotransformation of valdecoxib by plant cell cultures. Appl Biochem Biotechnol **144**, 201-212.

Morein, B., Sundquist, B., Höglund, S., Dalsgaard, K., and Osterhaus, A. (1984). Iscom, a novel structure for antigenic presentation of membrane proteins from enveloped viruses. Nature **308**, 457-460.

Morita, M., Shibuya, M., Kushiro, T., Masuda, K., and Ebizuka, Y. (2000). Molecular cloning and functional expression of triterpene synthases from pea (*Pisum sativum*) - New α-amyrin-producing enzyme is a multifunctional triterpene synthase. Eur J Biochem **267**, 3453-3460.

Morita, M., Shitan, N., Sawada, K., Van Montagu, M.C.E., Inzé, D., Rischer, H., Goossens, A., Oksman-Caldentey, K.M., Moriyama, Y., and Yazaki, K. (2009). Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. Proc Natl Acad Sci U S A **106**, 2447-2452.

Morreel, K., Goeminne, G., Storme, V., Sterck, L., Ralph, J., Coppieters, W., Breyne, P., Steenackers, M., Georges, M., Messens, E., and Boerjan, W. (2006). Genetical metabolomics of flavonoid biosynthesis in *Populus*: a case study. Plant J **47**, 224-237.

Murata, J., Roepke, J., Gordon, H., and De Luca, V. (2008). The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. Plant Cell **20**, 524-542.

Nagaki, M., Imaruoka, H., Kawakami, J., Saga, K., Kitahara, H., Sagami, H., Oba, R., Ohya, N., and Koyama, T. (2007). Biotransformation of prenyl alcohols by cultured cells of *Cucurbita maxima*. J Mol Catal, B Enzym **47**, 33-36.

Nagaraj, S., Youn, J.I., Weber, H., Iclozan, C., Lu, L., Cotter, M.J., Meyer, C., Becerra, C.R., Fishman, M., Antonia, S., Sporn, M.B., Liby, K.T., Rawal, B., Lee, J.H., and Gabrilovich, D.I. (2010). Anti-inflammatory triterpenoid blocks immune suppressive function of MDSCs and improves immune response in cancer. Clin Cancer Res **16**, 1812-1823.

Nakajima, N., Ishihara, K., and Hamada, H. (2001). Functional glucosylation of kojic acid and daidzein,with the *Eucalyptus* membrane-associated UDP-glucosyltransferase reaction system. J Biosci Bioeng **92**, 469-471.

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

Newman, D.J., and Cragg, G.M. (2007). Natural products as sources of new drugs over the last 25 years. J Nat Prod **70**, 461-477.

Ng, F., Yun, H., Lei, X.G., Danishefsky, S.J., Fahey, J., Stephenson, K., Flexner, C., and Lee, L. (2008). (3*R*,9*R*,10*R*)-Panaxytriol: a molecular-based nutraceutical with possible application to cancer prevention and treatment. Tetrahedron Lett **49**, 7178-7179.

O'Connor, S.E., and Maresh, J.J. (2006). Chemistry and biology of monoterpene indole alkaloid biosynthesis. Nat Prod Rep **23**, 532-547.

Oda, K., Matsuda, H., Murakami, T., Katayama, S., Ohgitani, T., and Yoshikawa, M. (2003). Relationship between adjuvant activity and amphipathic structure of soyasaponins. Vaccine **21**, 2145-2151.

Oksman-Caldentey, K.M., and Inzé, D. (2004). Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci **9**, 433-440.

Orden, A.A., Bisogno, F.R., Giordano, O.S., and Sanz, M.K. (2008). Comparative study in the asymmetric bioreduction of ketones by plant organs and undifferentiated cells. J Mol Catal, B Enzym **51**, 49-55.

Orden, A.A., Bisogno, F.R., Cifuente, D.A., Giordano, O.S., and Sanz, M.K. (2006). Asymmetric bioreduction of natural xenobiotic diketones by *Brassica napus* hairy roots. J Mol Catal, B Enzym **42**, 71-77.

Orians, C.M. (2000). The effects of hybridization in plants on secondary chemistry: Implications for the ecology and evolution of plant-herbivore interactions. Am J Bot **87**, 1749-1756.

Ortholand, J.Y., and Ganesan, A. (2004). Natural products and combinatorial chemistry: back to the future. Curr Opin Chem Biol **8**, 271-280.

Pacheco, A.D., Kagohara, E., Andrade, L.H., Comasseto, J.V., Crusius, I.H.S., Paula, C.R., and Porto, A.L.M. (2007). Biotransformations of nitro-aromatic compounds to amines and acetamides by tuberous roots of *Arracacia xanthorrhiza* and *Beta vulgaris* and associated microorganism (*Candida guilliermondii*). Enzyme Microb Technol **42**, 65-69.

Pauwels, L., Inzé, D., and Goossens, A. (2009). Jasmonate-inducible gene: what does it mean? Trends Plant Sci **14**, 87-91.

Pearse, M.J., and Drane, D. (2005). ISCOMATRIX® adjuvant for antigen delivery. Adv Drug Deliv Rev **57**, 465-474.

Peng, C.X., Gong, J.S., Zhang, X.F., Zhang, M., and Zheng, S.Q. (2008). Production of gastrodin through biotransformation of p-hydroxybenzyl alcohol using hairy root cultures of *Datura tatula* L. Afr J Biotechnol **7**, 211-216.

Phillips, D.R., Rasbery, J.M., Bartel, B., and Matsuda, S.P.T. (2006). Biosynthetic diversity in plant triterpene cyclization. Curr Opin Plant Biol **9**, 305-314.

Phillips, M.A., Szabadi, E., and Bradshaw, C.M. (2000). Comparison of the effects of clonidine and yohimbine on pupillary diameter at different illumination levels. Br J Clin Pharmacol **50**, 65-68.

Pisha, E., Chai, H., Lee, I.S., Chagwedera, T.E., Farnsworth, N.R., Cordell, G.A., Beecher, C.W.W., Fong, H.S., Kinghorn, A.D., Brown, D.M., Wani, M.C., Wall, M.E., Hieken, T.J., Dasgupta, T.K., and Pezzuto, J.M. (1995). Discovery of betulinic acid as a selective inhibitor of human melanoma that functions by induction of apoptosis. Nat Med **1**, 1046-1051.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. J Nat Prod, **in press**.

Pyo, J.S., Roh, H., Kim, D.K., Lee, J.G., Lee, Y.Y., Hong, S.S., Kwon, S.W., and Park, J.H. (2009). Anti-cancer effect of betulin on a human lung cancer cell line: A pharmacoproteomic approach using 2D SDS PAGE coupled with nano-HPLC tandem mass spectrometry. Planta Med **75**, 127-131.

Qi, L.W., Wang, C.Z., and Yuan, C.S. (2010). American ginseng: Potential structure-function relationship in cancer chemoprevention. Biochem Pharmacol **80**, 947-954.

Qi, L.W., Wang, C.Z., and Yuan, C.S. (2011). Isolation and analysis of ginseng: advances and challenges. Nat Prod Rep **28**, 467-495.

Rani, N., Joy, B., and Abraham, T.E. (2007). Cell suspension cultures of *Portulaca grandiflora* as potent catalysts for biotransformation of L-tyrosine into L-DOPA, an anti-Parkinson's drug. Pharm Biol **45**, 48-53.

Ravía, S., Garnenara, D., Schapiro, V., Bellomo, A., Adum, J., Seoane, G., and Gonzalez, D. (2006). Enantioselective reduction by crude plant parts: reduction of benzofuran-2-yl methyl ketone with carrot (*Daucus carota*) bits. J Chem Educ **83**, 1049-1051.

Reddy, A.S.N. (2001). Nuclear pre-mRNA splicing in plants. Critic Rev Plant Sci **20**, 523-571.

Rentsch, K.M. (2002). The importance of stereoselective determination of drugs in the clinical laboratory. J Biochem Biophys Methods **54**, 1-9.

Rinehart, K.L. (1977). Mutasynthesis of new antibiotics. Pure Appl Chem **49**, 1361-1384.

Rix, U., Fischer, C., Remsing, L.L., and Rohr, J. (2002). Modification of post-PKS tailoring steps through combinatorial biosynthesis. Nat Prod Rep **19**, 542-580.

Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R., and Keasling, J.D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature **440**, 940-943.

Roberts, S., and Kolewe, M. (2010). Plant natural products from cultured multipotent cells. Nat Biotechnol **28**, 1175-1176.

Runguphan, W., and O'Connor, S.E. (2009). Metabolic reprogramming of periwinkle plant culture. Nat Chem Biol **5**, 151-153.

Runguphan, W., Maresh, J.J., and O'Connor, S.E. (2009). Silencing of tryptamine biosynthesis for production of nonnatural alkaloids in plant culture. Proc Natl Acad Sci U S A **106**, 13673-13678.

Runguphan, W., Qu, X.D., and O'Connor, S.E. (2010). Integrating carbon-halogen bond formation into medicinal plant metabolism. Nature **468**, 461-467.

Ruppert, M., Ma, X.Y., and Stöckigt, J. (2005). Alkaloid biosynthesis in *Rauvolfia* - cDNA cloning of major enzymes of the ajmaline pathway. Curr Org Chem **9**, 1431-1444.

Ryan, A.J., Squires, S., Strutt, H.L., and Johnson, R.T. (1991). Camptothecin cytotoxicity in mammalian-cells is associated with the induction of persistent double strand breaks in replicating DNA. Nucleic Acids Res **19**, 3295-3300.

Rzeski, W., Stepulak, A., Szymański, M., Juszczak, M., Grabarska, A., Sifringer, M., Kaczor, J., and Kandefer-Szerszeń, M. (2009). Betulin elicits anti-cancer effects in tumour primary cultures and cell lines *in vitro*. Basic Clin Pharmacol Toxicol **105**, 425-432.

Saito, K., Sudo, H., Yamazaki, M., Koseki-Nakamura, M., Kitajima, M., Takayama, H., and Aimi, N. (2001). Feasible production of camptothecin by hairy root culture of *Ophiorrhiza pumila*. Plant Cell Rep **20**, 267-271.

Sakamaki, H., Itoh, K., Kitanaka, S., Sawada, S., and Horiuchi, C.A. (2008). Biotransformation of zerumbone by *Caragana chamlagu*. Biotechnol Lett **30**, 2025-2029.

Sakamaki, H., Itoh, K., Chai, W., Hayashida, Y., Kitanaka, S., and Horiuchi, C.A. (2004). Biotransformation of (±)-α-ionone and β-ionone by cultured cells of *Caragana chamlagu*. J Mol Catal, B Enzym **27**, 177-181.

Sakamaki, H., Itoh, K., Taniai, T., Kitanaka, S., Takagi, Y., Chai, W., and Horiuchi, C.A. (2005). Biotransformation of valencene by cultured cells of *Gynostemma pentaphyllum*. J Mol Catal, B Enzym **32**, 103-106.

Sakamoto, S., Shimoda, K., Ishihara, K., and Hamada, H. (2007). Biotransformation of monoterpene alcohols by plant cultured cells. J Mol Catal, B Enzym **48**, 115-116.

Salas, J.A., and Méndez, C. (2009). Indolocarbazole antitumour compounds by combinatorial biosynthesis. Curr Opin Chem Biol **13**, 152-160.

Sánchez, C., Méndez, C., and Salas, J.A. (2006). Indolocarbazole natural products: occurrence, biosynthesis, and biological activity. Nat Prod Rep **23**, 1007-1045.

Sánchez, C., Zhu, L.L., Braña, A.F., Salas, A.P., Rohr, J., Méndez, C., and Salas, J.A. (2005). Combinatorial biosynthesis of antitumor indolocarbazole compounds. Proc Natl Acad Sci U S A **102**, 461-466.

Sandoval, C., Méndez, J.M., Sánchez-Obregón, R., Alpízar, C.B., and Barrios, H. (2009). Reduction and cyclization in biotransformation of carbonyl compounds by cultured cells of *Taxus* species. Biocatal Biotransform **27**, 36-44.

Sato, N., Ishihara, K., Nakajima, N., and Hamada, H. (2007). Biotransformation of daidzein by plant cultured cells. J Mol Catal, B Enzym **48**, 117.

Scarpi, D., Occhiato, E.G., and Guarna, A. (2005). Selectivity of *Daucus carota* roots and baker's yeast in the enantioselective reduction of γ-nitroketones. Tetrahedron Asymmetry **16**, 1479-1483.

Schwab, W. (2003). Metabolome diversity: too few genes, too many metabolites? Phytochemistry **62**, 837-849.

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., Akashi, T., Aoki, T., Saito, K., and 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 U S A **105**, 14204-14209.

Shendure, J., and Ji, H.L. (2008). Next-generation DNA sequencing. Nat Biotechnol **26**, 1135-1145.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X.X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science **309**, 1728-1732.

Sherman, D.H. (2005). The Lego-ization of polyketide biosynthesis. Nat Biotechnol 23, 1083-1084.

Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T., and Ebizuka, Y. (2006). Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. FEBS J **273**, 948-959.

Shimoda, K., and Hamada, H. (2008). Glycosylation of hesperetin by plant cell cultures. Phytochemistry **69**, 1135-1140.

Shimoda, K., Kubota, N., Hirata, T., and Kawano, T. (2004). Asymmetric transformation of enol acetates with esterases from *Marchantia polymorpha*. J Mol Catal, B Enzym **29**, 123-127.

Shimoda, K., Kondo, Y., Abe, K., and Hamada, H. (2006a). Formation of water-soluble vitamin derivatives from lipophilic vitamins by cultured plant cells. Tetrahedron Lett **47**, 2695-2698.

Shimoda, K., Harada, T., Hamada, H., and Nakajima, N. (2007a). Biotransformation of raspberry ketone and zingerone by cultured cells of *Phytolacca americana*. Phytochemistry **68**, 487-492.

Shimoda, K., Kobayashi, T., Akagi, M., and Hamada, H. (2008a). Synthesis of oligosaccharides of genistein and quercetin as potential anti-inflammatory agents. Chem Lett **37**, 876-877.

Shimoda, K., Sakamoto, S., Nakajima, N., Hamada, H., and Hamada, H. (2008b). Synthesis of unnatural mono- and oligosaccharides of farnesol, geraniol, and (*S*)-perillyl alcohol by biocatalytic glycosylations. Chem Lett **37**, 556-557.

Shimoda, K., Sato, N., Kobayashi, T., and Hamada, H. (2008c). Glycosylation of daidzein by the *Eucalyptus* cell cultures. Phytochemistry **69**, 2303-2306.

Shimoda, K., Kondo, Y., Nishida, T., Hamada, H., and Nakajima, N. (2006b). Biotransformation of thymol, carvacrol, and eugenol by cultured cells of *Eucalyptus perriniana*. Phytochemistry **67**, 2256-2261.

Shimoda, K., Kondo, Y., Akagi, M., Abe, K., and Hamada, H. (2007b). Glycosylation of tocopherols by cultured cells of *Eucalyptus perriniana*. Phytochemistry **68**, 2678-2683.

Shimoda, K., Ishimoto, H., Kamiue, T., Kobayashi, T., and Hamada, H. (2009a). Glycosylation of sesamol by cultured plant cells. Phytochemistry **70**, 207-210.

Shimoda, K., Kubota, N., Kondo, Y., Sato, D., and Hamada, H. (2009b). Glycosylation of fluorophenols by plant cell cultures. Int J Mol Sci **10**, 1942-1949.

Shimoda, K., Kubota, N., Taniuchi, K., Sato, D., Nakajima, N., and Hamada, H. (2010). Biotransformation of naringin and naringenin by cultured *Eucalyptus perriniana* cells. Phytochemistry **71**, 201-205.

Shimoda, K., Kwon, S., Utsuki, A., Ohiwa, S., Katsuragi, H., Yonemoto, N., and Hamada, H. (2007c). Glycosylation of capsaicin and 8-nordihydrocapsaicin by cultured cells of *Catharanthus roseus*. Phytochemistry **68**, 1391-1396.

Shin, H.R., Kim, J.Y., Yun, T.K., Morgan, G., and Vainio, H. (2000). The cancer-preventive potential of *Panax ginseng*: a review of human and experimental evidence. Cancer Causes Control **11**, 565-576.

Sjölander, A., Cox, J.C., and Barr, I.G. (1998). ISCOMs: an adjuvant with multiple functions. J Leukoc Biol **64**, 713-723.

Skene, C.D., and Sutton, P. (2006). Saponin-adjuvanted particulate vaccines for clinical use. Methods **40**, 53-59.

Song, X.M., and Hu, S.H. (2009). Adjuvant activities of saponins from traditional Chinese medicinal herbs. Vaccine **27**, 4883-4890.

Song, X.M., Zang, L.M., and Hu, S.H. (2009). Amplified immune response by ginsenoside-based nanoparticles (ginsomes). Vaccine **27**, 2306-2311.

Sparg, S.G., Light, M.E., and van Staden, J. (2004). Biological activities and distribution of plant saponins. J Ethnopharmacol **94**, 219-243.

Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat Prod Rep **18**, 380-416.

Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R. (1998). Sesquiterpene synthases from grand fir (*Abies grandis*) - Comparison of constitutive and wound-induced activities, and cdna isolation,

characterization and bacterial expression of δ-selinene synthase and γ-humulene synthase. J Biol Chem **273**, 2078-2089.

Sun, H.X., Xie, Y., and Ye, Y.P. (2009a). Advances in saponin-based adjuvants. Vaccine **27**, 1787-1796.

Sun, H.X., Xie, Y., and Ye, Y.P. (2009b). ISCOMs and ISCOMATRIX^{TM}. Vaccine **27**, 4388-4401.

Suzuki, H., Reddy, M.S.S., Naoumkina, M., Aziz, N., May, G.D., Huhman, D.V., Sumner, L.W., Blount, J.W., Mendes, P., and Dixon, R.A. (2005). Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. Planta **220**, 696-707.

Tafur, S., Nelson, J.D., Delong, D.C., and Svoboda, G.H. (1976). Antiviral components of *Ophiorrhiza mungos*. Isolation of camptothecin and 10-methoxycamptothecin. Lloydia **39**, 261-262.

Tatsunami, R., and Yoshioka, T. (2006). Enzymatic and mechanistic studies on the formation of N-phenylglycolohydroxamic acid from nitrosobenzene and pyruvate in spinach leaf homogenate. J Agric Food Chem **54**, 590-596.

Teng, R.W., McManus, D., Mau, S.L., and Bacic, A. (2007). Biotransformation of podophyllotoxin by *Hordeum vulgare* cell suspension cultures. Biocatal Biotransform **25**, 1-8.

Teng, R.W., McManus, D., Aylward, J., Ogbourne, S., Armstrong, D., Mau, S.L., Johns, J., and Bacic, A. (2009). Biotransformation of ingenol-3-angelate in four plant cell suspension cultures. Biocatal Biotransformation **27**, 186-194.

Terrier, N., Torregrosa, L., Ageorges, A., Vialet, S., Verriès, C., Cheynier, V., and Romieu, C. (2009). Ectopic expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. Plant Physiol **149**, 1028-1041.

Thomas, C.J., Rahier, N.J., and Hecht, S.M. (2004). Camptothecin: current perspectives. Bioorg Med Chem **12**, 1585-1604.

Tong, L.P., Cui, J.N., Ren, W.M., Wang, X.Y., and Qian, X.H. (2008). Asymmetric bioreduction of substituted acenaphthenequinones using plant enzymatic systems: A novel strategy for the preparation of (+)- and (-)-mono hydroxyacenaphthenones. Chin Chem Lett **19**, 1179-1182.

Tung, N.H., Song, G.Y., Park, Y.J., and Kim, Y.H. (2009). Two new dammarane-type saponins from the leaves of *Panax ginseng*. Chem Pharm Bull **57**, 1412-1414.

Tung, N.H., Song, G.Y., Kim, J.A., Hyun, J.H., Kang, H.K., and Kim, Y.H. (2010). Dammarane-type saponins from the flower buds of *Panax ginseng* and their effects on human leukemia cells. Bioorg Med Chem Lett **20**, 309-314.

Utsukihara, T., Okada, S., Kato, N., and Horiuchi, C.A. (2007a). Biotransformation of α-bromo and α,α'-dibromo alkanone to α-hydroxyketone and α-diketone by *Spirulina platensis*. J Mol Catal, B Enzym **45**, 68-72.

Utsukihara, T., Watanabe, S., Tomiyama, A., Chai, W., and Horiuchi, C.A. (2006a). Stereoselective reduction of ketones by various vegetables. J Mol Catal, B Enzym **41**, 103-109.

Utsukihara, T., Misumi, O., Kato, N., Kuroiwa, T., and Horiuchi, C.A. (2006b). Reduction of various ketones by red algae. Tetrahedron Asymmetry **17**, 1179-1185.

Utsukihara, T., Sato, M., Kawamoto, M., Itoh, K.I., Sakamaki, H., Kuniyoshi, M., and Horiuchi, C.A. (2007b). Biotransformation of aromatic heterocyclic compounds by *Caragana chamlagu* and *Wasabia japonica*. J Mol Catal, B Enzym **48**, 59-63.

Utsukihara, T., Misumi, O., Nakajima, K., Koshimura, M., Kuniyoshi, M., Kuroiwa, T., and Horiuchi, C.A. (2008). Stereoinversion of 1-arylethanols by *Cyanidioschyzon merolae* NEIS-1332. J Mol Catal, B Enzym **51**, 19-23.

van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallared, D., and Verpoorte, R. (2004). The *Catharanthus* alkaloids: pharmacognosy and biotechnology. Curr Med Chem **11**, 607-628.

Vaněk, T., Halík, J., Vaňková, R., and Valterová, I. (2005). Formation of *trans*-verbenol and verbenone from α-pinene catalysed by immobilised *Picea abies* cells. Biosci Biotechnol Biochem **69**, 321-325.

Vereecken, N.J., Cozzolino, S., and Schiestl, F.P. (2010). Hybrid floral scent novelty drives pollinator shift in sexually deceptive orchids. BMC Evol Biol **10**, 103-114.

Villa, R., and Molinari, F. (2008). Reduction of carbonylic and carboxylic groups by plant cell cultures. J Nat Prod **71**, 693-696.

Wall, M.E., Wani, M.C., Cook, C.E., Palmer, K.H., McPhail, A.T., and Sim, G.A. (1966). Plant antitumor agents. I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from *Camptotheca acuminata*. J Am Chem Soc **88**, 3888-3890.

Walsh, C.T., and Fischbach, M.A. (2010). Natural products version 2.0: connecting genes to molecules. J Am Chem Soc **132**, 2469-2493.

Wang, L.S., Tu, Y.C., Lian, T.W., Hung, J.T., Yen, J.H., and Wu, M.J. (2006). Distinctive antioxidant and antiinflammatory effects of flavonols. J Agric Food Chem **54**, 9798-9804.

Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P., and McPhail, A.T. (1971). Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. J Am Chem Soc **93**, 2325-2327.

Watts, K.T., Lee, P.C., and Schmidt-Dannert, C. (2004). Exploring recombinant flavonoid biosynthesis in metabolically engineered *Escherichia coli*. ChemBioChem **5**, 500-507.

Weissman, K.J. (2007). Mutasynthesis - uniting chemistry and genetics for drug discovery. Trends Biotechnol **25**, 139-142.

Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat Rev Microbiol **3**, 925-936.

Weist, S., and Süssmuth, R.D. (2005). Mutational biosynthesis - a tool for the generation of structural diversity in the biosynthesis of antibiotics. Appl Microbiol Biotechnol **68**, 141-150.

Wellens, H.J.J., Bär, F.W., Gorgels, A.P., and Vanagt, E.J. (1980). Use of ajmaline in patients with the Wolff-Parkinson-White syndrome to disclose short refractory period of the accessory pathway. Am J Cardiol **45**, 130-133.

Welsch, M.E., Snyder, S.A., and Stockwell, B.R. (2010). Privileged scaffolds for library design and drug discovery. Curr Opin Chem Biol **14**, 347-361.

Wender, P.A., and Miller, B.L. (2009). Synthesis at the molecular frontier. Nature **460**, 197-201.

Werner, S.R., Chen, H., Jiang, H.X., and Morgan, J.A. (2010). Synthesis of non-natural flavanones and dihydrochalcones in metabolically engineered yeast. J Mol Catal, B Enzym **66**, 257-263.

Wu, S.Q., Schalk, M., Clark, A., Miles, R.B., Coates, R., and Chappell, J. (2006). Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nat Biotechnol **24**, 1441-1447.

Xie, B., Yang, J., Yang, Q., and Yuan, W.G. (2009a). Enantioselective reduction of fluorenones in surfactant-aqueous solution by fruits and vegetables. J Mol Catal, B Enzym **61**, 284-288.

Xie, Y., Xu, J.H., Lu, W.Y., and Lin, G.Q. (2009b). Adzuki bean: A new resource of biocatalyst for asymmetric reduction of aromatic ketones with high stereoselectivity and substrate tolerance. Bioresour Technol **100**, 2463-2468.

Yadav, J.S., Reddy, B.V.S., Sreelakshmi, C., Kumar, G., and Rao, A.B. (2008). Enantioselective reduction of 2-substituted tetrahydropyran-4-ones using *Daucus carota* plant cells. Tetrahedron Lett **49**, 2768-2771.

Yadav, J.S., Reddy, G., Sabitha, G., Krishna, A.D., Prasad, A.R., Hafeez, U.R.R., Rao, K.V., and Rao, A.B. (2007). *Daucus carota* and baker's yeast mediated bio-reduction of prochiral ketones. Tetrahedron Asymmetry **18**, 717-723.

Yan, C.Y., Yu, R.M., Zhang, Z., and Kong, L.Y. (2007). Biotransformation of 4-hydroxybenzen derivatives by hairy root cultures of *Polygonum multiflorum* thunb. J Integr Plant Biol **49**, 207-212.

Yang, L., Dai, J.G., Sakai, J., and Ando, M. (2005). Biotransformation of α-santonin by cell suspension cultures of five plants. Biotechnol Lett **27**, 793-797.

Yang, L., Dai, J., Sakai, J.I., and Ando, M. (2006). Biotransformation of α- and 6β-santonin by fungus and plant cell cultures. J Asian Nat Prod Res **8**, 317-326.

Yang, Z.G., Ye, Y.P., and Sun, H.X. (2007). Immunological adjuvant effect of ginsenoside Rh-4 from the roots of *Panax notoginseng* on specific antibody and cellular response to ovalbumin in mice. Chem Biodivers **4**, 232-240.

Yang, Z.H., Zeng, R., Yang, G., Wang, Y., Li, L.Z., Lv, Z.S., Yao, M., and Lai, B. (2008). Asymmetric reduction of prochiral ketones to chiral alcohols catalyzed by plants tissue. J Ind Microbiol Biotechnol **35**, 1047-1051.

Yates, M.S., Tauchi, M., Katsuoka, F., Flanders, K.C., Liby, K.T., Honda, T., Gribble, G.W., Johnson, D.A., Johnson, J.A., Burton, N.C., Guilarte, T.R., Yamamoto, M., Sporn, M.B., and Kensler, T.W. (2007). Pharmacodynamic characterization of chemopreventive triterpenoids as exceptionally potent inducers of Nrf2-regulated genes. Mol Cancer Ther **6**, 154-162.

Yoshikawa, M., Sugimoto, S., Nakamura, S., and Matsuda, H. (2007). Medicinal flowers. XI. Structures of new dammarane-type triterpene diglycosides with hydroperoxide group from flower buds of **Panax ginseng**. Chem Pharm Bull **55**, 571-576.

Yoshikuni, Y., Ferrin, T.E., and Keasling, J.D. (2006). Designed divergent evolution of enzyme function. Nature **440**, 1078-1082.

Zhan, Y.L., Zou, H.H., Ma, X.J., and Dai, J.G. (2005). Biotransformation of 14-deacetoxyl sinenxan A by *Ginkgo* cell suspension cultures and the cytotoxic activity evaluation. J Mol Catal, B Enzym **36**, 43-46.

Zhou, B.N., Hoch, J.M., Johnson, R.K., Mattern, M.R., Eng, W.K., Ma, J., Hecht, S.M., Newman, D.J., and Kingston, D.G.I. (2000). Use of COMPARE analysis to discover new natural product drugs: Isolation of camptothecin and 9-methoxycamptothecin from a new source. J Nat Prod **63**, 1273-1276.

Zhou, H., Xie, X.K., and Tang, Y. (2008). Engineering natural products using combinatorial biosynthesis and biocatalysis. Curr Opin Biotechnol **19**, 590-596.

Zhou, L.G., and Wu, J.Y. (2006). Development and application of medicinal plant tissue cultures for production of drugs and herbal medicinals in China. Nat Prod Rep **23**, 789-810.

Zhu, J.H., Yu, R.M., Yang, L., Hu, Y.S., Song, L.Y., Huang, Y.J., Li, W.M., and Guan, S.X. (2010) Novel biotransformation processes of dihydroartemisinic acid and artemisinic acid to their hydroxylated derivatives by two plant cell culture systems. Process Biochem **45**, 1652-1656.

Zou, J.H., Du, H.X., Zhang, Y., Dai, J.G., Yin, D.L., and Chen, X.G. (2008). Biotransformation and tumor multidrug resistance reversal potency of polyoxygenated taxadienes. J Mol Catal, B Enzym **55**, 12-18.

Zou, K., Zhu, S., Tohda, C., Cai, S.Q., and Komatsu, K. (2002). Dammarane-type triterpene saponins from *Panax japonicus*. J Nat Prod **65**, 346-351.

Chapter 2

A functional genomics approach to predict genes involved in triterpene saponin biosynthesis in five different plants.

Jacob Pollier*, Miguel González-Guzmán, Lander Ingelbrecht, Mart Theunis, Kenn Foubert, Sandra Apers, Luc Pieters, and Alain Goossens

* Author contribution: Plant cultivation and elicitation, (q)RT-PCR and cDNA-AFLP transcript profiling, data analysis and writing of the chapter.

Abstract

Saponins are an important class of plant natural products with an enormous structural variety and diverse bioactivities. Despite their potential, several of the steps involved in the saponin biosynthesis remain uncharacterized at the molecular level. Here, we used a functional genomics approach to identify candidate genes involved in the saponin biosynthesis of five different plants. After elicitation with jasmonate and targeted metabolite profiling, a genome-wide cDNA-AFLP transcript profiling was carried out for the five plants. Taking into account the putative functional annotation and the expression pattern of the visualized transcript tags, a set of 259 candidate genes involved in hydroxylation, glycosylation and other decorations of the saponin backbone were identified. The generated gene list provides the basis for a combinatorial biosynthesis platform that targets triterpene saponins in plants. Furthermore, the identified transcription factors and other regulators are lead candidates for studies investigating the control of saponin biosynthesis *in planta*.

1. Introduction

Saponins compose a structurally diverse group of amphipathic glycosides, with a steroid, steroidal alkaloid or triterpenoid aglycone backbone. The aglycone or sapogenin can be covalently linked to one (monodesmosidic) or more (di- or tridesmosidic) sugar chains via a glycosidic bond. The word saponin is derived from “sapo”, the Latin word for soap, and is given to this class of molecules because they are surface active agents, that, when dissolved in water, form a colloidal solution that foams upon shaking. Accordingly, they are used as emulsifier and foaming agents in the food and beverage industries. Importantly, many saponins also possess pharmacological properties, such as anti-inflammatory, antimicrobial, anticancer, or adjuvant activities, and are therefore commonly used in phytotherapy and the cosmetic industry (Dixon and Sumner, 2003; Sparg et al., 2004; Sun et al., 2009; Yendo et al., 2010; Augustin et al., 2011).

Although triterpene saponins are an important class of plant secondary metabolites, knowledge of their biosynthesis remains incomplete. The precursor molecule of all triterpene saponins is 2,3-oxidosqualene, a precursor they share with the sterols. 2,3-oxidosqualene is synthesized via the isoprenoid pathway, relying on the cytosolic mevalonate pathway for the isopentenyl diphosphate supply (Figure 1). The first committed step in the saponin biosynthesis, the cyclization of 2,3-oxidosqualene, forms the branch point between the primary sterol metabolism and the secondary saponin metabolism. The ancestor enzyme of all plant oxidosqualene cyclases (OSCs) involved in secondary metabolism is cycloartenol synthase (CAS), involved in primary steroid metabolism (Phillips et al., 2006). Most plants contain several OSCs that catalyze the formation of different triterpene skeletons. For instance, the genome of the model plant *Arabidopsis thaliana* contains up to 13 predicted OSCs encoding genes, producing various triterpenes of which the function *in planta* remains mostly unknown (Field and Osbourn, 2008). The OSC leading to the saponin backbone has been characterized for several saponin producing plants, including dammarenediol-II synthase (DDS) in *P. ginseng* (Tansakul et al., 2006), and β-amyrin synthase (BAS) in *M. truncatula* (Suzuki et al., 2002) and *G. glabra* (Hayashi et al., 2001). Usually, two distinct types of additional modifications of the cyclization products complete the saponin biosynthesis. In a first step, cytochrome P450 mediated oxidations of the cyclization products give rise to different sapogenins. Thus far, only two specific cytochrome P450 enzymes have been characterized. The first, β-amyrin and sophoradiol 24-hydroxylase (CYP93E1) was

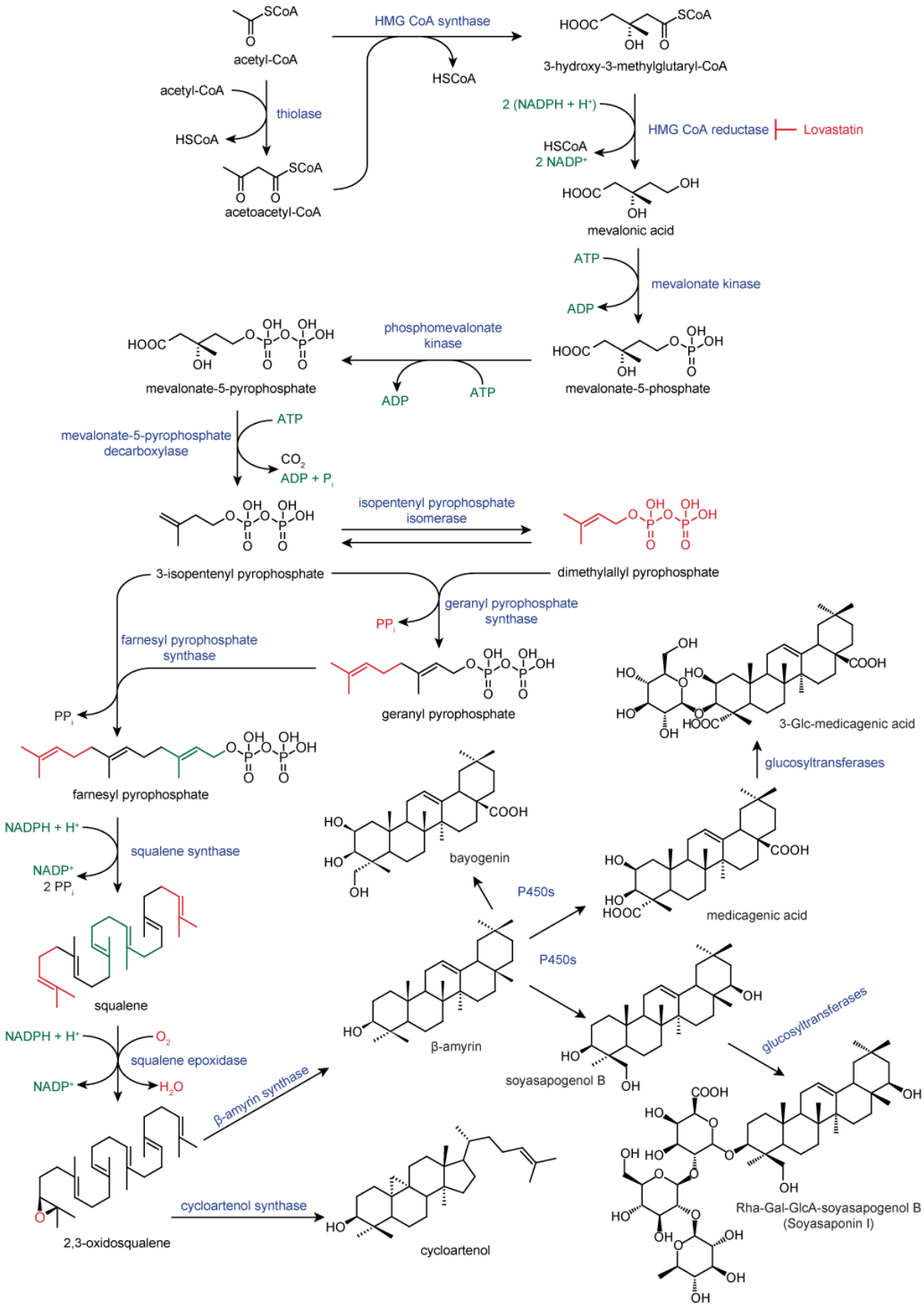


Figure 1. The triterpene saponin biosynthesis pathway in *M. truncatula*. HMG, 3-hydroxy-3-methylglutaryl.

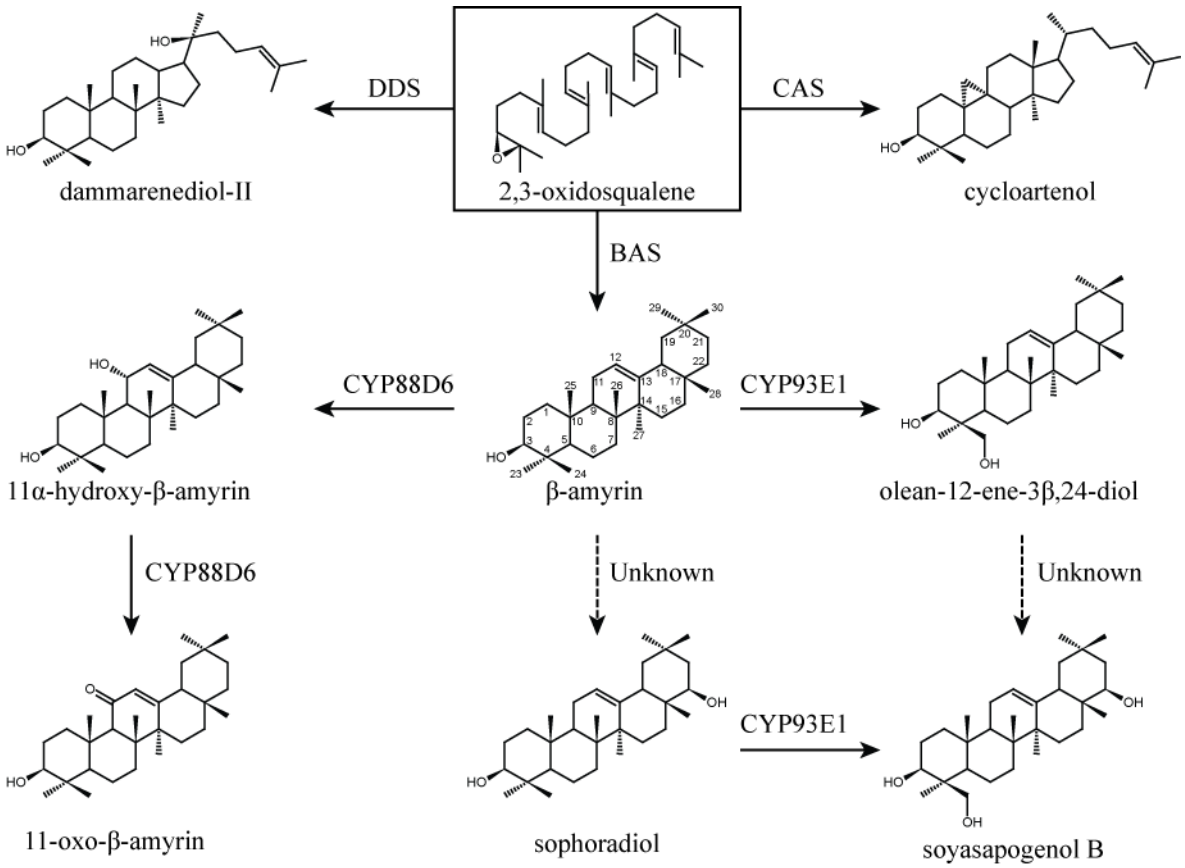


Figure 2. Cyclization of 2,3-oxidosqualene and characterized CytP450s in saponin biosynthesis.

identified in *Glycine max*, and catalyzes the oxidation of position 24 of the β-amyrin backbone (Figure 2) (Shibuya et al., 2006). The second, β-amyrin 11-oxidase (CYP88D6) catalyzes the sequential two-step oxidation of β -amyrin at the C-11 position to produce 11-oxo-β-amyrin in *G. glabra* (Figure 2) (Seki et al., 2008). Some more candidate genes of other plant species have been identified, such as Cyp51H10 from oat (Qi et al., 2006) and several cytochrome P450s in *M. truncatula* (Naoumkina et al., 2010), however, functional characterization of these genes at the molecular level is lacking. Hence, most of the enzymes catalyzing the specific oxidation reactions on the various triterpene skeletons remain unknown to date.

In the second distinct step of the saponin biosynthesis, the different sapogenins are glycosylated on different positions. The attached sugar chains can be linear or branched and can contain different monosaccharides, such as glucose, galactose, glucuronic acid, xylose, or rhamnose. Other, less common, modifications, such as acylations or malonylations can further increase the structural variability of the saponins. Even within a single plant species, an enormously rich mixture of saponins can occur. For example, in *P. ginseng* alone already

over 40 ginsenosides have been reported (Lü et al., 2009). Similar to the oxidation reactions, only few of the enzymes catalyzing the various decorations have been identified. Most progress has been made in the model legume *M. truncatula*, for which three glucosyltransferases have been characterized to date. UGT71G1 catalyzes the glycosylation of medicagenic acid *in vitro*, however, the enzyme also catalyzes the glycosylation of certain isoflavones and the flavonol quercetin with higher efficiency than triterpenes, hence, it is probably not a saponin-specific biosynthesis enzyme (Achnine et al., 2005). UGT73K1 was shown to specifically catalyze the glycosylation of hederagenin and soyasapogenols B and E *in vitro*. Furthermore, it was shown to be strongly upregulated in *M. truncatula* cell cultures that showed an increased saponin accumulation after jasmonate treatment (Achnine et al., 2005). Recently, another jasmonate responsive glucosyltransferase, UGT73F3, was shown to catalyze the glycosylation of hederagenin at the C28-position, and genetic loss-of-function studies in *M. truncatula* confirmed the *in vivo* function of this glucosyltransferase in saponin biosynthesis (Naoumkina et al., 2010). Furthermore, in *Saponaria vaccaria*, the triterpene-specific glucosyltransferase UGT74M1 was shown to catalyze the glycosylation of the C28-position (Meesapyodsuk et al., 2007).

To identify candidate genes involved in the biosynthesis of secondary metabolism, a functional genomics approach can be applied. Elicitors, such as the phytohormone jasmonate, induce the production of a compound of interest in plants or plant cell cultures and most often this effect is caused by the transcriptional activation of the genes involved in the biosynthesis of these compounds. Hence, comparing transcriptomes at various stages before and after elicitor treatment will allow pinpointing the desired subsets of the total gene pools as candidate genes responsible for the biosynthesis of the secondary metabolites (Oksman-Caldentey and Inzé, 2004; Pauwels et al., 2009). Several studies have already demonstrated the feasibility of this approach. For instance, treatment of tobacco Bright yellow 2 (BY-2) cell cultures with jasmonate induces nicotine biosynthesis in the cells. Genome-wide transcript profiling of the jasmonate elicited tobacco cell cultures resulted in the establishment of a tobacco gene compendium (Goossens et al., 2003) from which several new genes, encoding enzymes, regulators and transporters involved in nicotine biosynthesis were identified in subsequent functional screens (De Sutter et al., 2005; Häkkinen et al., 2007; Morita et al., 2009; De Boer et al., 2011; Lackman et al., 2011). In a similar series of studies in the triterpene saponin producing model legume *M. truncatula*, the response of the metabolome and transcriptome to different elicitors was investigated (Suzuki et al., 2005).

Analysis of unknown genes co-regulated with known triterpene saponin biosynthesis genes lead to the identification and characterization of the above-mentioned glucosyltransferases involved in triterpene saponin biosynthesis (Achnine et al., 2005; Naoumkina et al., 2010).

Here, a functional genomics approach to identify candidate genes involved in saponin biosynthesis of six plants, *Medicago truncatula*, *Glycyrrhiza glabra*, *Panax ginseng*, *Bupleurum falcatum*, *Centella asiatica*, and *Maesa lanceolata* was applied. Targeted metabolite analysis revealed that treatment with methyl jasmonate (MeJA) induced the production of triterpene saponins in *M. truncatula* cell cultures, *G. glabra* seedlings, *P. ginseng* hairy roots, and *C. asiatica* leaves. Subsequent RT-PCR analysis confirmed the transcriptional response of several genes involved in saponin biosynthesis. Furthermore, hydroponically grown *B. falcatum* seedlings and axenic *M. lanceolata* shoot cultures were treated with methyl jasmonate, and transcriptional response to the jasmonate treatment was confirmed. Subsequently, genome-wide cDNA-AFLP transcript profiling of five of the plants allowed identifying of a set of candidate saponin biosynthesis genes.

2. Results and Discussion

2.1.MeJA elicitation of the six saponin producing plants

2.1.1. MeJA elicitation stimulates saponin production in *M. truncatula* suspension cells

Saponin biosynthesis in *M. truncatula* has been reported to be MeJA inducible in suspension-cultured cells (Suzuki et al., 2002; Suzuki et al., 2005). Therefore, an elicitation experiment was set up on the same cell line and under similar conditions as described by Suzuki et al. (2005). As shown in Table 1, mock-treated cell cultures produced only two known saponins in detectable amounts: 3-Glc-28-Glc-medicagenic acid and 3-Rha-Gal-Glc-soyasapogenol B. Cultures that were stimulated with MeJA synthesized a more complex saponin mixture in which in total 13 different saponins could be detected (Table 1). This number was lower than that of Suzuki et al. (2005), and also fluctuated slightly between independent biological repeats of the experiments. Similar qualitative differences between identical elicitation experiments were also reported by Suzuki et al. (2005) and were attributed to the fact that different passages of the culture or accumulating epigenetic changes as a function of culturing time may to some extent affect the elicitation response. However, most importantly, the capacity of the *M. truncatula* cell culture to efficiently activate triterpene saponin biosynthesis following MeJA elicitation seemed to be intact. Therefore, we assumed that the

Table 1. Saponin profiling in elicited *M. truncatula* suspension cell cultures.

Saponin	EtOH	MeJA
Hex-Bayogenin		
Hex-Hex-Bayogenin		
Hex-Hex-Hex-Bayogenin		X
Hex-Hex-Rha-Bayogenin		X
Hex-hederagenin		X
Pen-hederagenin		
GlcA-hederagenin		X
Ara-Hex-hederagenin		
3-Glc-Ara-28-Glc-hederagenin		
Rha-Hex-Hex-hederagenin		X
Rha-Hex-Hex-Hex-hederagenin		
3-Glc-medicagenic acid		
3-Glc-28-Glc-medicagenic acid	X	X
3-Glc-malonyl-medicagenic acid		
3-Glc-28-Glc-malonyl-medicagenic acid		X
3-Glc-28-Ara-Rha-Xyl-medicagenic acid		X
3-Rha-Gal-Glc-soyasapogenol B	X	X
Hex-soyasapogenol E		X
Hex-Hex-soyasapogenol E		
Rha-Gal-GlcA-soyasapogenol E		X
Rha-Hex-Hex-soyasapogenol E		X
Rha-Hex-Hex-Hex-soyasapogenol E		
Hex-unknown aglycon		
Hex-Hex-unknown aglycon		X

Presence of a particular saponin in mock (EtOH) or MeJA treated cultures, 5 days after treatment, is indicated by X.

molecular mechanisms required in the JA-activated signalling cascade leading to the onset of the saponin biosynthesis pathway were largely unaffected by possible epigenetic changes.

To address the response to the MeJA treatment on the transcript level, induction of saponin biosynthesis was investigated with qRT-PCR taking advantage of the knowledge of some genes encoding early and late steps in the pathway. As such, it was shown that *BAS* (Suzuki et al., 2002), and the glucosyltransferases *UGT73K1* (Achnine et al., 2005) and *UGT73F3* (Naoumkina et al., 2010) were MeJA responsive and showed co-regulation (Figure 3). In contrast, the expression of *UGT71G1* (Achnine et al., 2005) was not significantly responsive to MeJA (Figure 3). However, in contrast to the other two glucosyltransferases, UGT71G1 also recognizes isoflavones and flavonols as substrates, and glycosylates these compounds with higher efficiencies than triterpenes (Achnine et al., 2005), suggesting it might not correspond to a saponin-specific biosynthesis gene. Together, these results suggest that the

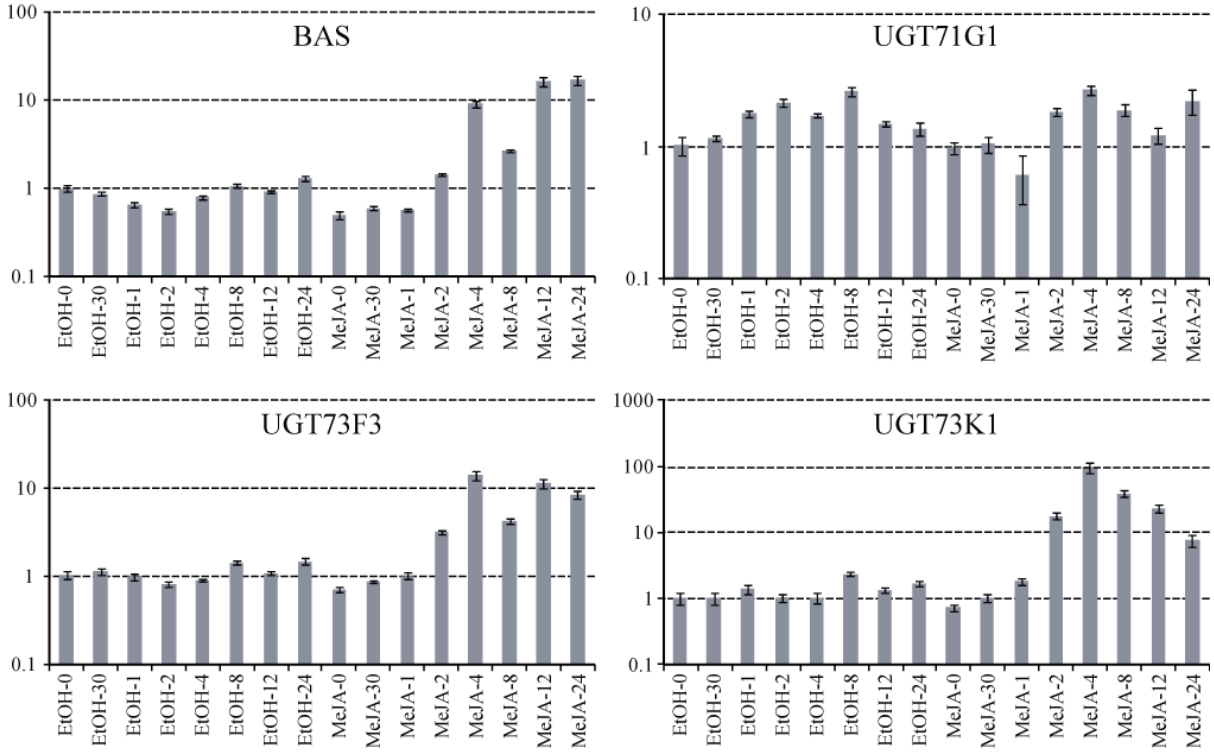


Figure 3. qRT-PCR of saponin biosynthesis genes from MeJA treated *M. truncatula* suspension cell cultures. Treatments and time-points are indicated at the bottom of the graphs.

cell culture was apt for a functional genomics-based gene discovery program in *M. truncatula* secondary metabolism.

2.1.2. MeJA elicitation stimulates glycyrrhizin production in roots but not leaves of liquorice plants

Although soyasaponin biosynthesis can be upregulated by MeJA in cultured cells of *G. glabra* (liquorice) (Hayashi et al., 2003), it might be more appropriate to evaluate the MeJA effect on saponin accumulation by functional genomics in whole plant tissues than in cultured cells because the latter lack the capacity to produce the economically more important saponin compound glycyrrhizin (Hayashi et al., 2003). Therefore, first, 7-month-old greenhouse-grown plants were sprayed and irrigated with MeJA (100 µM) or a mock solution, and the content of glycyrrhizin was measured in roots and leaves 15 days after treatments. In control roots, in which it was expected to accumulate, glycyrrhizin was indeed detected at levels of 0.5% dry weight. These levels are lower than those in the literature for liquorice plants (2-8%; (Hayashi et al., 1993)) or in own measurements with pulverized liquiritiae radix (European Pharmacopoeia approved; 5.7%). This observation was not unexpected because the plants analyzed were relatively young and glycyrrhizin levels are known to substantially

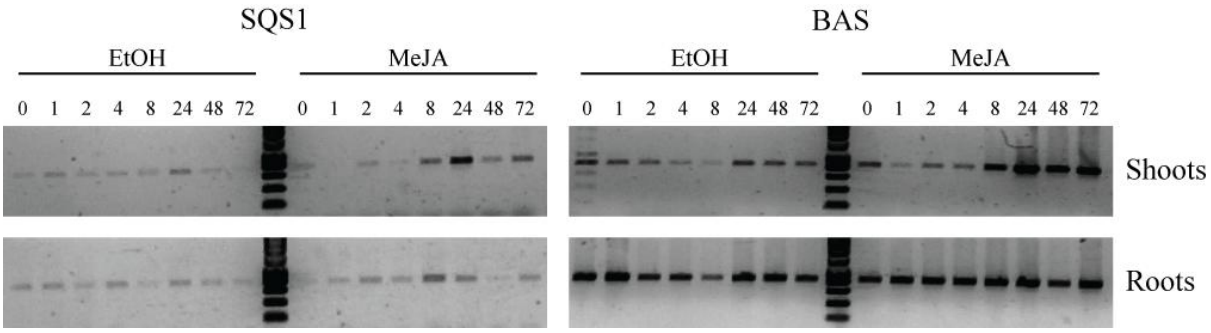


Figure 4. RT-PCR analysis of *SQS1* and *BAS* in MeJA treated shoot and root tissue of *G. glabra* seedlings. Treatment and time-points (in hours) are indicated on top of the gel images.

increase with plant age and concordant root thickening. Most importantly, glycyrrhizin in roots increased by 40% after 15 days of MeJA treatment to 0.7% of root dry weight. As expected, in the leaves of greenhouse-grown liquorice plants glycyrrhizin was not detected, even not after treatment with MeJA. Elicitation on transcript level was investigated via Reverse Transcription (RT)-PCR analysis of *squalene synthase* (*SQS*) *1* (Hayashi et al., 1999) and *BAS* (Hayashi et al., 2001), both saponin biosynthesis genes shown to be MeJA inducible in cultured *G. glabra* cells (Hayashi et al., 2003). RT-PCR analysis, performed on leaves and roots of 1-month-old seedlings of liquorice, grown in liquid medium in Erlenmeyer flasks, and treated or not with 100 μM MeJA, showed that these genes were expressed and MeJA inducible in both shoots and roots tissues (Figure 4). Together these results ensured that whole-plant material of liquorice was apt for a functional genomics-based gene discovery program.

2.1.3. MeJA elicitation stimulates ginsenoside production in 2 different *P. ginseng* hairy root lines

Ginsenoside production has been reported to be MeJA inducible in *P. ginseng* hairy roots, both on the transcript level (Choi et al., 2005; Kim et al., 2009; Zhao et al., 2010) and the metabolite level (Kim et al., 2009), hence, *P. ginseng* hairy root cultures can be useful for gene discovery for saponin biosynthesis. Two different *P. ginseng* hairy root cultures were acquired, called “Cusido” and “Liu”, and were treated with 100 μM MeJA. Subsequently, the response of the hairy root cultures to the MeJA elicitation was verified by RT-PCR and saponin profiling. RT-PCR analysis showed that the transcription of both *SQS* (GenBank: AB115496) and *squalene epoxidase* (*SQE)* (GenBank: AB003516) was upregulated upon MeJA treatment in both hairy root cultures (Figure 5A). Saponin profiling of the hairy root cultures revealed that two weeks after MeJA elicitation, the saponin content in both hairy root

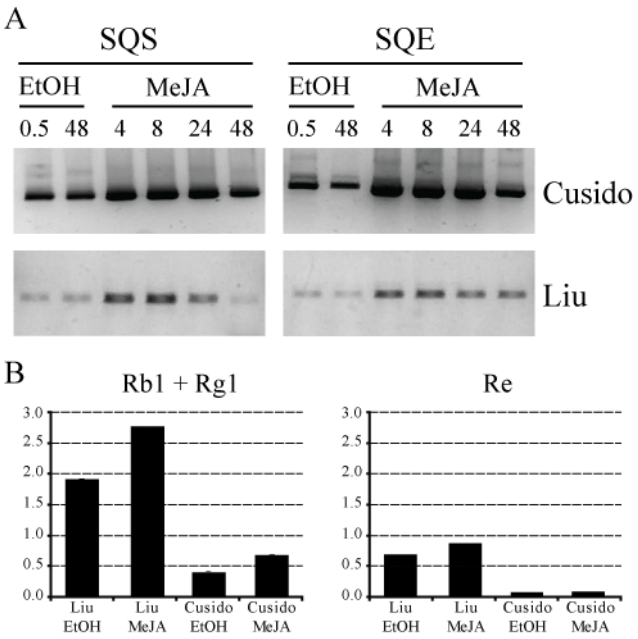


Figure 5. MeJA response of *P. ginseng* hairy roots. A. RT-PCR analysis of *SQS* and *SQE* in MeJA treated "Cusido" and "Liu" hairy root cultures. Treatment and time-points (in hours) are indicated on top of the gel images. B. Percentage of saponin content of ginsenosides Rb1+Rg1 and Re in "Liu" and "Cusido" hairy root cultures.

cultures was increased and that the "Liu" culture had a higher basal saponin level than the "Cusido" culture (Figure 5B). Since also a large publicly available EST sequence collection was derived from the "Liu" cultures (Choi et al., 2005), this hairy root culture was the preferred model system for the genome-wide cDNA-AFLP transcript profiling.

2.1.4. Elicitation of saponin biosynthesis in *C. asiatica* plants treated with MeJA.

Also for *C. asiatica*, it was reported that MeJA induces saponin production (Kim et al., 2004), hence, a jasmonate elicitation experiment was set up with two different types of *C. asiatica* plant materials: i) greenhouse grown plants grown from seeds obtained from a commercial source (www.SandMountainHerbs.com) and ii) callus cultures obtained from the lab of Prof. Rosa Cusido (Barcelona, Spain) (Figure 6A).

Based on the available public sequence information, primer sets were developed for the *BAS* (GenBank: AY520818), *SQS* (GenBank: AY787628) and *farnesyl pyrophosphate synthase* (*FPS*) (GenBank: AY787627) genes and a RT-PCR was performed to verify their expression pattern in all treated tissues. RT-PCR indicated opposite and tissue-dependent gene expression. On the one hand *FPS* and *SQS* expression was clearly induced by MeJA in greenhouse grown plants but not in cell suspension cultures. On the other hand, the *BAS* gene

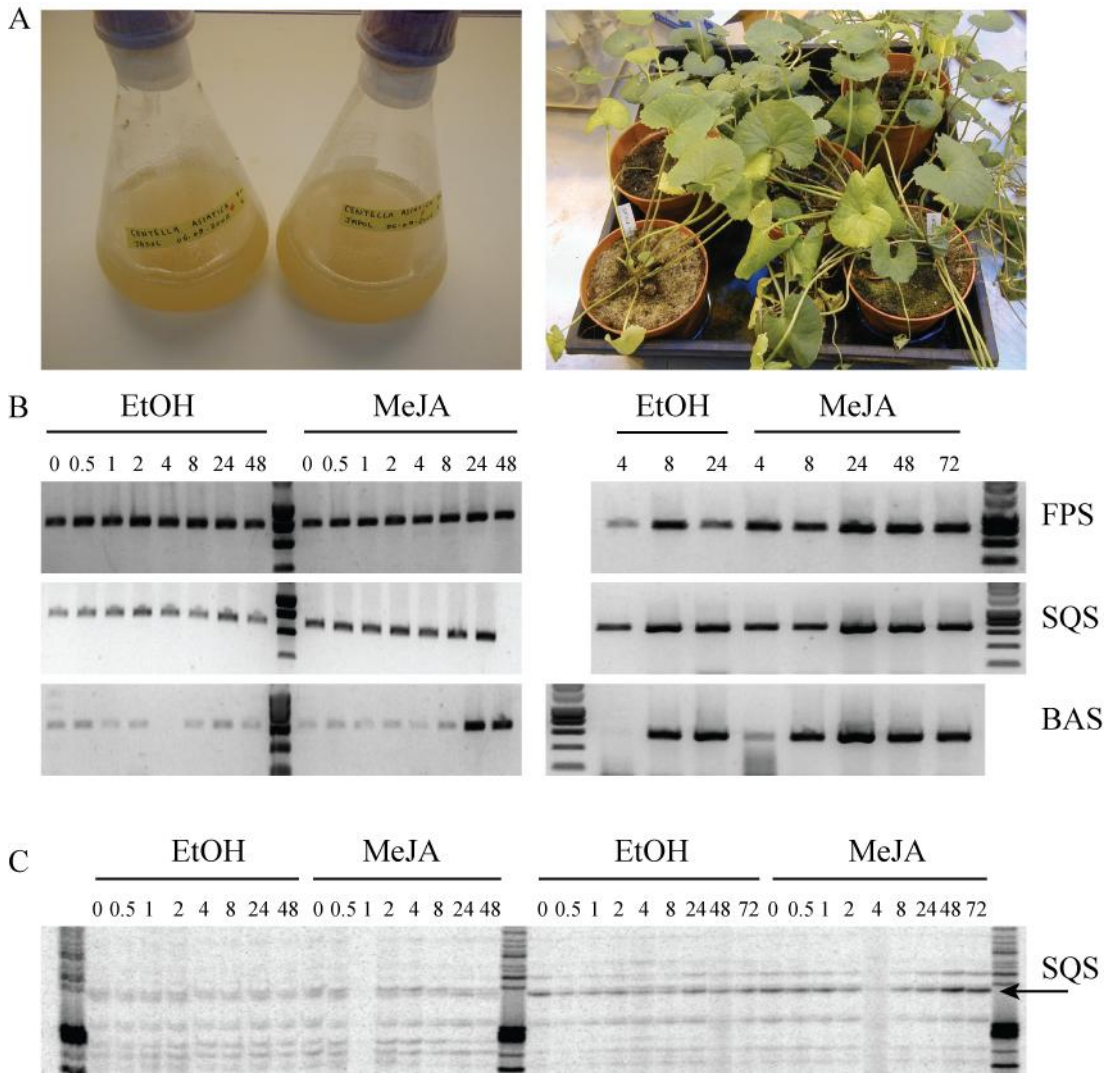


Figure 6. *C. asiatica* plant material and MeJA response. (A) *C. asiatica* cell cultures (left) and plants (right). (B) RT-PCR analysis of *FPS*, *SQS*, and *BAS* in MeJA treated cell cultures (left) and plants (right). Treatment and time-points (in hours) are indicated on top of the gel images. (C) Pilot cDNA-AFLP analysis, showing *SQS* induction in MeJA treated leaves (right), but not in MeJA treated cell cultures (left). Treatment and time-points (in hours) are indicated on top of the gel image.

was clearly upregulated in cell suspension cultures but not in greenhouse grown plants (Figure 6B). Identical results for the *FPS* and *SQS* gene tags were obtained when the cDNA-AFLP pilot was performed (Figure 5C).

Next, a new elicitation, but larger-scale, experiment was set up to allow measuring saponin content in elicited *C. asiatica* materials. In this experiment the cell suspension samples were harvested at 5, 7, and 9 days after elicitation, whereas the greenhouse grown plants were harvested 14 days after elicitation. This analysis indicated that only in greenhouse grown plants MeJA was capable of increasing saponin production, which means that the full

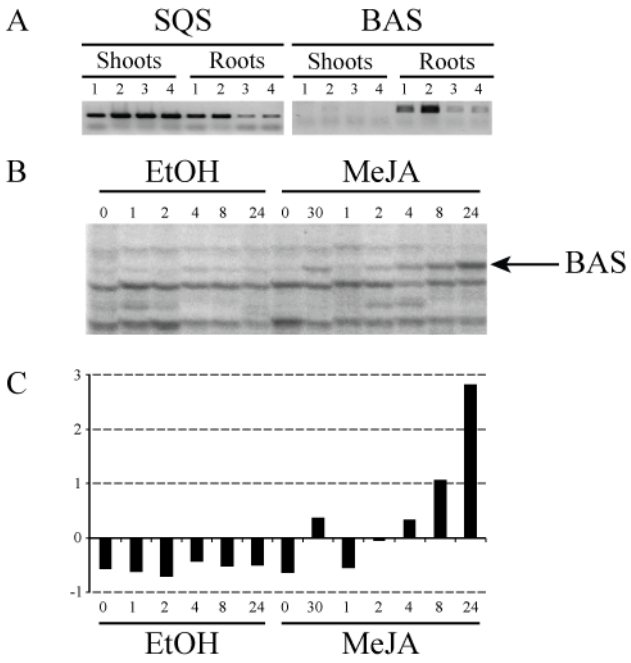


Figure 7. MeJA response of BAS in *B. falcatum* roots. A. RT-PCR analysis of *SQS* and *BAS* on 4 pools of hydroponically grown *B. falcatum* plants, pool numbers are indicated on top of the gel images. B. Detail of a cDNA-AFLP gel, the arrow marks the gene tag corresponding to *BAS*. Treatments and time-points are indicated on top of the gel image. C. Graphical representation of the expression of *BAS*, after normalization. For normalization details, see materials and methods. Treatments and time-points are indicated at the bottom of the graph.

genome-wide cDNA-AFLP analysis should be exclusively conducted on the latter material for saponin gene discovery.

2.1.5. Upregulation of genes involved in saikosaponin production in *B. falcatum* roots after MeJA treatment

The roots of *B. falcatum* are the traditional source of saikosaponins, and it was reported that the saponin production in *B. falcatum* root fragments was increased by MeJA treatment (Aoyagi et al., 2001). Available sequences of *B. falcatum SQS* and *BAS* (Kim et al., 2006) allowed performing an RT-PCR analysis on hydroponically grown *B. falcatum* plants, showing that, contrary to *SQS*, the expression of *BAS* is restricted to the roots of the plant (Figure 7A). The hydroponics system allowed efficient growing, elicitation and harvesting of the roots. Elicitation of *B. falcatum* plants was performed by adding MeJA to the hydroponics medium, at a final concentration of 50 µM. Subsequently, a pilot cDNA-AFLP experiment with a limited number of primer combinations was performed, which showed the response of the *B. falcatum* roots to the MeJA treatment. Notably, gene tags corresponding to *SQS* and *BAS* were found to be strongly MeJA responsive in *B. falcatum* root tissue (Figure 7B,C).

Together, these results suggest *B. falcatum* roots were apt for a functional genomics-based gene discovery program.

2.1.6. MeJA elicitation of *M. lanceolata* axenic shoot cultures

In Rwandan traditional medicine, aqueous extracts of *M. lanceolata* leaves and fruits are used for the treatment of various diseases, such as infectious hepatitis and bacillary dysentery, and it was shown that saponins isolated from the leaves were responsible for the antiviral, haemolytic and molluscicidal effects (Sindambiwe et al., 1996; Sindambiwe et al., 1998). Based on this information, a gene discovery program was initiated on *M. lanceolata* axenic shoot cultures. Since no gene sequence resources were available for *M. lanceolata*, the effect of the MeJA treatment was immediately investigated with a pilot cDNA-AFLP analysis, which confirmed the response of the axenic shoot cultures to the MeJA treatment on the transcript level. In parallel, accumulation of saponins in greenhouse-grown plants sprayed and irrigated with MeJA or a mock solution, and MeJA or mock-treated axenic shoot cultures was investigated. Contrary to *M. truncatula*, *G. glabra*, *C. asiatica*, and *P. ginseng*, no significantly increased saponin accumulation in the MeJA treated material could be observed. However, the total saponin content in the control greenhouse-grown plants was 4.9% (± 0.8%; n = 3), which was much higher than normally observed in leaves of in nature-grown, older trees (Theunis et al., 2007). Furthermore, also in the axenic shoot cultures, a very high saponin concentration was observed. This may indicate that the investigated tissues already accumulate maximal amounts of saponins and that it is impossible to further increase the saponin content in this material. Despite this, the transcriptional response was still present, and thus, the gene discovery program was performed on the MeJA treated axenic shoot cultures.

2.1.7. Conclusions on MeJA elicitations

Plants or plant tissue cultures were treated with either EtOH (as a control) or MeJA, and samples were harvested at different time points during 48 hours. Table 2 gives an overview of the used plant material, the elicitation conditions and the time selected time points for each of the target plants.

Table 2. Elicitation of the target plants.

Species	Plant material (harvested)	[MeJA]	Time points (hours)
M. truncatula	Cell cultures	100 µM	0, 0.5, 1, 2, 4, 8, 12, 24
G. glabra	Seedling cultures (roots & shoots)	100 µM	0, 0.5, 1, 2, 4, 8, 24, 48
M. lanceolata	Axenic shoot cultures (leaves)	500 µM	0, 0.5, 1, 2, 4, 8, 24, 48
P. ginseng	Hairy roots	100 µM	0, 0.5, 1, 2, 4, 8, 24, 48
C. asiatica	Plants (leaves) and cell cultures	100 µM	0, 0.5, 1, 2, 4, 8, 24, 48, 72
B. falcatum	Plants (hydroponics) (roots & shoots)	50 µM	0, 0.5, 1, 2, 4, 8, 24, 48

For all plants, elicitation was confirmed on the transcript level, via qRT-PCR, RT-PCR or a pilot cDNA-AFLP analysis. Furthermore, for four plants (*M. truncatula*, *G. glabra*, *C. asiatica* and *P. ginseng*), the elicitation was confirmed by metabolite profiling. Together, this allowed for the selection of tissue and time points for subsequent transcript profiling (Table 3). Because of sufficient gene resources from the other 5 plant species, a full *C. asiatica* cDNA-AFLP profiling was put on hold.

Table 3. Selected samples for genome-wide transcript profiling via cDNA-AFLP.

Species	Plant material	# Samples	Time points (hours)
M. truncatula	Cell cultures	16	EtOH: 0, 0.5, 1, 2, 4, 8, 12, 24 MeJA: 0, 0.5, 1, 2, 4, 8, 12, 24
G. glabra	Seedlings (roots & shoots)	32	EtOH: 0, 0.5, 1, 2, 4, 8, 24, 48 MeJA: 0, 0.5, 1, 2, 4, 8, 24, 48
M. lanceolata	Axenic shoots (leaves)	16	EtOH: 0, 0.5, 1, 2, 4, 8, 24, 48 MeJA: 0, 0.5, 1, 2, 4, 8, 24, 48
P. ginseng	Hairy roots	16	EtOH: 0, 0.5, 1, 2, 4, 8, 24, 48 MeJA: 0, 0.5, 1, 2, 4, 8, 24, 48
C. asiatica	Plants (leaves)	18	EtOH: 0, 0.5, 1, 2, 4, 8, 24, 48, 72 MeJA: 0, 0.5, 1, 2, 4, 8, 24, 48, 72
B. falcatum	Plants (hydroponics): roots	13	EtOH: 0, 1, 2, 4, 8, 24 MeJA: 0, 0.5, 1, 2, 4, 8, 24

2.2. Gene discovery for combinatorial biosynthesis by genome-wide transcript profiling

2.2.1. Transcript profiling of five plants by cDNA-AFLP

For the five target plants, a full, genome-wide transcript profiling was performed, using the complete set of PCRs and gel runs for a 128 *Bst*YI+1-*Mse*I+2 primer-combination (PC)

analysis. Together, a total of ≈71,000 transcript tags were visualized (Table 4). In addition to the *in silico* analysis of the gene expression patterns, visual inspection of the expression patterns on the gels allowed for the isolation of 4854 (6.8%) MeJA responsive transcript tags (Table 4). Remarkably, a much lower number of visualized tags and MeJA responsive tags were encountered in *M. truncatula*, as compared to the other plant species. The reason for this is unclear but it may be attributed to a high degree of ploidy or nucleotide polymorphisms in the seed populations acquired from a commercial source, leading to the visualization and apparent response of a higher number of redundant genes in the other plant species. Direct sequencing of the reamplified tags gave good-quality sequences for 3278 (67.5%) of the fragments (Table 4). To the remaining 1576 tags (32.5%), no unique sequence could be attributed unambiguously, indicating that they might not represent unique gene tags, hence, these tags were not considered for further analysis. Similar performance rates were obtained in previous cDNA-AFLP based transcript profiling studies (Goossens et al., 2003; Rischer et al., 2006; Maes et al., 2011). Blasting of the 3278 unique sequence tags, revealed that 2256 (68.8%) displayed sequence similarity to known plant sequences (Table 4). The percentage of hits in *M. truncatula* is remarkably high, as compared to the other plant species, due to the extensive genome and EST information available for this plant. Also for *G. glabra* and *P. ginseng*, a higher number of hits as compared to *M. lanceolata* or *B. falcatum* was observed. The homology between the two legume species, *M. truncatula* and *G. glabra*, may account for the higher number of hits in *G. glabra*, whereas the extensive EST database of *P. ginseng* (Choi et al., 2005) may account for the higher number of hits in *P. ginseng*. Hence, the low number of hits in *M. lanceolata* and *B. falcatum* likely was due to the low similarity between those plants and any plant species with an extensive genome or EST sequence database.

Table 4. Overview of the cDNA-AFLPs.

Species	**# PCs***	**Tags visualized**	**Differential tags**	**Unique tags**	**Hits**
M. truncatula	128	8462	282 (3,3%)	196 (69,5%)	191 (97,4 %)
G. glabra	128	14360	1009 (7,2%)	710 (70,4%)	574 (80,8 %)
M. lanceolata	128	13558	733 (5,4 %)	545 (74,4%)	312 (57,2 %)
P. ginseng	128	15840	1059 (6,7 %)	699 (66,0%)	501 (72,0 %)
B. falcatum	128	18800	1771 (9,4 %)	1128 (63,7%)	678 (60,1 %)
Total		71020	4854 (6,8 %)	3278 (67,5%)	2256 (68,8 %)

*PCs: Primer Combinations

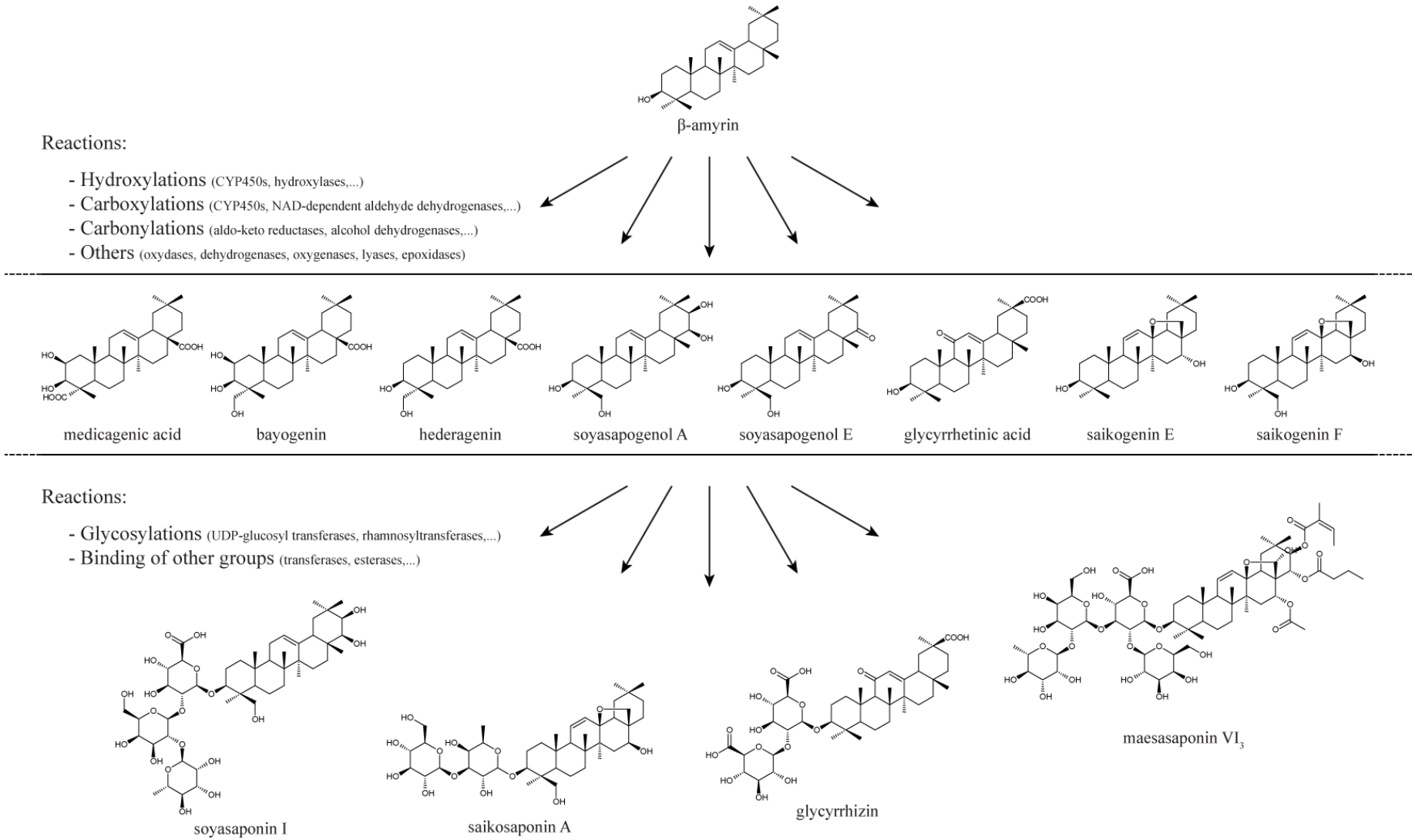


Figure 8. Distinct steps in the saponin biosynthesis. In a first step, the β-amyrin skeleton is oxidized on various positions to obtain the sapogenins. In a second distinct step, the sapogenins undergo glycosylation and binding of other groups. The predicted reactions are used to select candidate gene tags from the cDNA-AFLPs.

2.2.2. Selection of candidate genes

The selection of genes potentially involved in the biosynthesis of triterpene saponins in the investigated plants was based on two different criteria. First, only genes with an annotated activity that may be necessary for the saponin biosynthesis were considered. It is generally believed that, after cyclization of 2,3-oxidosqualene, two distinct types of additional modifications of the cyclization products complete the saponin biosynthesis and dramatically increase the structural diversity. Firstly, CytP450 mediated oxidations of the cyclization products give rise to different sapogenins. Subsequently (and/or concomitantly), the different sapogenins can be glycosylated on different positions, and other, less common, modifications, such as acylations and malonylations can further increase the structural variability of the saponins. Thus, genes potentially involved in any of these distinct steps (i.e. cyclization of oxidosqualene, oxidation of the cyclization products, glycosylation of the different sapogenins, and possible additional modifications) were selected (Figure 8).

Although it is generally believed that the oxidation of β-amyrin and dammarenediol are carried out by cytochrome P450 enzymes, we did not exclude that other types of oxidases can catalyze this step. A second criterion that was applied in the selection of the candidate genes was their expression pattern upon MeJA treatment. Genes that belong to the same metabolic pathway are often co-regulated, and as such, genes with a putative function that may be involved in later steps of the saponin biosynthesis and, if applicable, were co-regulated with the known saponin biosynthesis genes were selected as candidate genes. Additionally, for *M. truncatula*, *G. glabra* and *M. lanceolata*, a set of transcription factors with a MeJA response prior to or concomitant with the response of the biosynthesis genes were identified and selected as potential regulators of the saponin biosynthesis.

2.2.2.1. Selection of *M. truncatula* genes

Average linkage hierarchical clustering showed that upon MeJA treatment, the MeJA-responsive genes are either transcriptionally activated (73%, clusters II-V, Figure 9) or transcriptionally repressed (27%, cluster I, Figure 9). The activated gene tags can be further divided into 3 subclusters, based on the timing of the response (Figure 9). A first cluster consists of genes with a response immediately after (30 minutes or less) the MeJA treatment (cluster II, Figure 9). This cluster is highly enriched in putative transcription factors. In a second cluster (cluster III, Figure 9), genes roughly respond 1 to 4 hours after MeJA

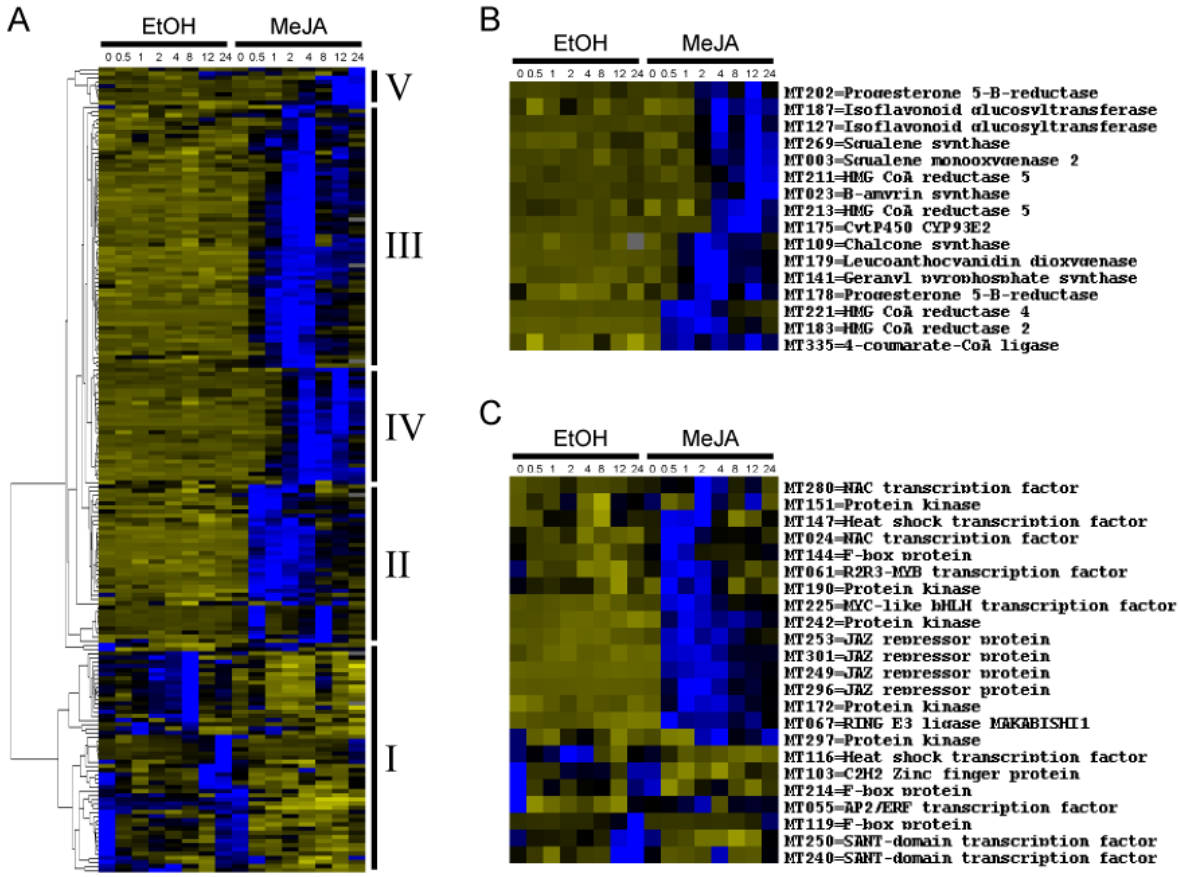


Figure 9. Transcriptome of jasmonate-elicited *M. truncatula* cells. A. General view on the average linkage hierarchical cluster and tree of all 196 unique *M. truncatula* gene tags. B. Subcluster of the *M. truncatula* transcriptome comprising tags corresponding to genes reported to be involved in saponin biosynthesis, or with high sequence similarity to such genes. C. Subcluster of the *M. truncatula* transcriptome comprising tags corresponding to genes encoding putative transcription factors or proteins involved in posttranslational protein modification and ubiquitin-mediated proteasomal degradation. Treatments and time points (in h) are indicated at the top. Blue and yellow boxes reflect transcriptional activation and repression relative to the average expression level, respectively. Gray boxes correspond to missing time points.

treatment. In a third group (clusters IV and V, Figure 9), genes respond 4 hours or later after the MeJA treatment. Interestingly, in the latter group, gene tags can be found corresponding to *SQS*, *SQE2* and *BAS*, all genes encoding enzymes catalyzing triterpene saponin biosynthesis. These tags displayed an almost identical expression pattern, thus suggesting a tight co-regulation, and reached maximum levels of expression 12 to 24 hours post elicitation (Figure 9). Thus, candidate saponin biosynthesis genes should be enzyme encoding genes with a potential function in saponin biosynthesis that cluster within this third group. The most promising candidate of this list is the gene corresponding to the tag Mt175. A blast search of the Mt175 tag in the *M. truncatula* EST database (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=medicago), revealed 100% homology to TC173919, corresponding to

Table 5. Overview of the selected *M. truncatula* cDNA-AFLP tags.

Tag	TC-code	Class	(putative) annotation
MT023	TC172516	Prior steps	β-amyrin synthase (Suzuki et al., 2002)
MT141	-	Prior steps	geranyl diphosphate synthase
MT173	TC173832	CytP450	9/13 hydroperoxide lyase
MT175	TC173919	CytP450	β-amyrin and sophoradiol 24-hydroxylase
MT004	TC183962	Ox/Red	mannitol dehydrogenase
MT019	TC174444	Ox/Red	pathogenesis-related protein
MT042	TC174665	Ox/Red	aldo/keto reductase
MT123	TC173670	Ox/Red	1-aminocyclopropanecarboxylic acid oxidase
MT138	TC173742	Ox/Red	β-glucosidase G2 (Naoumkina et al., 2007)
MT149	TC199332	Ox/Red	quinone oxidoreductase-like protein
MT178	TC180599	Ox/Red	progesterone 5-β-reductase
MT202	TC183591	Ox/Red	progesterone 5-β-reductase
MT260	TC185299	Ox/Red	alpha-dioxygenase
MT270	TC185273	Ox/Red	amine oxidase
MT308	TC180073	Ox/Red	probable mannitol dehydrogenase
MT309	TC186395	Ox/Red	2OG-Fe(II) oxygenase
MT038	TC178291	UGT	glucosyltransferase-13
MT127	TC179830	UGT	isoflavonoid glucosyltransferase
MT187	BG582596	UGT	isoflavonoid glucosyltransferase
-	-	UGT	UGT71G1 (Achnine et al., 2005)
-	-	UGT	UGT73F3 (Naoumkina et al., 2010)
-	-	UGT	UGT73K1 (Achnine et al., 2005)
MT024	TC188018	TF	NAC family transcription factor
MT061	TC173581	TF	MYB family transcription factor
MT067	TC180525	TF	RING domain-containing protein, E3 Ubiquitin ligase
MT097	TC190450	TF	hypothetical protein
MT148	TC176048	TF	CCR4-NOT transcription complex protein
MT185	TC184599	TF	hypothetical protein, calmodulin binding?
MT200	TC184840	TF	hypothetical protein
MT203	TC185509	TF	LOB domain-containing protein
MT225	TC186327	TF	MYC2 transcription factor
MT274	TC184516	TF	hypothetical protein
MT280	TC185226	TF	NAC family transcription factor
MT296	TC174418	TF	JAZ domain-containing protein

CytP450: Cytochrome P450s; Ox/Red: Oxidases and reductases; UGT: Glycosyltransferases; TF: Transcription factors and other regulatory proteins.

CYP93E2, the *M. truncatula* CytP450 gene with the highest homology to the soybean CytP450 gene encoding the β-amyrin and sophoradiol 24-hydroxylase (Figure 2) (Shibuya et al., 2006). This enzyme was also identified as a candidate gene involved in triterpene saponin biosynthesis in another screen (Naoumkina et al., 2010). Next to Mt175, one more gene tag corresponding to a CytP450 enzyme showed co-regulation with the known saponin

biosynthesis genes, Mt173. This gene tag of 385 bp is 100% homologous to TC173832, a TC annotated as 9/13 hydroperoxide lyase. However, this annotation is only based on homology with other hydroperoxide lyases, and thus, this gene tag was selected as a candidate gene as well. Several other co-regulated gene tags were selected, having a predicted oxidase/reductase or glucosyltransferase activity. Furthermore, 12 transcription factors and proteins of unknown function with a jasmonate response prior to (or concurrently with) the saponin biosynthesis genes were selected, such as Mt067 (See Chapter 6). An overview of the selected candidate gene tags is given in Table 5.

2.2.2.2.Selection of *G. glabra* genes

For *G. glabra*, average linkage hierarchical clustering based expression analysis showed that for 13% and 33% of the Gg tags, the MeJA response was restricted to leaves and roots respectively, whereas for the remaining 54% MeJA modulated differential expression was observed both in roots and leaves. This observation points to a distinct JA response, dependent on the tissue where the signal is perceived or transmitted. Both in leaves and roots MeJA predominantly stimulated gene expression, which is the case for 80% of the Gg tags, rather than repressing gene expression. Interestingly, cDNA-AFLP tag Gg0003, which corresponds to the *G. glabra BAS* (Hayashi et al., 2001), showed a differential response in the roots and the leaves of the plant. In roots, the expression of *BAS* was induced within 1 hour after MeJA treatment, whereas in shoots, the gene expression was induced much later (starting at 24 hours). A tight co-regulated expression pattern with the *BAS* tag is also observed for the tags corresponding to genes involved in the cytosolic isoprene precursor or mevalonate (MVA) pathway (i.e. tags Gg0577 and Gg0869, Figure 9). Also tag Gg0161 showed MeJA modulated expression predominantly in roots, concomitant in time with root induction of Gg0003 (Figure 10). Based on sequence similarity Gg0161 could be considered as an orthologue of the *M. truncatula UGT73K1* gene (Achnine et al., 2005), encoding the uridine diphosphate glucosyltransferase that catalyses glycosylation of e.g. soyasapogenol B, an aglycone that is also encountered in liquorice. Thus, the candidate saponin biosynthesis genes that were eventually selected had an expression pattern that showed an early induction in roots, and a late induction in shoots. As such, a total of 49 candidate biosynthesis genes were selected. Additionally, 20 candidate transcription factors were selected.

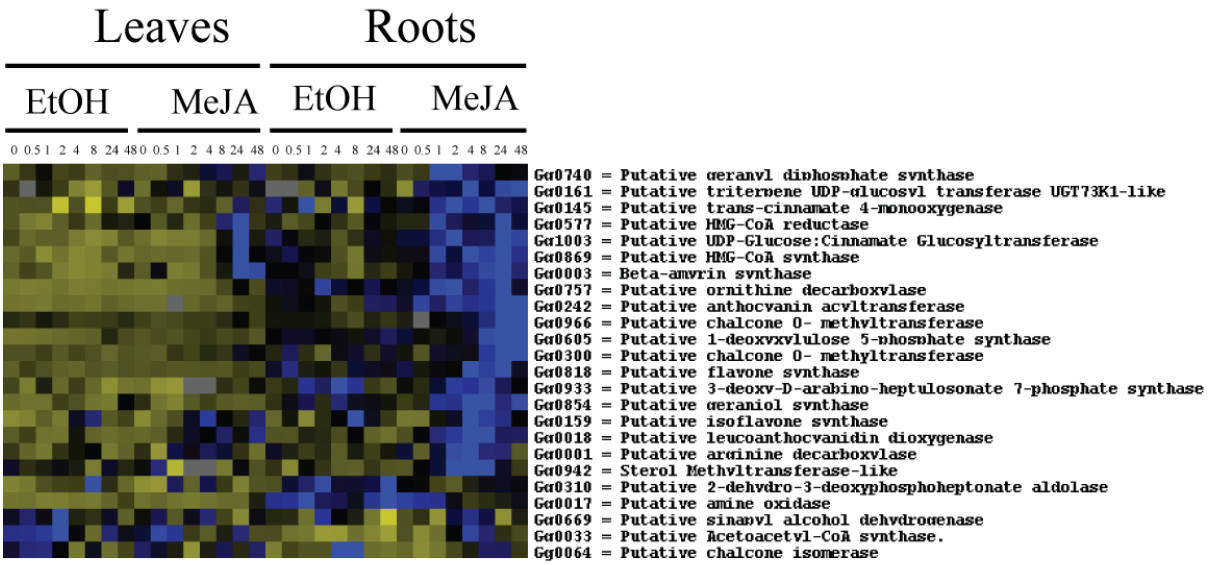


Figure 10. Transcriptome of jasmonate-elicited *G. glabra* seedlings. Subcluster of the *G. glabra* transcriptome comprising tags corresponding to genes reported to be involved in terpene or phenylpropanoid biosynthesis, or with high sequence similarity to such genes. Treatments and time points (in h) are indicated at the top. Blue and yellow boxes reflect transcriptional activation and repression relative to the average expression level, respectively. Gray boxes correspond to missing time points.

2.2.2.3.Gene selection for other plants

For the three other plant species, a similar approach to select candidate biosynthesis genes was applied. Average linkage hierarchical clustering of the *B. falcatum* cDNA-AFLP tags showed the presence of a cluster with a late induced expression, starting at 4 hours, that includes gene tags corresponding to enzymes of both the MVA and MEP isoprenoid pathways, as well as tags corresponding to the known *B. falcatum* enzymes *SQS* and *BAS* (Kim et al., 2006). Furthermore, a gene tag with homology to the *M. truncatula UGT73K1* gene (Achnine et al., 2005) clustered within this group. Hence, gene tags that have an annotated function that is possibly involved in the *B. falcatum* saponin biosynthesis, and that have a late induced expression were selected as candidate genes. As such, a set of 64 *B. falcatum* cDNA-AFLP tags was selected.

Also for *P. ginseng*, several known saponin biosynthesis genes were shown to respond to the MeJA treatment. The expression of Pg0408, a gene tag corresponding to *P. ginseng DDS* (Tansakul et al., 2006) was induced within 2 hours after MeJA treatment and showed co-regulation with cDNA-AFLP tags corresponding to *FPS* (Kim et al., 2010), *SQS2* (GenBank: GQ468527) and *SQE2* (Han et al., 2010). Furthermore, several genes from both the MVA

and MEP pathways were co-regulated with this gene set. Finally, 41 cDNA-AFLP tags with an expression profile similar to this gene set, and a putative function involved in ginsenoside biosynthesis were selected.

Within the MeJA modulated *M. lanceolata* tags, we could identify gene tags corresponding to early terpene precursor biosynthetic pathway genes, as was also the case with the other plant species, and additionally also some genes potentially involved in terpene biosynthesis branches leading to the monoterpenes and triterpenes, pointing to a true MeJA modulated redirection of terpenoid biosynthesis in leaves of *M. lanceolata* axenic shoot cultures. Based on the gene expression pattern and the putative gene annotation, a set of 36 gene tags was selected as potentially involved in the saponin biosynthesis pathway. Additionally, 16 putative transcription factors were selected. A gene tag corresponding to *BAS* was obtained via degenerate oligonucleotide primers based on known plant genes.

2.2.3. Conclusions on cDNA-AFLP transcript profiling

In conclusion, a total of 259 cDNA-AFLP tags were selected as candidate genes involved in saponin biosynthesis (Table 6).

Table 6. Overview of the selected genes.

	Mt	Gg	Ml	Pg	Bf
Previous steps	2	0	2	2	2
Oxidases/reductases	14	22	9	16	32
CytP450s	2	12	14	7	10
Glucosyltransferases	5	10	6	10	12
Transferases/esterases	0	5	6	6	8
Transcription factors	9	20	16	0	0
Total selected	32	69	53	41	64

3. Materials and Methods

3.1. Plant cultivation and elicitation

M. truncatula cell cultures (kindly provided by Richard A Dixon, Ardmore, USA), were maintained as described (Suzuki et al., 2002). For elicitation, 7 days after inoculation of 75 mL of a 14-day-old suspension culture into 175 mL fresh medium, cells were treated with

100 µM MeJA (dissolved in EtOH) or an equivalent amount of EtOH as a control. Samples were harvested, vacuum filtrated, and frozen at -80°C. For transcript profiling, samples were taken 0, 0.5, 1, 2, 4, 8, 12, 24 h after elicitor or mock treatments and for metabolite profiling 0 and 5 days after treatments.

G. glabra seeds, obtained from a commercial source (www.SandMountainHerbs.com), were surface sterilized by a short ethanol treatment (70%, v/v), followed by a 7% (v/v) bleach treatment for 20 min. Subsequently, sterilized seeds were washed thrice with sterile water and allowed to germinate on wet, sterile 3 mm Whatman paper. After germination, seedlings were transferred to 50 mL Schenk and Hildebrandt medium in 250 mL Erlenmeyer flasks, and grown on a shaker under 16h/8h light/dark regime, at 22°C, and 60 rpm. Three to 4 weeks after germination, seedlings were treated with 100 µM MeJA or an equivalent amount of EtOH as a control. Samples for transcript profiling were taken 0, 0.5, 1, 2, 4, 8, 24, and 48 h after treatment, and for metabolite profiling 0 and 5 days after treatment. Roots and leaves were harvested separately, and frozen at -80°C.

M. lanceolata axenic shoot cultures (kindly provided by Prof. Danny Geelen, Gent, Belgium) were maintained as described (Faizal et al., 2011). For elicitation, each pot of shoot cultures was sprayed with 2.0 mL deionized water, containing 0.05% (v:v) Tween-20 and 500 µM MeJA (10 µl 100 mM stock (dissolved in EtOH) per 2 mL water) or an equivalent amount of EtOH as a control. For transcript profiling, samples were taken 0, 0.5, 1, 2, 4, 8, 24, 48 h after elicitor or mock treatments and for metabolite profiling 0 and 5 days after treatments. For each sample, 3 different plants of three different pots were pooled.

P. ginseng hairy root cultures (kindly provided by Prof. Rosa Cusido, Barcelona, Spain and Prof. Jang-Ryol Liu, Taejon, South-Korea) were maintained as described (Choi et al., 2005). Two weeks after subculturing into fresh medium, they were treated with 100 µM MeJA (dissolved in EtOH) or an equivalent amount of EtOH as a control. Samples were harvested by vacuum filtration and frozen in liquid nitrogen 0, 0.5, 1, 2, 4, 8, 24, 48 h after elicitor or mock treatments for transcript profiling, and 14 days after treatments for metabolite profiling.

B. falcatum seeds, obtained from a commercial source (www.SandMountainHerbs.com), were sown in soil, and 2 weeks after germination, seedlings were transferred to aerated hydroponics medium containing 1 g/L 10-30-20 salts (Scotts, Ohio, USA), pH 6.5. Plants were grown at 16h/8h light/dark regime, at 21°C, and pH was monitored and daily adjusted to

6.5, by adding KOH to the hydroponics medium. Three weeks after the plants were transferred to the hydroponics medium, they were treated with 50 µM MeJA (dissolved in EtOH) or an equivalent amount of EtOH as a control, by adding the EtOH or MeJA solution to the hydroponics medium. For transcript profiling, root and leaf samples were harvested separately, 0, 0.5, 1, 2, 4, 8, 24, and 48 h after treatment, frozen in liquid nitrogen, and stored at -70°C. For each sample, 3 individual plants were pooled.

C. asiatica callus cultures (kindly provided by Prof. Rosa Cusido, Barcelona, Spain) were maintained as described (Mangas et al., 2008). For elicitation, 7 days after inoculation of 75 mL of a 14-day-old suspension culture into 175 mL fresh medium, cells were treated with 100 µM MeJA (dissolved in EtOH) or an equivalent amount of EtOH as a control. Samples were harvested, vacuum filtrated, and frozen at -80°C. For transcript profiling, samples were taken 0, 0.5, 1, 2, 4, 8, 24, and 48 h after elicitor or mock treatments and for metabolite profiling 0, 5, 7, and 9 days after treatments. *C. asiatica* seeds, obtained from a commercial source (www.SandMountainHerbs.com), were sown in soil, and after germination, a single plant was multiplied by cutting the creeping stolons. Elicitation was performed by watering and spraying the plants with a 100 µM MeJA solution in water. For transcript profiling, samples were taken 0, 0.5, 1, 2, 4, 8, 24, 48, and 72 h after elicitor or mock treatments and for metabolite profiling 0 and 14 days after treatments.

3.2.*Medicago* saponin profiling

To screen *M. truncatula* suspension cells for the presence of known *Medicago* saponins (Huhman and Sumner, 2002; Huhman et al., 2005; Kapusta et al., 2005a; Kapusta et al., 2005b), 0.5 g of lyophilized material was crushed in a mortar. Saponins were extracted with 40 ml MeOH (80%, v/v) by overnight stirring. After centrifugation (45 min, 3000×g), the supernatant was filtered and dried under vacuum. To clean up the sample, it was dissolved in 6 ml H_2O and brought on an Chromabond® SPE C18 cartridge (500 mg) (Machery-Nagel, Germany). The column was successively rinsed with 1 volume H_2O, MeOH (10% v/v) and MeOH (40% v/v). Finally the saponins were eluted with 12 mL MeOH (100% v/v). This fraction was dried under vacuum and dissolved in 3 mL acetonitrile (ACN)/H_2O (1:1). Saponins were identified by LC-MS and LC-MS/MS (Esquire Ion Trap 3000 plus). The chromatographic analysis was performed with a LiChroSpher 5 µ (250 x 4 mm; Merck, Germany) with a gradient (from 18 to 36% B in 55 min; from 36 to 95% B in 30 min ([A = 0.05% HAc; B = ACN + 0.05% HAc]). The flow rate was 0.5 mL min^{-1} and the UV detector

was set at 210 nm. This method (based on (Kapusta et al., 2005a)) for *Medicago* saponin detection was optimized with the Soyasaponin B mixture (kindly provided by Mark A. Berhow, National Center for Agricultural Utilization Research Center, USA).

3.3.*G. glabra* saponin profiling

To determine the quantity of glycyrrhizin in licorice tissues, frozen samples were lyophilized and powdered. An HPLC method for the determination of glycyrrhizin was developed based on an existing method (Liquiritiae extractum fluidum normatum, German Pharmacopoeia (DAB)) and optimized. The extraction procedure, the extraction solvent and the extraction yield were investigated. For extraction, 0.5 g of plant material was sonicated in EtOH (70%; 2 x 15 mL) for 1 h. After centrifugation (15 min, 3000×g), the supernatant was dried under vacuum and dissolved in 10 mL MeOH (80% v/v). Before measurement on LC-MS, this solution was diluted 10 fold. The chromatographic analysis was performed with a LiChroSpher 5 µ (250 x 4 mm; Merck, Germany) with a gradient (from 15 to 60% B in 50 min; from 60 to 95% B in 10 min [A = 0.05% HAc; B = ACN + 0.05% HAc]). The flow rate was 0.8 mL min^{-1} and the UV detector was set at 254 nm. A reversed-phase HPLC-UV system coupled to an ESI-MS detector was used to evaluate the samples, with glycyrrhizic acid-NH_4 salt (Serva, Germany) as external standard. The compound identity was verified with negative ESI-MS (Esquire Ion Trap 3000 plus, Bruker Daltonics, Germany). This method was fully validated according to the ICH guidelines (Text on Validation of Analytical Procedures,1994; Validation of Analytical procedures: Methodology,1996). The linearity of glycyrrhizin was investigated: graphical inspection of the residuals, the LOF test and the correlation 19 coefficients proved the method to be linear in the range tested (25 – 400 µg mL^{-1}). The *t*-test on the intercepts revealed that point (0,0) falls within the calibration curve. The precision (n=18; six replicates on 3 days; RSD between days = 4.22%) and the accuracy (n=9; three replicates of three levels; recovery = 87%; RSD% = 11%) are proven to be acceptable with the European Pharmacopoeia approved samples of liquiritiae radix that contained a mean content of glycyrrhizin of 5.7%.

3.4.*C. asiatica* saponin profiling

0.5 g of lyophilized material was crushed in a mortar. Saponins were extracted twice with 40 ml MeOH (90%, v/v) for 1 hour, followed by filtration and evaporation of the MeOH under vacuum. To clean up the sample, it was dissolved in 15 ml MeOH and water was added till a

volume of 50 ml. subsequently, 5 ml of the sample was brought on an Chromabond® SPE C18 cartridge (500 mg) (Machery-Nagel, Germany). The column was successively rinsed with 1 volume H_2O, MeOH (10% v/v) and MeOH (40% v/v). Finally the saponins were eluted with 12 mL MeOH (100% v/v). This fraction was dried under vacuum and dissolved in 3 mL acetonitrile (ACN)/H_2O (1:1). Saponins were identified by LC-MS and LC-MS/MS (Esquire Ion Trap 3000 plus). The chromatographic analysis was performed with a LiChroSpher 5 µ (250 x 4 mm; Merck, Germany) with a gradient (from 20 to 36% B in 15 min; from 36 to 70% B in 7 min ([A = 0.05% HAc; B = ACN + 0.05% HAc]). The flow rate was 0.5 mL min^{-1} and the UV detector was set at 214 nm.

3.5.*M. lanceolata* and *P. ginseng* saponin profiling

M. lanceolata saponin profiling was performed according to Theunis et al. (2007).

The extraction and HPLC method for ginsenosides from the European Pharmacopoeia (2002) was used, with Rb1 and Rg1 as external standards.

3.6.Transcript profiling

For reverse transcription PCR (RT-PCR), cDNA was prepared with SuperScript™ II Reverse Transcriptase (Invitrogen), and primers were designed on the available sequences using the Primer3 program (Rozen and Skaletsky, 2000).

For quantitative Real Time PCR (qRT-PCR), total RNA was extracted with the RNeasy mini kit (Qiagen), and cDNA prepared with SuperScript™ II Reverse Transcriptase (Invitrogen). Primers were designed with Beacon Designer version 4.0 (Premier Biosoft International, Palo Alto, CA, USA). qRT-PCR was carried out using a Lightcycler 480 (Roche) and SYBR Green QPCR Master Mix (Stratagene). For reference genes, 40S ribosomal protein S8 (40S) (TC160725 of the MTGI from TIGR) and translation elongation factor 1α (ELF1α) (TC148782 of the MTGI from TIGR) were used. Reactions were performed in triplicate, and relative quantification with multiple reference genes was performed with qBase (Hellemans et al., 2007).

For cDNA-AFLP, after preparation of total RNA, double-stranded cDNA was prepared as described (Vuylsteke et al., 2007). After appropriate sample preparation, cDNA-AFLP based transcript profiling was performed as described, with all 128 possible *Bst*YI+1/*Mse*I+2

primer combinations (Vuylsteke et al., 2007). Gel images were analyzed with the AFLP-QUANTARPRO software (Keygene, Wageningen, The Netherlands), allowing accurate quantification of band intensities. The intensity of all individual bands was determined, and the obtained raw expression data were corrected for lane variations (due to PCR or loading differences) by dividing the raw data by a correction factor. The correction factor was calculated by dividing the sum of the expression levels of all fragments within one lane by the highest sum of all lanes within a primer combination. Subsequently, the standard deviation (SD) and the average were calculated for each individual band. Individual gene expression profiles were variance normalized by subtracting the calculated average from each individual data point, after which the obtained value was divided by the SD. A coefficient of variation (CV) was obtained by dividing the SD by the calculated average. Gene tags displaying expression values with a CV $\geq$0.5 were considered as differentially expressed. Based on this cut-off value, together with visual inspection of cDNA-AFLP gels, differentially expressed gene tags were selected for further analysis. Cluster analysis, sequencing, and BLAST analysis were performed as described (Rischer et al., 2006).

Acknowledgements

We thank Wilson Ardiles-Diaz for sequencing the cDNA-AFLP tags, Richard A. Dixon (Samuel Roberts Noble Foundation, Ardmore, USA) for generously providing the *M. truncatula* cell line, Mark A. Berhow, (National Center for Agricultural Utilization Research Center, USA) for providing the soyasaponin B mixture, Prof. Rosa Cusido, (Barcelona, Spain) and Prof. Jang-Ryol Liu, (Taejon, South-Korea) for the *C. asiatica* cell cultures, and *P. ginseng* "Cusido" and "Liu" hairy roots, respectively, and Prof. Danny Geelen (Gent, Belgium) for the *Maesa* plants. This work was supported by the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" project SBO040093).

References

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

Aoyagi, H., Kobayashi, Y., Yamada, K., Yokoyama, M., Kusakari, K., and Tanaka, H. (2001). Efficient production of saikosaponins in *Bupleurum falcatum* root fragments combined with signal transducers. Appl Microbiol Biotechnol **57**, 482-488.

Augustin, J.M., Kuzina, V., Andersen, S.B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry **72**, 435-457.

Choi, D.W., Jung, J.D., Ha, Y.I., Park, H.W., In, D.S., Chung, H.J., and Liu, J.R. (2005). Analysis of transcripts in methyl jasmonate-treated ginseng hairy roots to identify genes involved in the biosynthesis of ginsenosides and other secondary metabolites. Plant Cell Rep **23**, 557-566.

De Boer, K., Tilleman, S., Pauwels, L., Vanden Bossche, R., De Sutter, V., Vanderhaeghen, R., Hilson, P., Hamill, J.D., and Goossens, A. (2011). AP2/ERF and bHLH tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. Plant J, **in press** doi: 10.1111/j.1365-313X.2011.04566.x.

De Sutter, V., Vanderhaeghen, R., Tilleman, S., Lammertyn, F., Vanhoutte, I., Karimi, M., Inzé, D., Goossens, A., and Hilson, P. (2005). Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J **44**, 1065-1076.

Dixon, R.A., and Sumner, L.W. (2003). Legume natural products: Understanding and manipulating complex pathways for human and animal health. Plant Physiol **131**, 878-885.

European Pharmacopoeia (2002). Council of Europe; Strasbourg, Vol. **4**, 1244.

Faizal, A., Lambert, E., Foubert, K., Apers, S., and Geelen, D. (2011). *In vitro* propagation of four saponin producing *Maesa* species. Plant Cell Tiss Organ Cult, **in press** doi: 10.1007/s11240-010-9909-z.

Field, B., and Osbourn, A.E. (2008). Metabolic diversification - independent assembly of operon-like gene clusters in different plants. Science **320**, 543-547.

Goossens, A., Häkkinen, S.T., Laakso, I., Seppänen-Laakso, T., Biondi, S., De Sutter, V., Lammertyn, F., Nuutila, A.M., Söderlund, H., Zabeau, M., Inzé, D., and Oksman-Caldentey, K.M. (2003). A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proc Natl Acad Sci U S A **100**, 8595-8600.

Häkkinen, S.T., Tilleman, S., Świątek, A., De Sutter, V., Rischer, H., Vanhoutte, I., Van Onckelen, H., Hilson, P., Inzé, D., Oksman-Caldentey, K.M., and Goossens, A. (2007). Functional characterisation of genes involved in pyridine alkaloid biosynthesis in tobacco. Phytochemistry **68**, 2773-2785.

Han, J.Y., In, J.G., Kwon, Y.S., and Choi, Y.E. (2010). Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng*. Phytochemistry **71**, 36-46.

Hayashi, H., Fukui, H., and Tabata, M. (1993). Distribution pattern of saponins in different organs of *Glycyrrhiza glabra*. Planta Med **59**, 351-353.

Hayashi, H., Huang, P.Y., and Inoue, K. (2003). Up-regulation of soyasaponin biosynthesis by methyl jasmonate in cultured cells of *Glycyrrhiza glabra*. Plant Cell Physiol **44**, 404-411.

Hayashi, H., Hirota, A., Hiraoka, N., and Ikeshiro, Y. (1999). Molecular cloning and characterization of two cDNAs for *Glycyrrhiza glabra* squalene synthase. Biol Pharm Bull **22**, 947-950.

Hayashi, H., Huang, P.Y., Kirakosyan, A., Inoue, K., Hiraoka, N., Ikeshiro, Y., Kushiro, T., Shibuya, M., and Ebizuka, Y. (2001). Cloning and characterization of a cDNA encoding β-Amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. Biol Pharm Bull **24**, 912-916.

Huhman, D.V., and Sumner, L.W. (2002). Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry **59**, 347-360.

Huhman, D.V., Berhow, M.A., and Sumner, L.W. (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. J Agric Food Chem **53**, 1914-1920.

Kapusta, I., Janda, B., Stochmal, A., and Oleszek, W. (2005a). Determination of saponins in aerial parts of barrel medic (*Medicago truncatula*) by liquid chromatography-electrospray ionization/mass spectrometry. J Agric Food Chem **53**, 7654-7660.

Kapusta, I., Stochmal, A., Perrone, A., Piacente, S., Pizza, C., and Oleszek, W. (2005b). Triterpene saponins from barrel medic (*Medicago truncatula*) aerial parts. J Agric Food Chem **53**, 2164-2170.

Kim, O.T., Kim, M.Y., Hong, M.H., Ahn, J.C., and Hwang, B. (2004). Stimulation of asiaticoside accumulation in the whole plant cultures of *Centella asiatica* (L.) Urban by elicitors. Plant Cell Rep **23**, 339-344.

Kim, O.T., Bang, K.H., Kim, Y.C., Hyun, D.Y., Kim, M.Y., and Cha, S.W. (2009). Upregulation of ginsenoside and gene expression related to triterpene biosynthesis in ginseng hairy root cultures elicited by methyl jasmonate. Plant Cell Tiss Organ Cult **98**, 25-33.

Kim, O.T., Bang, K.H., Jung, S.J., Kim, Y.C., Hyun, D.Y., Kim, S.H., and Cha, S.W. (2010). Molecular characterization of ginseng farnesyl diphosphate synthase gene and its up-regulation by methyl jasmonate. Biol Plant **54**, 47-53.

Kim, Y.S., Cho, J.H., Ahn, J., and Hwang, B. (2006). Upregulation of isoprenoid pathway genes during enhanced saikosaponin biosynthesis in the hairy roots of *Bupleurum falcatum*. Mol Cells **22**, 269-274.

Lackman, P., González-Guzmán, M., Tilleman, S., Carqueijeiro, I., Cuéllar Pérez, A., Moses, T., Seo, M., Kanno, Y., Häkkinen, S.T., Van Montagu, M.C.E., Thevelein, J.M., Maaheimo, H., Oksman-Caldentey, K.M., Rodriguez, P.L., Rischer, H., and Goossens, A. (2011). Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in Arabidopsis and tobacco. Proc Natl Acad Sci U S A **108**, 5891-5896.

Lü, J.M., Yao, Q.Z., and Chen, C.Y. (2009). Ginseng compounds: an update on their molecular mechanisms and medical applications. Curr Vasc Pharmacol **7**, 293-302.

Maes, L., Van Nieuwerburgh, F.C.W., Zhang, Y.S., Reed, D.W., Pollier, J., Casteele, S., Inzé, D., Covello, P.S., Deforce, D.L.D., and Goossens, A. (2011). Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants. New Phytol **189**, 176-189.

Mangas, S., Moyano, E., Osuna, L., Cusido, R.M., Bonfill, M., Palazón, J. (2008). Triterpenoid saponin content and the expression level of some related genes in calli of *Centella asiatica*. Biotechnol Lett **10**, 1853-1859.

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

Morita, M., Shitan, N., Sawada, K., Van Montagu, M.C.E., Inzé, D., Rischer, H., Goossens, A., Oksman-Caldentey, K.M., Moriyama, Y., and Yazaki, K. (2009). Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. Proc Natl Acad Sci U S A **106**, 2447-2452.

Naoumkina, M.A., Farag, M.A., Sumner, L.W., Tang, Y., Liu, C.J., and Dixon, R.A. (2007). Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc Natl Acad Sci U S A **104**, 17909-17915.

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

Oksman-Caldentey, K.M., and Inzé, D. (2004). Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci **9**, 433-440.

Pauwels, L., Inzé, D., and Goossens, A. (2009). Jasmonate-inducible gene: what does it mean? Trends Plant Sci **14**, 87-91.

Phillips, D.R., Rasbery, J.M., Bartel, B., and Matsuda, S.P.T. (2006). Biosynthetic diversity in plant triterpene cyclization. Curr Opin Plant Biol **9**, 305-314.

Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A., Melton, R., and Osbourn, A. (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. Proc Natl Acad Sci U S A **103**, 18848-18853.

Rischer, H., Orešič, M., Seppänen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M.C.E., Inzé, D., Oksman-Caldentey, K.M., and Goossens, A. (2006). Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc Natl Acad Sci U S A **103**, 5614-5619.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132**, 365-386.

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., Akashi, T., Aoki, T., Saito, K., and 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 U S A **105**, 14204-14209.

Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T., and Ebizuka, Y. (2006). Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. FEBS J **273**, 948-959.

Sindambiwe, J.B., Calomme, M., Geerts, S., Pieters, L., Vlietinck, A.J., and Vanden Berghe, D.A. (1998). Evaluation of biological activities of triterpenoid saponins from *Maesa lanceolata*. J Nat Prod **61**, 585-590.

Sindambiwe, J.B., Balde, A.M., Debruyne, T., Pieters, L., VanDenHeuvel, H., Claeys, M., VanDenBerghe, D.A., and Vlietinck, A.J. (1996). Triterpenoid saponins from *Maesa lanceolata*. Phytochemistry **41**, 269-277.

Sparg, S.G., Light, M.E., and van Staden, J. (2004). Biological activities and distribution of plant saponins. J Ethnopharmacol **94**, 219-243.

Sun, H.X., Xie, Y., and Ye, Y.P. (2009a). Advances in saponin-based adjuvants. Vaccine **27**, 1787-1796.

Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P.T., and Dixon, R.A. (2002). A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. Plant J **32**, 1033-1048.

Suzuki, H., Reddy, M.S.S., Naoumkina, M., Aziz, N., May, G.D., Huhman, D.V., Sumner, L.W., Blount, J.W., Mendes, P., and Dixon, R.A. (2005). Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. Planta **220**, 696-707.

Tansakul, P., Shibuya, M., Kushiro, T., and Ebizuka, Y. (2006). Dammarenediol-II synthase, the first dedicated enzyme for ginsenoside biosynthesis, in *Panax ginseng*. FEBS Lett **580**, 5143-5149.

Theunis, M., Foubert, K., Pollier, J., González-Guzmán, M., Goossens, A., Vlietinck, A.J., Pieters, L.A.C., and Apers, S. (2007). Determination of saponins in *Maesa lanceolata* by LC-UV: Development and validation. Phytochemistry **68**, 2825-2830.

Vuylsteke, M., Peleman, J.D., and van Eijk, M.J.T. (2007). AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat Protoc **2**, 1399-1413.

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

Zhao, C.L., Cui, X.M., Chen, Y.P., and Liang, Q.A. (2010). Key enzymes of triterpenoid saponin biosynthesis and the induction of their activities and gene expressions in plants. Nat Prod Comm **5**, 1147-1158.

Chapter 3

An integrated PCR colony hybridization approach to screen cDNA libraries for full-length coding sequences

Jacob Pollier*, Miguel González-Guzmán, Wilson Ardiles-Diaz, and Alain Goossens

* Author contribution: design and optimization of the method, screening of the *M. lanceolata* cDNA-library, development and use of multiple probes and writing of the manuscript.

Publication status:

A manuscript was submitted to PLoS ONE (IF: 4.351).

Abstract

Background

The cDNA-Amplified Fragment Length Polymorphism (cDNA-AFLP) technique is a commonly used method for genome-wide expression analysis that does not require prior sequence knowledge. Typically with cDNA-AFLP, quantitative expression data and sequence information is obtained for a large number of gene tags that are differentially expressed. However, most of the gene tags do not correspond to full-length (FL) coding sequences, which is a prerequisite for subsequent functional analysis.

Results

A medium-throughput screening strategy, based on an approach that integrates PCR and colony hybridization, was developed that allows in parallel screening of a cDNA library for FL clones corresponding to incomplete cDNAs. The method was applied to screen for the FL open reading frames of a selection of 163 cDNA-AFLP tags from three different medicinal plants. This led to the identification of 109 (67%) FL clones. Furthermore, the protocol allows for the use of multiple probes in a single hybridization event, thus significantly increasing the throughput of the method when screening for rare transcripts.

Conclusions

The presented strategy offers an efficient method for the conversion of incomplete expressed sequence tags (ESTs), such as cDNA-AFLP tags, to FL coding sequences.

1. Background

cDNA-AFLP is a widely used, robust, and reproducible tool for genome-wide expression analysis in any organism, without the need for prior sequence knowledge (Breyne et al., 2003; Vuylsteke et al., 2007). The technique is derived from AFLP (Vos et al., 1995) and is based on the selective PCR amplification of restriction fragments from a double-stranded cDNA template. To this end, the double-stranded cDNA template is digested with restriction enzymes, followed by ligation of specific adapters to the sticky ends of the digested cDNA. Subsequently, a subset of the restriction fragments is amplified by PCR by using primers with a few selective nucleotides in addition to the sequence complementary to the adaptor and restriction site sequences. The amplified cDNA fragments are visualized on high-resolution polyacrylamide gels, on which the intensity of the fragments reflects the relative abundance (copy number) of the corresponding genes across the samples (Vuylsteke et al., 2006; Vuylsteke et al., 2007). To identify the differentially expressed genes, the corresponding cDNA-AFLP tags are purified from the polyacrylamide gel, reamplified, and sequenced. However, the resulting cDNA-AFLP tag sequences most often do not correspond to FL-coding sequences and, thus, do not provide sufficient sequence information for functional characterization.

Obtaining a FL-coding sequence of a gene of interest from a non-model species without genome information might often prove to be a laborious and difficult task. One of the commonly used methods to obtain a FL cDNA clone starting from a partial clone such as a cDNA-AFLP tag or any other type of EST, is the rapid amplification of cDNA ends (RACE)-PCR strategy, in which adaptors ligated to the 3' and 5' ends of the cDNA are used to selectively amplify these 3' or 5' cDNA fragments by PCR through a combination of gene-specific and adaptor-specific primers (Chenchik et al., 1996). A common problem associated with RACE-PCRs is the amplification of non-specific PCR products due to the presence of one of the primer sequences in all cDNAs (Bower and Johnston, 2010). Several methods have been developed to solve this problem, but most of them rely on extra enzymatic steps after completion of the first strand cDNA synthesis, which may introduce mistakes (Schramm et al., 2000). An alternative to the RACE-PCR strategy is the screening of cDNA libraries. To identify clones from a cDNA library that contain the gene of interest, labeled gene-specific DNA probes are hybridized to colonies or plaques transferred to a nylon membrane. This traditional manner of library screening is labour intensive, especially when screening for rare

transcripts (Wan et al., 2006). A more rapid approach to obtain a clone of interest from a cDNA library is by PCR screening of pooled clones. This approach requires arraying of individual clones into microtiter wells, and is therefore only practical for abundant transcripts (Munroe et al., 1995). Alternatively, Self-Ligation of Inverse PCR Products (SLIP) allows for the simultaneous screening for low abundant transcripts. In this technique, the plasmid of interest is amplified from a small aliquot of the cDNA library via inverse PCR, using primers designed in opposite orientation on the incomplete cDNA. The obtained PCR fragments are self-ligated, transformed into *E. coli* and plated. Subsequently, individual colonies are picked for DNA isolation and sequencing (Hoskins et al., 2005; Wan et al., 2006). However, similar to various RACE-PCR approaches, this PCR-based library screening strategy is susceptible to amplification artefacts. Hence, because of the time-consuming and labor-intensive iterative screening of a number of cDNA libraries or the need for directed strategies to process individual clones in a RACE-PCR approach, to obtain the FL cDNA clone of large numbers of incomplete cDNAs in parallel, may prove to be a real challenge.

To overcome this bottleneck, we developed a medium-throughput cDNA library screening strategy that integrates PCR and colony hybridization. The fast and easy PCR methodology was combined with the robust colony hybridization technique, allowing the screening of approximately 100,000 cDNA clones for individual cDNA clones, in a rapid and straightforward manner. The screening strategy can be divided into three basic steps. First, 100,000 cDNA colony-forming units or clones were divided in 12 pools of approximately 8,000 cDNA clones each. The pools (inoculated from the primary transformants) were grown overnight in liquid medium to increase the cell mass and each culture was then split into two halves. In the second step, plasmid DNA was extracted of one half of each of the overnight-grown cultures. The extracted pool plasmid DNA was used as PCR template to identify the pool containing the cDNA clone corresponding to the incomplete cDNA tag. In step three, the remaining half of the overnight-grown culture was plated out on Petri dishes for classical colony hybridization, using the incomplete cDNA tag as probe. The developed screening strategy was used to screen cDNA libraries of three medicinal plants, *Maesa lanceolata* (false assegai), *Glycyrrhiza glabra* (liquorice), and *Panax ginseng* (ginseng) and allowed to pick up over 100 FL clones in two screening rounds, with a success rate of 67%.

2. Results

2.1.Identification of MlJAZ1

In our research on secondary metabolism, we investigated the effect of methyl jasmonate (MeJA) elicitation on the transcriptome of medicinal plants by cDNA-AFLP transcript profiling. In one such study, the transcriptome of MeJA-elicited axenic shoot cultures of *M. lanceolata* was analyzed. A collection of MeJA-responsive gene tags was obtained, of which ML074 was strongly transcriptionally activated within 30 minutes after MeJA treatment (Figure 1A,B). The cDNA-AFLP tag was purified from the gel, reamplified as described (Rischer et al., 2006; Vuylsteke et al., 2007), and sequenced, revealing an EST of 449 nucleotides that was closely homologous to the JA signaling repressor JAZ1/TIFY10A from *Arabidopsis thaliana*. To obtain the FL cDNA corresponding to the ML074 tag, a *M. lanceolata* Uncut Nanoquantity cDNA library (Invitrogen, Carlsbad, CA, USA) was screened with our approach. The primers for the PCR screening (forward and reverse, 5'-TTTATTCCCCCAGCACTCTG-3' and 5'-TCGGAGCTTGCCTTACTAGC-3', respectively) (Figure 1D) were developed on the cDNA-AFLP tag with the Primer3 program (Rozen and Skaletsky, 2000). Subsequent screening of the pool plasmid DNA revealed that the clone was present in pools 1, 6, 7, 8, and 12 (Figure 1C). After colony hybridization on the membrane of pool 7, three candidate colonies could be observed on the resulting autoradiogram (Figure 1E). By means of colony PCR with the above-mentioned primers, two of the three candidate colonies were found to contain the clone of interest (Figure 1F). Sequencing of the two positive clones demonstrated that in both clones an identical FL open reading frame (ORF) of 819 nucleotides occurred, encoding a protein of 273 amino acids, hereafter referred to as MlJAZ1. The 5' untranslated region (UTR) present in clone 1 was 153 nucleotides long, whereas the 5' UTR of clone 2 was only 36 nucleotides long. The sequence of clone 1 was deposited in the GenBank with accession number JF313904. Analysis of the obtained MlJAZ1 sequence showed that the protein contained the characteristic tify and Jas domains of the JAZ protein family (Figure 1D,G). The tify domain and the C-terminal Jas domain are characterized by the highly conserved TIF[F/Y]XG and $SLX_2FX_2KRX_2RX_5PY$ amino acid sequences, respectively (Chini et al., 2007; Thines et al., 2007; Vanholme et al., 2007; Yan et al., 2007; Staswick, 2008; Grunewald et al., 2009; Pauwels et al., 2010).

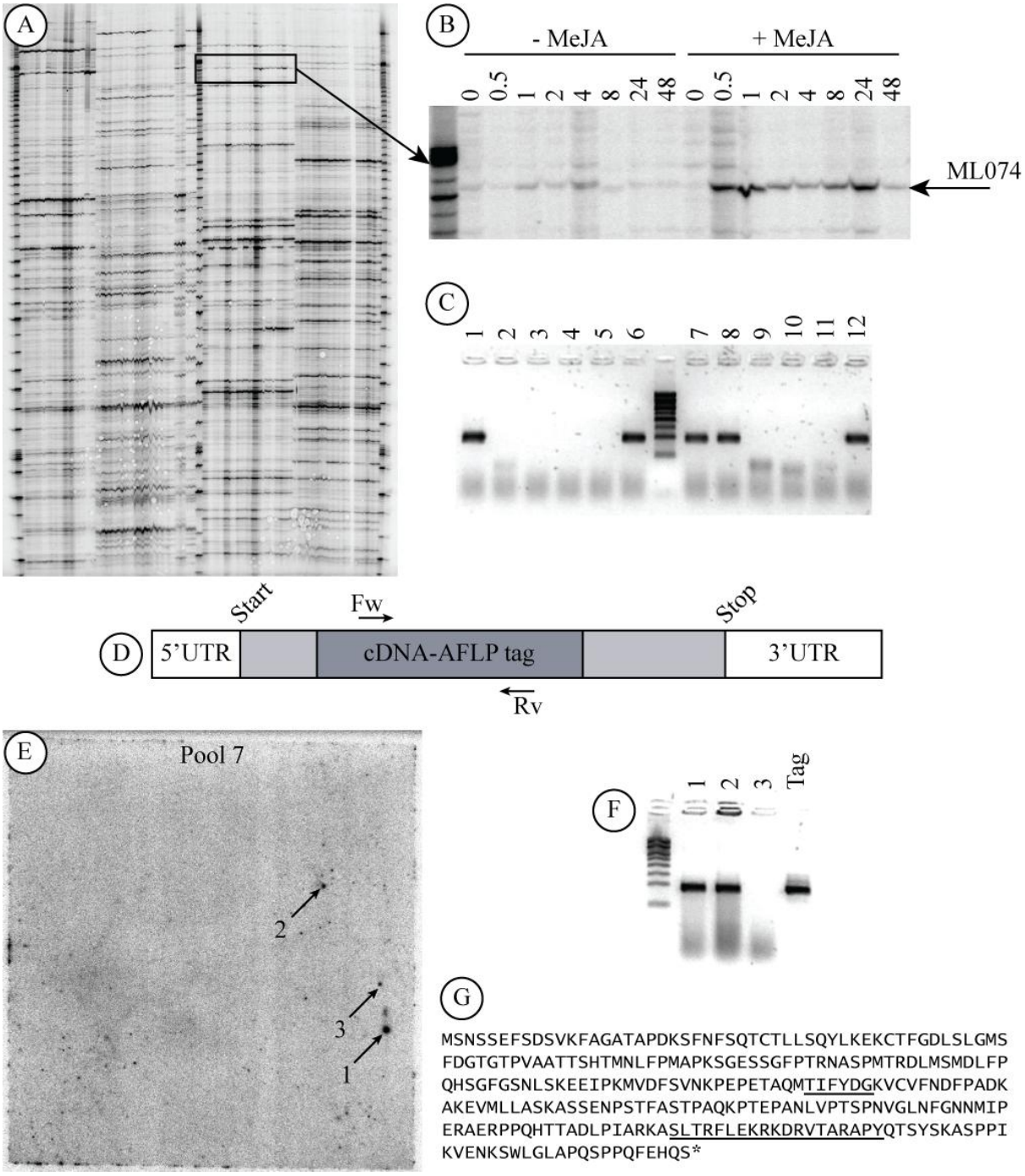


Figure 1. Identification of *MlJAZ1*. A. cDNA-AFLP gel. B. Detail of the cDNA-AFLP gel showing the MeJA response of the ML074 tag. C. Result of the PCR screening, showing the presence of ML074 in pools 1, 6, 7, 8, and 12. D. Schematic representation of *MlJAZ1*. E. Autoradiogram of pool 7 after hybridization with the radioactively labeled ML074 tag; arrows mark the three candidate colonies. F. Colony PCR on the candidate colonies. The cDNA-AFLP tag was used as template for the positive control. G. Identification of the tify and Jas domains in the MlJAZ1 amino acid sequence.

2.2.In parallel screening for *Maesa lanceolata* genes

Besides ML074, 52 other *M. lanceolata* cDNA-AFLP tags were selected for further analysis. PCR screening of the pool plasmid DNA indicated that clones corresponding to 52 of the 53 selected tags were present in at least one of the 12 library pools (Figure 2). Furthermore, clones corresponding to 23 tags were present in all the pools, suggesting a high representation in the library. Subsequently, a first round of colony hybridizations were performed, in which the 52 radioactively labeled cDNA-AFLP tags were hybridized on one membrane of a pool for which the presence of a clone corresponding to the cDNA-AFLP tag was confirmed by the PCR screening. For 45 tags, at least one candidate colony was identified, but colony PCR revealed that for 15 tags, all identified candidate colonies were false positives. Thus, for 30 cDNA-AFLP tags, confirmed colonies were obtained and subsequently sequenced, until a clone with a FL-coding sequence was identified, resulting in the identification of FL sequences for 19 of the 53 initially selected cDNA-AFLP tags (36%). For the remaining 33 cDNA-AFLP tags that occurred in one or more of the pools, and for which no corresponding FL-coding sequence was identified, a second round of hybridizations was performed (Figure 2), until for every cDNA-AFLP tag a candidate colony was identified, or until all the pools with a hit in the PCR screen were exhausted. For all, but one, of the 33 remaining tags, candidate colonies were identified. Subsequent colony PCR showed that for six tags, all identified colonies were again false positives and, thus, colony PCR confirmed clones were found for 45 of the 53 initially selected cDNA-AFLP tags. Sequencing of the candidate colonies allowed us to identify 21 more FL clones in the second hybridization round, ending with a total of 40 confirmed FL clones out of the 53 initially selected cDNA-AFLP tags (75%); colony PCR confirmed clones corresponding to the five remaining cDNA-AFLP tags all contained truncated versions of the corresponding ORF. A flow-chart of the screening for FL clones of the *M. lanceolata* cDNA-AFLP tags is given in Figure 2.

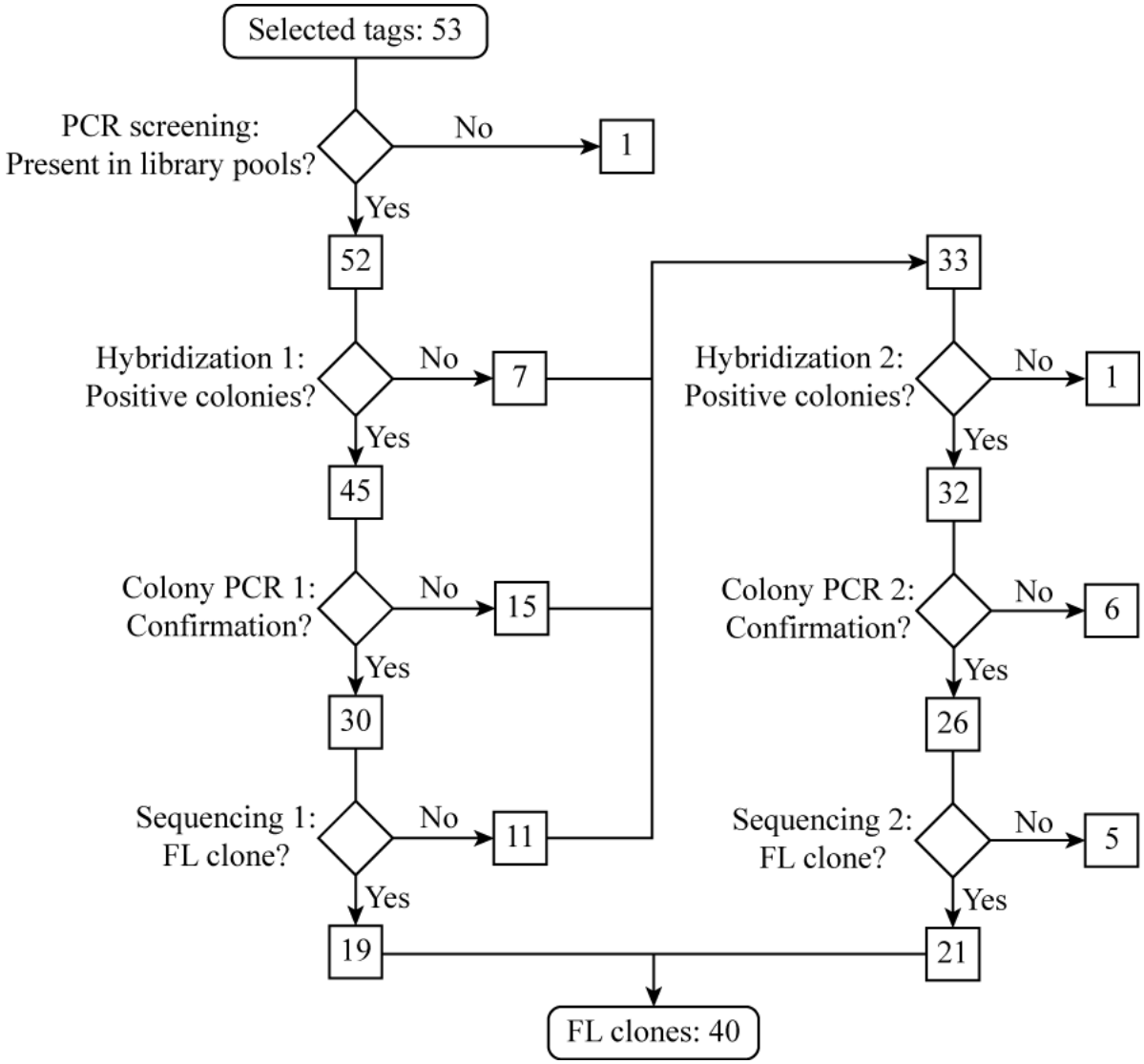


Figure 2. Flow-chart of the screening for FL clones corresponding to the selected *M. lanceolata* cDNA-AFLP tags. For details, see text.

2.3.*Panax ginseng* and *Glycyrrhiza glabra* cDNA library screening

Besides *M. lanceolata*, cDNA-AFLP tags were selected from two other medicinal plants treated with MeJA, *P. ginseng* and *G. glabra*. For *P. ginseng*, a total of 41 cDNA-AFLP tags were selected, of which the screening PCR confirmed the presence of 38 corresponding clones in the approximately 100,000 screened clones of the cDNA library. For *G. glabra*, a total of 69 tags were selected and the PCR screening confirmed the presence of 59 clones. FL genes of *G. glabra* and *P. ginseng* were screened for as for *M. lanceolata* and 47 and 22 FL genes were obtained out of the 69 (68%) and the 42 (54%) initially selected tags, respectively. In conclusion, by combining the three screenings, we identified a FL clone corresponding to 109 cDNA-AFLP tags out of a selection of 163 tags, reaching an overall

Table 1. Overview of the cloning efficiency.

Species	Enzymes	Transcription factors	Total
M. lanceolata	32/37 (86%)	8/16 (50%)	40/53 (75%)
G. glabra	36/49 (73%)	11/20 (55%)	47/69 (68%)
P. ginseng	22/41 (54%)	Not selected	22/41 (54%)

success rate of 67%. The overview (Table 1) illustrates that the success rate is higher for enzyme-encoding genes than for transcription factors, possibly because the latter are rarer represented in the library than the former.

3. Discussion

The designed "EST-to-FL conversion" strategy described here has proven efficient in the screening for FL clones corresponding to a large number of cDNA-AFLP tags. During the screening process, two hybridization rounds could be done per week, each including 12 membranes, for which the following schedule was applied. On day 1, probes were labeled, and hybridizations were performed overnight. On day 2, the membranes were washed thrice to remove excess probe and exposed to the autoradiography films for 2 days. On day 4, the autoradiography films were developed and the membranes were stripped prior to a second round of overnight hybridizations. On day 5, the membranes were washed and exposed to the autoradiography films for 2 days. On day 8, the second batch of films were developed and the membranes stripped, and another round of hybridizations could be started. This system allowed us to use the membranes for up to six hybridization rounds without significant loss of hybridization potential.

Furthermore, with this hybridization system, multiple probes can be utilized in a single hybridization event. Therefore, 50 ng of each probe was mixed, prior to labeling, in a double volume of the probe labeling reaction mixture and so, up to six probes on a single membrane could be hybridized at once. Although the hybridization throughput is highly increased and, thus, less handling with radiolabeled substrates is necessary, the subsequent colony PCR step becomes more elaborate as a consequence, because every identified candidate colony has to be checked with all of the primer combinations corresponding to all of the probes used in the hybridization. Nonetheless, despite this drawback, the use of multiple probes proved to be beneficial in the second round of hybridizations (Figure 2), when transcripts without positive colonies in the first hybridization round were screened.

After hybridizations, candidate colonies were confirmed by colony-PCR and subsequently sequenced. For instance, for the 53 selected *M. lanceolata* tags, a total of 167 plasmids were sent for sequencing. Subsequent sequence analysis showed that 73/167 (43.7%) plasmids corresponded to FL clones, and at least one FL clone was obtained for 40 of the 53 selected tags. Although the applied strategy is not specific for FL clones, it allows a medium-throughput isolation of candidate clones. Hence, the better the quality of the screened cDNA library, the higher the chance of encountering a FL clone for a particular tag. To increase the chance of picking a FL clone from the candidate colonies, colony-PCR could be performed using a forward primer corresponding to the vector, at the 5' end of the cDNA, and a reverse primer corresponding to the tag or EST, and subsequently, the clone giving the largest PCR fragment could be sequenced. Alternatively, RACE-PCR-based approaches may have a higher chance of finding a FL clone for a particular tag, however, these methods are susceptible to amplification artefacts and are definitely more labour intensive and thus not applicable to obtain the FL clone of large numbers of incomplete cDNAs in parallel. Similar to our strategy, SLIP allows in parallel screening of a cDNA library for FL clones (Hoskins et al., 2005; Wan et al., 2006). Eliminating polyacrylamide gels, radioactivity and colony hybridization steps, the SLIP strategy allows the recovery of FL clones from plasmid cDNA libraries in 5 days (Wan et al., 2006). However, SLIP is basically an enzymatic method and is therefore sensitive to PCR artefacts, and tends to have a bias towards recovery of shorter, incomplete clones (Hoskins et al., 2005), disadvantages that are all omitted by colony hybridization strategies.

Here, the usefulness of our "EST-to-FL conversion" strategy has been shown for one class of incomplete cDNAs, i.e. cDNA-AFLP tags, but, principally, it can be applied to any type of incomplete tag sequence. Indeed, several commonly used techniques generate often very large collections of incomplete cDNAs or ESTs. For instance, in the latest GenBank database of EST release (dbEST release 030111, http://www.ncbi.nlm.nih.gov/dbEST/index.html), a total of 69 million ESTs were released. Similarly to cDNA-AFLP tags, these ESTs are typically between 200 and 500 nucleotides long. Contigs can be generated from large collections of ESTs from a single organism and allow the identification of the FL ORFs of the genes of interest. However, for most of the organisms, the EST collection is too small to generate such contigs. The presented strategy might be applied to pick up the FL ORFs corresponding to the published ESTs. Primers designed on the released EST sequences can be used for the PCR screening and the resulting PCR product can be purified and labeled to use

it as a probe to pick up the full clone via the colony hybridization step of the presented strategy.

4. Conclusions

A screening strategy, based on an integrated PCR colony hybridization approach was developed that allows the simultaneous screening for FL clones corresponding to a large number of incomplete ESTs obtained by cDNA-AFLP transcript profiling. The designed method was used to screen for FL clones of an initial selection of 163 cDNA-AFLP tags, leading to the identification of 109 FL genes (67%) from three different medicinal plants. This strategy will be applicable for the conversion of any type of EST to FL clones for functional analysis.

5. Methods

5.1.Pool preparation

As starting material, Uncut Nanoquantity cDNA libraries (custom-made by Invitrogen, Carlsbad, CA, USA) were used. Approximately 120,000 clones of the original library were transferred to a 2-mL tube and Lysogeny Broth (LB) medium containing 50 µg/mL kanamycin (LBKAN50) was added to a total volume of 1.5 mL. This library dilution was used to inoculate 12 flasks, each containing 10 mL of LBKAN50 medium with approximately 8,000 cDNA clones each (100 µL of the prepared library dilution per flask). The resulting Escherichia coli cultures were incubated for 12-16 hours at 37°C with shaking at 300 rpm.

5.2.PCR screening

For the plasmid DNA extraction, 5 mL of each 10-mL culture was used. To identify the pools that contained the cDNA clone corresponding to the target cDNA-AFLP tag, 50 ng of the pool plasmid DNA was taken as PCR template. Specific PCR primers were designed based on the cDNA-AFLP tag sequences with the Primer3 program (Rozen and Skaletsky, 2000).

5.3.Pool plating

The titer of the remaining 5 mL of each 10-mL culture was estimated by plating a serial dilution of the cultures on LBKAN50 plates, followed by overnight incubation at 37°C.

Subsequently, of each pool, approximately 10,000 cDNA clones were plated on LBKAN50 medium containing 15 g/L agar in 24 x 24 cm Petri dishes and grown overnight at 37°C. Prior to colony lifting, the plates were pre-cooled at 4°C.

5.4.Colony lifting

The colonies grown on the 24 x 24 cm Petri dishes were lifted on Hybond-N^+ membranes (GE-Healthcare, Little Chalfont, UK), according to the manufacturer's instructions. The DNA was cross-linked to the membranes by UV irradiation.

5.5.Probe labeling

Probes were re-amplified by PCR from the cDNA-AFLP tag (Vuylsteke et al., 2007) and subsequently labeled with ^{32}P via a random priming reaction by incubating 25 µL of the following mixture at 37°C for 6 hours: 50 ng denatured probe, 1x bovine serum albumin (0.1 mg/mL), 5 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT), 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 20 µM of dATP, dGTP, and dTTP each, 12.5 µCi deoxycytidine α-^{32}P (500 µCi/mmol dCTP Easy Tide) (Perkin Elmer, Waltham, MA, USA), 1 ng hexanucleotide random primer (Roche Applied Science, Brussels, Belgium), and 5 U Klenow Fragment of DNA Polymerase I (GE-Healthcare). The labeled probes were stored at -20°C until used for hybridization.

5.6.Hybridization

The membranes were pre-wetted for 30 min at 65°C in preheated hybridization buffer (0.4 M Na_2HPO_4, pH 7.2, 1 mM ethylenediaminetetraacetic acid (EDTA), 7% (w/v) sodium dodecyl sulfate (SDS)), after which the denatured radioactive probe was added. After overnight hybridization at 65°C, the membranes were washed thrice with preheated washing buffer (0.04 M Na_2HPO_4, pH 7.2, 1 mM EDTA, and 5% (w/v) SDS) at 65°C. Subsequently, excess buffer was drained off. The membranes were packed in SaranWrap and exposed to Kodak™ BioMax films (Perkin Elmer) for 2 days. The membranes can be re-used for hybridization after stripping by washing them thrice with preheated stripping solution (1 mM EDTA and 0.1% (w/v) SDS) at 75°C.

5.7.Clone confirmation

Candidate colonies were picked from the 24 x 24 cm Petri dishes and checked by colony PCR with the primers designed on the cDNA-AFLP tag. Positive clones were sequenced with the primers designed on the pENTR222 vector of the Uncut Nanoquantity cDNA library, 5'-ACGACGGCCAGTCTTAAGCTCGG-3' and 5'-ACCATGTAATACGACTCACTATAGG-3'.

Acknowledgements

The authors thank Martine De Cock for help in preparing the manuscript. This work was supported by the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" project SBO040093).

References

Bower, N.I., and Johnston, I.A. (2010). Targeted rapid amplification of cDNA ends (T-RACE)-an improved RACE reaction through degradation of non-target sequences. Nucleic Acids Res **38**, e194.

Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inzé, D., and Zabeau, M. (2003). Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol Genet Genomics **269**, 173-179.

Chenchik, A., Diachenko, L., Moqadam, F., Tarabykin, V., Lukyanov, S., and Siebert, P.D. (1996). Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. Biotechniques **21**, 526-534.

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J.M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F.M., Ponce, M.R., Micol, J.L., and Solano, R. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. Nature **448**, 666-671.

Grunewald, W., Vanholme, B., Pauwels, L., Plovie, E., Inzé, D., Gheysen, G., and Goossens, A. (2009). Expression of the *Arabidopsis* jasmonate signalling repressor *JAZ1/TIFY10A* is stimulated by auxin. EMBO Rep **10**, 923-928.

Hoskins, R.A., Stapleton, M., George, R.A., Yu, C., Wan, K.H., Carlson, J.W., and Celniker, S.E. (2005). Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP). Nucleic Acids Res **33**, e185.

Munroe, D.J., Loebbert, R., Bric, E., Whitton, T., Prawitt, D., Vu, D., Buckler, A., Winterpacht, A., Zabel, B., and Housman, D.E. (1995). Systematic screening of an arrayed cDNA library by PCR. Proc Natl Acad Sci U S A **92**, 2209-2213.

Pauwels, L., Fernández Barbero, G., Geerinck, J., Tilleman, S., Grunewald, W., Cuéllar Pérez, A., Chico, J.M., Vanden Bossche, R., Sewell, J., Gil, E., García-Casado, G., Witters, E., Inzé, D., Long, J.A., De Jaeger, G., Solano, R., and Goossens, A. (2010). NINJA connects the co-repressor TOPLESS to jasmonate signalling. Nature **464**, 788-791.

Rischer, H., Orešič, M., Seppänen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M.C.E., Inzé, D., Oksman-Caldentey, K.M., and Goossens, A. (2006). Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc Natl Acad Sci U S A **103**, 5614-5619.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132**, 365-386.

Schramm, G., Bruchhaus, I., and Roeder, T. (2000). A simple and reliable 5'RACE approach. Nucleic Acids Res **28**, e96.

Staswick, P.E. (2008). JAZing up jasmonate signaling. Trends Plant Sci **13**, 66-71.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G.H., Nomura, K., He, S.Y., Howe, G.A., and Browse, J. (2007). JAZ repressor proteins are targets of the SCF^{COI1} complex during jasmonate signalling. Nature **448**, 661-665.

Vanholme, B., Grunewald, W., Bateman, A., Kohchi, T., and Gheysen, G. (2007). The tify family previously known as ZIM. Trends Plant Sci **12**, 239-244.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995). AFLP - a new technique for DNA fingerprinting. Nucleic Acids Res **23**, 4407-4414.

Vuylsteke, M., Peleman, J.D., and van Eijk, M.J.T. (2007). AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat Protoc **2**, 1399-1413.

Vuylsteke, M., Van Den Daele, H., Vercauteren, A., Zabeau, M., and Kuiper, M. (2006). Genetic dissection of transcriptional regulation by cDNA-AFLP. Plant J **45**, 439-446.

Wan, K.H., Yu, C., George, R.A., Carlson, J.W., Hoskins, R.A., Svirskas, R., Stapleton, M., and Celniker, S.E. (2006). High-throughput plasmid cDNA library screening. Nat Protoc **1**, 624-632.

Yan, Y.X., Stolz, S., Chételat, A., Reymond, P., Pagni, M., Dubugnon, L., and Farmer, E.E. (2007). A downstream mediator in the growth repression limb of the jasmonate pathway. Plant Cell **19**, 2470-2483.

Chapter 4

Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry.

Jacob Pollier*, Kris Morreel, and Alain Goossens

* Author contribution: Generation and cultivation of the hairy roots, metabolite extractions, LC ESI FT-ICR MS profiling, analysis of the data, and writing of the manuscript.

Published in the Journal of Natural Products (IF: 3.159):

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. J Nat Prod, **in press**, doi: 10.1021/np200218r.

Abstract

Triterpenes are one of the largest classes of plant natural products, with an enormous variety in structure and bioactivities. Here, triterpene saponins from hairy roots of the model legume *Medicago truncatula* were profiled with reversed-phase liquid chromatography coupled to negative-ion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (LC ESI FT-ICR MS). Owing to the accuracy of the FT-ICR MS, reliable molecular formulas of the detected compounds could be predicted, which, together with the generated MS^n spectra, allowed the tentative identification of 79 different saponins, of which 61 had not been detected previously in *M. truncatula*. Upon collision-induced dissociation of saponins that contain a uronic acid residue in the sugar chain, fragment ions resulting from cross-ring cleavages of the uronic acid residues were observed. The identified saponins are glycosides of 10 different sapogenins, of which three were not detected before in *M. truncatula*. Zanhic acid glycosides, which are prevalent in the aerial parts of *M. truncatula*, were absent in the hairy root extracts. This metabolite compendium will facilitate future functional genomic studies of triterpene saponin biosynthesis in *M. truncatula*.

1. Introduction

Legumes (Fabaceae) are important agricultural crops, characterized by the presence of root nodules that contain symbiotic nitrogen-fixing bacteria. Fixation of atmospheric nitrogen allows reduced fertilizer costs and makes legumes a key player in crop rotation to replenish nitrogen-depleted soil. Furthermore, legumes serve as a major source of proteins and oil for human and animal nutrition. Leguminous crops with a significant economical value include oilseed crops such as soybean (*Glycine max*) and peanut (*Arachis hypogaea*), forage and soil-conditioning crops such as alfalfa (*Medicago sativa*) and clover (*Trifolium* spp.), and pulses such as beans (*Phaseolus* spp.), peas (*Pisum* spp.), and lentils (*Lens* spp.).

Themore than 19,400 leguminous species, distributed over 730 genera, produce a vast array of metabolites with diverse structures and biological activities. An important group of legume natural products are triterpene saponins, a class of secondary metabolites that display a wide range of biological activities (Dixon and Sumner, 2003). In plants they serve as allelopathic, antipalatability, anti-insect, or antifungal agents (Huhman and Sumner, 2002). Saponins also have a broad range of pharmaceutical properties, such as anti-inflammatory, antimicrobial, anticancer, or adjuvant activities (Sparg et al., 2004; Sun et al., 2009). Besides saponins, isoflavonoids, alkaloids, and nonprotein amino acids commonly occur as natural products in several legume species (Dixon and Sumner, 2003; Dixon and Pasinetti, 2010).

As a model species for legume biology, barrel medic (*Medicago truncatula*), a close relative of alfalfa, is an excellent system to study secondary metabolism of legumes (Dixon and Sumner, 2003; Farag et al., 2009). Consequently, several studies about the metabolite composition of *M. truncatula* have been published over the past decade. Metabolite profiling using LC-MS-based methods revealed the presence of a complex mixture of triterpene saponins in various *M. truncatula* tissues (Huhman and Sumner, 2002; Huhman et al., 2005; Kapusta et al., 2005a; Kapusta et al., 2005b). In accordance with other legumes from the Papilionoideae subfamily, *M. truncatula* is rich in isoflavonoids (Farag et al., 2007; Schliemann et al., 2008).

Hairy roots are generated by an infection of healthy plant tissue, not necessarily roots, with *Agrobacterium rhizogenes*. They are characterized by fast growth rates and do not require growth regulators for their cultivation. Furthermore, they have a high genetic stability and a high biosynthetic potential for a longer time than natural roots (Guillon et al., 2006a;

Georgiev et al., 2007). Their potential to produce compounds that are also naturally occurring in wild-type roots in combination with their tolerance to cultivation in large-scale bioreactors makes hairy roots an appealing alternative for the destructive harvesting of wild-type roots, especially those of endangered plant species, to obtain valuable phytochemicals (Sevón and Oksman-Caldentey, 2002; Guillon et al., 2006b). Hairy roots are also a valuable tool to investigate the secondary metabolism in *M. truncatula*. A prerequisite for the study of secondary metabolism and its application in metabolic engineering projects is the possibility to generate transgenic plant tissue. Although fertile transgenic *M. truncatula* plants can be obtained via *Agrobacterium tumefaciens* mediated transformation (Chabaud et al., 1996; Cook, 1999; Trieu et al., 2000; Araújo et al., 2004), the published methods are time-consuming and lack the required throughput for simultaneous analysis of a significant number of genes. In contrast, *A. rhizogenes*-mediated transformation of *M. truncatula* seedlings allows a fast and efficient production of transgenic hairy roots (Boisson-Dernier et al., 2001); the generation time needed from seed to a transgenic *M. truncatula* hairy root culture, ready for profiling or other analysis, is only 3 months.

For any metabolic engineering project targeting triterpene saponin biosynthesis, a detailed analysis of the saponin composition of the generated hairy roots is required. In the past decade, a few metabolic profiling studies on saponins in *M. truncatula* have been reported. In a first study, the presence of a large number of saponins in root extracts of *M. truncatula* was shown: based on fragmentation data under negative ionization, 27 *M. truncatula* saponins were tentatively identified (Huhman and Sumner, 2002). Optimization of the latter method resulted in an extended list of 31 tentatively identified saponins (Huhman et al., 2005). In a later study, 12 additional saponins from the aerial parts of *M. truncatula* were characterized via spectroscopic methods, and their fragmentation behavior under negative ionization reported (Kapusta et al., 2005a; Kapusta et al., 2005b). Nevertheless, *M. truncatula* was postulated to synthesize many more triterpene saponins than identified to date (Huhman and Sumner, 2002).

For identification purposes, LC-MS-based metabolomics studies benefit from the use of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). FT-ICR MS provides highly accurate mass measurements, allowing a reliable prediction of the molecular formula of the detected ions. Here, we applied reversed-phase liquid chromatography (LC) coupled to negative-ion electrospray ionization (ESI) FT-ICR MS to investigate the saponin

composition of hairy roots of *M. truncatula*. As a result, 79 saponins were detected and tentatively identified, of which soyasaponin I and 3-Glc-28-Glc-medicagenic acid were the most abundant compounds. The identified saponins are glycosides of 10 different sapogenins, of which three had hitherto not been detected before in *M. truncatula*.

2. Results and discussion

2.1.General methodology

For metabolite profiling of triterpene saponins, the compounds from a methanol extract of *M. truncatula* hairy roots were separated by reversed-phase LC, followed by negative-ion ESI FT-ICR MS. The accurate *m*/*z* measurements of the FT-ICR MS, combined with the isotope abundances, allowed us to predict the molecular formula of the detected saponins. The structural information needed for the identification of the saponins was obtained by MS^n fragmentation using the linear ion trap (IT). Full FT-MS spectra were interchanged with dependent IT-MS^n scan events, existing of an MS^2 and two MS^3 scans. In the MS^2 scan, the most abundant ion in the previous full MS scan was fragmented. The two most abundant daughter ions in every MS^2 scan event were subjected to a dependent MS^3 scan event. In addition, in-source fragmentation data were included in the identification process.

2.2.Identification of saponins

Saponins are amphipathic glycosides consisting of a lipophilic aglycone or sapogenin covalently linked to one (monodesmosidic) or more (di- or tridesmosidic) sugar chains via an ether or ester glycosidic bond. The sugar residues are mainly hexoses (e.g., glucose and galactose), 6-deoxyhexoses (e.g., rhamnose and fucose), pentoses (e.g., arabinose and xylose), and uronic acids (e.g., glucuronic acid and galacturonic acid). As reported before, upon collision-induced dissociation (CID) under negative ESI-MS, saponins undergo glycosidic cleavages, retaining the charge at the reducing terminus (i.e., the fragment that contains the aglycone) (Domon and Costello, 1988; Liu et al., 2004). The detected (grand)daughter ions in the MS^n spectra provide structural information about the sugar residues and aglycone of the fragmented saponin.

Figure 1 illustrates the identification of a hitherto unknown saponin. As indicated in Figure 1A, the unknown saponin showed an $[M-H]^-$ anion at *m*/*z* 945.50965. Because saponins consist of the elements C, H, and O, three different molecular formulas are possible

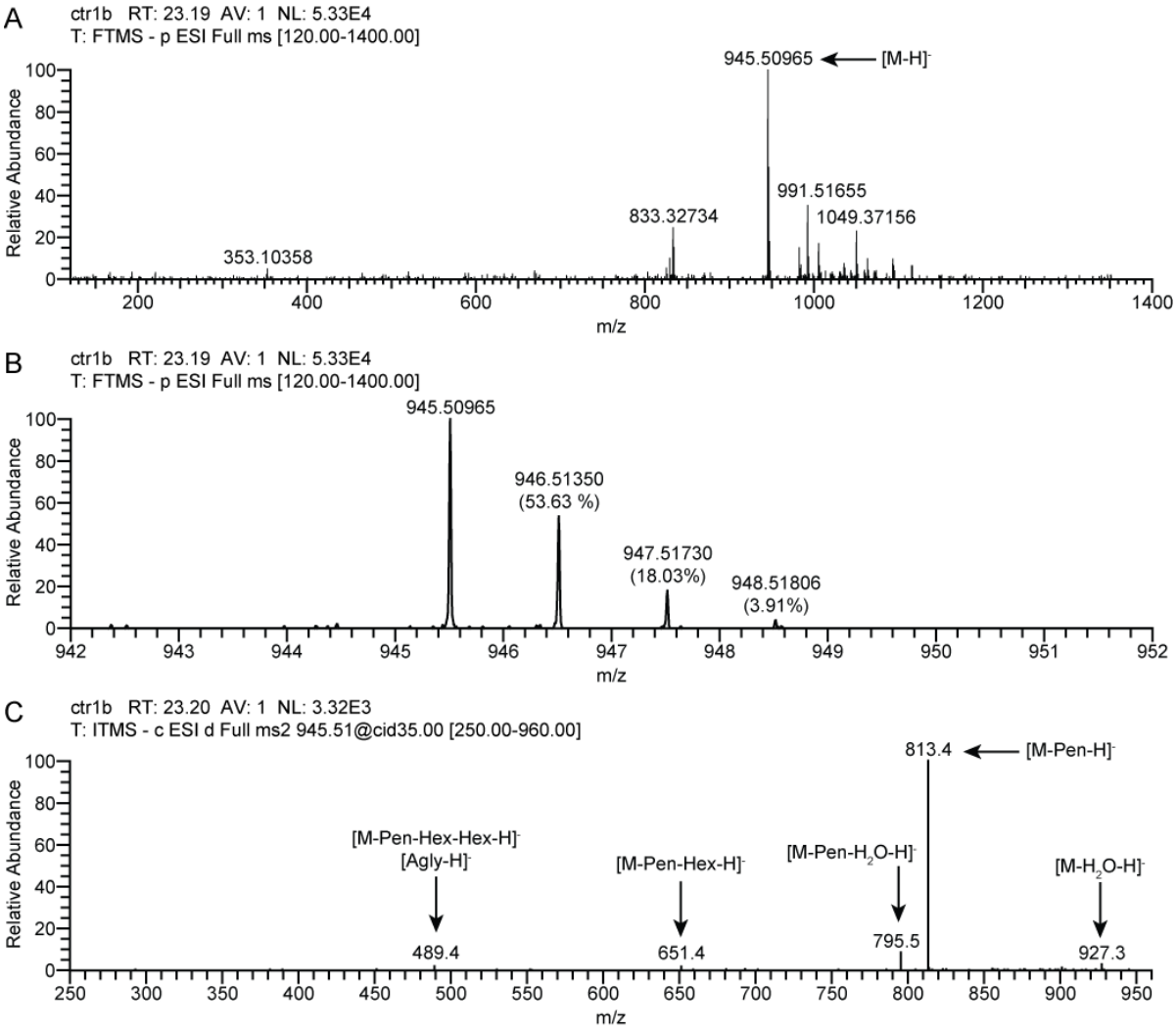


Figure 1. LC ESI FT-ICR MS chromatograms. (A) MS scan of peak at t_R 23.19 min. (B) Zoom of the MS scan of panel A, showing that the quasi-molecular ion is accompanied by additional isotope ions. The relative intensity of the isotope ions is indicated between brackets. (C) MS^2 fragmentation of the $[M-H]^-$ ion at *m/z* 945.51 (Pen-Hex-Hex-aglycone D).

within an error range of 5 ppm: $C_{65}H_{69}O_6$ (δ ppm = -0.331), $C_{40}H_{81}O_{24}$ (δ ppm = -2.830), and $C_{47}H_{77}O_{19}$ (δ ppm = 3.380). To determine the correct molecular formula, the abundance of the isotopes was considered. In nature, stable elemental isotopes occur in a stable ratio. For instance, ^{13}C occurs at approximately 1.11% of the most frequent carbon isotope, ^{12}C. Similarly, ^{2}H and ^{17}O occur at 0.015 and 0.038% of ^{1}H and ^{16}O, respectively. Hence, the observed monoisotopic quasi-molecular ions (M_0) are always accompanied by additional isotope ions (M+1,M+2, ...) (Figure 1B), of which the relative abundance depends on the elemental composition of the compound. The predicted relative abundance of the M + 1 ions of the three possible molecular formulas was 73.41, 46.53, and 54.05%, respectively. For the unknown saponin, the observed relative abundance of the M + 1 ion was 53.63% (Figure 1B),

indicating that the molecular formula corresponding to the $[M-H]^-$ anion is $C_{47}H_{77}O_{19}$ (δ ppm = 3.380). The sugar residues and the aglycone could be identified from the MS^n spectra. For this saponin, MS^2 fragmentation led to the generation of five daughter ions, resulting from the loss of water (-18 Da) and/or one or more hexose (-162 Da) and/or pentose (-132 Da) residues. The smallest daughter ion, at *m*/*z* 489, represents the aglycone ion, $[Agly-H]^-$, resulting from the loss of a pentose and two hexose moieties (Figure 1C). Based on this rationale, tens of other saponins present in the extract of *M. truncatula* hairy roots could tentatively be identified. All identified saponins and their MS^n data are listed in Table 1.

2.3.Cross-ring cleavage of uronic acid residues.

Saponins are reported to undergo only glycosidic cleavages upon CID under negative ESI MS (Liu et al., 2004). However, in this study, we observed a recurrent fragmentation pattern for saponins that contain a uronic acid residue, indicating a cross-ring cleavage of the uronic acid residue upon CID, as illustrated with the MS^2 spectrum observed for 3-Rha-Gal-GlcA-Soyasapogenol B (Figure 2A,B). The anions with the highest relative intensity, observed at *m*/*z* 923 and *m*/*z* 879, correlated to the neutral losses of 18 (H_2O) and 62 Da (H_2O + CO_2), respectively.

A similar pattern was observed after the loss of the rhamnose residue (*m*/*z* 795) and the ions observed at *m*/*z* 751 and 733 correlated to an additional loss of 44 (CO_2) and 62 Da (H_2O + CO_2), respectively. Furthermore, three more characteristic peaks were observed at *m*/*z* 615, 597, and 525, corresponding to a loss of the galactose residue, and an additional loss of 18 (H_2O), 36 (H_2O + H_2O), and 108 Da. The latter loss can only be explained by a cross-ring cleavage of the uronic acid residue. Cross-ring cleavages of glycosides are charge-remote fragmentations, i.e. fragmentations that are not initiated by the charge position. Because of the low internal energy of negative ions, cross-ring cleavages of glycosides do not readily occur (Liu et al., 2004). An exception are those that are linked via an ester bond as in, for example, feruloylated oligosaccharides (Levigne et al., 2004; Quéméner and Ralet, 2004) or sinapoyl glucose (Vanholme et al., 2010). Sinapoyl glucose serves as an UV protectant in *Arabidopsis*, and upon CID in the negative ionization mode, losses of 60, 90 and 120 Da are observed (Vanholme et al., 2010), which correspond with a $^{0,4}X$, $^{0,3}X$, and $^{0,2}X$ cross-ring cleavage, respectively (see legend of Figure 2 for explanation on nomenclature). However, in this study, the cross-ring cleavage of the uronic acid residue should be charge-driven, because the carboxylic group of the uronic acid bears the negative charge. The loss of 108 Da might

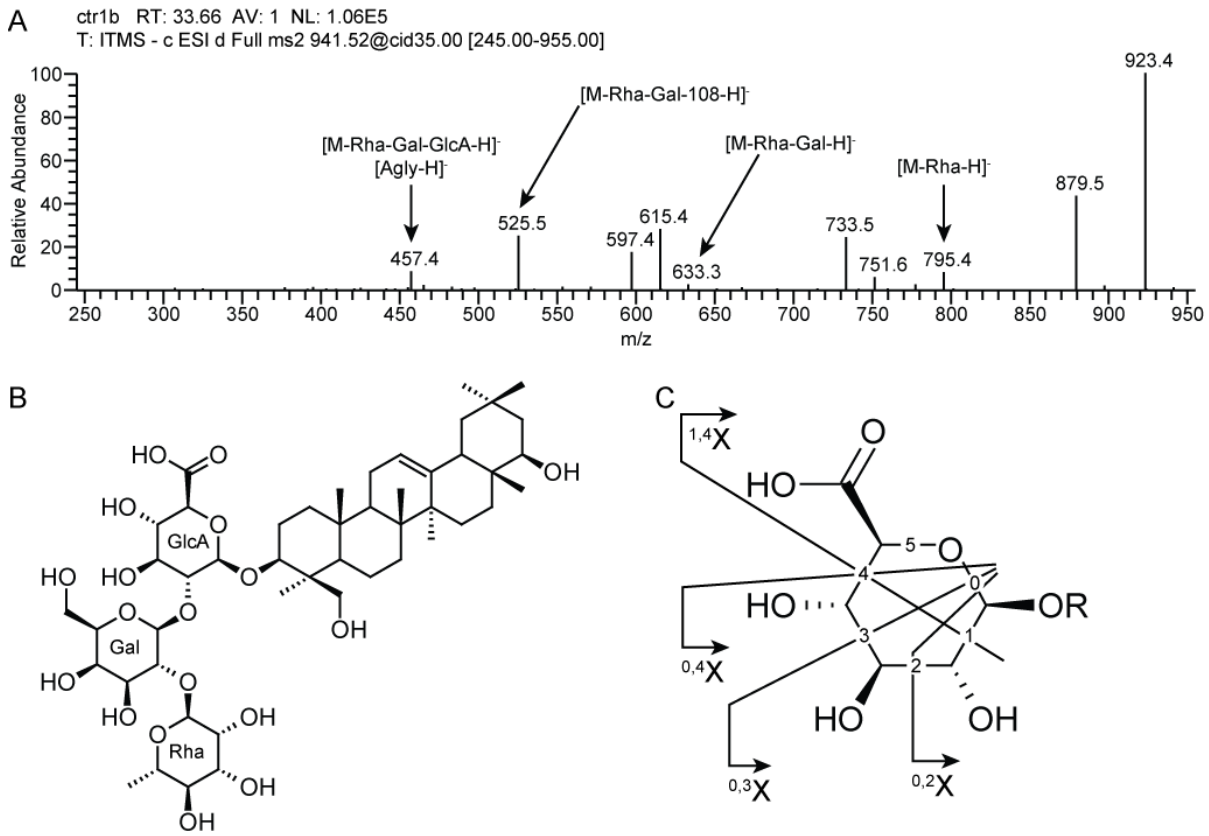


Figure 2. Cross-ring cleavage of uronic acid residues. (A) MS^2 fragmentation of 3-Rha-Gal-GlcA-soyasapogenol B. (B) Structure of 3-Rha-Gal-GlcA-soyasapogenol B. (C) Nomenclature of the cross-ring fragmentation of glucuronic acid.

be due to the loss of water followed by an additional loss of a $C_3H_6O_3$ fragment of 90 Da or by the loss of a $C_3H_8O_4$ fragment involving a $^{1,4}X$ cross-ring cleavage of the uronic acid residue (Figure 2C). This typical fragmentation pattern was observed in 20 compounds, and aided in the tentative identification of hitherto unknown saponins. By means of this methodology, a total of 79 saponins present in the *M. truncatula* hairy root extracts were identified (Table 1). However, MS^n does not provide enough information for absolute structural characterization of the metabolites; hence, the identifications remain tentative.

2.4.Most abundant compounds in *M. truncatula* hairy roots

In the metabolite extracts of *M. truncatula* hairy roots, the major peak, at a retention time (t_R) of 33.63 min, yielded an $[M-H]^-$ anion at *m/z* 941.51540 ($C_{48}H_{78}O_{18}$, δ ppm = 4.101) (Figure 3A). In *M. truncatula*, two saponins have been described with the same molecular formula, Rha-Hex-Hex-hederagenin and soyasaponin I (3-Rha-Gal-GlcA-soyasapogenol B) (Huhman et al., 2005). The MS^2 spectrum (Figure 2A and Table 1) corresponded to previously observed MS/MS data of soyasaponin I (Jin et al., 2007) , allowing us to conclude

that soyasaponin I is the most abundant saponin in the *M. truncatula* hairy roots extract. This is consistent with the previous observation that the major saponins in the roots of *M. truncatula* are soyasapogenol conjugates (Huhman et al., 2005). The second most abundant compound, eluting at 25.80 min, yielded an $[M-H]^-$ anion at *m*/*z* 825.42987 ($C_{42}H_{66}O_{16}$, δ ppm = 2.494) (Figure 3B). In the MS^2 spectrum (Figure 3C), the base peak occurs at *m*/*z* 439, a fragment ion resulting from the decarboxylation and dehydration of the medicagenic acid aglycone ion (Huhman and Sumner, 2002). Additional anions were observed at *m*/*z* 663, *m*/*z* 619, and *m*/*z* 601, correlating to the sequential loss of a hexose (162 Da), a CO_2 (44 Da) and a H_2O (18 Da) molecule from the parent ion. This fragmentation pattern corresponded to the previously reported MS data of 3-Glc-28-Glc-medicagenic acid and 3-Glc-Glc-medicagenic acid (Huhman and Sumner, 2002). To annotate the detected compound, we scanned for other peaks containing an anion with an *m*/*z* value between 825.40 and 825.45. Four additional isomers were detected, but only one (t_R 32.15 min) yielded a fragment ion at *m*/*z* 439 associated with medicagenic acid. As 3-Glc-28-Glc-medicagenic acid has been reported to have a lower retention time than 3-Glc-Glc-medicagenic acid (Huhman and Sumner, 2002), the peak at t_R 25.80 min has been annotated as 3-Glc-28-Glc-medicagenic acid, and the peak at t_R 32.15 min as 3-Glc-Glc-medicagenic acid.

2.5.Absence of zanhic acid glycosides

A remarkable observation was the absence of zanhic acid glycosides in the list of identified saponins. In previous studies, zanhic acid was shown to be one of the major aglycone moieties in aerial parts of *M. truncatula* (Kapusta et al., 2005a; Kapusta et al., 2005b), and it was also reported to be present in other *M. truncatula* plant tissues, including roots (Huhman et al., 2005). In a more recent study, however, zanhic acid could not be detected in *M. truncatula* roots (Confalonieri et al., 2009). Similarly, no zanhic acid-containing compounds were detected in the hairy roots. Hairy root cultures are known to accumulate secondary metabolites that are normally produced in "natural" roots of differentiated plants (Guillon et al., 2006a; Guillon et al., 2006b; Hu and Du, 2006). Hence, the absence of zanhic acid residues in hairy roots may indicate that this compound is produced only in the aerial parts of the plant. Consequently, the occurrence of the enzyme responsible for the 16α-hydroxylation of the saponin backbone (Figure 4) might be restricted to the aerial parts of the plant. This finding might further support a tissue-specific role for the saponins *in planta*. Indeed, the saponin mixture from the aerial parts of the plant has a protective effect against herbivores,

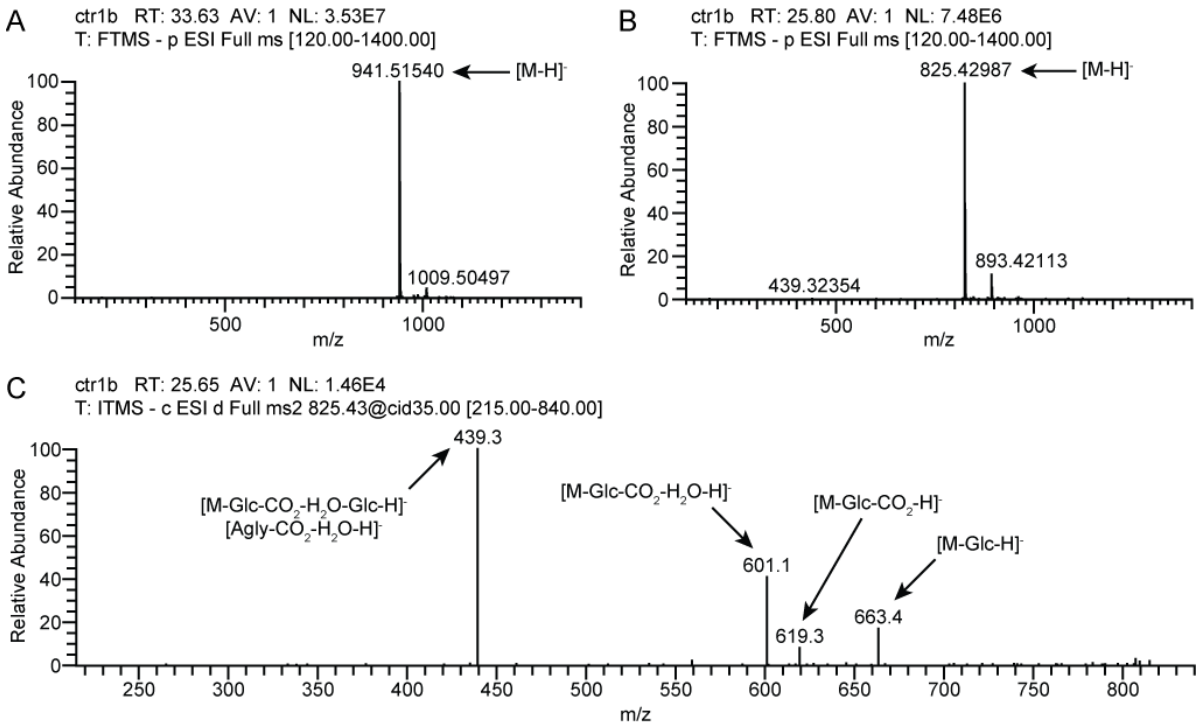


Figure 3. LC ESI FT-ICR MS chromatograms. (A) MS scan of peak at t_R 33.63 min (3-Rha-Gal-GlcA-soyasapogenol B). (B) MS scan of peak at t_R 25.80 min (3-Glc-28-Glc-medicagenic acid). (C) MS^2 fragmentation of 3-Glc-28-Glc-medicagenic acid.

whereas that from the roots rather shows allelopathic and antimicrobial activities (Tava and Avato, 2006). Thus, the *in vivo* role of the zanhic acid glycosides might be related to antinutritional effects in the aerial parts of the plant. In addition to zanhic acid glycosides, aerial parts are particularly rich in medicagenic acid glycosides (Kapusta et al., 2005a; Kapusta et al., 2005b; Huhman et al., 2005). Notably, none of the major medicagenic acid glycosides shown to occur in the aerial parts of the plant were found in the hairy roots, further underscoring the unique saponin composition of roots and aerial parts. In root extracts of greenhouse-grown *M. truncatula* plants, the presence of 31 saponins, mainly soyasapogenol conjugates, was shown (Huhman et al., 2005). Here, we could confirm the presence of 18 of these 31 previously described saponins. For some of the remaining saponins of this list, we could detect the corresponding nominal masses, but the observed accurate masses and fragmentation data did not unambiguously support the proposed structures. Hence, we did not include them in Table 1. However, in addition to the 18 previously described saponins, 61 other saponins were encountered in this study, all of which had not been detected before in *M. truncatula*. This result underscores the power of the applied FT-ICR MS platform in metabolome analysis.

Table 1. Observed saponins from *M. truncatula* hairy roots, tentative identification, and FT-ICR MS^n characteristics.

Peak numbers and tentative identification of the observed saponins		[M-H]$^-$	Formula (δ ppm)	FT-ICR MS^n: *m/z* (% base peak)
1	dHex-Hex-HexA-Aglycone D	973.50386	$C_{48}H_{78}O_{20}$ (2.560)	**MS^2 [973.50]**: 955 (100)[M-H_2O-H]$^-$, 929 (10)[M-CO_2-H]$^-$, 911 (53)[M-H_2O-CO_2-H]$^-$, 827 (3)[M-dHex-H]$^-$, 809 (7)[M-dHex-H_2O-H]$^-$, 783 (6)[M-dHex-CO_2-H]$^-$, 765 (19)[M-dHex-H_2O-CO_2-H]$^-$, 665 (2)[M-dHex-Hex-H]$^-$, 647 (26)[M-dHex-Hex-H_2O-H]$^-$, 629 (14)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 603 (7)[M-dHex-Hex-H_2O-CO_2-H]$^-$, 557 (13)[M-dHex-Hex-108-H]$^-$, 489 (11)[Agly-H]$^-$ **MS^3 [973.50→955]**: 911 (100)[M-H_2O-CO_2-H]$^-$, 765 (21)[M-dHex-H_2O-CO_2-H]$^-$, 679 (9), 629 (49)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 557 (55)[M-dHex-Hex-108-H]$^-$, 489 (7)[Agly-H]$^-$
2	Pen-Hex-Hex-Aglycone D	945.50965	$C_{47}H_{78}O_{19}$ (3.380)	**MS^2 [945.51]**: 927 (3)[M-H_2O-H]$^-$, 813 (100)[M-Pen-H]$^-$, 795 (8)[M-Pen-H_2O-H]$^-$, 651 (2)[M-Pen-Hex-H]$^-$, 489 (2)[Agly-H]$^-$ **MS^3 [945.51→813]**: 651 (64)[M-Pen-Hex-H]$^-$, 489 (100)[Agly-H]$^-$
3	Hex-Hex-HexA-Bayogenin	987.48370	$C_{48}H_{76}O_{21}$ (3.106)	**MS^2 [987.48]**: 825 (4)[M-Hex-H]$^-$, 807 (100)[M-Hex-H_2O-H]$^-$, 645 (10)[M-Hex-Hex-H_2O-H]$^-$, 601 (18)[M-Hex-Hex-H_2O-CO_2-H]$^-$, 487 (2)[Agly-H]$^-$ **MS^3 [987.48→807]**: 645 (85)[M-Hex-Hex-H_2O-H]$^-$, 601 (100)[M-Hex-Hex-H_2O-CO_2-H]$^-$, 487 (12)[Agly-H]$^-$
4	Hex-dHex-Hex-Hex-Bayogenin dHex-Hex-Hex-Bayogenin*	1119.56348 957.51011*	$C_{54}H_{88}O_{24}$ (3.754) $C_{48}H_{78}O_{19}$ (3.818)	**MS^2 [1119.56]**: 957 (100)[M-Hex-H]$^-$, 811 (54)[M-Hex-dHex-H]$^-$, 793 (9)[M-Hex-dHex-H_2O-H]$^-$, 649 (33)[M-Hex-dHex-Hex-H]$^-$, 631 (13)[M-Hex-dHex-Hex-H_2O-H]$^-$, 487 (9)[Agly-H]$^-$ **MS^3 [1119.56→957]**: 811 (100)[M-Hex-dHex-H]$^-$, 793 (6)[M-Hex-dHex-H_2O-H]$^-$, 649 (11)[M-Hex-dHex-Hex-H]$^-$, 631 (3)[M-Hex-dHex-Hex-H_2O-H]$^-$, 487 (2)[Agly-H]$^-$ **MS^3 [1119.56→811]**: 649 (100)[M-Hex-dHex-Hex-H]$^-$,631 (8)[M-Hex-dHex-Hex-H_2O-H]$^-$, 487 (68)[Agly-H]$^-$
5	Malonyl-Hex-Hex-HexA-Bayogenin	1073.48680	$C_{51}H_{78}O_{24}$ (5.378)	**MS^2 [1073.49]**: 1029 (100)[M-CO_2-H]$^-$ **MS^3 [1073.49→1029]**: 867 (15)[M-CO_2-Hex-H]$^-$, 849 (100)[M-CO_2-Hex-H_2O-H]$^-$, 807 (7)[M-malonyl-Hex-H_2O-H]$^-$, 789 (40)[M-malonyl-Hex-H_2O-H_2O-H]$^-$, 763 (11)[M-malonyl-Hex-H_2O-CO_2-H]$^-$, 643 (32)[M-CO_2-Hex-Hex-H_2O-CO_2-H]$^-$, 601 (45)[M-malonyl-Hex-Hex-H_2O-CO_2-H]$^-$, 583 (38)[M-malonyl-Hex-Hex-H_2O-H_2O-CO_2-H]$^-$, 565 (22), 515 (83), 487 (14)[Agly-H]$^-$
6	Hex-Hex-Hex-Bayogenin	973.50421	$C_{48}H_{78}O_{20}$ (2.919)	**MS^2 [973.50]**: 955 (10)[M-H_2O-H]$^-$,811 (100)[M-Hex-H]$^-$, 793 (2)[M-Hex-H_2O-H]$^-$, 649 (20)[M-Hex-Hex-H]$^-$, 631 (7)[M-Hex-Hex-H_2O-H]$^-$, 487 (8) [Agly-H]$^-$ **MS^3 [973.50→811]**: 649 (100)[M-Hex-Hex-H]$^-$, 631 (20)[M-Hex-Hex-H_2O-H]$^-$, 487 (36) [Agly-H]$^-$ **MS^3 [973.50→649]**: 487 (100) [Agly-H]$^-$
7	dHex-Hex-HexA-Soyasapogenol A	957.50926	$C_{48}H_{78}O_{19}$ (2.931)	**MS^2 [957.51]**: 939 (100)[M-H_2O-H]$^-$, 895 (49)[M-H_2O-CO_2-H]$^-$, 811 (5)[M-dHex-H]$^-$, 767 (3)[M-dHex-CO_2-H]$^-$, 749 (21)[M-dHex-H_2O-CO_2-H]$^-$, 631 (25)[M-dHex-Hex-H_2O-H]$^-$, 613 (18)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 541 (18)[M-dHex-Hex-108-H]$^-$, 473 (11)[Agly-H]$^-$, 453 (5) **MS^3 [957.51→939]**: 895 (83)[M-H_2O-CO_2-H]$^-$, 749 (19)[M-dHex-H_2O-CO_2-H]$^-$, 613 (61)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 541 (100)[M-dHex-Hex-108-H]$^-$, 473 (4)[Agly-H]$^-$, 453 (2)
8	Hex-Hex-HexA-Aglycone A	985.46884	$C_{48}H_{74}O_{21}$ (3.914)	**MS^2 [985.47]**: 823 (100)[M-Hex-H]$^-$, 779 (2)[M-Hex-CO_2-H]$^-$, 761 (2)[M-Hex-CO_2-H_2O-H]$^-$, 643 (4)[M-Hex-Hex-H_2O-H]$^-$, 617 (2)[M-Hex-Hex-CO_2-H]$^-$, 599 (2)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 485 (3)[Agly-H]$^-$ **MS^3 [985.47→823]**: 805 (1)[M-Hex-H_2O-H]$^-$, 779 (10)[M-Hex-CO_2-H]$^-$, 761 (3)[M-Hex-CO_2-H_2O-H]$^-$, 661 (2)[M-Hex-Hex-H]$^-$, 643 (100)[M-Hex-Hex-H_2O-H]$^-$, 617 (42)[M-Hex-Hex-CO_2-H]$^-$, 599 (28)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 581 (3)[M-Hex-Hex-CO_2-H_2O-H_2O-H]$^-$, 567 (18), 485 (65)[Agly-H]$^-$, 467 (19)
9	Hex-Hex-Hex-Bayogenin	973.50407	$C_{48}H_{78}O_{20}$ (2.776)	**MS^2 [973.50]**: 955 (10)[M-H_2O-H]$^-$,811 (100)[M-Hex-H]$^-$, 649 (20)[M-Hex-Hex-H]$^-$, 631 (7)[M-Hex-Hex-H_2O-H]$^-$, 487 (8) [Agly-H]$^-$

				MS^3 [973.50→811]: 649 (100)[M-Hex-Hex-H]$^-$, 631 (42)[M-Hex-Hex-H_2O-H]$^-$, 487 (86) [Agly-H]$^-$ **MS^3 [973.50→649]**: 487 (100) [Agly-H]$^-$ 409 (1), 393 (1), 391 (1)
10	Hex-Hex-Aglycone C	823.41585	$C_{42}H_{64}O_{16}$ (4.481)	**MS^2 [823.42]**: 805 (4)[M-H_2O-H]$^-$, 661 (2)[M-Hex-H]$^-$, 617 (3)[M-Hex-CO_2-H]$^-$, 599 (7)[M-Hex-CO_2-H_2O-H]$^-$, 437 (100)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [823.42→437]**: 391 (100), 375 (6)
11	Hex-HexA-Bayogenin	825.43079	$C_{42}H_{66}O_{16}$ (3.610)	**MS^2 [825.43]**: 807 (10)[M-H_2O-H]$^-$, 781 (3)[M-CO_2-H]$^-$, 765 (3), 763 (2)[M-H_2O-CO_2-H]$^-$, 705 (7), 663 (100)[M-Hex-H]$^-$, 645 (10)[M-Hex-H_2O-H]$^-$, 619 (41)[M-Hex-CO_2-H]$^-$, 601 (41)[M-Hex-CO_2-H_2O-H]$^-$, 587 (6)[M-Hex-CO_2-H_2O-H_2O-H]$^-$, 529 (6), 487 (34)[Agly-H]$^-$, **MS^3 [825.43→663]**: 587 (4)[M-Hex-CO_2-H_2O-H_2O-H]$^-$, 487 (100)[Agly-H]$^-$, 467 (5), 455 (6), 439 (5)
12	Hex-Hex-HexA-Hederagenin	971.48934	$C_{48}H_{76}O_{20}$ (3.728)	**MS^2 [971.49]**: 851 (4), 809 (100)[M-Hex-H]$^-$, 791 (7)[M-Hex-H_2O-H]$^-$, 747 (3)[M-Hex-CO_2-H_2O-H]$^-$, 629 (20)[M-Hex-Hex-H_2O-H]$^-$, 603 (3)[M-Hex-Hex-CO_2-H]$^-$, 585 (6)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 539 (1)[M-Hex-Hex-108-H]$^-$, 471 (4) [Agly-H]$^-$ **MS^3 [971.49→809]**: 747 (4)[M-Hex-CO_2-H_2O-H]$^-$, 647 (3)[M-Hex-Hex-H]$^-$, 629 (100)[M-Hex-Hex-H_2O-H]$^-$, 603 (9)[M-Hex-Hex-CO_2-H]$^-$, 585 (15)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 553 (5), 539 (4)[M-Hex-Hex-108-H]$^-$, 471 (15)[Agly-H]$^-$, 469 (6), 439 (6)
13	Hex-Hex-Hex-Medicagenic acid	987.48516	$C_{48}H_{76}O_{21}$ (4.584)	**MS^2 [987.49]**: 825 (100)[M-Hex-H]$^-$, 781 (5)[M-Hex-CO_2-H]$^-$, 711 (3), 663 (4)[M-Hex-Hex-H]$^-$, 601 (4)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 439 (53)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [987.49→825]**: 439 (100)[Agly-CO_2-H_2O-H]$^-$
14	Malonyl-Hex-Hex-HexA-Hederagenin	1057.49199 1013.50072*	$C_{51}H_{78}O_{23}$ (5.558)	**MS^2 [1013.50]**: 971 (5)[M-malonyl-H]$^-$, 833 (100)[M-Hex-CO_2-H_2O-H]$^-$, 791 (8)[M-malonyl-Hex-H_2O-H]$^-$, 773 (20)[M-malonyl-Hex-H_2O-H_2O-H]$^-$, 729 (3)[M-malonyl-Hex-H_2O-CO_2-H]$^-$, 713 (3), 671 (37)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 633 (4), 627 (25)[M-Hex-Hex-CO_2-CO_2-H_2O-H]$^-$, 611 (2), 601 (7), 585 (9)[M-malonyl-Hex-Hex-CO_2-H_2O-H]$^-$, 567 (43)[M-malonyl-Hex-Hex-CO_2-H_2O-H_2O-H]$^-$, 549 (6), 499 (70)[Hex-Hex-HexA-H]$^-$, 471 (2)[Agly-H]$^-$, 469 (5) **MS^3 [1013.50→833]**: 773 (11)[M-malonyl-Hex-H_2O-H_2O-H]$^-$, 671 (100)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 633 (11), 585 (15)[M-malonyl-Hex-Hex-CO_2-H_2O-H]$^-$, 567 (10)[M-malonyl-Hex-Hex-CO_2-H_2O-H_2O-H]$^-$, 471 (4)[Agly-H]$^-$
15	dHex-Hex-Hex-Hederagenin	941.51456	$C_{48}H_{78}O_{18}$ (3.209)	**MS^2 [941.51]**: 795 (100)[M-dHex-H]$^-$, 777 (10)[M-dHex-H_2O-H]$^-$, 633 (12)[M-dHex-Hex-H]$^-$, 615 (2)[M-dHex-Hex-H_2O-H]$^-$, 471 (1)[Agly-H]$^-$ **MS^3 [941.51→795]**: 633 (100)[M-dHex-Hex-H]$^-$, 615 (6)[M-dHex-Hex-H_2O-H]$^-$, 471 (25)[Agly-H]$^-$ **MS^3 [941.51→633]**: 471 (100)[Agly-H]$^-$, 405 (1), 393 (1)
16	Malonyl-Hex-HexA-Bayogenin	911.43120 867.44138*	$C_{45}H_{68}O_{19}$ (3.288)	**MS^2 [867.44]**: 825 (2)[M-malonyl-H]$^-$, 807 (37)[M-malonyl-H_2O-H]$^-$, 705 (3)[M-CO_2-Hex-H]$^-$, 645 (100)[M-malonyl-Hex-H_2O-H]$^-$, 601 (6)[M-malonyl-Hex-H_2O-CO_2-H]$^-$ **MS^3 [867.44→645]**: 469 (100)[Agly-H_2O-H]$^-$, 423 (10), 409 (40), 391 (8), 379 (21) **MS^3 [867.44→807]**: 601 (100)[M-malonyl-Hex-H_2O-CO_2-H]$^-$, 487 (2)[Agly-H]$^-$
17	Hex-Hex-HexA-Aglycone A	985.46931	$C_{48}H_{74}O_{21}$ (4.391)	**MS^2 [985.47]**: 823 (6)[M-Hex-H]$^-$, 805 (100)[M-Hex-H_2O-H]$^-$, 643 (10)[M-Hex-Hex-H_2O-H]$^-$, 599 (33)[M-Hex-Hex-H_2O-CO_2-H]$^-$, **MS^3 [985.47→805]**: 685 (4), 643 (45)[M-Hex-Hex-H_2O-H]$^-$, 599 (100)[M-Hex-Hex-H_2O-CO_2-H]$^-$, **MS^3 [985.47→599]**: 527 (100), 485 (2)[Agly-H]$^-$
18	Hex-Hex-Bayogenin	811.45241	$C_{42}H_{68}O_{15}$ (4.763)	**MS^2 [811.45]**: 649 (100)[M-Hex-Hex-H]$^-$, 631 (8)[M-Hex-Hex-H_2O-H]$^-$, 487 (62) [Agly-H]$^-$ **MS^3 [811.45→649]**: 487 (100) [Agly-H]$^-$ **MS^3 [811.45→487]**: 409 (100), 403 (21), 393 (38), 391 (74)
19	Hex-Hex-HexA-Hederagenin	971.48997	$C_{48}H_{76}O_{20}$ (4.377)	**MS^2 [971.49]**: 851 (3), 809 (100)[M-Hex-H]$^-$, 603 (5)[M-Hex-Hex-CO_2-H]$^-$, 585 (9)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 569 (9), 499 (1), 471 (10)[Agly-H]$^-$

				MS^3 [971.49→809]: 647 (32)[M-Hex-Hex-H]$^-$, 629 (27)[M-Hex-Hex-H_2O-H]$^-$, 611 (10)[M-Hex-Hex-H_2O-H_2O-H]$^-$, 603 (25)[M-Hex-Hex-CO_2-H]$^-$, 587 (12), 585 (15)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 567 (58)[M-Hex-Hex-CO_2-H_2O-H_2O-H]$^-$, 553 (14), 499 (78), 471 (100)[Agly-H]$^-$, 469 (15), 439 (17) **MS^3 [971.49→471]**: 439 (37), 405 (30), 393 (100)
20	Malonyl-Hex-Hex-HexA-Aglycone A	1071.47137 1027.48023*	$C_{51}H_{76}O_{24}$ (5.593)	**MS^2 [1071.47]**: 1027 (100)[M-CO_2-H]$^-$ **MS^2 [1027.48]**: 865 (14)[M-CO_2-Hex-H]$^-$, 847 (37)[M-CO_2-Hex-H_2O-H]$^-$, 823 (10)[M-malonyl-Hex-H]$^-$, 805 (7)[M-malonyl-Hex-H_2O-H]$^-$, 787 (100)[M-malonyl-Hex-H_2O-H_2O-H]$^-$, 779 (7), 761 (11), 641 (77)[M-CO_2-Hex-CO_2-H_2O-H]$^-$, 625 (9)[M-malonyl-Hex-Hex-H_2O-H_2O-H]$^-$, 581 (51)[M-malonyl-Hex-Hex-H_2O-H_2O-CO_2-H]$^-$, 513 (44), 485 (9)[Agly-H]$^-$ **MS^3 [1027.48→787]**: 625 (54)[M-malonyl-Hex-Hex-H_2O-H_2O-H]$^-$, 581 (100)[M-malonyl-Hex-Hex-H_2O-H_2O-CO_2-H]$^-$, 513 (22), 485 (45)[Agly-H]$^-$ **MS^3 [1027.48→641]**: 581 (100)[M-malonyl-Hex-Hex-H_2O-H_2O-CO_2-H]$^-$, 569 (6), 513 (11), 485 (38)[Agly-H]$^-$, 439 (2)
21	3-Glc-28-Glc-Medicagenic acid	825.42987	$C_{42}H_{66}O_{16}$ (2.494)	**MS^2 [825.43]**: 663 (5)[M-Glc-H]$^-$, 619 (3)[M-Glc-CO_2-H]$^-$, 601 (35)[M-Glc-CO_2-H_2O-H]$^-$, 439 (100)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [825.43→439]**: 393 (51), 391 (100), 375 (8) **MS^3 [825.43→601]**: 439 (100)[Agly-CO_2-H_2O-H]$^-$
22	Malonyl-Hex-HexA-Bayogenin	911.43371 867.44133*	$C_{45}H_{68}O_{19}$ (6.041)	**MS^2 [867.44]**: 825 (34)[M-malonyl-H]$^-$, 807 (25)[M-malonyl-H_2O-H]$^-$, 705 (9)[M-CO_2-Hex-H]$^-$, 645 (77)[M-malonyl-Hex-H_2O-H]$^-$, 601 (100)[M-malonyl-Hex-H_2O-CO_2-H]$^-$ **MS^3 [867.44→601]**: 583 (3), 529 (100), 487 (7)[Agly-H]$^-$
23	dHex-Hex-Hex-Hederagenin	941.51536	$C_{48}H_{78}O_{18}$ (4.058)	**MS^2 [941.51]**: 795 (100)[M-dHex-H]$^-$, 777 (12)[M-dHex-H_2O-H]$^-$, 633 (13)[M-dHex-Hex-H]$^-$, 615 (2)[M-dHex-Hex-H_2O-H]$^-$ **MS^3 [941.51→795]**: 633 (100)[M-dHex-Hex-H]$^-$, 615 (5)[M-dHex-Hex-H_2O-H]$^-$, 471 (20)[Agly-H]$^-$ **MS^3 [941.51→633]**: 471 (100)[Agly-H]$^-$, 393 (3)
24	Hex-Hex-HexA-Bayogenin	987.48378	$C_{48}H_{76}O_{21}$ (3.187)	**MS^2 [987.48]**: 969 (3)[M-H_2O-H]$^-$, 825 (100)[M-Hex-H]$^-$, 645 (2)[M-Hex-Hex-H_2O-H]$^-$, 601 (2)[M-Hex-Hex-H_2O-CO_2-H]$^-$, 487 (3)[Agly-H]$^-$ **MS^3 [987.48→825]**: 781 (4)[M-Hex-CO_2-H]$^-$, 645 (88)[M-Hex-Hex-H_2O-H]$^-$, 619 (24)[M-Hex-Hex-CO_2-H]$^-$, 601 (29)[M-Hex-Hex-H_2O-CO_2-H]$^-$, 569 (23), 487 (100)[Agly-H]$^-$, 467 (3)
25	Hex-dHex-Hex-HexA-Hederagenin	1117.54979	$C_{54}H_{86}O_{24}$ (5.515)	**MS^2 [1117.55]**: 1099 (5)[M-H_2O-H]$^-$, 955 (100)[M-Hex-H]$^-$, 937 (4)[M-Hex-H_2O-H]$^-$, 893 (4)[M-Hex-H_2O-CO_2-H]$^-$, 791 (1)[M-Hex-dHex-H_2O-H]$^-$, 747 (1)[M-Hex-dHex-H_2O-CO_2-H]$^-$, 629 (2)[M-Hex-dHex-Hex-H]$^-$, 539 (2)[M-Hex-dHex-Hex-108-H]$^-$ **MS^2 [1117.55→955]**: 937 (88)[M-Hex-H_2O-H]$^-$, 893 (67)[M-Hex-H_2O-CO_2-H]$^-$, 747 (67)[M-Hex-dHex-H_2O-CO_2-H]$^-$, 729 (21)[M-Hex-dHex-H_2O-H_2O-CO_2-H]$^-$, 629 (100)[M-Hex-dHex-Hex-H]$^-$, 611 (15)[M-Hex-dHex-Hex-H_2O-H]$^-$, 585 (14)[M-Hex-dHex-Hex-CO_2-H]$^-$, 539 (16)[M-Hex-dHex-Hex-108-H]$^-$, 471 (17)[Agly-H]$^-$, 453 (31)
26	3-Glc-malonyl-28-Glc-Medicagenic acid	911.43233	$C_{45}H_{68}O_{19}$ (4.522)	**MS^2 [911.43]**: 867 (100)[M-CO_2-H]$^-$, 705 (3)[M-Glc-CO_2-H]$^-$ **MS^3 [911.43→867]**: 705 (4)[M-Glc-CO_2-H]$^-$, 661 (7)[M-Glc-CO_2-CO_2-H]$^-$, 601 (36)[M-Glc-malonyl-CO_2-H_2O-H]$^-$, 439 (100)[Agly-CO_2-H_2O-H]$^-$
27	Hex-Bayogenin	649.39774	$C_{36}H_{58}O_{10}$ (3.108)	**MS^2 [649.40]**: 487 (100)[Agly-H]$^-$, 469 (15), 441 (28), 409 (36)
28	Hex-Medicagenic acid	663.37701	$C_{36}H_{56}O_{11}$ (3.051)	**MS^2 [663.38]**: 645 (5)[M-H_2O-H]$^-$, 501 (13)[Agly-H]$^-$, 439 (100)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [663.38→439]**: 393 (83), 391 (100)
29	Hex-HexA-Aglycone A	823.41492	$C_{42}H_{64}O_{16}$ (3.352)	**MS^2 [823.41]**: 661 (100)[M-Hex-H]$^-$, 617 (33)[M-Hex-CO_2-H]$^-$, 599 (52)[M-Hex-H_2O-CO_2-H]$^-$, 485 (15)[Agly-H]$^-$

				MS^3 [823.41→661]: 585 (5), 485 (100)[Agly-H]$^-$, 467 (34) **MS^3 [823.41→599]**: 527 (100), 485 (3)[Agly-H]$^-$, 467 (2)
30	Malonyl-Hex-Hex-Medicagenic acid	911.43152 867.44123*	$C_{45}H_{68}O_{19}$ (3.638)	**MS^2 [867.44]**: 825 (33)[M-malonyl-H]$^-$, 807 (16)[M-malonyl-H_2O-H]$^-$, 705 (12)[M-Hex-CO_2-H]$^-$, 663 (10)[M-malonyl-Hex-H]$^-$, 645 (63)[M-malonyl-Hex-H_2O-H]$^-$, 601 (100)[M-malonyl-Hex-H_2O-CO_2-H]$^-$, 487 (9), 439 (82)[Agly-CO_2-H_2O-H]$^-$
31	dHex-Hex-HexA-Hederagenin	955.49433	$C_{48}H_{76}O_{19}$ (3.690)	**MS^2 [955.49]**: 937 (100)[M-H_2O-H]$^-$, 911 (7)[M-CO_2-H]$^-$, 893 (35)[M-H_2O-CO_2-H]$^-$, 809 (8)[M-dHex-H]$^-$, 747 (24)[M-dHex-H_2O-CO_2-H]$^-$, 629 (28)[M-dHex-Hex-H_2O-H]$^-$, 611 (11)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 539 (22)[M-dHex-Hex-108-H]$^-$, 471 (4)[Agly-H]$^-$ **MS^3 [955.49→937]**: 893 (100)[M-H_2O-CO_2-H]$^-$, 859 (6), 747 (20)[M-dHex-H_2O-CO_2-H]$^-$, 611 (66)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 539 (74)[M-dHex-Hex-108-H]$^-$, 471 (13)[Agly-H]$^-$, 451 (4), 441 (9) **MS^3 [955.49→893]**: 747 (100)[M-dHex-H_2O-CO_2-H]$^-$, 729 (2), 585 (3), 567 (5), 471 (12)[Agly-H]$^-$
32	Hex-dHex-Hex-HexA-Soyasapogenol E	1101.55386	$C_{54}H_{86}O_{23}$ (4.673)	**MS^2 [1101.55]**: 1083 (8)[M-H_2O-H]$^-$, 939 (100)[M-Hex-H]$^-$, 921 (6)[M-Hex-H_2O-H]$^-$, 877 (4)[M-Hex-H_2O-CO_2-H]$^-$, 731 (1)[M-Hex-dHex-H_2O-CO_2-H]$^-$, 613 (2)[M-Hex-dHex-Hex-H_2O-H]$^-$, 523 (3) **MS^3 [1101.55→939]**: 921 (100)[M-Hex-H_2O-H]$^-$, 877 (36)[M-Hex-H_2O-CO_2-H]$^-$, 793 (5), 731 (43)[M-Hex-dHex-H_2O-CO_2-H]$^-$, 713 (12)[M-Hex-dHex-H_2O-H_2O-CO_2-H]$^-$, 613 (47)[M-Hex-dHex-Hex-H_2O-H]$^-$, 595 (26)[M-Hex-dHex-Hex-H_2O-H_2O-H]$^-$, 523 (58)[M-Hex-dHex-Hex-108-H]$^-$, 455 (7)[Agly-H]$^-$, 453 (4) **MS^3 [1101.55→1083]**: 921 (100)[M-Hex-H_2O-H]$^-$, 877 (43)[M-Hex-H_2O-CO_2-H]$^-$, 757 (22), 523 (50)[M-Hex-dHex-Hex-108-H]$^-$
33	Hex-Hex-HexA-Hederagenin	971.48921	$C_{48}H_{76}O_{20}$ (3.594)	**MS^2 [971.49]**: 809 (100)[M-Hex-H]$^-$, 791 (1)[M-Hex-H_2O-H]$^-$, 765 (1)[M-Hex-CO_2-H]$^-$, 629 (2)[M-Hex-Hex-H_2O-H]$^-$, 603 (1)[M-Hex-Hex-CO_2-H]$^-$, 585 (1)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 539 (1), 471 (1)[Agly-H]$^-$ **MS^3 [971.49→809]**: 765 (6)[M-Hex-CO_2-H]$^-$, 747 (6)[M-Hex-CO_2-H_2O-H]$^-$, 647 (1)[M-Hex-Hex-H]$^-$, 629 (100)[M-Hex-Hex-H_2O-H]$^-$, 611 (2)[M-Hex-Hex-H_2O-H_2O-H]$^-$, 603 (27)[M-Hex-Hex-CO_2-H]$^-$, 585 (24)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 567 (5)[M-Hex-Hex-CO_2-H_2O-H_2O-H]$^-$, 553 (17), 539 (9)[M-Hex-Hex-108-H]$^-$, 471 (100)[Agly-H]$^-$, 453 (17)
34	dHex-Hex-Hex-HexA-Soyasapogenol B	1103.57097	$C_{54}H_{88}O_{23}$ (5.987)	**MS^2 [1103.57]**: 1085 (100)[M-H_2O-H]$^-$, 1041 (60)[M-H_2O-CO_2-H]$^-$, 957 (9)[M-dHex-H]$^-$, 913 (7)[M-dHex-CO_2-H]$^-$, 895 (33)[M-dHex-H_2O-CO_2-H]$^-$, 795 (2)[M-dHex-Hex-H]$^-$, 777 (31)[M-dHex-Hex-H_2O-H]$^-$, 759 (16)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 620 (2), 571 (1), 457 (1)[Agly-H]$^-$ **MS^3 [1103.57→1085]**: 1067 (18)[M-H_2O-H_2O-H]$^-$, 1041 (100)[M-H_2O-CO_2-H]$^-$, 895 (19)[M-dHex-H_2O-CO_2-H]$^-$, 759 (54)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 619 (11)
35	Malonyl-Hex-malonyl-Hex-Bayogenin	983.45391	$C_{48}H_{72}O_{21}$ (4.654)	**(MS)**: 983 (21)[M-H]$^-$, 895 (21)[M-CO_2-CO_2-H]$^-$, 734 (11)[M-CO_2-CO_2-Hex-H]$^-$, 691 (100)[M-malonyl-Hex-CO_2-H]$^-$, 649 (18)[M-malonyl-Hex-malonyl-H]$^-$, 631 (62)[M-malonyl-Hex-malonyl-H_2O-H]$^-$, 487 (39)[Agly-H]$^-$
36	Malonyl-Hex-malonyl-Hex-Medicagenic acid	997.4345	$C_{48}H_{70}O_{22}$ (5.918)	**MS^2 [997.43]**: 953 (51)[M-CO_2-H]$^-$, 909 (100)[M-CO_2-CO_2-H]$^-$, 733 (6), 705 (24)[M-malonyl-Hex-CO_2-H]$^-$, 691 (26) **MS^3 [997.43→909]**: 891 (4)[M-CO_2-CO_2-H_2O-H]$^-$, 865 (6)[M-CO_2-CO_2-CO_2-H]$^-$, 849 (3)[M-CO_2-CO_2-malonyl-H_2O-H]$^-$, 703 (4)[M-CO_2-CO_2-CO_2-Hex-H]$^-$, 661 (6)[M-CO_2-CO_2-malonyl-Hex-H]$^-$, 643 (15)[M-CO_2-CO_2-malonyl-Hex-H_2O-H]$^-$, 439 (100)[Agly-CO_2-H_2O-H]$^-$
37	dHex-Hex-HexA-Soyasapogenol A	957.51003	$C_{48}H_{78}O_{19}$ (3.735)	**MS^2 [957.51]**: 939 (100)[M-H_2O-H]$^-$, 895 (50)[M-H_2O-CO_2-H]$^-$, 811 (8)[M-dHex-H]$^-$, 767 (3)[M-dHex-CO_2-H]$^-$, 749 (29)[M-dHex-H_2O-CO_2-H]$^-$, 649 (1)[M-dHex-Hex-H]$^-$, 631 (29)[M-dHex-Hex-H_2O-H]$^-$, 613 (15)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 541 (23)[M-dHex-Hex-108-H]$^-$, 473 (8)[Agly-H]$^-$ **MS^3 [957.51→939]**: 895 (100)[M-H_2O-CO_2-H]$^-$, 749 (14)[M-dHex-H_2O-CO_2-H]$^-$, 613 (61)[M-dHex-

				Hex-H_2O-H_2O-H]$^-$, 541 (62)[M-dHex-Hex-108-H]$^-$ **MS^3 [957.51→895]**: 749 (100)[M-dHex-H_2O-CO_2-H]$^-$, 587 (5)[M-dHex-Hex-H_2O-CO_2-H]$^-$, 473 (13)[Agly-H]$^-$
38	Malonyl-Hex-HexA-Hederagenin	895.43802 851.44613*	$C_{45}H_{68}O_{18}$ (5.283) $C_{44}H_{68}O_{16}$ (3.136)	**MS^2 [851.45]**: 809 (29)[M-malonyl-H]$^-$, 791 (27)[M-malonyl-H_2O-H]$^-$, 689 (82)[M-Hex-CO_2-H]$^-$, 629 (84)[M-malonyl-Hex-H_2O-H]$^-$, 585 (100)[M-malonyl-Hex-CO_2-H_2O-H]$^-$, 471 (4)[Agly-H]$^-$ **MS^3 [851.45→585]**: 513 (100), 471 (13)[Agly-H]$^-$ **MS^3 [851.45→629]**: 629 (100)[M-malonyl-H_2O-H]$^-$, 513 (54), 471 (78)[Agly-H]$^-$, 453 (45), 405 (64), 393 (74), 387 (22)
39	dHex-Hex-HexA-Hederagenin	955.49392	$C_{48}H_{76}O_{19}$ (3.261)	**MS^2 [955.49]**: 937 (100)[M-H_2O-H]$^-$, 911 (34)[M-CO_2-H]$^-$, 893 (68)[M-H_2O-CO_2-H]$^-$, 809 (30)[M-dHex-H]$^-$, 747 (20)[M-dHex-H_2O-CO_2-H]$^-$, 689 (8), 629 (48)[M-dHex-Hex-H_2O-H]$^-$, 611 (8)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 539 (14)[M- dHex-Hex-108-H]$^-$, 471 (17)[Agly-H]$^-$
40	Hex-HexA-malonyl-Aglycone A	909.41850 865.42579*	$C_{45}H_{66}O_{19}$ (6.539) $C_{44}H_{66}O_{17}$ (3.543)	**MS^2 [865.43]**: 823 (3)[M-malonyl-H]$^-$, 821 (4)[M-CO_2-CO_2-H]$^-$, 805 (28)[M-malonyl-H_2O-H]$^-$, 761 (4)[M-malonyl-CO_2-H_2O-H]$^-$, 703 (8)[M-CO_2-Hex-H]$^-$, 643 (100)[M-malonyl-Hex-H_2O-H]$^-$, 599 (35)[M-malonyl-Hex-CO_2-H_2O-H]$^-$, 485 (1)[Agly-H]$^-$ **MS^3 [865.43→643]**: 509 (100), 485 (64)[Agly-H]$^-$, 467 (41), 421 (41)
41	Hex-Hederagenin	633.40242	$C_{36}H_{58}O_9$ (2.547)	**MS^2 [633.40]**: 615 (33)[M-H_2O-H]$^-$, 513 (15), 471 (100)[Agly-H]$^-$, 453 (24), 405 (24), 393 (39)
42	Hex-Hex-HexA-Soyasapogenol E	955.49420	$C_{48}H_{76}O_{19}$ (3.554)	**MS^2 [955.49]**: 937 (2)[M-H_2O-H]$^-$, 835 (2), 793(100)[M-Hex-H]$^-$, 775 (2)[M-Hex-H_2O-H]$^-$, 731 (2)[M-Hex-H_2O-CO_2-H]$^-$, 613 (4)[M-Hex-Hex-H_2O-H]$^-$, 523 (5)[M-Hex-Hex-108-H]$^-$ **MS^3 [955.49→793]**: 749 (3)[M-Hex-CO_2-H]$^-$, 731 (29)[M-Hex-H_2O-CO_2-H]$^-$, 631 (2)[M-Hex-Hex-H]$^-$, 613 (100)[M-Hex-Hex-H_2O-H]$^-$, 595 (8)[M-Hex-Hex-H_2O-H_2O-H]$^-$, 587 (5)[M-Hex-Hex-CO_2-H]$^-$, 569 (8)[M-Hex-Hex-H_2O-CO_2-H]$^-$, 551 (6)[M-Hex-Hex-H_2O-H_2O-CO_2-H]$^-$, 537 (5), 523 (54)[M-Hex-Hex-108-H]$^-$, 455 (14)[Agly-H]$^-$, 453 (6)
43	Malonyl-Hex-malonyl-Hex-Hederagenin	967.45877 879.47798*	$C_{48}H_{72}O_{20}$ (4.498)	**MS^2 [879,47]**: 861 (2)[M-CO_2-CO_2-H_2O-H]$^-$, 819 (6)[M-malonyl-CO_2-H_2O-H]$^-$, 777 (5)[M-malonyl-malonyl-H_2O-H]$^-$, 717 (27)[M-Hex-CO_2-CO_2-H]$^-$, 675 (100)[M-Hex-malonyl-CO_2-H]$^-$, 657 (3)[M-Hex-malonyl-CO_2-H_2O-H]$^-$, 615 (72)[M-Hex-malonyl-malonyl-H_2O-H]$^-$ **MS^3 [879.47→675]**: 657 (3)[M-Hex-malonyl-CO_2-H_2O-H]$^-$, 633 (20)[M-Hex-malonyl-malonyl-H]$^-$, 615 (100)[M-Hex-malonyl-malonyl-H_2O-H]$^-$, 513 (6)[M-Hex-malonyl-Hex-CO_2-H]$^-$, 485 (5), 471 (2)[Agly-H]$^-$, 453 (2), 425 (6), 393 (2) **MS^3 [879.47→615]**: 585 (26), 543 (6), 499 (6), 495 (8), 471 (23)[Agly-H]$^-$, 453 (9), 405 (16), 393 (100), 373 (7)
44	Hex-Pen-Hederagenin	765.44543	$C_{71}H_{66}O_{13}$ (3.088)	**MS^2 [765.45]**: 603 (100)[M-Hex-H]$^-$, 585 (7)[M-Hex-H_2O-H]$^-$, 471 (6)[Agly-H]$^-$ **MS^3 [765.45→603]**: 471 (100)[Agly-H]$^-$, 423 (14)
45	dHex-Hex-Hex-Soyasapogenol E	925.51905	$C_{48}H_{78}O_{17}$ (2.620)	**MS^2 [925.52]**: 779 (100)[M-dHex-H]$^-$, 761 (13)[M-dHex-H_2O-H]$^-$, 617 (7)[M-dHex-Hex-H]$^-$, 599 (1)[M-dHex-Hex-H_2O-H]$^-$, 455 (1)[Agly-H]$^-$ **MS^3 [925.52→779]**: 617 (100)[M-dHex-Hex-H]$^-$, 599 (4)[M-dHex-Hex-H_2O-H]$^-$, 455 (9)[Agly-H]$^-$
46	dHex-Hex-HexA-Hederagenin	955.49417	$C_{48}H_{76}O_{19}$ (3.523)	**MS^2 [955.49]**: 937 (100)[M-H_2O-H]$^-$, 911 (7)[M-CO_2-H]$^-$, 893 (50)[M-H_2O-CO_2-H]$^-$, 809 (15)[M-dHex-H]$^-$, 767 (8), 747 (8)[M-dHex-H_2O-CO_2-H]$^-$, 629 (23)[M-dHex-Hex-H_2O-H]$^-$, 611 (28)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 539 (36)[M- dHex-Hex-108-H]$^-$, 471 (11)[Agly-H]$^-$
47	Malonyl-Hex-Hex-HexA-Soyasapogenol E	1041.49582 997.50288*	$C_{51}H_{78}O_{22}$ (4.438)	**MS^2 [997.50]**: 835 (3)[M-Hex-CO_2-H]$^-$, 817 (51)[M-Hex-H_2O-CO_2-H]$^-$, 775 (12)[M-Hex-H_2O-malonyl-H]$^-$, 757 (4)[M-Hex-H_2O-malonyl-H_2O-H]$^-$, 679 (3), 655 (100)[M-Hex-Hex-H_2O-CO_2-H]$^-$, 611 (5)[M-Hex-Hex-H_2O-CO_2-CO_2-H]$^-$, 595 (2)[M-Hex-Hex-H_2O-malonyl-H_2O-H]$^-$, 551 (3)[M-Hex-Hex-H_2O-CO_2-malonyl-H_2O-H]$^-$ **MS^3 [997.50→655]**: 595 (4)[M-Hex-Hex-H_2O-malonyl-H_2O-H]$^-$,551 (20)[M-Hex-Hex-H_2O-CO_2-

				malonyl-H_2O-H]$^-$,539 (100), 509 (65), 455 (15)[Agly-H]$^-$
48	dHex-Hex-Hex-Aglycone A	955.49454	$C_{48}H_{76}O_{19}$ (3.910)	**MS^2 [955.49]**: 809 (100)[M-dHex-H]$^-$, 791 (8)[M-dHex-H_2O-H]$^-$, 647 (5)[M-dHex-Hex-H]$^-$ **MS^3 [955.49→809]**: 647 (100)[M-dHex-Hex-H]$^-$, 485 (5)[Agly-H]$^-$
49	dHex-Hex-HexA-Bayogenin	971.48812	$C_{48}H_{76}O_{20}$ (2.473)	**MS^2 [971.49]**: 953 (66)[M-H_2O-H]$^-$, 927 (15)[M-CO_2-H]$^-$, 909 (100)[M-H_2O-CO_2-H]$^-$, 825 (4)[M-dHex-H]$^-$, 763 (6)[M-dHex-H_2O-CO_2-H]$^-$, 663 (3)[M-dHex-Hex-H]$^-$, 645 (47)[M-dHex-Hex-H_2O-H]$^-$, 627 (5)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 601 (5)[M-dHex-Hex-H_2O-H_2O-CO_2-H]$^-$, 555 (15)[M-dHex-Hex-108-H]$^-$, 527 (5), 487 (5)[Agly-H]$^-$, 483 (3), 469 (6) **MS^3 [971.49→909]**: 837 (7), 763 (100)[M-dHex-H_2O-CO_2-H]$^-$, 745 (3)[M-dHex-H_2O-H_2O-CO_2-H]$^-$, 601 (24)[M-dHex-Hex-H_2O-H_2O-CO_2-H]$^-$, 583 (2), 573 (10), 487 (6)[Agly-H]$^-$, 485 (1), 459 (4), 439 (3) **MS^3 [971.49→953]**: 909 (100)[M-H_2O-CO_2-H]$^-$, 627 (11)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 555 (40)[M-dHex-Hex-108-H]$^-$, 527 (2), 487 (2)[Agly-H]$^-$, 469 (6), 441 (1)
50	dHex-Hex-HexA-Hederagenin	955.49245	$C_{48}H_{76}O_{19}$ (1.723)	**MS^2 [955.49]**: 937 (100)[M-H_2O-H]$^-$, 893 (35)[M-H_2O-CO_2-H]$^-$, 747 (30)[M-dHex-H_2O-CO_2-H]$^-$, 629 (41)[M-dHex-Hex-H_2O-H]$^-$, 611 (22)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 539 (21)[M-dHex-Hex-108-H]$^-$, 471 (18)[Agly-H]$^-$, 451 (4), 307 (4)
51	dHex-Hex-Hex-Bayogenin	957.50866	$C_{48}H_{78}O_{19}$ (2.304)	**MS^2 [957.50]**: 939(2)[M-H_2O-H]$^-$, 811 (100)[M-dHex-H]$^-$, 793 (6)[M-dHex-H_2O-H]$^-$, 649 (12)[M-dHex-Hex-H]$^-$, 631 (2)[M-dHex-Hex-H_2O-H]$^-$, 487 (2)[Agly-H]$^-$ **MS^3 [957.50→811]**: 649 (100)[M-dHex-Hex-H]$^-$, 631 (12)[M-dHex-Hex-H_2O-H]$^-$, 487 (56)[Agly-H]$^-$
52	Hex-HexA-Bayogenin	825.43024	$C_{42}H_{66}O_{16}$ (3.162)	**MS^2 [825.43]**: 645 (100)[M-Hex-H_2O-H]$^-$, 619 (28)[M-Hex-CO_2-H]$^-$, 601 (20)[M-Hex-H_2O-CO_2-H]$^-$, 569 (11), 487 (78)[Agly-H]$^-$, 455 (8) **MS^3 [825.43→645]**: 557 (11), 487 (11)[Agly-H]$^-$, 469 (47), 439 (100) **MS^3 [825.43→487]**: 467 (67), 455 (72), 437 (15), 409 (100), 393 (10), 391 (15)
53	Malonyl-dHex-Hex-HexA-Soyasapogenol B (Pisumsaponin I)	1027.51636	$C_{51}H_{80}O_{21}$ (4.308)	**MS^2 [1027.52]**: 983 (100)[M-CO_2-H]$^-$ **MS^3 [1027.52→983]**: 965 (100)[M-CO_2-H_2O-H]$^-$, 921 (54)[M-CO_2-CO_2-H_2O-H]$^-$, 837 (10)[M-CO_2-dHex-H]$^-$, 793 (10)[M-CO_2-dHex-CO_2-H_2O-H]$^-$, 775 (23)[M-CO_2-dHex-CO_2-H_2O-H_2O-H]$^-$, 675 (3)[M-CO_2-dHex-Hex-H]$^-$, 657 (24)[M-CO_2-dHex-Hex-H_2O-H]$^-$, 639 (20)[M-CO_2-dHex-Hex-H_2O-H_2O-H]$^-$, 613 (1)[M-CO_2-dHex-Hex-H_2O-CO_2-H]$^-$, 567 (21)[M-CO_2-dHex-Hex-108-H]$^-$, 499 (7)[M-CO_2-dHex-Hex-HexA-H]$^-$
54	Hex-HexA-Soyasapogenol E	793.43923	$C_{42}H_{66}O_{14}$ (1.575)	**MS^2 [793.44]**: 673 (8), 631 (100)[M-Hex-H]$^-$, 613 (13)[M-Hex-H_2O-H]$^-$, 569 (8)[M-Hex-H_2O-CO_2-H]$^-$ **MS^3 [793.44→631]**: 613 (4)[M-Hex-H_2O-H]$^-$, 555 (3), 497 (2), 455 (100)[Agly-H]$^-$
55	Malonyl-dHex-Hex-Medicagenic acid	895.43821 851.44654*	$C_{45}H_{68}O_{18}$ (5.496) $C_{44}H_{68}O_{16}$ (3.617)	**MS^2 [851.45]**: 809 (49)[M-malonyl-H]$^-$, 791 (8)[M-malonyl-H_2O-H]$^-$, 689 (37)[M-Hex-CO_2-H]$^-$, 629 (48)[M-malonyl-Hex-H_2O-H]$^-$, 585 (100)[M-malonyl-Hex-H_2O-CO_2-H]$^-$, 439 (9)[Agly-CO_2-H_2O-H]$^-$
56	Hex-Hex-Bayogenin	811.45028	$C_{42}H_{68}O_{15}$ (2.138)	**MS^2 [811.45]**: 649 (100)[M-Hex-H]$^-$, 631 (8)[M-Hex-H_2O-H]$^-$, 487 (58)[Agly-H]$^-$ **MS^3 [811.45→649]**: 487 (100)[Agly-H]$^-$, 439 (1)
57	dHex-Hex-Hex-Aglycone A	955.49230	$C_{48}H_{76}O_{19}$ (1.566)	**MS^2 [955.49]**: 937 (3)[M-H_2O-H]$^-$, 809 (100)[M-dHex-H]$^-$, 791 (6)[M-dHex-H_2O-H]$^-$, 647 (15)[M-dHex-Hex-H]$^-$, 629 (6)[M-dHex-Hex-H_2O-H]$^-$, 485 (8)[Agly-H]$^-$ **MS^3 [955.49→809]**: 647 (100)[M-dHex-Hex-H]$^-$, 629 (26)[M-dHex-Hex-H_2O-H]$^-$, 485 (92)[Agly-H]$^-$
58	Hex-Hex-Bayogenin	811.45207	$C_{42}H_{68}O_{15}$ (4.344)	**MS^2 [811.45]**: 649 (100)[M-Hex-H]$^-$, 631 (31)[M-Hex-H_2O-H]$^-$, 487 (98)[Agly-H]$^-$ **MS^3 [811.45→649]**: 487 (100)[Agly-H]$^-$, 439 (1)
59	Hex-HexA-Bayogenin	825.43143	$C_{42}H_{66}O_{16}$ (4.386)	**MS^2 [825.43]**: 645 (100)[M-Hex-H_2O-H]$^-$, 601 (17)[M-Hex-H_2O-CO_2-H]$^-$, 487 (6)[Agly-H]$^-$, 439 (1) **MS^3 [825.43→645]**: 487 (100)[Agly-H]$^-$, 467 (2), 455 (5), 439 (9) **MS^3 [825.43→601]**: 529 (100), 487 (8)[Agly-H]$^-$
60	3-Glc-Glc-Medicagenic acid	825.43148	$C_{42}H_{66}O_{16}$ (4.446)	**MS^2 [825.43]**: 663 (1)[M-Glc-H]$^-$, 601 (1)[M-Glc-H_2O-CO_2-H]$^-$, 439 (100)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [825.43→439]**: 393 (61), 391 (100), 375 (25)

61	dHex-Hex-Hex-Hederagenin	941.51621	$C_{48}H_{78}O_{18}$ (4.961)	**MS^2 [941.51]**: 795 (100)[M-dHex-H]$^-$, 777 (10)[M-dHex-H_2O-H]$^-$, 633 (13)[M-dHex-Hex-H]$^-$, 615 (2)[M-dHex-Hex-H_2O-H]$^-$, 471 (1)[Agly-H]$^-$ **MS^3 [941.51→795]**: 633 (100)[M-dHex-Hex-H]$^-$, 615 (5)[M-dHex-Hex-H_2O-H]$^-$, 471 (24)[Agly-H]$^-$ **MS^3 [941.51→633]**: 471 (100)[Agly-H]$^-$
62	Malonyl-Hex-Hex-Bayogenin	897.45213 853.46093*	$C_{45}H_{70}O_{18}$ (3.556) $C_{44}H_{70}O_{16}$ (2.132)	**MS^2 [853.46]**: 811 (100)[M-malonyl-H]$^-$, 793 (85)[M-malonyl-H_2O-H]$^-$, 691 (6)[M-CO_2-Hex-H]$^-$, 673 (2)[M-CO_2-H_2O-Hex-H]$^-$, 649 (3)[M-malonyl-Hex-H]$^-$, 631 (6)[M-malonyl-Hex-H_2O-H]$^-$, 613 (2)[M-malonyl-Hex-CO_2-H]$^-$, 487 (10)[Agly-H]$^-$ **MS^3 [853.46→811]**: 649 (100)[M-malonyl-Hex-H]$^-$, 631 (33)[M-malonyl-Hex-H_2O-H]$^-$, 487 (40)[Agly-H]$^-$ **MS^3 [853.46→793]**: 715 (2), 691 (6), 649 (8)[M-malonyl-Hex-H]$^-$, 631 (51)[M-malonyl-Hex-H_2O-H]$^-$, 613 (8)[M-malonyl-Hex-CO_2-H]$^-$, 595 (4)[M-malonyl-Hex-H_2O-CO_2-H]$^-$, 583 (2), 515 (1), 487 (100)[Agly-H]$^-$, 469 (4), 439 (10)
63	Hex-HexA-Hederagenin	809.43693	$C_{42}H_{66}O_{15}$ (4.985)	**MS^2 [809.44]**: 747 (5)[M-H_2O-CO_2-H]$^-$, 647 (2)[M-Hex-H]$^-$, 629 (100)[M-Hex-H_2O-H]$^-$, 603 (11)[M-Hex-CO_2-H]$^-$, 585 (14)[M-Hex-H_2O-CO_2-H]$^-$, 567 (2)[M-Hex-H_2O-H_2O-CO_2-H]$^-$, 553 (5), 539 (2)[M-Hex-108-H]$^-$, 471 (15)[Agly-H]$^-$, 469 (6), 453 (1), 451 (1), 439 (8), 423 (2)
64	Hex-HexA-Aglycone A	823.41390	$C_{42}H_{66}O_{16}$ (2.113)	**MS^2 [823.41]**: 779 (12)[M-CO_2-H]$^-$, 761 (5)[M-H_2O-CO_2-H]$^-$, 643 (100)[M-Hex-H_2O-H]$^-$, 617 (40)[M-Hex-CO_2-H]$^-$, 599 (34)[M-Hex-H_2O-CO_2-H]$^-$, 581 (4)[M-Hex-H_2O-H_2O-CO_2-H]$^-$, 567 (12), 485 (72)[Agly-H]$^-$, 467 (21)
65	Pen-Hex-Hex-Aglycone B	925.48149	$C_{47}H_{74}O_{18}$ (1.352)	**MS^2 [925.48]**: 907 (25)[M-H_2O-H]$^-$, 793 (100)[M-Pen-H]$^-$, 775 (12)[M-Pen-H_2O-H]$^-$, 631 (54)[M-Pen-Hex-H]$^-$, 613 (7)[M-Pen-Hex-H_2O-H]$^-$, 469 (5)[Agly-H]$^-$ **MS^3 [925.48→793]**: 775 (9)[M-Pen-H_2O-H]$^-$, 631 (100)[M-Pen-Hex-H]$^-$, 613 (10)[M-Pen-Hex-H_2O-H]$^-$, 469 (9)[Agly-H]$^-$ **MS^3 [925.48→631]**: 613 (29)[M-Pen-Hex-H_2O-H]$^-$, 469 (100)[Agly-H]$^-$, 451 (3)
66	Hex-Hex-Aglycone A	809.43647	$C_{42}H_{66}O_{15}$ (4.417)	**MS^2 [809.44]**: 647 (99)[M-Hex-H]$^-$, 629 (27)[M-Hex-H_2O-H]$^-$, 485 (100)[Agly-H]$^-$ **MS^3 [809.44→485]**: 467 (100), 455 (21), 439 (11), 423 (40), 421 (61), 403 (48), 391 (11), 363 (10), 347 (18) **MS^3 [809.44→647]**: 485 (100)[Agly-H]$^-$
67	Hex-Medicagenic acid	663.37661	$C_{36}H_{56}O_{11}$ (2.448)	**MS^2 [663.38]**: 487 (6)[Agly-H_2O-H]$^-$, 439 (100)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [663.38→439]**: 393 (61), 391 (100), 375 (11)
68	dHex-Hex-HexA-Soyasapogenol E	939.49979	$C_{48}H_{76}O_{18}$ (4.152)	**MS^2 [939.50]**: 921 (100)[M-H_2O-H]$^-$, 877 (24)[M-H_2O-CO_2-H]$^-$, 793 (11)[M-dHex-H]$^-$, 749 (5)[M-dHex-CO_2-H]$^-$, 731 (18)[M-dHex-H_2O-CO_2-H]$^-$, 613 (14)[M-dHex-Hex-H_2O-H]$^-$, 595 (17)[M-dHex-Hex-2H_2O-H]$^-$, 523 (11)[M-dHex-Hex-108-H]$^-$, 465 (24), 455 (8)[Agly-H]$^-$ **MS^3 [939.50→921]**: 877 (100)[M-H_2O-CO_2-H]$^-$, 731 (44)[M-dHex-H_2O-CO_2-H]$^-$, 595 (35)[M-dHex-Hex-2H_2O-H]$^-$, 523 (14)[M-dHex-Hex-108-H]$^-$, 465 (45)
69	dHex-Hex-HexA-Hederagenin	955.49569	$C_{48}H_{76}O_{19}$ (5.114)	**MS^2 [955.49]**: 937 (100)[M-H_2O-H]$^-$, 893 (63)[M-H_2O-CO_2-H]$^-$, 809 (14)[M-dHex-H]$^-$, 765 (23)[M-dHex-CO_2-H]$^-$, 747 (19)[M-dHex-H_2O-CO_2-H]$^-$, 629 (18)[M-dHex-Hex-H_2O-H]$^-$, 611 (19)[M-dHex-Hex-2H_2O-H]$^-$, 539 (21))[M-dHex-Hex-108-H]$^-$, 471 (8)[Agly-H]$^-$, 469 (3)
70	Hex-Bayogenin	649.39741	$C_{36}H_{58}O_{10}$ (2.599)	**MS^2 [649.39]**: 487 (100)[Agly-H]$^-$, 409 (3), 391 (2) **MS^3 [649.39→487]**: 467 (29), 455 (14), 421 (35), 409 (100), 403 (17), 393 (34), 391 (52), 379 (17)
71	Hex-HexA-Hederagenin	809.43534	$C_{42}H_{66}O_{15}$ (3.021)	**MS^2 [809.44]**: 629 (100)[M-Hex-H_2O-H]$^-$, 585 (11)[M-Hex-H_2O-CO_2-H]$^-$, 471 (2)[Agly-H]$^-$ **MS^3 [809.44→629]**: 499 (12), 471 (100)[Agly-H]$^-$, 469 (11), 439 (12) **MS^3 [809.44→585]**: 513 (100), 499 (2), 471 (5)[Agly-H]$^-$
72	Malonyl-Hex-Medicagenic acid	749.37838	$C_{39}H_{58}O_{14}$ (4.003)	**MS^2 [749.38]**: 705 (100)[M-CO_2-H]$^-$

		705.38773*	$C_{38}H_{58}O_{12}$ (3.089)	**MS^2 [705.39]**: 439 (100)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [705.39→439]**: 393 (70), 391 (100), 375 (9)
73	Pen-Hex-HexA-Soyasapogenol B	927.49925	$C_{47}H_{76}O_{18}$ (3.624)	**MS^2 [927.50]**: 909 (100)[M-H_2O-H]$^-$, 865 (14)[M-CO_2-H_2O-H]$^-$, 795 (50)[M-Pen-H]$^-$, 733 (34)[M-Pen-CO_2-H_2O-H]$^-$, 633 (3)[M-Pen-Hex-H]$^-$, 615 (44)[M-Pen-Hex-H_2O-H]$^-$, 597 (65)[M-Pen-Hex-H_2O-H_2O-H]$^-$, 571 (4)[M-Pen-Hex-CO_2-H_2O-H]$^-$, 525 (72)[M-Pen-Hex-108-H]$^-$, 457 (15)[Agly-H]$^-$
74	Hex-HexA-Aglycone A	823.41450	$C_{42}H_{66}O_{16}$ (2.842)	**MS^2 [823.41]**: 643 (100)[M-Hex-H_2O-H]$^-$, 599 (50)[M-Hex-H_2O-CO_2-H]$^-$, 485 (4)[Agly-H]$^-$, 467 (4) **MS^3 [823.41→643]**: 599 (26)[M-Hex-H_2O-CO_2-H]$^-$, 515 (13), 487 (12), 485 (87)[Agly-H]$^-$, 469 (12), 467 (100) **MS^3 [823.41→599]**: 581 (1), 527 (100), 485 (4)[Agly-H]$^-$, 439 (2)
75	Malonyl-Hex-Bayogenin	735.39924 691.40869*	$C_{39}H_{60}O_{13}$ (4.003) $C_{38}H_{60}O_{11}$ (3.477)	**MS^2 [691.41]**: 649 (10)[M-malonyl-H]$^-$, 631 (100)[M-malonyl-H_2O-H]$^-$, 487 (2)[Agly-H]$^-$ **MS^3 [691.41→631]**: 601 (100), 487 (10)[Agly-H]$^-$, 469 (20), 421 (6), 409 (18), 393 (10), 391 (10) **MS^3 [691.41→649]**: 487 (100)[Agly-H]$^-$, 439 (1)
76	Malonyl-Hex-HexA-Hederagenin	895.43874 851.44633*	$C_{45}H_{68}O_{18}$ (6.088) $C_{44}H_{68}O_{16}$ (3.371)	**MS^2 [851.45]**: 809 (1)[M-malonyl-H]$^-$, 791 (2)[M-malonyl-H_2O-H]$^-$, 671 (39)[M-Hex-C_2H_2O-H_2O-H]$^-$, 627 (14)[M-Hex-malonyl-H_2O-H]$^-$, 611 (11)[M-Hex-malonyl-CO_2-H]$^-$, 585 (9), 567 (13), 499 (100), 471 (4)[Agly-H]$^-$
77	3-Rha-Gal-GlcA-Soyasapogenol B	941.51540	$C_{48}H_{78}O_{18}$ (4.101)	**MS^2 [941.52]**: 923 (100)[M-H_2O-H]$^-$, 879 (43)[M-H_2O-CO_2-H]$^-$, 795 (8)[M-Rha-H]$^-$, 751 (5)[M-Rha-CO_2-H]$^-$, 733 (24)[M-Rha-H_2O-CO_2-H]$^-$, 633 (2)[M-Rha-Gal-H]$^-$, 615 (28)[M-Rha-Gal-H_2O-H]$^-$, 597 (17)[M-Rha-Gal-2H_2O-H]$^-$, 525 (25)[M-Rha-Gal-108-H]$^-$, 457 (8)[Agly-H]$^-$ **MS^3 [941.52→923]**: 879 (100)[M-H_2O-CO_2-H]$^-$, 733 (19)[M-Rha-H_2O-CO_2-H]$^-$, 597 (54)[M-Rha-Gal-2H_2O-H]$^-$, 525 (70)[M-Rha-Gal-108-H]$^-$, 457 (4)[Agly-H]$^-$ **MS^3 [941.52→879]**: 733 (100) [M-Rha-H_2O-CO_2-H]$^-$, 457 (7)[Agly-H]$^-$
78	Rha-Gal-GlcA-Soyasapogenol E	939.49939	$C_{48}H_{76}O_{18}$ (3.726)	**MS^2 [939.50]**: 921 (100)[M-H_2O-H]$^-$, 877 (42)[M-H_2O-CO_2-H]$^-$, 793 (8)[M-Rha-H]$^-$, 749 (1)[M-Rha-CO_2-H]$^-$, 731 (41)[M-Rha-H_2O-CO_2-H]$^-$, 613 (47)[M-Rha-Gal-H_2O-H]$^-$, 595 (29)[M-Rha-Gal-2H_2O-H]$^-$, 523 (64)[M-Rha-Gal-108-H]$^-$, 455 (5)[Agly-H]$^-$ **MS^3 [939.50→921]**: 877 (78)[M-H_2O-CO_2-H]$^-$, 731 (26)[M-Rha-H_2O-CO_2-H]$^-$, 595 (36)[M-Rha-Gal-2H_2O-H]$^-$, 523 (100)[M-Rha-Gal-108-H]$^-$, 455 (6)[Agly-H]$^-$ **MS^3 [939.50→523]**: 455 (91)[Agly-H]$^-$, 453 (100), 439 (13)
79	HexA-Hederagenin	647.38199	$C_{36}H_{56}O_{10}$ (3.726)	**MS^2 [647.38]**: 629 (5), 587 (7), 571 (18), 527 (4), 499 (3), 471 (100)[Agly-H]$^-$, 439 (7) **MS^3 [647.38→471]**: 439 (18), 405 (43), 393 (100)

* Ion resulting from in source fragmentation.

HexA, uronic acid, such as glucuronic acid or galacturonic acid; Hex, hexose, such as glucose or galactose; dHex, 6-deoxyhexose, such as rhamnose or fucose; Pen, pentose, such as arabinose or xylose; GlcA, glucuronic acid; Glc, glucose; Gal, galactose; Rha, rhamnose

2.6.Ten different aglycones

One prerequisite we applied to include the detected compounds in the list of tentatively identified saponins was the occurrence of an aglycone ion in the MS^n fragmentation data. As such, we could show that the 79 identified saponins present in the extract of *M. truncatula* hairy roots were glycosides of 10 different aglycones. The structures of the aglycones are given in Figure 4 and the observed aglycone ions, their predicted molecular formulas and tentative identifications are listed in Table 2. For five of the aglycones, i.e., hederagenin, bayogenin, medicagenic acid, and soyasapogenols B and E (Figure 4), saponins containing the aglycone have been described before (Huhman and Sumner, 2002; Kapusta et al., 2005a; Kapusta et al., 2005b; Huhman et al., 2005). A sixth aglycone yielded an $[Agly-H]^-$ ion at *m*/*z* 473 (Figure 5A). Although this aglycone was only detected as a (grand)daughter ion with nominal *m*/*z* values, its molecular formula could be predicted by the accurate prediction of the molecular formula of the parent ion and the observed loss of sugar residues. With this method, the molecular formula was predicted to be $C_{30}H_{50}O_4$, which corresponds to the molecular formula of soyasapogenol A. No saponins of *M. truncatula* that contain soyasapogenol A as the aglycone had been described, but the compound had been shown to occur in *M. truncatula* as a sapogenin (Confalonieri et al., 2009). Hence, the two saponins containing this aglycone were tentatively identified as soyasapogenol A glycosides.

Table 2. Observed aglycones in *M. truncatula*.

m/*z* [Observed ion]	Molecular Formula	Tentative identification
455 $[Agly-H]^-$	$C_{30}H_{48}O_3$	Soyasapogenol E
457 $[Agly-H]^-$	$C_{30}H_{50}O_3$	Soyasapogenol B
469 $[Agly-H]^-$	$C_{30}H_{46}O_4$	Aglycone B (3β-hydroxy-23-oxo-olean-12-en-28-oic acid)
471 $[Agly-H]^-$	$C_{30}H_{48}O_4$	Hederagenin
473 $[Agly-H]^-$	$C_{30}H_{50}O_4$	Soyasapogenol A
485 $[Agly-H]^-$	$C_{30}H_{46}O_5$	Aglycone A (2β,3β-dihydroxy-23-oxo-olean-12-en-28-oic acid)
487 $[Agly-H]^-$	$C_{30}H_{48}O_5$	Bayogenin
489 $[Agly-H]^-$	$C_{30}H_{50}O_5$	Aglycone D
437 $[Agly-CO_2-H_2O-H]^-$	$C_{30}H_{44}O_6$	Aglycone C
439 $[Agly-CO_2-H_2O-H]^-$	$C_{30}H_{46}O_6$	Medicagenic acid

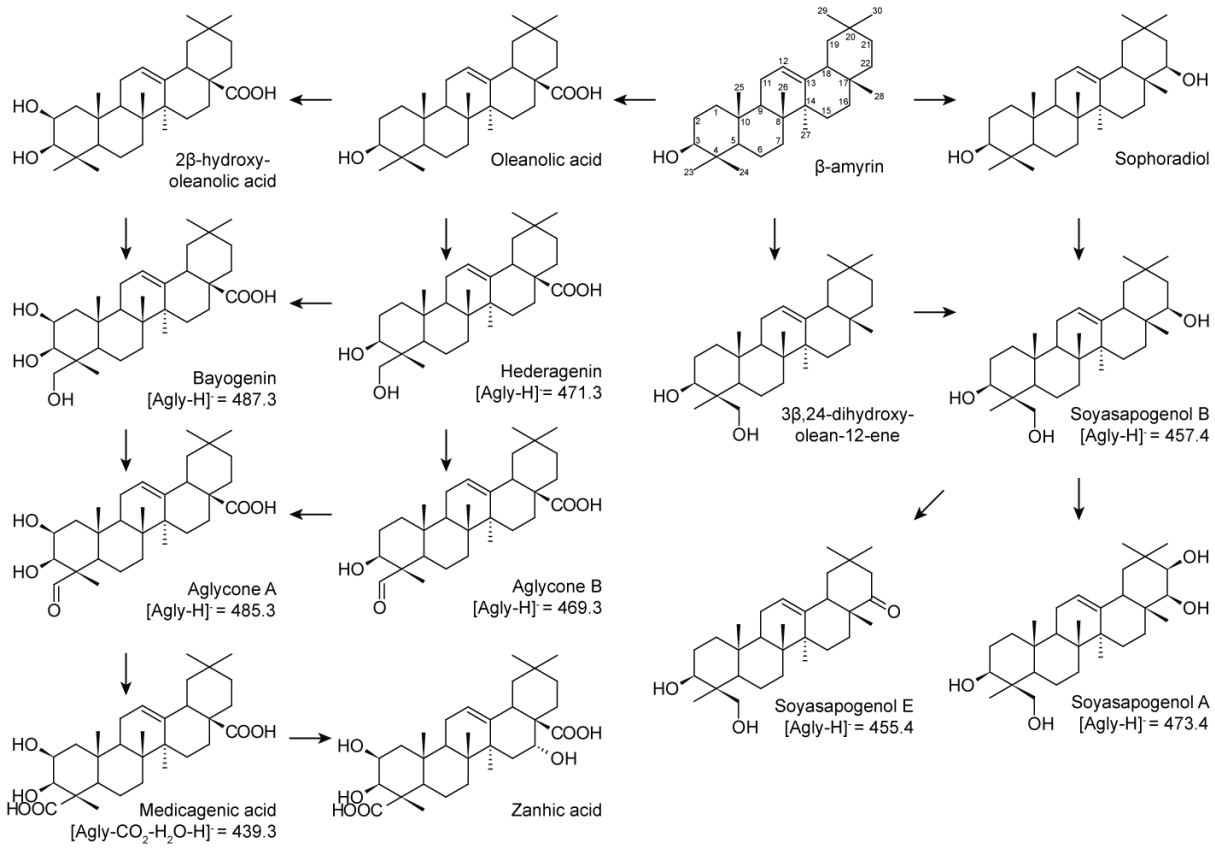


Figure 4. Theoretical biosynthetic pathway of sapogenins in *M. truncatula* and *m*/*z* values of the observed aglycones in the hairy roots extract.

Furthermore, four other, unknown, aglycones were observed. Aglycone A was shown to occur in 10 of the tentatively identified saponins. It had an observed $[Agly\text{-}H]^-$ ion at *m*/*z* 485 (Figure 5C), and its molecular formula was calculated to be $C_{30}H_{46}O_5$. An aglycone with this mass had been previously observed in *M. truncatula*, but its structure was not elucidated (Huhman and Sumner, 2002; Huhman et al., 2005; Naoumkina et al., 2010). Only one sapogenin with this molecular formula is known to occur in the genus *Medicago*, namely 2*β*,3*β*-dihydroxy-23-oxo-olean-12-en-28-oic acid. This aglycone contains C-23 formyl group, which is thought to be the biosynthetic intermediate in the enzymatic oxidation of the C-23 hydroxy group of bayogenin, eventually leading to the carboxylation of medicagenic acid (Figure 4) (Tava et al., 2010). Thus, compounds containing aglycone A as the sapogenin could be 2*β*,3*β*-dihydroxy-23-oxo-olean-12-en-28-oic acid glycosides. Another identified saponin yielded an aglycone anion at *m*/*z* 469 (Figure 5D) and a predicted molecular formula of $C_{30}H_{46}O_4$ (aglycone B). Similar to aglycone A, this compound may contain a C-23 formyl group and may be the result of an enzymatic oxidation of hederagenin, leading to 3*β*-hydroxy-23-oxo-olean-12-en-28-oic acid (Figure 4).

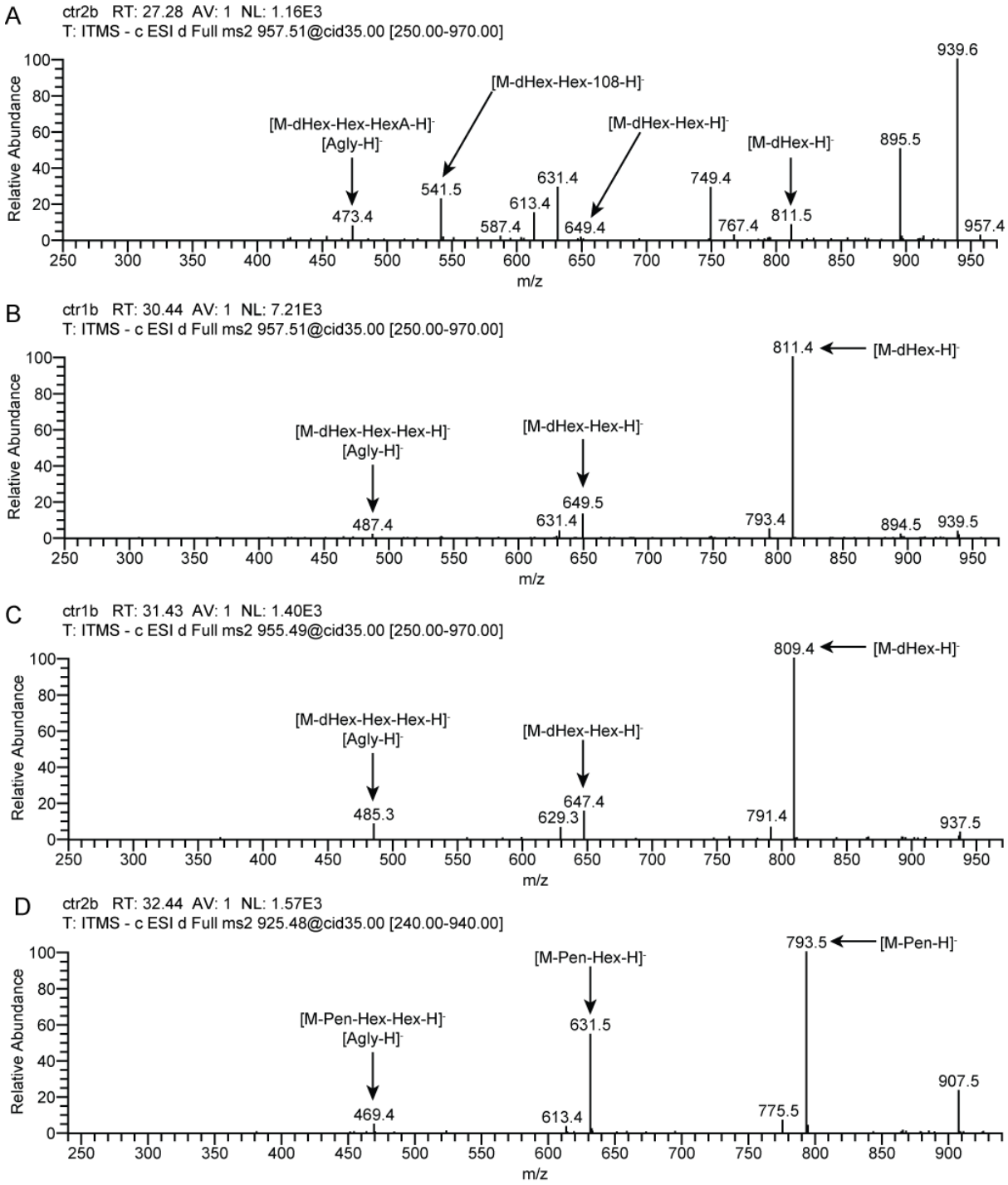


Figure 5. LC ESI FT-ICR MS chromatograms leading to the detection of the various aglycones. (A) MS^2 fragmentation of dHex-Hex-HexA-soyasapogenol A. (B) MS^2 fragmentation of dHex-Hex-Hex-bayogenin. (C) MS^2 fragmentation of dHex-Hex-Hex-aglycone A. (D) MS^2 fragmentation of Pen-Hex-Hex-aglycone B.

Aglycone C, a third unknown aglycone encountered in one of the saponins of *M. truncatula* hairy roots, yielded a first product ion at *m*/*z* 437. Similar to medicagenic acid, this ion is the result of the decarboxylation and dehydration of aglycone C, which has a calculated molecular formula of $C_{30}H_{44}O_6$. Decarboxylation and dehydration of aglycones are observed

only in medicagenic acid (Huhman and Sumner, 2002) and zanhic acid (Kapusta et al., 2005b) and, thus, could be related to the presence of a C-23 carboxylic group. Consequently, we assume that aglycone C contains a C-23 carboxylic group. Two saponins contain an aglycone (aglycone D) yielding an anion at *m*/*z* 489 (Figure 1C) and a calculated molecular formula of $C_{30}H_{50}O_5$. No aglycones with this molecular formula have been described in *Medicago* to date, but, assuming a *β*-amyrin-derived triterpene skeleton, the molecular formula points toward an aglycone containing five hydroxy groups. Thus, we postulate that aglycone D could be an oxidation product of soyasapogenol A. However, since we cannot postulate a structure based on these assumptions, we did not include the aglycones C and D in Figure 4.

3. Experimental section

3.1. Generation and cultivation of transgenic *M. truncatula* hairy roots

A. rhizogenes-mediated transformation of *M. truncatula* (ecotype Jemalong J5) hairy roots was carried out as reported (Boisson-Dernier et al., 2001) with the following modifications. After scarification by treatment with sulfuric acid for 5 min, seeds were surface sterilized with 12% NaOCl during 2 min and washed with sterile H_2O. Subsequently, seeds were treated with 1 μM 6-benzylaminopurine for 3 h and thereafter allowed to germinate on wet, sterile 3 mm Whatman paper at room temperature in the dark. After 2 days of germination, the seedlings were wounded by cutting approximately 2 mm from the root tip of the radicle. The wounded seedlings were infected with *A. rhizogenes* and placed on slanted agar plates containing Murashige and Skoog (MS) medium (pH 5.8) supplemented with vitamins (Duchefa). The plates were sealed with micropore tape and placed vertically. Cocultivation was allowed for 10 days under a 16 h/8 h light/dark regime at 22°C. After 10 days, the plants were transferred to new plates, containing 100 mg/L cefotaxime to prevent *A. rhizogenes* growth, and incubated under identical conditions. After 10 days, hairy roots were excised from the plants and transferred to liquid MS medium (pH 5.8) supplemented with vitamins, 1% (w/v) sucrose, and 100 mg/L cefotaxime to eliminate *A. rhizogenes* contamination. The hairy roots were grown for 7 days in the dark at 24°C and shaking at 300 rpm. Subsequently, the roots were transferred to horizontal Petri dishes containing solid MS medium (pH 5.8) supplemented with 1% (w/v) sucrose and 100 mg/L cefotaxime and grown in the dark at 24°C. After 3 weeks, young hairy root tissue was subcultured to solid medium without antibiotics. For maintenance, the axenic hairy roots cultures were subcultured every 3 weeks

onto fresh plates. Hairy roots for metabolic profiling were grown for 21 days in liquid MS medium (pH 5.8) supplemented with vitamins and 1% (w/v) sucrose.

3.2.Metabolite extractions

M. truncatula hairy roots, grown for 21 days in liquid medium, were harvested and rinsed with purified H_2O under vacuum filtration. Subsequently, the roots were frozen and ground in liquid N_2. Of the ground material, 400 mg was extracted with 1 mL of MeOH at room temperature for 10 min, followed by centrifugation for 10 min at 20800*g*. Under vacuum, 500 μL of the supernatant was evaporated to dryness, and the residue was dissolved in 600 μL of H_2O/cyclohexane (2:1, v/v). Samples were centrifuged (10 min at 20800*g*), and 200 μL of the aqueous phase was retained for analysis.

3.3.LC ESI FT-ICR MS

For reversed-phase LC, an Acquity UPLC BEH C18 column (150 x 2.1 mm, 1.7 μm; Waters, Milford, MA) was serially coupled to an Acquity UPLC BEH C18 column (100 x 2.1 mm, 1.7 μm) and mounted on an ultra-high-performance LC system consisting of a Accela pump (Thermo Electron Corporation, Waltham, MA, USA) and Accela autosampler (Thermo Electron Corporation). The Accela LC system was hyphenated to a LTQ FT Ultra (Thermo Electron Corporation) via an electrospray ionization source. The following gradient was run using water:MeCN (99:1, v:v) acidified with 0.1% (v:v) HOAc (solvent A) and MeCN:water (99:1, v:v) acidified with 0.1% (v:v) HOAc (solvent B): time 0 min, 5% B; 30 min, 55% B; 35 min, 100% B. The loop size, flow, and column temperature were 25 μL, 300 μL/min and 80°C, respectively. Full loop injection was applied. Negative ionization was obtained with the following parameter values: capillary temperature 150°C, sheath gas 25 (arbitrary units), aux. gas 3 (arbitrary units), and spray voltage 4.5 kV. Full FT-MS spectra between *m*/*z* 120–1400 were recorded at a resolution of 100,000. For identification, full MS spectra were interchanged with a dependent MS^2 scan event in which the most abundant ion in the previous full MS scan was fragmented, and two dependent MS^3 scan events in which the two most abundant daughter ions were fragmented. The collision energy was set at 35%.

Acknowledgements

The authors thank Martine De Cock for help in preparing the manuscript. This work was supported by the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" project SBO040093).

References

Araújo, S. D., Duque, A. S. R. L. A., dos Santos, D. M. M. F., and Fevereiro, M. P. S. (2004). An efficient transformation method to regenerate a high number of transgenic plants using a new embryogenic line of *Medicago truncatula* cv. Jemalong. Plant Cell Tiss Organ Cult **78**, 123-131.

Boisson-Dernier, A., Chabaud, M., Garcia, F., Bécard, G., Rosenberg, C., and Barker, D. G. (2001). *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. Mol Plant Microbe Interact **14**, 695-700.

Chabaud, M., Larsonneau, C., Marmouget, C., and Huguet, T. (1996). Transformation of barrel medic (*Medicago truncatula* Gaertn) by *Agrobacterium tumefaciens* and regeneration *via* somatic embryogenesis of transgenic plants with the *MtENOD12* nodulin promoter fused to the *gus* reporter gene. Plant Cell Rep **15**, 305-310.

Confalonieri, M., Cammareri, M., Biazzi, E., Pecchia, P., Fevereiro, M. P. S., Balestrazzi, A., Tava, A., and Conicella, C. (2009). Enhanced triterpene saponin biosynthesis and root nodulation in transgenic barrel medic (*Medicago truncatula* Gaertn.) expressing a novel β-amyrin synthase (*AsOXA1*) gene. Plant Biotechnol J **7**, 172-182.

Cook, D.R. (1999). *Medicago truncatula* - a model in the making! Curr Opin Plant Biol **2**, 301-304

Dixon, R.A., and Sumner, L.W. (2003). Legume natural products: Understanding and manipulating complex pathways for human and animal health. Plant Physiol **131**, 878-885.

Dixon, R.A., and Pasinetti, G.M. (2010). Flavonoids and isoflavonoids: From plant biology to agriculture and neuroscience. Plant Physiol **154**, 453-457.

Farag, M.A., Huhman, D.V., Lei, Z.T., and Sumner, L.W. (2007). Metabolic profiling and systematic identification of flavonoids and isoflavonoids in roots and cell suspension cultures of *Medicago truncatula* using HPLC-UV-ESI-MS and GC-MS. Phytochemistry **68**, 342-354

Farag, M.A., Deavours, B.E., de Fátima, A., Naoumkina, M., Dixon, R.A., and Sumner, L.W. (2009). Integrated metabolite and transcript profiling identify a biosynthetic mechanism for hispidol in *Medicago truncatula* cell cultures. Plant Physiol **151**, 1096-1113.

Georgiev, M.I., Pavlov, A.I., and Bley, T. (2007). Hairy root type plant in vitro systems as sources of bioactive substances. Appl Microbiol Biotechnol **74**, 1175-1185.

Guillon, S., Trémouillaux-Guiller, J., Pati, P.K., Rideau, M., and Gantet, P. (2006a). Harnessing the potential of hairy roots: dawn of a new era. Trends Biotechnol **24**, 403-409.

Guillon, S., Trémouillaux-Guiller, J., Pati, P.K., Rideau, M., and Gantet, P. (2006b). Hairy root research: recent scenario and exciting prospects. Curr Opin Plant Biol **9**, 341-346.

Hu, Z.B., and Du, M. (2006). Hairy root and its application in plant genetic engineering. J Integr Plant Biol **48**, 121-127.

Huhman, D.V., and Sumner, L.W. (2002). Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry **59**, 347-360.

Huhman, D.V., Berhow, M.A., and Sumner, L.W. (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. J Agric Food Chem **53**, 1914-1920.

Jin, M., Yang, Y., Su, B., and Ren, Q. (2007). Determination of soyasaponins Ba and Bb in human serum by high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry. J Chromatogr B **846**, 169-175.

Kapusta, I., Janda, B., Stochmal, A., and Oleszek, W. (2005a). Determination of saponins in aerial parts of barrel medic (*Medicago truncatula*) by liquid chromatography-electrospray ionization/mass spectrometry. J Agric Food Chem **53**, 7654-7660.

Kapusta, I., Stochmal, A., Perrone, A., Piacente, S., Pizza, C., and Oleszek, W. (2005b). Triterpene saponins from barrel medic (*Medicago truncatula*) aerial parts. J Agric Food Chem **53**, 2164-2170.

Levigne, S.V., Ralet, M.C.J., Quéméner, B.C., Pollet, B.N.L., Lapierre, C., and Thibault, J.F.J. (2004). Isolation from sugar beet cell walls of Arabinan oligosaccharides esterified by two ferulic acid monomers. Plant Physiol **134**, 1173-1180.

Liu, S.Y., Cui, M., Liu, Z.Q., Song, F.R., and Mo, W.J. (2004). Structural analysis of saponins from medicinal herbs using electrospray ionization tandem mass spectrometry. J Am Soc Mass Spectrom **15**, 133-141.

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

Quéméner, B., and Ralet, M.C. (2004). Evidence for linkage position determination in known feruloylated mono- and disaccharides using electrospray ion trap mass spectrometry. J Mass Spectrom **39**, 1153-1160.

Schliemann, W., Ammer, C., and Strack, D. (2008). Metabolite profiling of mycorrhizal roots of *Medicago truncatula*. Phytochemistry **69**, 112-146.

Sevón, N., and Oksman-Caldentey, K.M. (2002). *Agrobacterium rhizogenes*-mediated transformation: Root cultures as a source of alkaloids. Planta Med **68**, 859-868.

Sparg, S.G., Light, M.E., and van Staden, J. (2004). Biological activities and distribution of plant saponins. J Ethnopharmacol **94**, 219-243.

Sun, H.X., Xie, Y., and Ye, Y.P. (2009). Advances in saponin-based adjuvants. Vaccine **27**, 1787-1796.

Tava, A., and Avato, P. (2006). Chemical and biological activity of triterpene saponins from *Medicago* species. Nat Prod Comm **1**, 1159-1180.

Tava, A., Scotti, C., and Avato, P. (2010). Biosynthesis of saponins in the genus *Medicago*. Phytochem Rev **in press**, doi: 10.1007/s11101-010-9169-x.

Trieu, A.T., Burleigh, S.H., Kardailsky, I.V., Maldonado-Mendoza, I.E., Versaw, W.K., Blaylock, L.A., Shin, H.S., Chiou, T.J., Katagi, H., Dewbre, G.R., Weigel, D., and Harrison, M.J. (2000). Transformation of *Medicago truncatula* via infiltration of seedlings or flowering plants with *Agrobacterium*. Plant J **22**, 531-541.

Vanholme, R., Ralph, J., Akiyama, T., Lu, F.C., Pazo, J.R., Kim, H., Christensen, J.H., Van Reusel, B., Storme, V., De Rycke, R., Rohde, A., Morreel, K., Boerjan, W. (2010). Engineering traditional monolignols out of lignin by concomitant up-regulation of *F5H1* and down-regulation of *COMT* in Arabidopsis. Plant J. **64**, 885-897.

Chapter 5

Combinatorial biochemistry of triterpene saponins in plants.

Jacob Pollier*, Miguel González-Guzmán, Kris Morreel, Lander Ingelbrecht, and Alain Goossens

* Author contribution: molecular cloning, generation of *M. truncatula* hairy roots, RT-PCR transcript profiling, metabolite profiling, data interpretation, and writing of the manuscript.

Abstract

Plants are a rich source of natural products that feed various drug discovery programs. However, there is an ever increasing demand for novel compounds, due to, amongst others, the growing drug tolerance and resistance in microorganisms and newly emerging diseases. In microorganisms, combinatorial biosynthesis is a widely used tool to increase structural variation in several classes of (microbial) natural products. Because of the more complex genetic make-up of plants and the lack of tools to manipulate them, combinatorial biosynthesis in plants is still in its infancy. The recent technical advances in gene discovery and functional genomics have brought plants within the scope of combinatorial biosynthesis. Here, we demonstrate the establishment of a combinatorial biosynthesis platform in plants, with the metabolite class of the triterpene saponins as the target. By ectopic expression of candidate saponin biosynthesis genes of several medicinal plants in *M. truncatula* hairy roots, transgenic lines were generated, some of which accumulated a set of novel compounds. Structural elucidation of the novel compounds revealed that new saponin aglycones were obtained for oleanane- and dammarane-type saponins were obtained.

1. Introduction

Plants are an enormously rich source of natural products that are used in a variety of therapeutics and in various drug discovery programs (Li and Vederas, 2009). Due to problems such as growing drug tolerance and resistance in microorganisms and newly emerging diseases, the demand for a continuous supply of novel molecules with novel or superior pharmaceutical activities is ever increasing. Screening of medicinal plant extracts may lead to the discovery of novel compounds, however, these methods are often time consuming and expensive (Littleton et al., 2005). Hence, despite the potential of plants in terms of novel compounds, the pharmaceutical industry has shifted focus to the high throughput screening of combinatorial chemistry libraries. This shift may have been premature, since the output in terms of new drugs is far below the initial expectations (Ortholand and Ganesan, 2004).

An alternative approach to generate new molecules is combinatorial biosynthesis, that can be defined as an approach in which gene products from different organisms are combined to produce new bioactive compounds. In microorganisms, various combinatorial biosynthesis techniques have been successfully applied to increase structural variation in several classes of microbial natural products, such as polyketides and indolocarbazoles (Floss, 2006; Sánchez et al., 2006; Lam, 2007; Menzella and Reeves, 2007; Zhou et al., 2008; Salas and Méndez, 2009). Until recently, due to their more complex genetic makeup and the lack of tools to manipulate them, plants remained out of scope of combinatorial biosynthesis. However, with the recent technical advances in gene discovery and functional genomics, the large-scale implementation of different combinatorial biosynthesis techniques in plants may become feasible. Indeed, some pioneering studies have already shown the feasibility of this approach. For instance, RNA silencing of tryptophan decarboxylase in *Catharanthus roseus* hairy roots effectively blocked the biosynthesis of tryptamine in the transgenic roots. As a consequence, accumulation of terpenoid indole alkaloids (TIAs), a class of pharmacologically important secondary metabolites for which tryptamine is used as a building block, was eliminated in these cultures. Subsequent feeding of these ‘mutant’ hairy root cultures with an unnatural analog of tryptamine yielded several novel TIAs. Importantly, the new compounds were not ‘contaminated’ by the TIAs naturally present in ‘wild-type’ *C. roseus* (Runguphan et al., 2009). Furthermore, by introducing prokaryotic halogenases that chlorinate the indole ring of tryptophan into *C. roseus* hairy roots, chlorinated tryptophan was produced *in planta*, and

used by the downstream TIA biosynthesis enzymes to produce halogenated alkaloids *in planta* (Runguphan et al., 2010). The latter pioneering study can be considered as the first example of combinatorial biosynthesis in plants and underscores its huge potential.

Here, we successfully demonstrate the establishment of a combinatorial biosynthesis platform in plants, with the secondary metabolite class of the triterpene saponins as the target. To this end, candidate saponin biosynthesis genes of different medicinal plants were expressed in hairy roots of the saponin-producing model legume *Medicago truncatula*. Several of the introduced enzymes proved to be functional and generated substrates in the transgenic hairy roots that are normally not present in *M. truncatula*. Furthermore, these substrates were incorporated by downstream biosynthesis enzymes, and ultimately led to the accumulation of novel compounds in the transgenic *M. truncatula* hairy roots. Structural elucidation of the novel compounds revealed that new aglycones were obtained for oleanane-type saponins and that new, dammarane-type saponins, a class of pharmacologically important saponins naturally restricted to the genus *Panax*, were produced in the *M. truncatula* hairy roots. And although the target of the study was triterpene saponins, we also detected novel, non-saponin compounds in one of the overexpression lines, indicating that the applied method is applicable to generate structural variation in other classes of plant natural products as well.

2. Results

2.1. Heterologous expression of 52 candidate genes in *M. truncatula* hairy roots

From the genome-wide cDNA-AFLP transcript profiling of MeJA treated plants or plant cell cultures, a list of 259 MeJA responsive gene tags were selected as candidate genes involved in saponin biosynthesis or its regulation (see chapter 2). Subsequently, the FL-ORFs of these candidate genes were cloned (*M. truncatula*) or screened for in a cDNA-library (*M. lanceolata*, *P. ginseng*, *G. glabra*, and *B. falcatum*), leading to a set of 140 FL genes consisting of candidate genes involved in saponin biosynthesis or its regulation (see chapter 3). To obtain proof of concept of combinatorial biosynthesis of triterpene saponins in plants, a subset of this list, 52 genes, were selected to be transformed into *M. truncatula* hairy roots. This list consisted solely of candidate biosynthesis genes from *P. ginseng* (20), *M. lanceolata* (24), and *B. falcatum* (8). CaMV 35S-mediated overexpression constructs of the FL-ORFs of the candidate genes were transformed into *M. truncatula* hairy roots, and for each construct, hairy roots of a minimum of 5 independent transformation events were

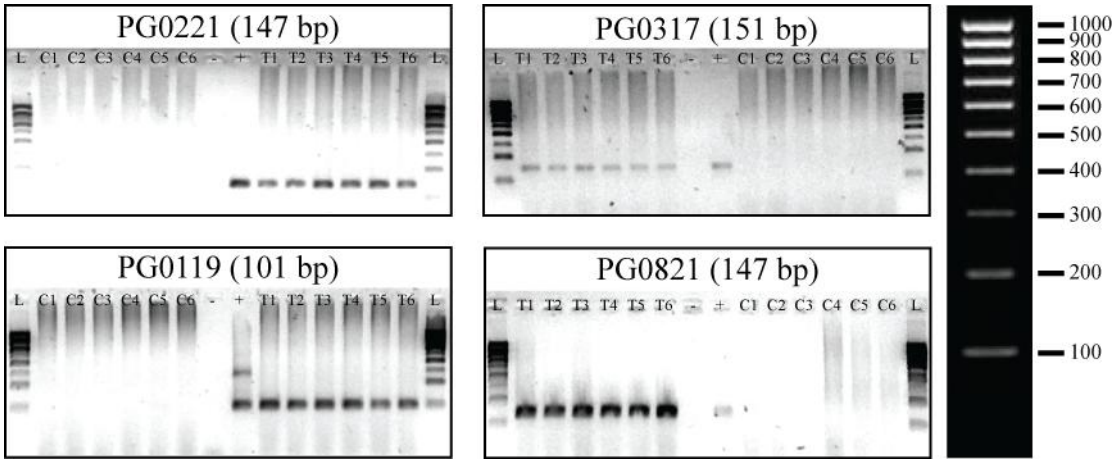


Figure 1. RT-PCR screening of *M. truncatula* hairy roots for the presence of the transgene in *PG0221*, *PG0119*, *PG0317*, and *PG0821* overexpression lines. The expected fragment size is indicated between brackets. As positive (+) and negative (-) controls, the cDNA-AFLP fragment and water were used, respectively. T1-T6, transgenic lines; C1-C6, control lines; L, DNA ladder (SmartLadder, Eurogentec), the size of the fragments of the ladder are indicated on the right.

obtained. Subsequently, the obtained hairy roots were screened for the presence of the transgene by RT-PCR (Figure 1). As such, for all of the 52 genes, at least three independent hairy root lines expressing the transgene were obtained.

2.2.Metabolite profiling reveals the presence of dammarane-type triterpene saponins in transgenic hairy roots expressing *dammarenediol synthase* of *P. ginseng*

Since all saponins of *M. truncatula* possess an oleanane-type triterpene backbone derived from β-amyrin (Tava et al., 2010), overexpression of the gene encoding dammarenediol-II synthase (PgDDS) of *P. ginseng* (Tansakul et al., 2006), may lead to the production of dammarenediol-II in the *M. truncatula* hairy roots. Subsequently, this compound may be incorporated in the natural *M. truncatula* saponin biosynthesis, and as such lead to a new class of aglycones in *M. truncatula* (Figure 2). Furthermore, oxidation of the triterpene backbone in *P. ginseng* and *M. truncatula* occurs on different positions of the backbone, and thus, compounds new to both parental species may be produced in the transgenic lines.

In our genome-wide transcript transcript profiling, a gene tag corresponding to *PgDDS* (PG0408) was encountered (see chapter 2). Subsequent cloning of the FL-ORF of the gene (see chapter 3) allowed for ectopic expression of *PgDDS* in *M. truncatula* hairy roots. After confirmation of overexpression with RT-PCR (Figure 2B), five biological repeats of three independent transgenic and control lines were profiled with liquid chromatography

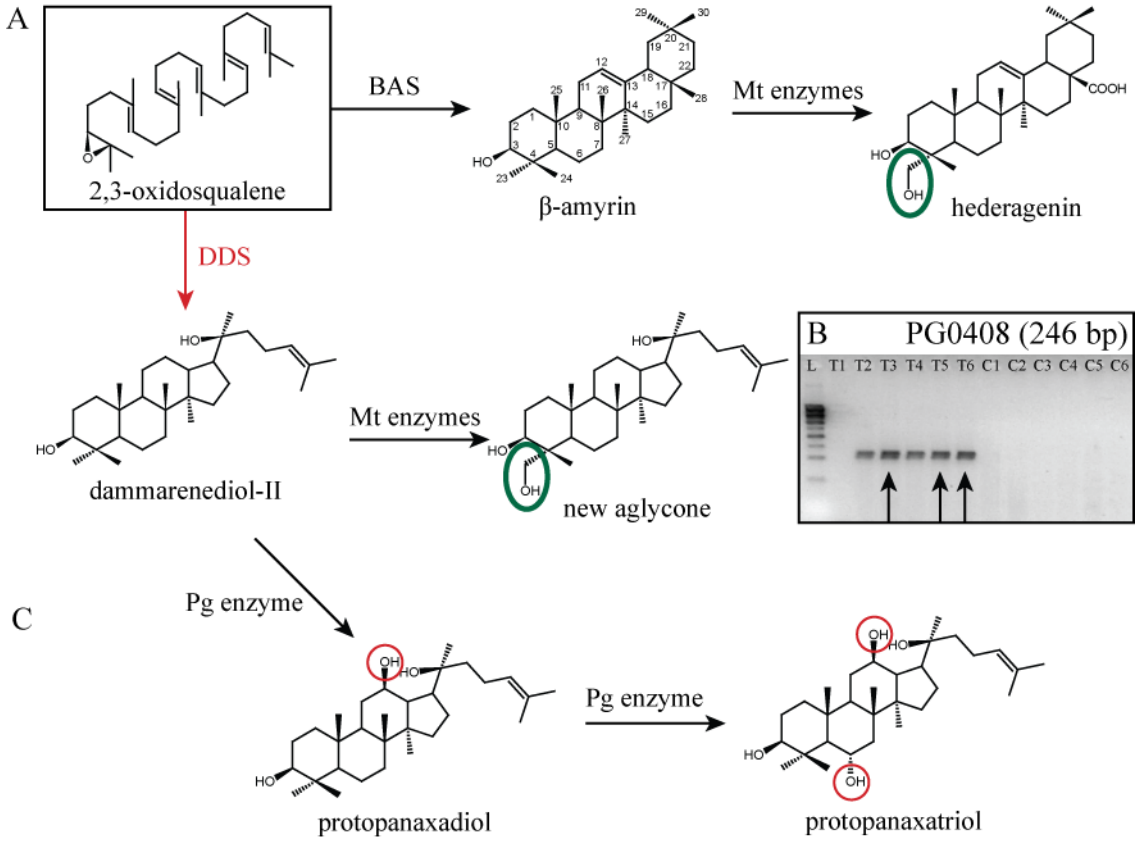


Figure 2. Overexpression of *PgDDS* in *M. truncatula* hairy roots. A. Presence of PgDDS could lead to a new class of aglycones in *M. truncatula* hairy roots. B. RT-PCR screening of *M. truncatula* hairy roots for the presence of the transgene in *PG0408* overexpression lines. The lines indicated with an arrow were selected for further analysis. T1-T6, transgenic lines; C1-C6, control lines; L, DNA ladder. C. Biosynthesis of protopanaxatriol in *P. ginseng*.

electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (LC-ESI-FT-ICRMS) to compare the saponin composition of overexpression roots with that of control roots. Comparative analysis of the root extracts of control and PG0408 overexpression roots yielded a total of 19,778 *m/z* peaks. Notably, 280 discrete peaks between PG0408 overexpression roots and control roots were observed, corresponding to 39 discretely present compounds. These compounds were present in all of the PG0408 overexpression lines, whereas they were absent in all of the control lines. To (tentatively) identify the new compounds, the accurate *m/z* measurements of the FT-ICR MS were combined with the isotope abundances to predict the molecular formula of the detected compounds. The structural information needed for the identification of the saponins was obtained by MS^n fragmentation using the linear ion trap (IT). Full FT-MS spectra were interchanged with dependent IT-MS^n scan events, existing of a MS^2 and two MS^3 scans. In the MS^2 scan, the most abundant ion in the previous full MS scan was fragmented. The two

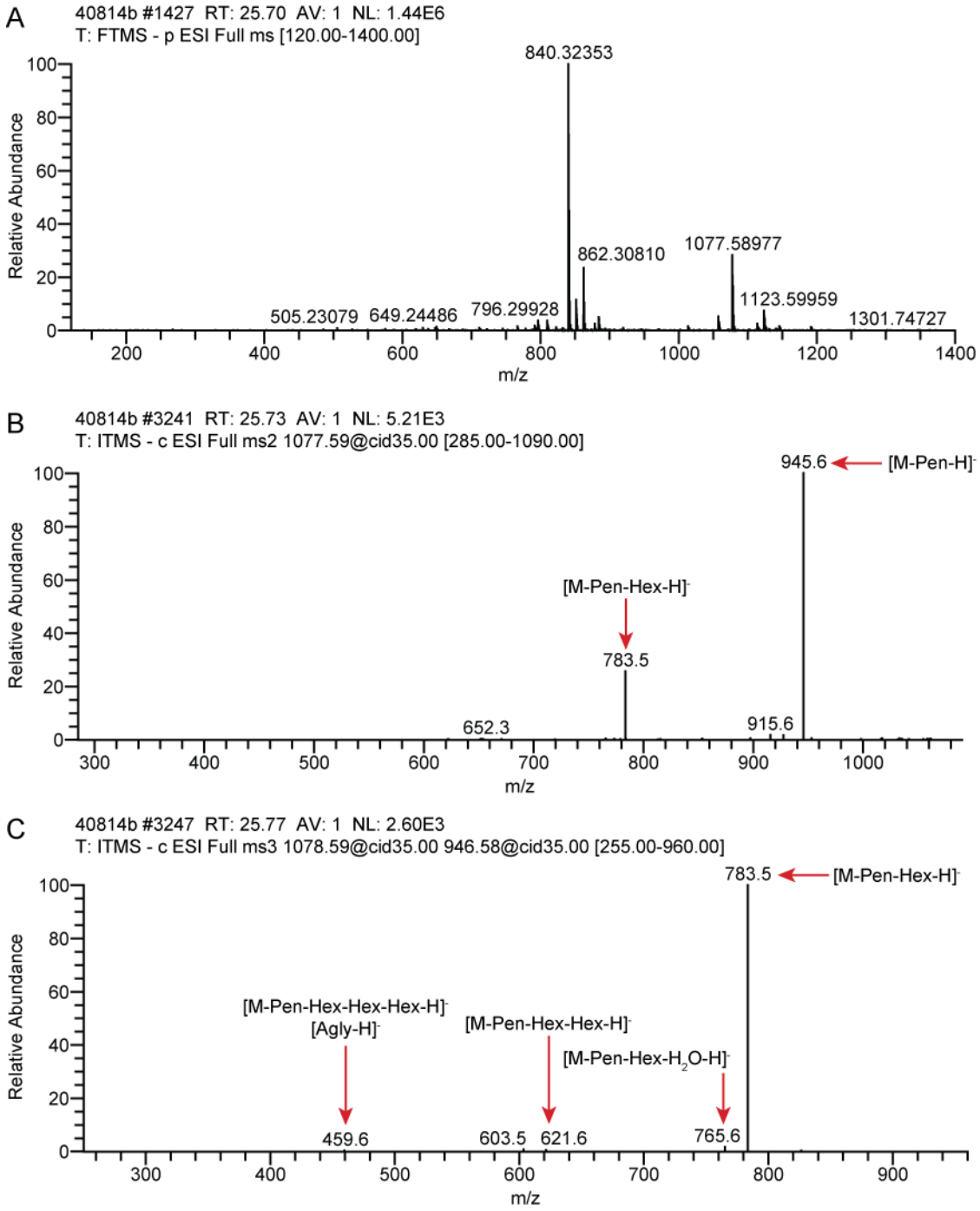


Figure 3. LC ESI FT-ICR MS chromatograms. (A) MS scan of peak at t_R 25.70 min. (B) MS^2 fragmentation of the $[M-H]^-$ ion at m/z 1077.59 (C) MS^3 fragmentation of the most abundant MS^2 daughter ion of the $[M-H]^-$ ion at m/z 1077.59.

most abundant daughter ions in every MS^2 scan event were subjected to a MS^3 scan event. As such, for some of the most abundant new compounds, the aglycone backbone could be determined. For instance, the new compound at retention time (t_R) 25.70 yields a $[M-H]^-$ anion at *m/z* 1077.58977 (Figure 3A). Because saponins consist of the elements C, H, and O, four different molecular formulas are possible within an error range of 5 ppm: $C_{71}H_{81}O_9$

(δ ppm = 1.079), $C_{46}H_{93}O_{27}$ (δ ppm = -1.114), $C_{53}H_{89}O_{22}$ (δ ppm = 4.336), and $C_{64}H_{85}O_{14}$ (δ ppm = -4.371). To determine the correct molecular formula, the abundance of the isotopes was considered. The predicted relative abundance of the M+1 ions of the four possible molecular formulas were 80.37%, 53.48%, 61.00%, and 72.85%, respectively. For the unknown saponin, the observed relative abundance of the M+1 ion was 60.74%, indicating that the molecular formula corresponding to the $[M-H]^-$ anion is $C_{53}H_{89}O_{22}$ (δ ppm = 4.336). The sugar residues and the aglycone could be identified from the MS^n spectra. For this saponin, MS^2 fragmentation led to the generation of 2 major daughter ions, the first, at *m*/*z* 945, resulting from the loss of a pentose (-132 Da) residue and a second, at *m*/*z* 783 from the loss of the pentose and an extra hexose (-162 Da) residue (Figure 3B). MS^3 fragmentation of the most abundant MS^2 daughter ion ($[M-Pen-H]^-$) led to the generation of 5 daughter ions, the smallest of which, at *m*/*z* 459, represents the aglycone ion, $[Agly-H]^-$, resulting from the loss of a pentose and three hexose moieties (Figure 3C). Although this aglycone was only detected as a granddaughter ion with nominal *m*/*z* values, its molecular formula could be predicted by the accurate prediction of the molecular formula of the parent ion and the observed loss of sugar residues. With this method, the molecular formula of the aglycone was predicted to be $C_{30}H_{52}O_3$. In *M. truncatula*, no aglycones with this molecular formula have been described before, however, in *P. ginseng* a dammarane-type triterpene sapogenin with the same molecular formula, protopanaxadiol, had been described (Kim et al., 2009; Lü et al., 2009). This aglycone has three hydroxyl groups, two of which are generated by PgDDS, and a third by an unknown *P. ginseng* enzyme (Figure 2) (Yue et al., 2008). In *M. truncatula*,

Table 1. Tentatively annotated new compounds in hairy roots overexpressing PG0408.

Cpd #	$[M-H]^-$	Formula (δ ppm)	Aglycone	Decoration
4	1077.58977	$C_{53}H_{89}O_{22}$ (4.336)	$C_{30}H_{52}O_3$ hydroxydammarenediol	Pen-Hex-Hex-Hex-
7	809.47104	$C_{43}H_{70}O_{14}$ (2.174)	$C_{30}H_{52}O_2$ dammarenediol	Malonyl-Hex-Hex-
12	929.55140	$C_{48}H_{82}O_{17}$ (3.739)	$C_{30}H_{52}O_3$ hydroxydammarenediol	dHex-Hex-Hex-
15	885.48794	$C_{45}H_{74}O_{17}$ (2.955)	$C_{30}H_{52}O_4$ dihydroxydammarenediol	Malonyl-Hex-Hex-
16	929.51545	$C_{47}H_{78}O_{18}$ (4.208)	$C_{30}H_{50}O_4$ carboxydammarenediol	Pen-Hex-Hex-
34	933.54564	$C_{47}H_{82}O_{18}$ (3.001)	$C_{30}H_{54}O_4$ 25-hydroxyprotopanaxadiol	Pen-Hex-Hex-

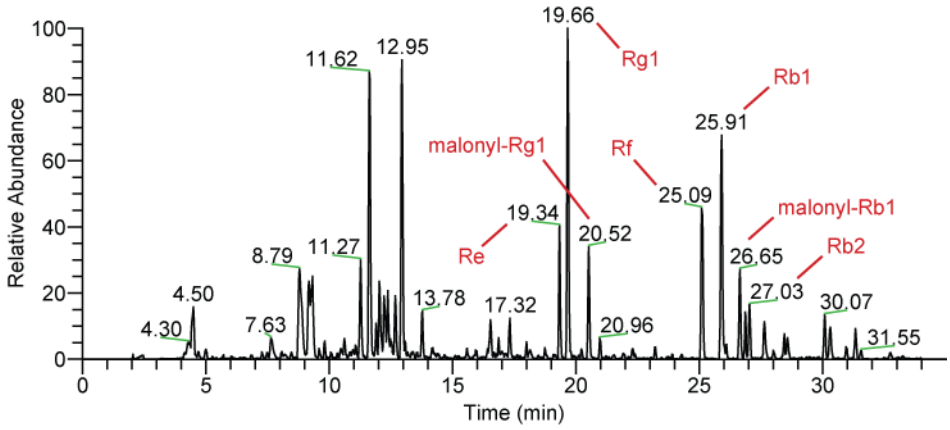


Figure 4. Chromatogram of a *P. ginseng* extract analyzed with the *M. truncatula* method. Several of the ginsenosides described in the literature are annotated in the chromatogram.

however, oxidation at the same position of the oleanane-type triterpene sapogenins is not occurring. Hence, the new aglycone may be dammarenediol-II, with an extra hydroxyl group. Analogously, for several of the other new compounds present in the metabolite extract of PG0408 overexpression lines, the aglycone could be determined via a similar analysis. All of them could be tentatively annotated as dammarane-type aglycones (Table 1).

In a strict sense, for combinatorial biosynthesis the produced compounds should be truly novel and not just ginsenosides that are also naturally produced in wild-type *P. ginseng* roots. Hence, also a *P. ginseng* sample was extracted and analyzed with the *M. truncatula* profiling method, and several of the major compounds of the extract were identified based on their similar masses and fragmentation pattern of *P. ginseng* saponins described in the literature (Figure 4). Comparison of the retention time, masses and the fragmentation patterns of dammarenediol derived compounds in the *M. truncatula* PG408 line, with the ones in the *P. ginseng* extract revealed that the compounds that are occurring in the transgenic *M. truncatula* hairy roots are not present in either the *M. truncatula* control roots nor the *P. ginseng* sample. Compound 15, for instance, has a predicted molecular formula of $C_{45}H_{74}O_{17}$, which is similar to the one of malonyl-Rg1 (Ma et al., 2005). Furthermore, their fragmentation pattern indicates that both Rg1 and compound 15 consist of an aglycone with a similar mass, on which 2 hexose moieties and a malonyl group are bound (Figure 5). However, the retention times of both compounds are significantly different: malonyl-Rg1 elutes at 20.52 minutes (Figure 4), whereas compound 15 elutes at 29.55 minutes. Since separation of the compounds was achieved using reversed phase liquid chromatography, the

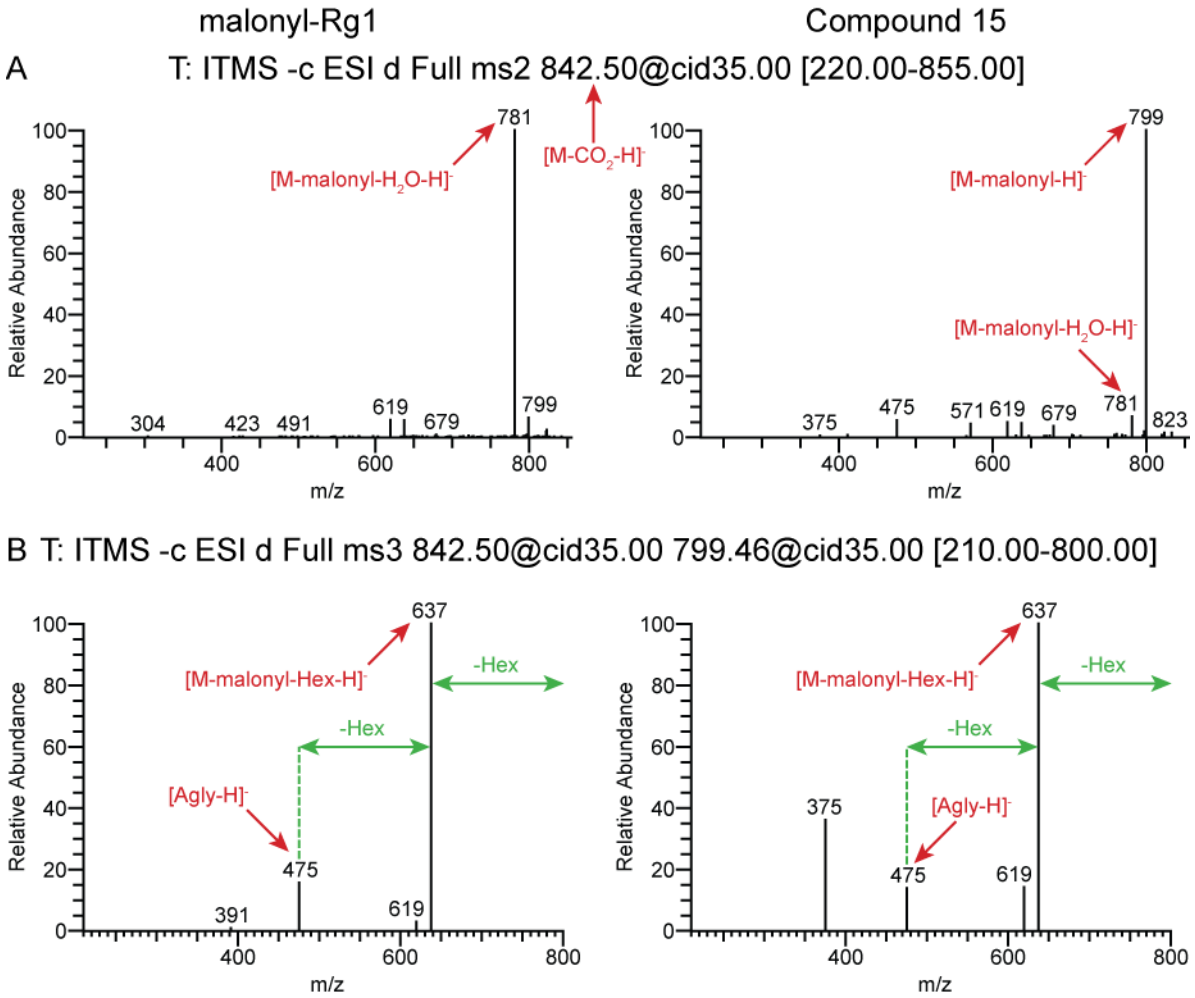


Figure 5. Comparison of the ginsenoside malonyl-Rg1 (left) and compound 15 (right). (A) MS^2 fragmentation of the $[M\text{-}CO_2\text{-}H]^-$ ion at *m/z* 842.50. (B) MS^3 fragmentation of the MS^2 daughter ion at *m/z* 799.

different elution times indicate a different hydrophobicity of the molecules, and thus different molecules. In conclusion, compound 15 is new to both species, and thus proof of concept of combinatorial biosynthesis in plants has been established.

2.3.New saponins in *M. truncatula* hairy roots overexpressing a *B. falcatum* CytP450 (BF0567).

Based on the putative annotation of the genes or the presence of a phenotype in the hairy roots, four other lines were prioritized for metabolite profiling as performed for the PG0408 overexpression lines. The results of the analysis of these 4 lines are summarized in Table 2.

Table 2. Top-candidate genes analyzed in a pilot LC ESI FT-ICR MS screening.

Gene	Annotation	# peaks	# new compounds
ML041	CytP450	18500	0
ML158	Glucosyltransferase	18825	0
PG0555	CytP450	19293	0
BF0567	CytP450	18689	19

In the BF0567 overexpression lines, the presence of 19 new, mostly low abundant compounds was confirmed. For 7 compounds of this list, essential MS^n information for the tentative identification of the compounds was generated (Table 3). This revealed that all new compounds were triterpene saponins, most of them having an aglycone brutoformula corresponding to known *Medicago* aglycones. However, BF0567 is a gene from *B. falcatum*, a plant that produces a set of saikogenins as aglycones of its saponins. These aglycones have the same brutoformula as the sapogenins from *M. truncatula*, and thus, distinguishing between the two types of aglycones based on brutoformula alone is impossible. However, the fragmentation pattern of the new saponins of the BF0567 overexpression lines, and the

Table 3. Tentatively identified new saponins in BF0567 overexpression lines.

Cpd	Name	$[M-H]^-$	Formula (δ ppm)	FT-ICR MS^n: *m/z* (% base peak)
1	Hex-Hex-Hex-Aglycone	1003.47939	$C_{48}H_{76}O_{22}$ (3.830)	**MS^2 [1003.48]**: 841 (100)[M-Hex-H]$^-$, 797 (17)[M-Hex-CO_2-H]$^-$, 455 (76) [Agly-CO_2-H_2O-H]$^-$
2	Hex-dHex-Hex-HexA-Aglycone	1133.54373	$C_{54}H_{86}O_{25}$ (4.578)	**MS^2 [1133.54]**: 971 (100)[M-Hex-H]$^-$ **MS^3 [1133.54→971]**: 953 (70)[M-Hex-H_2O-H]$^-$, 927 (10)[M-Hex-CO_2-H]$^-$, 909 (100)[M-Hex-H_2O-CO_2-H]$^-$, 763 (17)[M-Hex-dHex-CO_2-H_2O-H]$^-$, 645 (78)[M-Hex-dHex-Hex-H_2O-H]$^-$, 627 (18)[M-Hex-dHex-Hex-2H_2O-H]$^-$, 555 (13)[M-Hex-dHex-Hex-108-H]$^-$, 469 (15)[Agly-H_2O-H]$^-$
3	Hex-HexA-Aglycone	825.42998	$C_{42}H_{66}O_{16}$ (2.630)	**MS^2 [825.43]**: 705 (16), 663 (100)[M-Hex-H]$^-$, 645 (3)[M-Hex-H_2O-H]$^-$, 601 (9)[M-Hex-H_2O-CO_2-H]$^-$, 487 (3)[Agly-H]$^-$ **MS^3 [925.43→663]**: 645 (3)[M-Hex-H_2O-H]$^-$, 617 (11)[M-Hex-46-H]$^-$, 601 (100)[M-Hex-H_2O-CO_2-H]$^-$, 587 (6), 487 (49)[Agly-H]$^-$, 393 (10)
4	Hex-dHex-Hex-HexA-Aglycone	1117.54977	$C_{54}H_{86}O_{24}$ (5.497)	**MS^2 [1117.55]**: 955 (100)[M-Hex-H]$^-$, 937 (3)[M-Hex-H_2O-H]$^-$ **MS^3 [1117.55→955]**: 937 (100)[M-Hex-H_2O-H]$^-$, 893 (87)[M-Hex-H_2O-CO_2-H]$^-$, 809 (6)[M-Hex-dHex-H]$^-$, 747 (12)[M-Hex-dHex-CO_2-H_2O-H]$^-$, 629 (81)[M-Hex-dHex-Hex-H_2O-H]$^-$, 611 (34)[M-Hex-dHex-Hex-2H_2O-H]$^-$, 567 (5)[M-Hex-dHex-Hex-2H_2O-CO_2-H]$^-$, 555 (71)[M-Hex-dHex-Hex-108-H]$^-$, 471 (5)[Agly-H]$^-$, 409 (4), 407 (3)
5	dHex-Hex-Hex-Aglycone	957.50948	$C_{48}H_{78}O_{19}$ (3.161)	**MS^2 [957.51]**: 811 (100)[M-dHex-H]$^-$, 793 (9)[M-dHex-H_2O-H]$^-$, 649 (12)[M-dHex-Hex-H]$^-$, 631 (3)[M-dHex-Hex-H_2O-H]$^-$, 487 (1)[Agly-H]$^-$ **MS^3 [957.51→811]**: 749 (3)[M-dHex-H_2O-CO_2-H]$^-$, 649 (100)[M-dHex-Hex-H]$^-$, 587 (7)[M-dHex-Hex-H_2O-CO_2-H]$^-$, 487 (16)[Agly-H]$^-$ **MS^3 [957.51→649]**: 631 (4)[M-dHex-Hex-H_2O-H]$^-$, 603 (100)[M-dHex-Hex-HCOOH-H]$^-$, 587 (10)[M-dHex-Hex-H_2O-CO_2-H]$^-$, 571 (8), 487 (86)[Agly-H]$^-$, 425 (32)[Agly-H_2O-CO_2-H]$^-$
6	Hex-Aglycone	649.39731	$C_{36}H_{58}O_{10}$ (2.447)	**MS^2 [649.40]**: 631 (14)[M-H_2O-H]$^-$, 603 (100)[M-HCOOH-H]$^-$, 587 (12)[M-H_2O-CO_2-H]$^-$, 487 (32)[Agly-H]$^-$, 425 (26)[Agly-H_2O-CO_2-H]$^-$
7	Hex-???	781.43869	$C_{41}H_{66}O_{14}$ (0.909)	**MS^2 [781.44]**: 763 (15)[M-H_2O-H]$^-$, 737 (23)[M-CO_2-H]$^-$, 619 (100)[M-Hex-H]$^-$, 557 (14)[M-Hex-H_2O-CO_2-H]$^-$

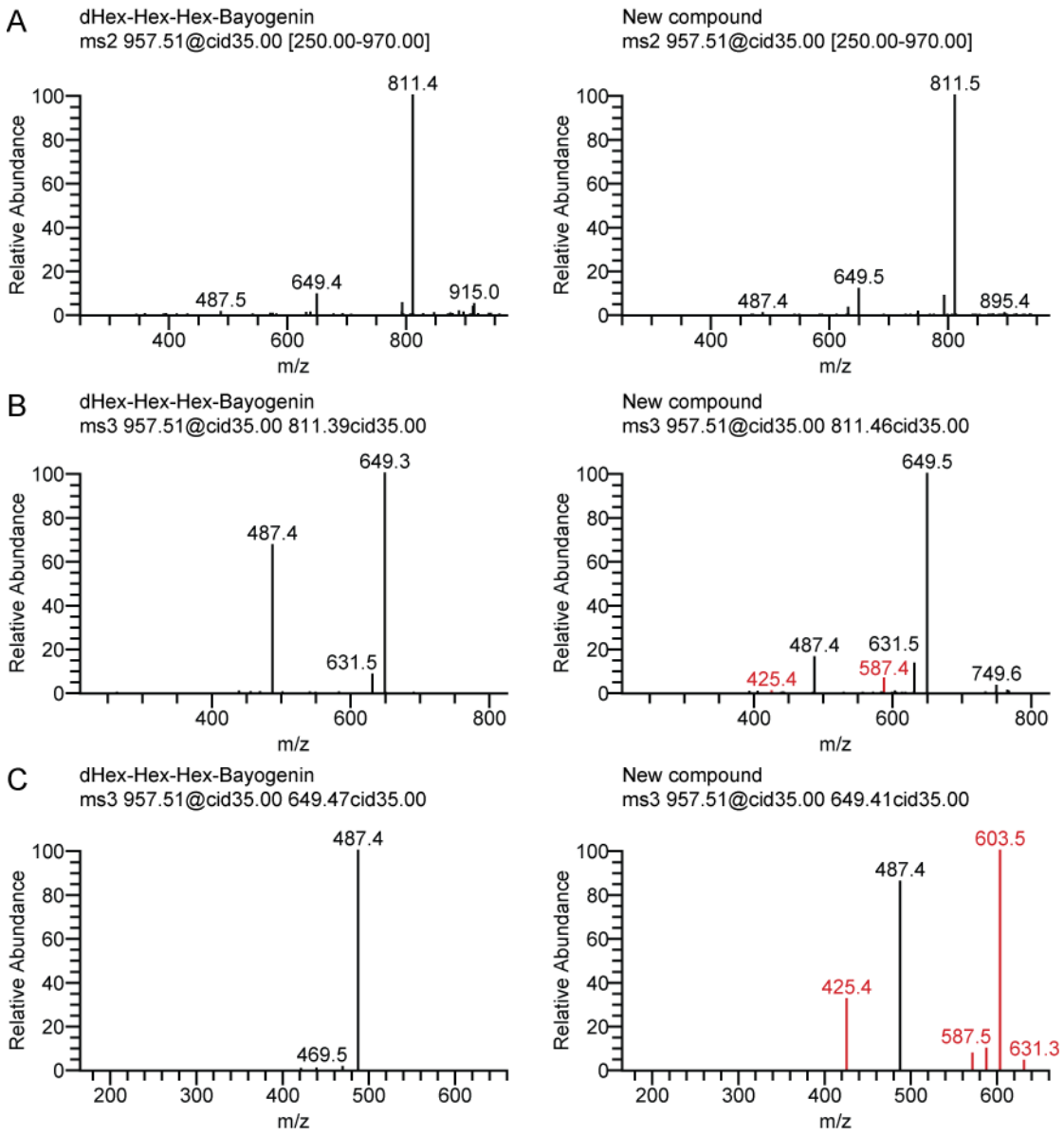


Figure 6. Comparison of the fragmentation behavior of dHex-Hex-Hex-Bayogenin (left) and dHex-Hex-Hex-New Aglycone (right). A. MS^2 fragmentation of the $[M-H]^-$ ion at *m/z* 957.51. B. MS^3 fragmentation of MS^2 daughter ion at *m/z* 811. C. MS^3 fragmentation of MS^2 daughter ion at *m/z* 649. The different fragment ions are indicated in red.

corresponding *M. truncatula* saponins are significantly different. This is illustrated with the fragmentation pattern of the new compound with an $[M-H]^-$ ion at *m/z* 957.50948, leading to an aglycone with the $[Agly-H]^-$ anion at *m/z* 487. In *M. truncatula* control roots, a similar compound could be detected, dHex-Hex-Hex-bayogenin, either as an in-source fragment ion of compound 4 or as compound 51 (see chapter 4). In Figure 6, the fragmentation pattern of dHex-Hex-Hex-bayogenin is compared with that of the new compound, showing that, although they have the same sugar moieties, there are clear differences in the MS^3

fragmentation. Notably, when the daughter ion at *m/z* 649 is fragmented, as in case of the new compound, the most abundant granddaughter ion is not the aglycone ion, as is the case for Hex-bayogenin, but an ion corresponding with a loss of 46 Da, which may be the result of the loss of formic acid during the fragmentation process. Furthermore, another high abundant peak, at *m/z* 425 was observed, corresponding to a loss of 62 Da from the aglycone ion. Similar to medicagenic acid (see chapter 4), this may be due to the loss of CO_2 and H_2O from the aglycone ion, hence, the aglycone should be hydroxylated at position 3, and carboxylated at position 24. This would indicate that 2 more hydroxyl groups are present on this aglycone, however their exact positions remain unknown. Hence, carboxylation at position 28 is impossible, and thus, hydroxylation, followed by formation of an ether bridge remains an option. In conclusion, several new saponins were detected in the BF0567 overexpression lines, and although they have aglycones with similar masses as the *M. truncatula* sapogenins, they show a different fragmentation pattern, and therefore may represent novel aglycones.

2.4. Overexpression of a *M. lanceolata* CytP450 leads to accumulation of new, unknown compounds

In a second selection, based on bioactivity screens with extracts of the transgenic hairy roots (Moses et al., unpublished results), several more lines were screened. Within this set, in the lines that overexpress a *M. lanceolata* CytP450, *ML385*, accumulation of very high amounts of a few new compounds was observed (Figure 7).

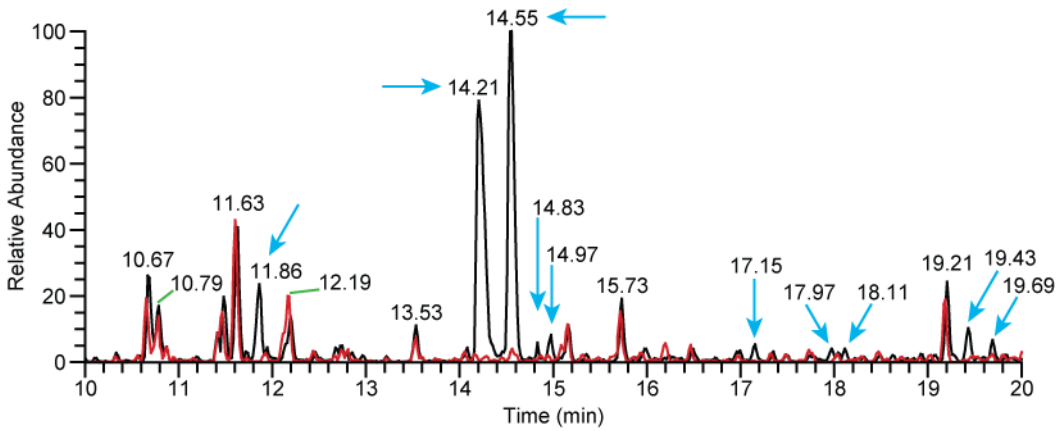


Figure 7. Comparison of the chromatogram of a control line (red) and a ML385 overexpression line (black). Several new peaks (indicated with blue arrow) can be observed in the raw chromatogram.

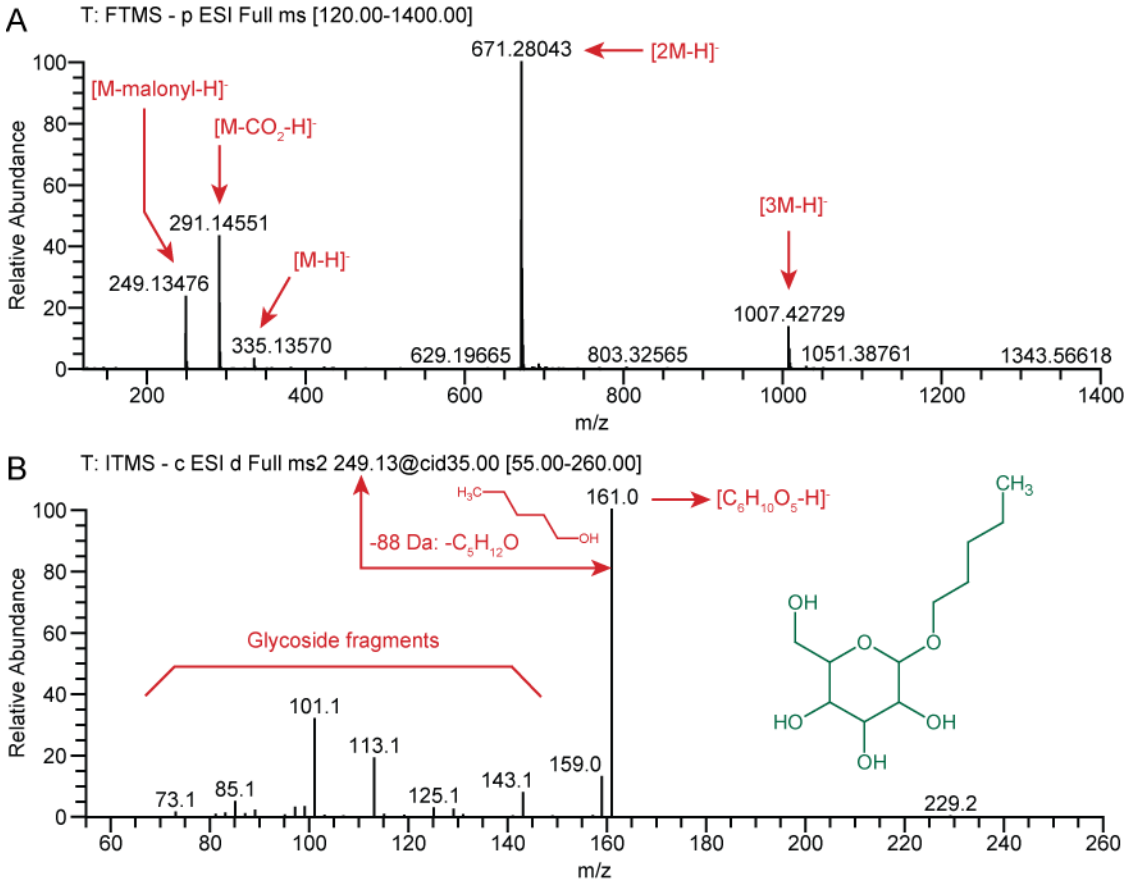


Figure 8. Tentative identification of the compound corresponding to the peak at t_R 14.21. (A) MS scan of peak at t_R 14.21 min. (B) MS^2 fragmentation of the [M-malonyl-H]$^-$ ion at *m/z* 249.13 that could correspond to pentyl hexose (green).

The retention time of the new compounds is much shorter than the retention time of saponins. Furthermore, also the generated MS^n data indicated that the new compounds were not saponins. Tentative identification of the compounds revealed that at least some of them are most likely glycosides. For instance, the new compound eluting at t_R 14.21 (Figure 7) has a [M-H]$^-$ anion at *m/z* 335.13570 (Figure 8A). Due to in source fragmentation, additional ions were observed at *m/z* 291.14551 and *m/z* 249.13476, correlating to the loss of 44 and 86 Da from the molecular species, respectively. These ions are identified as the loss of a CO_2 and a malonyl group, respectively. This indicates that the new compound is malonylated, which is reported to occur in glycosides of *M. truncatula* saponins and flavonoids as well (Huhman and Sumner, 2002). The molecular formulas corresponding to the [M-H]$^-$ anion and the [M-malonyl-H]$^-$ anion are $C_{14}H_{24}O_9$ (δ ppm = 2.818), and $C_{11}H_{22}O_6$ (δ ppm = 1.599), respectively, confirming the loss of a malonyl group ($C_3H_2O_3$). MS^2 fragmentation of the [M-malonyl-H]$^-$ ion at *m/z* 249.13 led to the generation of 1 major daughter ion at *m/z* 161, resulting from the loss of 88 Da from the parent ion. Furthermore, several less abundant

daughter ions were present, most of which were characteristic for the fragmentation of a glycoside residue (Figure 8B). Hence, the daughter ion at *m*/*z* 161 results from a hexose residue ($C_6H_{10}O_5$), whereas the neutral loss of 88 Da corresponds to the loss of pentanol ($C_5H_{12}O$). Thus, the compound at *m*/*z* 249.13 could be pentyl hexose, and as such the parental compound at *m*/*z* 335 could be malonylated pentyl hexose. Similarly, structural elucidation of the other new compounds was attempted (Table 4), showing that some of the other compounds are also possible glycosides.

Table 4. Tentative identification of the new compounds in the ML385 overexpression lines.

t_R	$[M-H]^-$	Formula (δ ppm)	Tentative annotation
11.86	381.17720	$C_{16}H_{30}O_{10}$ (1.521)	pentose + pentyl + hexose
14.21	355.13570	$C_{14}H_{24}O_9$ (2.818)	malonyl + pentyl + hexose
14.55	467.17822	$C_{19}H_{32}O_{13}$ (2.581)	malonyl + pentose + pentyl + hexose
14.83	369.12011	$C_{17}H_{22}O_9$ (2.722)	cinnamate + acetyl + hexose
14.97	501.16258	$C_{22}H_{30}O_{13}$ (2.426)	malonyl + ???
17.15	573.22059	$C_{26}H_{38}O_{14}$ (2.985)	vanillic acid + ???
17.97	763.27025	$C_{33}H_{48}O_{20}$ (4.760)	malonyl + hexose + hydroxyphenylacetic acid + ???
18.11	763.27023	$C_{33}H_{48}O_{20}$ (4.734)	malonyl + hexose + hydroxyphenylacetic acid + ???
19.43	501.19884	$C_{23}H_{34}O_{12}$ (2.176)	???
19.69	529.19380	$C_{24}H_{34}O_{13}$ (2.146)	Phthalic acid + ???

3. Discussion

3.1.Proof of concept: novel dammarane-type saponins in *M. truncatula*

The aim of this study was to obtain proof of concept of combinatorial biosynthesis in plants, with the secondary metabolite class of the triterpene saponins as a target. To this end, candidate saponin biosynthesis genes of different medicinal plants were expressed in hairy roots of the saponin-producing model legume *M. truncatula*, thereby creating the possibility of new enzyme-substrate combinations in *M. truncatula*. The products thereof may be used by the downstream saponin biosynthesis genes and ultimately lead to novel saponins in the *M. truncatula* hairy roots. Proof of concept of combinatorial biosynthesis was first obtained by the overexpression of dammarenediol-II synthase of *P. ginseng* (Tansakul et al., 2006), which led to the accumulation of 39 new compounds in the transgenic *M. truncatula* hairy

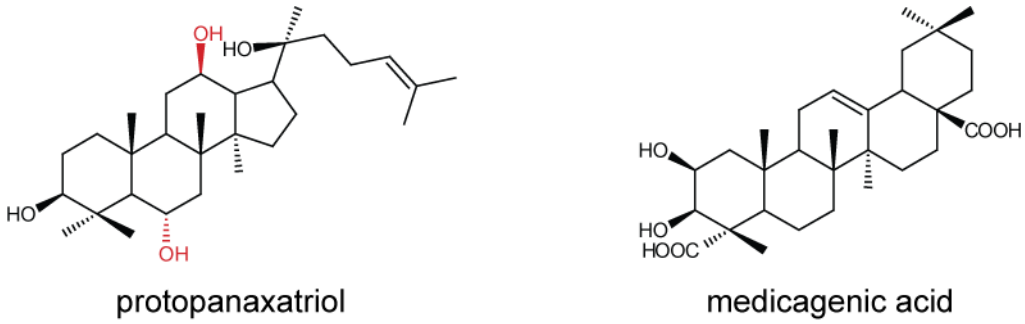


Figure 9. Structures of protopanaxatriol and medicagenic acid.

roots. Structural elucidation of some of these compounds showed that they were saponins with dammarane-type aglycones, a new type of aglycones for *M. truncatula*. Furthermore, comparison with a *P. ginseng* extract showed that these saponins are not only new to *M. truncatula*, but are also new to *P. ginseng*. Although for some of the compounds isomers with a similar fragmentation pattern could be found in both *P. ginseng* and *M. truncatula*, the retention times of the isomers were significantly different, indicating a different hydrophobicity of the molecules that may be due to a different position of hydroxylation on the dammarane backbone. In *P. ginseng*, hydroxylation of dammarenediol occurs preferentially on positions 6 and 12 (Figure 9). In *M. truncatula*, however, oxidation of the β-amyrin backbone never occurs on these positions as illustrated with the most oxidized sapogenin in *M. truncatula* hairy roots, medicagenic acid (Figure 9). Hence, the dammarenediol molecule produced by the dammarenediol synthase in the *M. truncatula* hairy roots is most probably hydroxylated on a different position.

3.2.Novel oxidations on the oleanane skeleton

Next to the *PG0408* lines, new compounds have also been detected in *M. truncatula* hairy roots transformed with the *BF0567* gene that encodes a CytP450 enzyme from *B. falcatum*. In saponin biosynthesis, CytP450 enzymes are involved in the oxidation of the cyclization products of 2,3-oxidosqualene on various positions, leading to the different sapogenins. Since the new aglycones were encountered in the *M. truncatula* hairy roots overexpressing *BF0567*, it can be expected that the corresponding enzyme acts on the oleanane backbone on a position that is normally not attacked by *M. truncatula* CytP450 enzymes. Comparison of the sapogenins of *M. truncatula* and *B. falcatum* may address the function of the enzyme. Firstly, unlike the *M. truncatula* sapogenins (Chapter 4), the three most abundant sapogenins of *B. falcatum* contain a C13-O-CH_2-C17 ether linkage (Figure 10A) (Ashour and Wink, 2011).

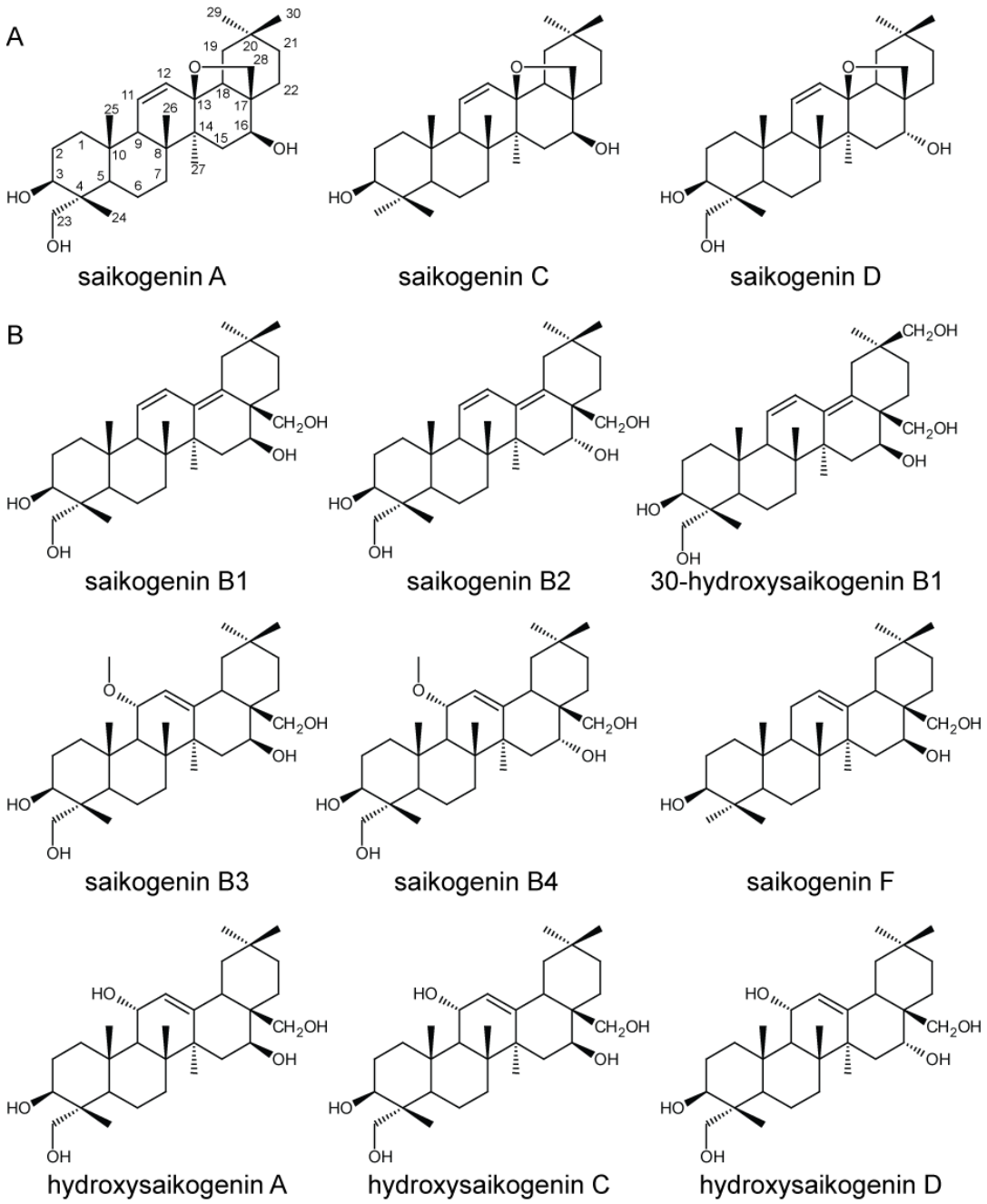


Figure 10. Structures of all saikogenins described in the literature (Ashour and Wink, 2011). (A) Saikogenins with a C13-O-CH_2-C17 ether bridge. (B) All other described saikogenins.

This oxygen bridge may be formed by a dehydration reaction between a C13-OH group and a C17-CH_2OH group (Vincken et al., 2007). The hydroxyl group at the C13-position is never encountered in a free form, and its formation also requires isomerisation of the C12-C13 double bond towards unsaturation at position C11-C12. Thus, by hydroxylation at the C13-position and formation of the ether linkage as a consequence, isomerisation of the double bond may occur in parallel. Since the molecular formula of the saikogenins with an ether bridge is the same as some of the predicted molecular formulas of the detected new

aglycones, the formation of a 13,28 ether bridge by BF0567 is a valid hypothesis for the detected new aglycones.

A second difference between the *B. falcatum* saikogenins and *M. truncatula* sapogenins is the occurrence of a hydroxyl group at the C16-position in all saikogenins of *B. falcatum* (Figure 10). Although present in zanhic acid, a main sapogenin of *M. truncatula* leaf tissue (Kapusta et al., 2005a; Kapusta et al., 2005b), this modification does not occur in *M. truncatula* hairy roots (see chapter 4), and would thus be a possible activity of BF0567. If so, zanhic acid conjugates are expected to be detectable in the *BF0567* expressing hairy roots. Indeed, fragmentation of compound 1 (Table 3) leads to the detection of a possible zanhic acid derived fragment ion at *m*/*z* 455, resulting from the decarboxylation and dehydration of the aglycone ion. Furthermore, also some of the other observed aglycone masses could be explained by this type of modification. Hence, also oxidation at the C16-position remains a possible explanation. To confirm this type of activity, the saponin profile of the BF0567 expressing roots could be compared with that of a *M. truncatula* leaf sample.

Another difference between the *M. truncatula* sapogenins and the *B. falcatum* saikogenins is the oxidation at the C11-position. Like the oxidation at the C16-position, this could also be a possible mode of action of the *BF0567* gene. For instance, the fragmentation of compound 1 (Table 4) leads to a fragment ion at *m*/*z* 455, resulting from the decarboxylation and dehydration of the aglycone ion. Although a zanhic acid aglycone is a possible explanation for this, it cannot be excluded that the observed aglycone is derived from medicagenic acid (Figure 9) that is oxidized at the C11-position instead of the C16-position as in zanhic acid. In conclusion, the overexpression of the BF0567 CytP450 enzyme in *M. truncatula* hairy roots, led to the accumulation of saponins with new aglycones, thr exact structure of which cannot be elucidated yet with the applied methodology.

3.3.Combinatorial biosynthesis of non-saponin compounds

In *M. truncatula* hairy root lines overexpressing the *ML385* gene, the accumulation of high amounts of novel, likely non-saponin compounds was observed. An extract of *M. lanceolata* was screened for these compounds, and indicated that similar compounds were not encountered in wild-type *M. lanceolata* plants, confirming the specificity of the detected molecules in *M. truncatula*, and thus confirming combinatorial biosynthesis. Structural elucidation revealed these compounds were not saponins, but glycosides of small molecules

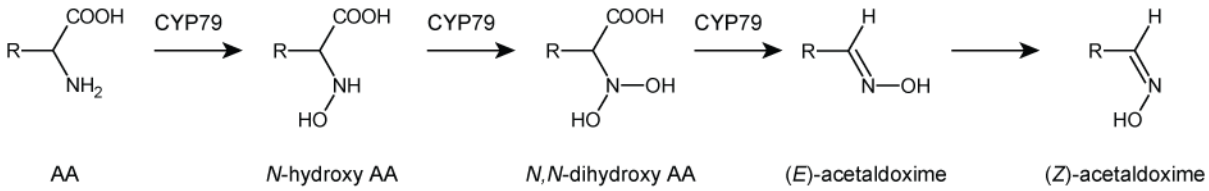


Figure 11. Conversion of amino acids (AA) to the corresponding acetaldoximes by enzymes of the CYP79 clan.

such as vanillic acid, phthalic acid, cinnamic acid, hydroxyphenylacetic acid or pentanol (Table 4), however, their exact structure remains unknown. Hence, despite the focus on putative saponin biosynthesis genes, at least one of the selected genes seems to be involved in another biosynthesis pathway. A blast search of the FL-ORF of *ML385* revealed the enzyme belongs to the CYP79 clan. Several enzymes of this clan have been characterized on the molecular level, showing they are all involved in *N*-hydroxylation of amino acids (Andersen et al., 2000; Bak et al., 2000; Hull et al., 2000; Mikkelsen et al., 2000; Nielsen and Møller, 2000; Wittstock and Halkier, 2000). Via a multi-step reaction, the characterized enzymes converse the amino acids to the corresponding acetaldoximes (Figure 11), which are subsequently used as input for the biosynthesis of amino acid-derived secondary metabolites. The closest (characterized) homolog of ML385 is the *Sorghum bicolor* CYP79A1 (tyrosine *N*-monooxygenase) that catalyzes the first step in the biosynthesis of dhurrin, a cyanogenic glucoside. Ectopic expression of this enzyme in *Arabidopsis* and tobacco (*Nicotiana tabacum*) plants led to the accumulation of several *p*-hydroxyphenylacetaldoxime-derived glycosides in the transgenic plants (Bak et al., 2000). Similar here, we detected the accumulation of a set of glycosides in the *M. truncatula* hairy roots. Hence, the *M. lanceolata* CYP79, ML385, may act by converting tyrosine or other amino acids to the corresponding acetaldoximes. Subsequently, enzymes present in the *M. truncatula* hairy roots may convert these molecules to yield the detected glucosides.

3.4.Conclusions

Heterologous expression of candidate saponin biosynthesis genes in hairy roots of *M. truncatula* led to the accumulation of a set of compounds in some of the transgenic lines. These compounds were shown to be novel to both *M. truncatula* and the plant of which the expressed gene was derived. Hence, proof of concept of combinatorial biosynthesis in plants was established. The targets of this study were compounds belonging to the secondary metabolite class of the triterpene saponins. Novel saponin aglycones were obtained for

oleanane-type saponins and the pharmacologically more important dammarane-type saponins. Furthermore, we also detected novel, non-saponin compounds in one of the overexpression constructs, showing that the applied method is applicable to generate structural variation in other classes of plant natural products as well.

4. Materials and Methods

4.1.Generation of DNA constructs

For stable overexpression, the identified FL-clones entry clones were Gateway recombined into the pK7WG2D binary vector (Karimi et al., 2002), and the resulting clones were transformed into *A. rhizogenes* by electroporation.

4.2.Generation of *M. truncatula* hairy roots

M. truncatula hairy roots were generated as described (Pollier et al., 2011).

4.3.Transcriptional analysis of *M. truncatula* hairy roots

For Reverse Transcription PCR (RT-PCR), total RNA was extracted with TRI Reagent® (Ambion), and cDNA prepared with SuperScript™ II Reverse Transcriptase (Invitrogen). Specific PCR primers were designed based on the cDNA-AFLP tag sequences with the Primer3 program (Rozen and Skaletsky, 2000).

4.4.Metabolite profiling of *M. truncatula* hairy roots

Metabolite profiling of *M. truncatula* hairy roots was performed as described (Pollier et al., 2011).

4.5.Identification of new peaks

Per transgene construct, 5 five biological repeats of three independent transgenic lines were analyzed. The resulting chromatograms were grouped per construct and each group was integrated and aligned versus the control group with the XCMS package (Smith et al., 2006) in R version 2.6.1. The following parameter values were applied: xcmsSet(fwhm=8, max=300, snthresh=5, mzdiff=0.5), group(bw=8, max=300), retcor(method=loess, family=symmetric). A second grouping was done with the same parameter values. In this analysis, discretely present *m*/*z* peaks correspond to novel or disappeared compounds in the

transgenic lines. Due to in-source fragmentation, multiple *m*/*z* peaks for each compound were often observed. For identification, full MS spectra were interchanged with a dependent MS^2 scan event in which the most abundant ion in the previous full MS scan was fragmented, and two dependent MS^3 scan events of the two most abundant daughter ions. The collision energy was set at 35%. Elucidation of the MS^n spectra was based on (Pollier et al., 2011).

Acknowledgements

This work was supported by the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" project SBO040093).

References

Andersen, M.D., Busk, P.K., Svendsen, I., and Møller, B.L. (2000). Cytochromes P-450 from cassava (*Manihot esculenta* Crantz) catalyzing the first steps in the biosynthesis of the cyanogenic glucosides linamarin and lotaustralin. J Biol Chem **275**, 1966-1975.

Ashour, M.L., and Wink, M. (2011). Genus *Bupleurum*: a review of its phytochemistry, pharmacology and modes of action. J Pharm Pharmacol **63**, 305-321.

Bak, S., Olsen, C.E., Halkier, B.A., and Møller, B.L. (2000). Transgenic tobacco and Arabidopsis plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. Plant Physiol **123**, 1437-1448.

Floss, H.G. (2006). Combinatorial biosynthesis - Potential and problems. J Biotechnol **124**, 242-257.

Huhman, D.V., and Sumner, L.W. (2002). Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry **59**, 347-360.

Hull, A.K., Vij, R., and Celenza, J.L. (2000). Arabidopsis cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. Proc Natl Acad Sci U S A **97**, 2379-2384.

Kapusta, I., Janda, B., Stochmal, A., and Oleszek, W. (2005a). Determination of saponins in aerial parts of barrel medic (*Medicago truncatula*) by liquid chromatography−electrospray ionization/mass spectrometry. J Agric Food Chem **53**, 7654-7660.

Kapusta, I., Stochmal, A., Perrone, A., Piacente, S., Pizza, C., and Oleszek, W. (2005b). Triterpene saponins from barrel medic (*Medicago truncatula*) aerial parts. J Agric Food Chem **53**, 2164-2170.

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci **7**, 193-195.

Kim, O.T., Bang, K.H., Kim, Y.C., Hyun, D.Y., Kim, M.Y., and Cha, S.W. (2009). Upregulation of ginsenoside and gene expression related to triterpene biosynthesis in ginseng hairy root cultures elicited by methyl jasmonate. Plant Cell Tiss Organ Cult **98**, 25-33.

Lam, K.S. (2007). New aspects of natural products in drug discovery. Trends Microbiol **15**, 279-289.

Li, J.W.H., and Vederas, J.C. (2009). Drug discovery and natural products: end of an era or an endless frontier? Science **325**, 161-165.

Littleton, J., Rogers, T., and Falcone, D. (2005). Novel approaches to plant drug discovery based on high throughput pharmacological screening and genetic manipulation. Life Sci **78**, 467-475.

Lü, J.M., Yao, Q., and Chen, C. (2009). Ginseng compounds: an update on their molecular mechanisms and medical applications. Curr Vasc Pharmacol **7**, 293-302.

Ma, X.Q., Liang, X.M., Xu, Q., Zhang, X.Z., and Xiao, H.B. (2005). Identification of ginsenosides in roots of *Panax ginseng* by HPLC-APCI/MS. Phytochem Anal **16**, 181-187.

Menzella, H.G., and Reeves, C.D. (2007). Combinatorial biosynthesis for drug development. Curr Opin Microbiol **10**, 238-245.

Mikkelsen, M.D., Hansen, C.H., Wittstock, U., and Halkier, B.A. (2000). Cytochrome P450 CYP79B2 from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. J Biol Chem **275**, 33712-33717.

Nielsen, J.S., and Møller, B.L. (2000). Cloning and expression of cytochrome P450 enzymes catalyzing the conversion of tyrosine to p-hydroxyphenylacetaldoxime in the biosynthesis of cyanogenic glucosides in *Triglochin maritima*. Plant Physiol **122**, 1311-1321.

Ortholand, J.Y., and Ganesan, A. (2004). products and combinatorial chemistry: back to the future. Curr Opin Chem Biol **8**, 271-280.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. J Nat Prod, **in press**.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132**, 365-386.

Runguphan, W., Maresh, J.J., and O'Connor, S.E. (2009). Silencing of tryptamine biosynthesis for production of nonnatural alkaloids in plant culture. Proc Natl Acad Sci U S A **106**, 13673-13678.

Runguphan, W., Qu, X.D., and O'Connor, S.E. (2010). Integrating carbon-halogen bond formation into medicinal plant metabolism. Nature **468**, 461-467.

Salas, J.A., and Méndez, C. (2009). Indolocarbazole antitumour compounds by combinatorial biosynthesis. Curr Opin Chem Biol **13**, 1-9.

Sánchez, C., Méndez, C., and Salas, J.A. (2006). Indolocarbazole natural products: occurrence, biosynthesis, and biological activity. Nat Prod Rep **23**, 1007-1045.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem **78**, 779-787.

Tansakul, P., Shibuya, M., Kushiro, T., and Ebizuka, Y. (2006). Dammarenediol-II synthase, the first dedicated enzyme for ginsenoside biosynthesis, in *Panax ginseng*. FEBS lett **580**, 5143-5149.

Tava, A., Scotti, C., and Avato, P. (2010). Biosynthesis of saponins in the genus *Medicago*. Phytochem Rev, **in press**, doi: 10.1007/s11101-010-9169-x.

Vincken, J.P., Heng, L., de Groot, A., and Gruppen, H. (2007). Saponins, classification and occurrence in the plant kingdom. Phytochemistry **68**, 275-297.

Wittstock, U., and Halkier, B.A. (2000). Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of *L*-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. J Biol Chem **275**, 14659-14666.

Yue, C.J., Zhou, X., and Zhong, J.J. (2008). Protopanaxadiol 6-Hydroxylase and its role in regulating the ginsenoside heterogeneity in *Panax notoginseng* cells. Biotechnol Bioeng **100**, 933-940.

Zhou, H., Xie, X.K., and Tang, Y. (2008). Engineering natural products using combinatorial biosynthesis and biocatalysis. Curr Opin Biotechnol **19**, 590-596.

Chapter 6

The RING E3 ubiquitin ligase MAKIBISHI1 regulates triterpene saponin biosynthesis in *Medicago truncatula*.

Jacob Pollier*, Miguel González-Guzmán, Tessa Moses, Robin Vanden Bossche, Riet De Rycke, Peter Marhavý, Kris Morreel, Wieslaw Oleszek, Johan M. Thevelein, Alain Goossens

* Author contribution: maintenance and elicitation cell culture, cDNA-AFLP transcript profiling, phylogenetic analysis, onion transformation, cultivation of hairy roots, qPCR analysis, saponin profiling, analysis of the data, and writing of the manuscript.

Publication status:

Manuscript in preparation.

Abstract

Triterpene saponins are compounds that possess many biological activities and have a widespread occurrence in the plant kingdom. They share a common biogenic origin with the ubiquitous sterols. Both compounds are derived from the precursor oxidosqualene that relies on the mevalonate pathway for isopentenyl diphosphate supply. Saponin accumulation in plant cells is stimulated in the defense responses that are triggered by phytohormones, such as jasmonates, but the underlying molecular mechanisms remain unknown. To screen for candidate regulators of saponin biosynthesis, we used genome-wide transcript profiling of the jasmonate-response of the model legume *Medicago truncatula*. A jasmonate-inducible gene encoding a RING membrane-anchor E3 ubiquitin ligase, MAKIBISHI1 (MKB1), was identified. MKB1 targeted the 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), the enzyme catalyzing the rate-limiting step in the mevalonate pathway, for ubiquitin-mediated proteasomal degradation, thereby controlling saponin biosynthesis. The MKB1-mediated control apparatus from *M. truncatula* is equivalent to the endoplasmatic reticulum-associated degradation system that regulates HMGR stability and sterol synthesis in yeast and mammalian cells. Our findings demonstrate that the general mechanistic principle for the control of triterpene biosynthesis is conserved across eukaryotes, despite the divergent molecular structures of both the effectors and targets involved.

1. Introduction

The more than 19,000 species of legumes, distributed over 730 genera, display an overwhelming variety in type and number of natural products they can produce. Most renowned are the isoflavones. In contrast to the omnipresent flavonoids, the occurrence of these molecules is more restricted in the plant kingdom, in particular to the Papilionoideae subfamily of the Leguminosae, in which they function as phytoalexins or induce nodulation through symbiotic N_2-fixing bacteria (Dixon and Sumner, 2003). Another substantial group of secondary metabolites in legumes is formed by the triterpene saponins that serve, *in planta*, as allelopathic or anti-insect agents, but also display a wide range of biological activities. Many leguminous genera can produce triterpene saponins, including *Acacia*, *Glycyrrhiza*, *Lathyrus*, *Lupinus*, *Medicago*, *Trifolium*, and *Vigna*, with different pharmaceutical properties, such as antiinflammatory, antimicrobial, anticancer, or adjuvant activities (Dixon and Sumner, 2003; Sparg et al., 2004; Sun et al., 2009; Yendo et al., 2010; Augustin et al., 2011).

Leguminous natural products can be synthesized constitutively or inducibly, depending on species, developmental stage, or environmental conditions of biotic or abiotic nature. In particular, the phytohormone jasmonate (JA) and its derivatives, collectively called jasmonates (JAs), are powerful elicitors of secondary metabolite synthesis in a multitude of leguminous plants and have been demonstrated to induce leguminous natural products of a wide structural variety, including the isoflavones and triterpene saponins (Suzuki et al., 2002, 2005; Zhao et al., 2005; Pauwels et al., 2009; Yendo et al., 2010). The central module of the JA signaling cascade has been discovered in Arabidopsis (*Arabidopsis thaliana*) and found to be conserved across the plant kingdom (Pauwels et al., 2009; Fonseca et al., 2009a; Browse, 2009). In *Arabidopsis*, the F-box protein CORONATINE INSENSITIVE (COI1) fulfils a key role in JA signaling. COI1 participates in the SKP/CULLIN/F-BOX-COI1 (SCF^{COI1}) E3 ubiquitin ligase complex that perceives bioactive JAs such as jasmonoyl-isoleucine (JA-Ile). In the presence of JA-Ile, COI1 interacts with the JA ZIM-domain (JAZ) repressor proteins that are subsequently marked for degradation by the 26S proteasome (Chini et al., 2007; Thines et al., 2007; Fonseca et al., 2009b; Sheard et al., 2010). This event releases the bHLH-type transcriptional activators, such as MYC2, MYC3, and MYC4, thereby activating the first wave of JA-induced gene expression (Pauwels et al., 2009; Fernández-Calvo et al., 2011). In the absence of JA-Ile, the JAZ proteins bind the MYC factors and recruit a co-repressor

complex that blocks MYC activity and, hence, the JA response (Chini et al., 2007; Thines et al., 2007; Pauwels et al., 2010; Fernández-Calvo et al., 2011). In leguminous plants the efforts to understand JA signal transduction have remained rather limited. A soybean (*Glycine max*) *COI1* gene that shares significant homology with that of *Arabidopsis*, has been identified and can restore the defective JA responses, including plant defense and fertility, of the *Arabidopsis coi1-1* mutant (Wang et al., 2005), supporting the existence of a general JA pathway with signal components that are functionally conserved across the plant kingdom. Nevertheless, the action radius of JAs in different plant species also exhibits species-specific molecular fine-tuning, activating species-specific secondary metabolic pathways, for instance (Pauwels et al., 2009). To date, only a few other JA signaling factors or regulators of secondary metabolism from legumes have been identified. Comprehensive genome-wide transcriptome analyses of JA-responsive gene expression in *M. truncatula* cell suspension cultures have identified enzymes that catalyze triterpene saponin biosynthesis and a few elicitor-induced transcription factors involved in the regulation of the biosynthesis of phenolics (Suzuki et al., 2002, 2005; Achnine et al., 2005; Naoumkina et al., 2007, 2008, 2010). However, until now the molecular components of the signaling cascade(s) that steer triterpene saponin biosynthesis remain unknown.

Triterpene saponins share a common biogenic origin with sterols, molecules that are ubiquitous in eukaryotic cells. Both classes of metabolites are derived from oxidosqualene, the last common intermediate in triterpene saponin and sterol biosynthesis (Figure 1, Chapter 2). As such, it is generally accepted that triterpene saponin biosynthesis depends on the cytosolic pathway that involves mevalonate (MVA) as a key intermediate for isopentenyl diphosphate supply (Chappell, 2002). Saponin and sterol biosynthetic pathways have more points in common than mere precursors. The two pathways diverge after 2,3-oxidosqualene, which can be cyclized either by cycloartenol synthase (CAS) to yield cycloartenol, the sterol precursor, or by β-amyrin synthase (BAS) to yield β-amyrin, the precursor of many classes of triterpene saponins (Figure 1, Chapter 2). CAS and BAS share extensive sequence homology, suggesting an evolutionary recruitment of (at least some) triterpene saponin biosynthesis genes from sterol metabolism genes (Phillips et al., 2006).

In yeast and mammalian cells, the sterol content is continuously adjusted by the regulation of the levels of key synthetic enzymes, in particular of 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), by means of sterol-regulated ubiquitin-mediated protein degradation.

When cultured yeast or mammalian cells are grown in sterol-rich medium, the HMGR levels drop whereas in sterol-starved cells they increase. The constitutively active endoplasmatic reticulum (ER)-associated degradation (ERAD) quality control system directs the feedback regulated destruction of the HMGR enzyme, both in yeast and mammals (Hampton, 2002a and 2002b; Debose-Boyd, 2008; Hirsch et al., 2009; Jo and Debose-Boyd, 2010). The ERAD molecular machinery that recognizes and targets HMGR for destruction has first been characterized in yeast (*Saccharomyces cerevisiae*) and involves the RING E3-ligase HMGR DEGRADATION 1 (Hrd1) (Hampton, 2002a and 2002b). Later, two human proteins with homology to Hrd1 were identified, called HRD1 and gp78, that ubiquitinate and target HMGR for degradation, implying that the role of the ERAD system in sterol metabolism is broadly conserved (Hirsch et al., 2009). Both yeast and humans have other, distinct, E3-ligases, such as RING-FINGER PROTEIN WITH MEMBRANE ANCHOR 1 (RMA1) in humans that participate in ERAD but not in the regulation of HMGR stability (Hirsch et al., 2009).

Here, we monitored over time the methyl jasmonate (MeJA)-modulated transcriptome of *M. truncatula* cells and identified MAKIBISHI 1 (MKB1), a membrane-anchored RING E3-ligase homologous to RMA1. MKB1 operates in an ERAD-like manner to regulate the degradation of the HMGR enzymes and, thereby, steers the flux to the triterpene biosynthetic pathway in *M. truncatula*.

2. Results

2.1. Expression of regulatory genes is upregulated in elicited *M. truncatula* cells prior to the onset of secondary metabolism

In suspension-cultured cells of *M. truncatula* MeJA-inducible triterpene saponin biosynthesis has been reported (Suzuki et al., 2002; 2005; Naoumkina et al., 2007). A MeJA elicitation experiment was set up on the same cell line and under similar conditions as described by Suzuki et al. (2005), but, we covered an extensive time course sample and performed transcript profiling by means of cDNA-AFLP analysis. The expression of a total of 8,462 transcript tags was visualized and 282 MeJA-responsive tags, hereafter designated Mt tags, were identified (see chapter 2). Approximately 9% of the identified MeJA-responsive tags corresponded to genes potentially involved in transcriptional regulation of gene expression, including 14 transcription factors, and approximately 8% corresponded to signal transducing

proteins, including 5 protein kinases. 10% of the Mt tags was presumably involved in the regulation of protein synthesis and fate, including 10 proteins involved in the ubiquitin (Ub)-mediated targeting of proteins for proteasomal degradation (Figure 9C, chapter 2). From these three categories, the putative transcription factors, protein kinases, RING- and F-box-proteins were selected for further analysis. Fifteen of the selected tags had maximal transcriptional upregulation within 30 min to 2 h after elicitation, thus prior or concurrent to the onset of the genes involved in triterpene saponin and phenylpropanoid biosynthesis (Figure 9C, chapter 2). For all the 15 early induced gene tags, a full-matching EST was encountered in the DFCI Medicago Gene Index (http://compbio.dfci.harvard.edu/cgibin/tgi/gimain.pl?gudb=medicago), hence verification of the expression pattern could be achieved by consulting the *Medicago truncatula* Gene Expression Atlas (MtGEA; Benedito et al., 2008; He et al., 2009; http://bioinfo.noble.org/gene-atlas/), in which also data of MeJA-treated cell cultures are included (measured at 2 h and 24 h; Naoumkina et al., 2007). As such, for 12 of the cDNA-AFLP gene tags, the early MeJA response was confirmed.

2.2. A reverse genetics screen for novel genes involved in the regulation of legume saponin biosynthesis reveals the MAKIBISHI phenotype

To assess the function of the candidate regulatory genes, a reverse genetics screen was launched in which gain- and loss-of-function of the candidate genes was pursued by constitutive CaMV 35S-mediated overexpression of the full-length open reading frame (FL-ORF) or of a hairpin RNA-mediated interference (hpRNAi) construct, respectively, in *M. truncatula* hairy roots. When transgenic hairy roots were transferred to liquid medium for upscaling, one particular hpRNAi construct, corresponding to the gene tag Mt067, caused a striking phenotype in transgenic lines. In its most pronounced form, silencing of the *Mt067* gene, encoding a RING E3 Ub ligase, caused 'dissociation' of the hairy roots in 'caltrop-like' structures (Figure 1; Figure 2). Accordingly, the *Mt067* gene was renamed *MAKIBISHI1* (*MKB1*), referring to the Japanese term for caltrop. No such caltrop phenotypes were observed in *Mt067*/*MKB1* overexpression ($MKB1^{OE}$) lines or hairy root lines transformed with control constructs (Figure 1; Figure 2).

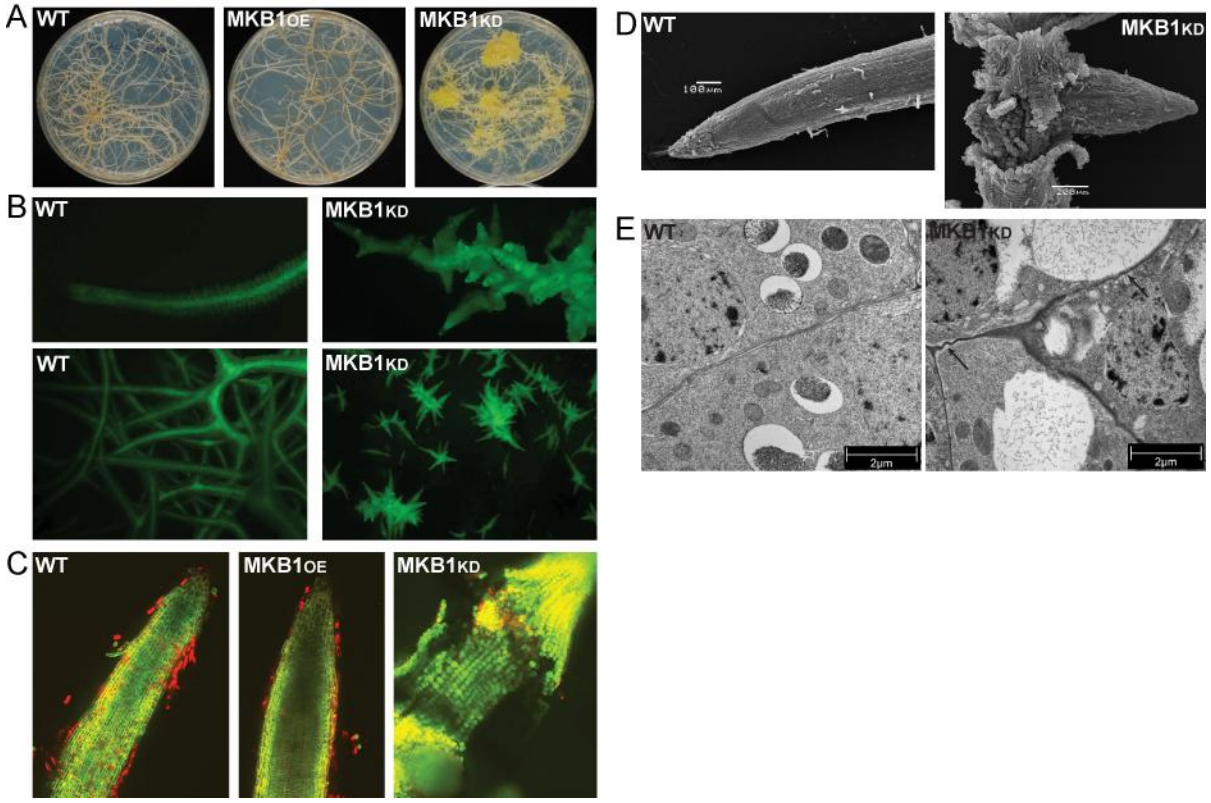


Figure 1. The MAKIBISHI phenotype. (A) Light microscopy analysis of roots of control (WT), MKB1^{OE}, and MKB1^{KD} lines grown on solid medium. (B) Light microscopy analysis of roots of control (WT) and MKB1^{KD} lines grown in liquid medium. (C) Longitudinal optical sections obtained by confocal microscopy of roots of WT, MKB1^{OE}, and MKB1^{KD} lines grown in liquid medium. (D) SEM analysis of WT and MKB1^{KD} root lines grown in liquid medium. (E) TEM analysis of cross sections from roots from WT and MKB1^{KD} lines grown in liquid medium. Arrows indicate wavy/thickened appearance of cell margins in MKB1^{KD} roots.

Microscopic analysis of the MAKIBISHI phenotype in the *MKB1* knock-down (MKB1^{KD}) lines revealed severe root growth and epidermal deficiencies as visualized by confocal microscopy on roots grown in liquid medium and taking advantage of the constitutive GFP expression provided by the T-DNA insert from the pK7GWIWG2D(II) expression vector used to generate the hpRNAi (Karimi et al., 2002). Although the MKB1^{KD} roots formed a root tip that was still quite similar to that of control roots, the root elongation zone was clearly aberrant and displayed marked loosening of the outer cell layers (Figure 1C). Interestingly, the MKB1^{KD} root phenotype strongly resembled the stunted root morphology and epidermal defects of the oat (*Avena sativa*) mutants *sad3* and *sad4*, that lack the glycosyl transferase activity needed to fully glycosylate the triterpene saponin avenacin A-1. Accumulation of incompletely glycosylated avenacin A-1 in oat roots disrupts membrane trafficking and provokes degeneration of the epidermis, with consequential effects on root hair formation and root growth (Mylona et al, 2008). The growth and epidermal defects of

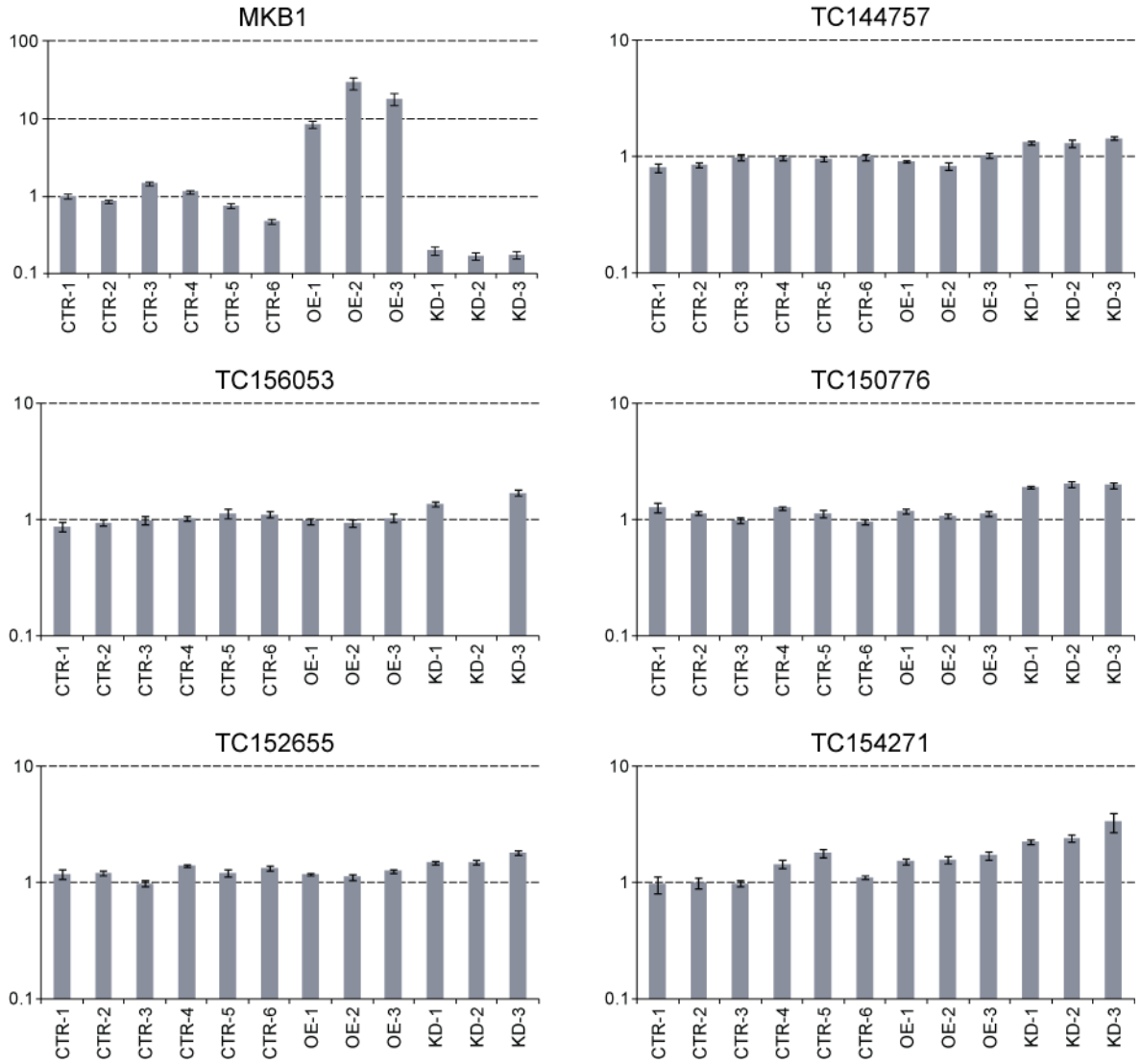


Figure 2. Expression of *MKB1* and its homologues in *M. truncatula* hairy roots. CTR, control lines; OE, MKB1^{OE} lines; KD, MKB1^{KD} lines.

MKB1^{KD} roots were further illustrated by scanning electron microscopy (SEM; Figure 1D) and transmission electron microscopy (TEM; Figure 1E). SEM analysis pointed towards a discontinuous epidermis and a resultant direct contact of cortical and other inner cells with the medium (Figure 1D). TEM analysis revealed that cell margins in MKB1^{KD} roots had a wavy appearance (Figure 1E).

2.3.RNAi-mediated silencing of *MAKIBISHI1* causes increased accumulation of monoglycosylated triterpene saponins

Considering the resemblance of MKB1^{KD} roots with oat roots deficient in saponin biosynthesis and the co-regulation of the *MKB1* and saponin biosynthesis genes in *M.*

truncatula, we performed metabolite profiling by liquid chromatography electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (LC-ESI-FT-ICRMS) to compare saponin accumulation in MKB1^{KD} and MKB1^{OE} roots with that of control roots (Figure 3; Supplementary Table S1, p216). Comparative analysis of the root extracts of control and MKB1^{OE} roots yielded a total of 19,584 *m*/*z* peaks, but notably, without significant differences between the two lines. Comparative analysis of control and MKB1^{KD} roots yielded 20,637 *m*/*z* peaks that were distributed over 269 peak groups (see Methods), indicating that at least 269 distinct compounds had been profiled. In addition to discretely present peaks (141 and 292 *m*/*z* peaks in MKB1^{KD} and control samples, respectively), a Student's *t*-test with Welch correction ($\alpha=10^{-5}$) pinpointed 15 significantly up- and 343 down-regulated *m*/*z* peaks in MKB1^{KD} roots, respectively. Whenever a peak group contained at least two discretely present or significantly changed peaks (26 up- and 89 down-regulated peak groups in MKB1^{KD} when compared to the control), they were further structurally elucidated (Supplementary Table S1, p216). Of the 26 up-regulated peak groups, six corresponded to phenolics and eight to saponins, whereas 24 and 12 of the 89 down-regulated peak groups corresponded to phenolics and saponins, respectively (Supplementary Table S1, p216). The remainder of the peak groups could not be annotated. Triglycosylated saponins were more represented among the down-regulated peaks, whereas half of all saponins that were up-regulated were monoglycosylated saponins, such as Hex-medicagenic acid, Hex-bayogenin, and Hex-hederagenin (Figure 3; Table 1).

Remarkably, the LC-ESI-FT-ICR-MS analysis of the growth medium of MKB1^{KD} roots revealed the presence of tens of compounds, whereas in the chromatograms of the growth medium from control or MKB1^{OE} roots metabolites were totally absent (Figure 3D), indicating release or leakage of compounds from the MKB1^{KD} roots. This release might (partially) account for the high number of down-regulated peaks in the MKB1^{KD} roots. However, some of the compounds that were up-regulated in the MKB1^{KD} roots occurred also in the medium, such as the monoglycosylated saponins Hex-medicagenic acid, Hex-bayogenin, and Hex-hederagenin (Figure 3). Collectively, these findings suggest that silencing of *MKB1* primarily causes deficient, ectopic, or de novo biosynthesis of triterpene saponins, leading to the overaccumulation and release of incompletely glycosylated saponins.

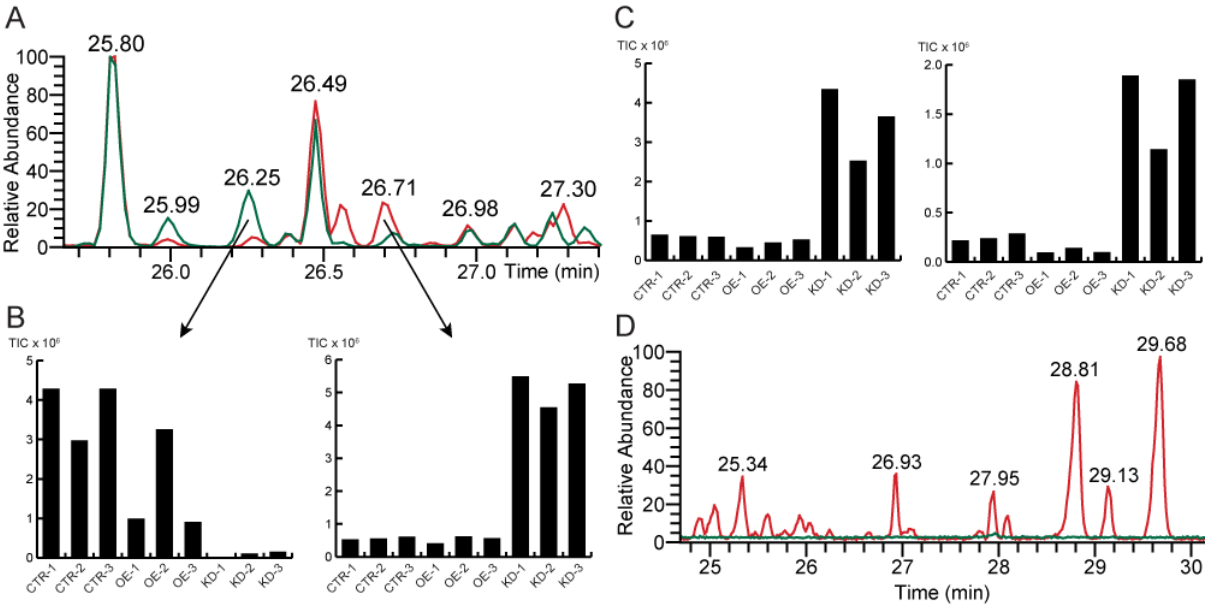


Figure 3. Saponin profiling of MKB1^{OE} and MKB1^{KD} roots. (A) Detail of the full MS scan of a control line (green) and a MKB1^{KD} line (red). (B) Average Total Ion Current (TIC) of the main masses corresponding to the peaks at retention time 26.25 (unknown compound) and 26.71 (Hex-medicagenic acid). (C) Average Total Ion Current (TIC) of the main masses corresponding to Hex-bayogenin (left) and Hex-hederagenin (right). (D) Detail of the full MS scan of the medium from a control line (green) and a MKB1^{KD} line (red). The peak at t_R 27.95 represents Hex-medicagenic acid.

2.4. Accumulation of monoglycosylated triterpene saponins triggers the MAKIBISHI phenotype

By crossing with an oat mutant blocked in the first committed step in the avenacin pathway, the presence of incompletely glycosylated avenacin A-1 was shown to be the cause rather than the consequence of the membrane trafficking, epidermal, and root defects in the *sad3* and *sad4* oat mutants (Mylona et al., 2008). To evaluate whether this was also the case in the *M. truncatula* transgenic roots we tried to phenocopy the MKB1^{KD} phenotype in the absence of suitable mutant *M. truncatula* backgrounds. Therefore, aliquots of the culture medium of MKB1^{KD} roots that also contained the incompletely glycosylated saponins were applied to control root cultures. Assessment of the subsequent root growth demonstrated that the MKB1^{KD} root culture medium could cause epidermal loosening of control roots (Figure 4A). The strength of the phenotype varied depending on the concentration and the age of the MKB1^{KD} root culture from which the medium was collected. The epidermal defects were never obtained by supplementing the culture medium with similar or higher doses of medium from control roots that is devoid of saponins (Figure 4A). The phenocopy was still observable when an MKB1^{KD} root culture aliquot was applied that had passed through a protein filter

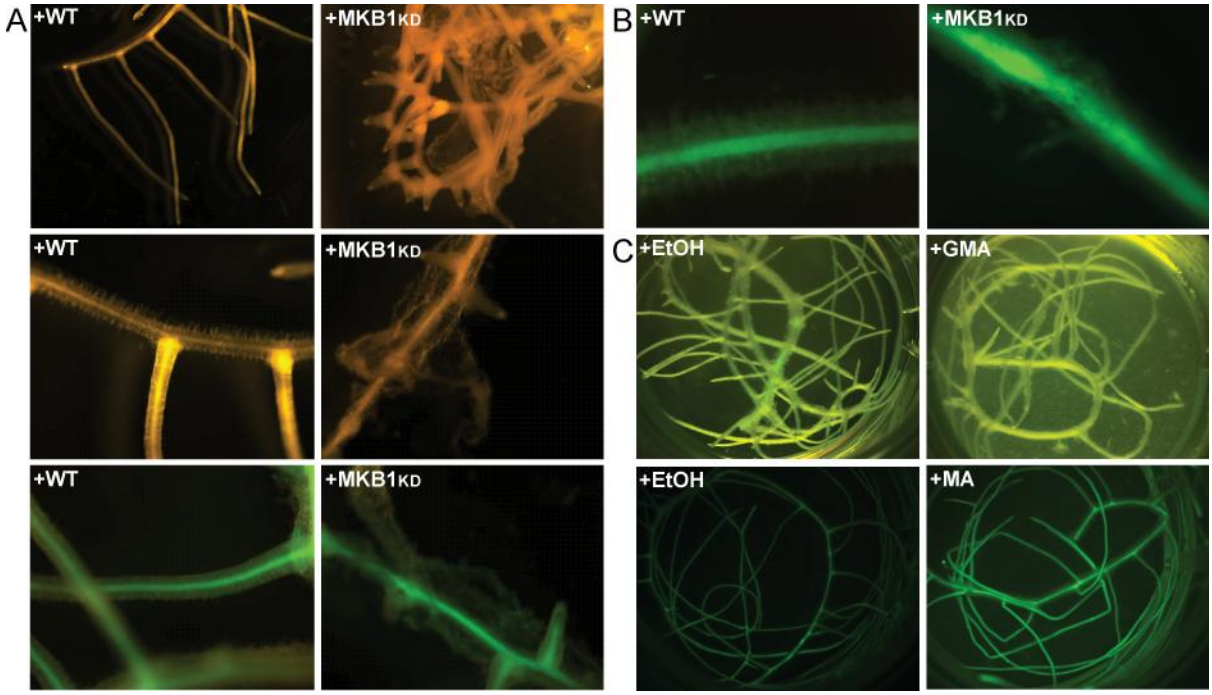


Figure 4. Phenocopy of the MKB1^{KD} phenotype. (A) Light microscopy analysis of control hairy roots incubated for 1 week in medium supplemented with medium from control (WT; left) or MKB1^{KD} (right) root lines. (B) Light microscopy analysis of control hairy roots incubated for 1 week in medium supplemented with medium from control (WT; left) or MKB1^{KD} (right) root lines that was passed through a protein filter with a cut-off of 5 kDa. (C) Light microscopy analysis of control hairy roots incubated for 1 day in medium supplemented with purified 3-O-Glc-medicagenic acid (GMA), medicagenic acid (MA) or an equivalent amount of the solvent (EtOH).

before application, implying that small molecule(s), such as incompletely glycosylated saponins might cause the MAKIBISHI phenotype (Figure 4B). No phenocopy was observed after treatment with individual hormones (e.g., gibberellins, cytokinins, JAs, auxins, or anti-auxins) or with commercially available compounds, such as the flavonoid quercetin or the fully glycosylated soyasaponin I, the most abundant compound in control hairy roots and abundantly present in the medium of MKB1^{KD} roots. However, a transient phenocopy was obtained after addition of purified 3-*O*-Glc-medicagenic acid (Oleszek et al., 1990) to the growth medium of control root cultures, but not of purified medicagenic acid (Supplemental Figure 2C). Collectively, these findings suggest that the ectopic biosynthesis of one or more monoglycosylated saponins accounts, at least partially, for the root and epidermis defects in the MKB1^{KD} roots.

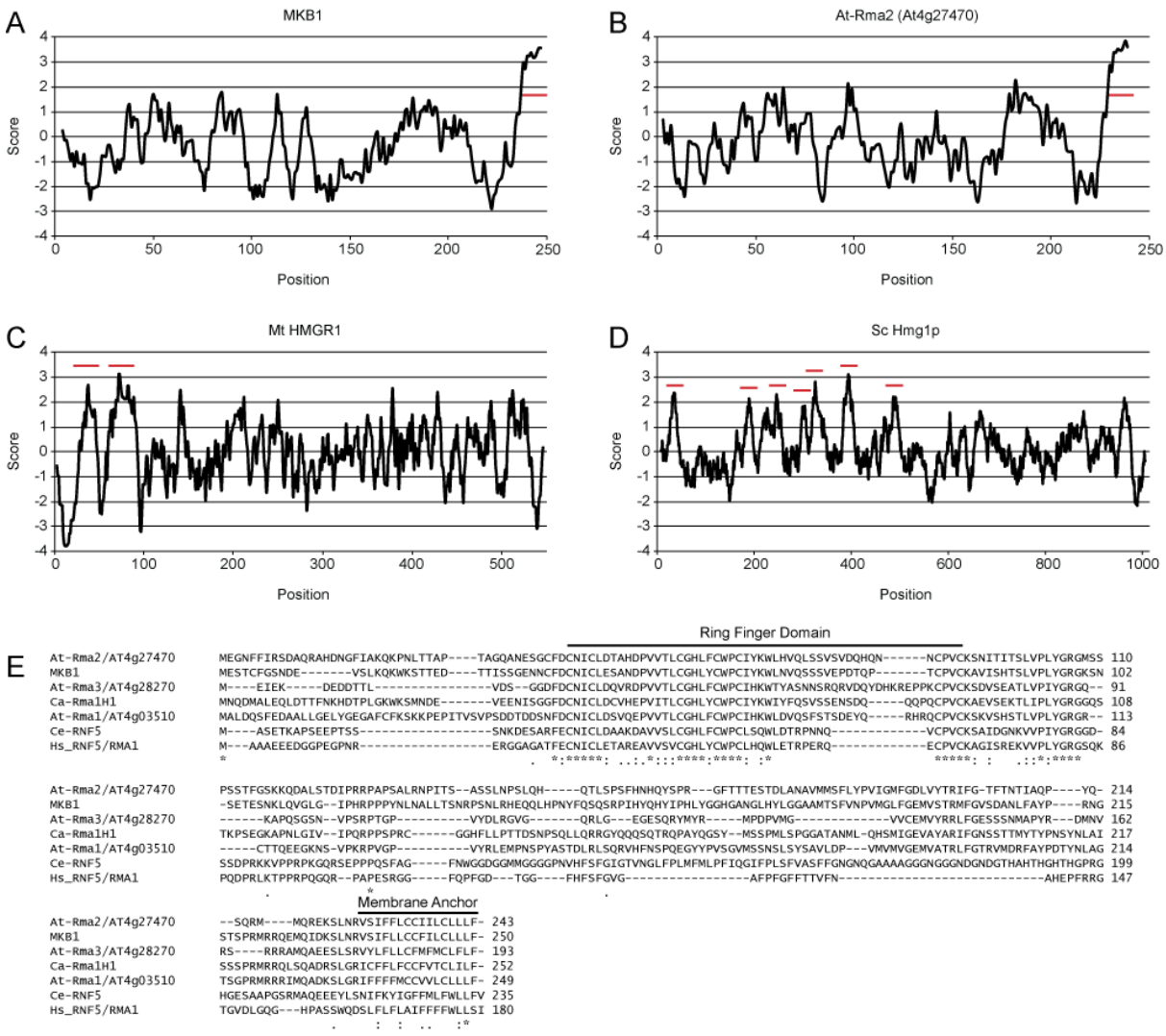


Figure 5. Sequence and structural comparison of eukaryote RMA and HMGR proteins. (A) Kyte & Doolittle hydropathy plot of MKB1 with window size 7. The red bar indicates the membrane anchor. (B) Kyte & Doolittle hydropathy plot of Arabidopsis Rma2 (At4g27470) with window size 7. The red bar indicates the membrane anchor. (C) Kyte & Doolittle hydropathy plot of *M. truncatula* HMGR1 with window size 7. The two red bars indicate the hydrophobic transmembrane domains. (D) Kyte & Doolittle hydropathy plot of *S. cerevisiae* Hmg1p with window size 15. The seven red bars indicate the hydrophobic transmembrane domains. (E) Comparison of the amino acid sequence of MKB1 with the Arabidopsis Rma 1 (At4g03510), Rma2 (At4g28270), and Rma3 (At4g27470), the hot pepper Rma1H1, the *C. elegans* RNF5, and the human RNF5/Rma1 proteins. Conserved amino acids that are identical in the 7 proteins are indicated with an asterisk. The solid lines indicate the N-terminal RING Finger Domain and the C-terminal Membrane Anchor, respectively.

2.5.MAKIBISHI1 encodes a RING Membrane-Anchor E3 Ubiquitin Ligase Homolog

MKB1 encodes a protein of 250 amino acids, with a calculated mass of 27.97 kDa and a predicted pI of 7.61. The MKB1 protein contains a single N-terminal C3HC4-type RING

domain (Kosarev et al., 2002) and hydropathy analysis (Kyte and Doolittle, 1982) revealed the presence of a short hydrophobic domain, a putative membrane anchor, at the C-terminus of the protein (Figure 5; Figure 6A). BLAST searches and phylogenetic analysis showed that MKB1 is closely related to the *Arabidopsis* RMA proteins, with RMA2 (At4g27470) being the closest homolog (49% identity at the amino acid level). *Arabidopsis* RMA proteins also correspond to membrane-anchored RING E3 Ub ligases (Matsuda et al., 2001; Lee et al., 2009) and are conserved in higher eukaryotes, but absent in yeast (Matsuda et al., 2001). Sequence analysis showed that MKB1 shares 24% and 22% homology with human RMA1 and nematode RNF5 RMA-type RING E3 Ub ligases, respectively (Figure 5; Figure 6B). Screening of the sequences present in the DFCI Medicago Gene Index (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago) indicated that *M. truncatula* encodes at least five other RMA-type proteins that are closely related to MKB1 (Figure 6B). qRT-PCR analysis of these homologues revealed silencing in the MKB^{KD} lines is specific for MKB1 (Figure 2).

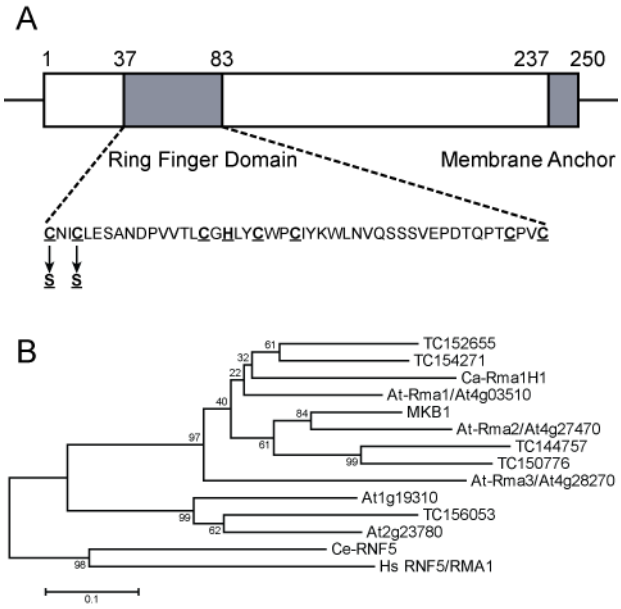


Figure 6. Phylogenetic analysis of MKB1 and other membrane anchored RING E3 Ub ligases. (A) Schematic representation of the MKB1 protein and its domain structure. (B) Phylogenetic tree of membrane anchored RING E3 Ub ligases. The percentage of replicate trees that clustered together in the bootstrap test is shown next to the branches.

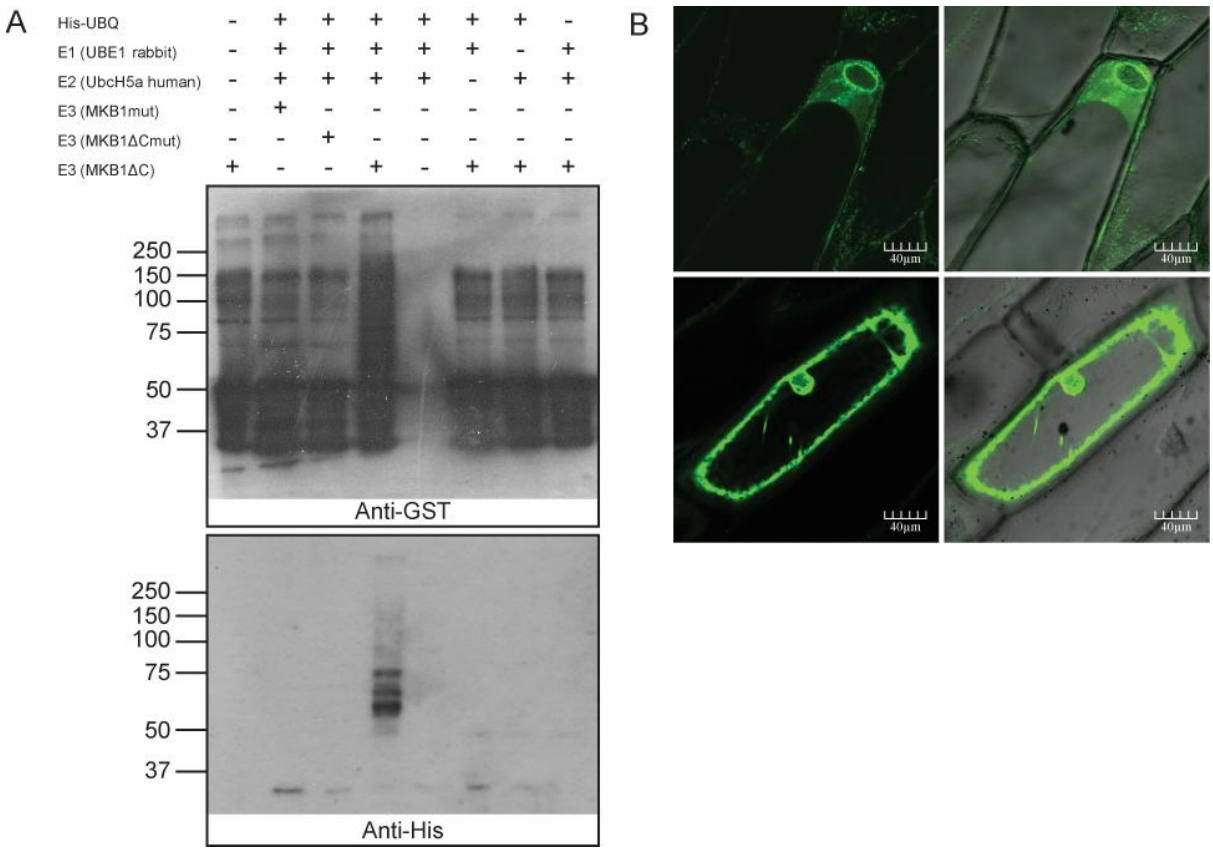


Figure 7. MKB1 is an ER-localized protein with auto-ubiquitination activity. (A) *In vitro* auto-ubiquitination assay of MKB1. The bacterially expressed GST-MKB1 constructs were incubated with ATP in the presence or absence of his-tagged Ub, E1 (rabbit UBE1), and E2 (human UbcH5a). Samples were resolved by 8% SDS-PAGE, followed by protein immunoblot analysis using anti-GST (top) or anti-His (bottom) antibodies. (B) Subcellular localization of MKB1 in bombarded onion cells. The pictures show the GFP signal and the GFP-bright light merged image (left and right, respectively) of GFP-MKB1 and GFP-MKB1ΔC (top and bottom, respectively).

2.6.MAKIBISHI1 localizes to the ER and possesses E3 Ub ligase activity *in vitro*

Arabidopsis and mammalian RMA proteins localize to the ER and possess E3 Ub ligase activity in vitro (Matsuda et al., 2001; Lee et al., 2009). To verify the ubiquitination activity of MKB1, a truncated version of the MKB1 protein, lacking the membrane anchor domain (MKB1ΔC) was fused to the glutathione S-transferase (GST) tag, produced and purified from *Escherichia coli* cells. The recombinant protein was incubated with His-tagged Ub, ATP, and E1 (rabbit UBE1), and E2 (human Ubc5Ha) enzymes and subjected to immunoblot analysis with anti-GST and anti-His antibodies. GST-MKB1 E3 Ub ligase activity was demonstrated by the high molecular mass smear ladders, in which discrete bands corresponding to ubiquitinated GST-MKB1ΔC could be observed. No such signal was obtained when E1, E2, or Ub were omitted from the reaction mixture (Figure 7A). Next, a double amino-acid-

mutated version of the GST-MKB1ΔC fusion protein was created in which the amino acid residues Cys^{37} and Cys^{40}, essential within the RING domain (Kosarev et al., 2002), were substituted by Ser residues (Figure 6A). As expected, the corresponding recombinant protein (GST-MKB1ΔCmut) did not exhibit ubiquitination activity anymore (Figure 7A), thus indicating that MKB1 represents a bona fide E3 Ub ligase.

To verify the presumed ER localization of MKB1, we generated GFP-tagged versions of both the FL MKB1 and the truncated MKB1ΔC. The fusion proteins were produced in bombarded onion (*Allium cepa*) cells that were analyzed by confocal microscopy. The GFP-MKB1 protein was visible in a network pattern, reminiscent of ER-targeted proteins (Lee et al., 2009). This pattern was lost and replaced by a cytosolic distribution pattern when the GFP-MKB1ΔC protein was produced in onion cells (Figure 7B). Thus, like its homologs in other organisms, MKB1 corresponds to an active RMA-type ER-localized E3 Ub RING ligase.

2.7.MAKIBISHI1 targets HMGR for degradation

Arabidopsis RMA1 protein has recently been shown to regulate aquaporin levels by ubiquitination and subsequent proteasomal degradation (Lee et al., 2009). Similarly, human RMA proteins are involved in the ERAD quality control system that ensures that only properly folded proteins are released to their appropriate destinations (Hirsch et al., 2009). Furthermore, humans and yeasts have other ER-localized RING E3 Ub ligases, i.e., the HRD-type, that play a crucial role in the control of sterol synthesis through the regulation of HMGR levels (Hampton, 2002a and 2002b; DeBose-Boyd, 2008; Hirsch et al., 2009; Jo and Debose-Boyd, 2010). To our knowledge, a role for RMA proteins in the latter process has not been reported yet. Nonetheless, despite the apparent lack of sequence similarity between the two types of RING ligases and the different membrane topology of the HMGR enzymes from plants, yeasts and mammals (Figure 8; Figure 9A), we reasoned that the *M. truncatula* MKB1 homolog of the RMA proteins might target the HMGR, just like the HRD proteins. Two observations indirectly supported this hypothesis: (i) $MKB1^{KD}$ roots accumulated incompletely glycosylated saponins, pointing to a defect in the saponin biosynthetic pathway or the transport chain and (ii) mining of the gene expression data in the MtGEA revealed an extraordinarily high expression correlation between *MKB1* and *M. truncatula HMGR1* (Figure 8) but not between *HMGR1* and the putative *M. truncatula* homolog of yeast *HRD1* (Figure 8).

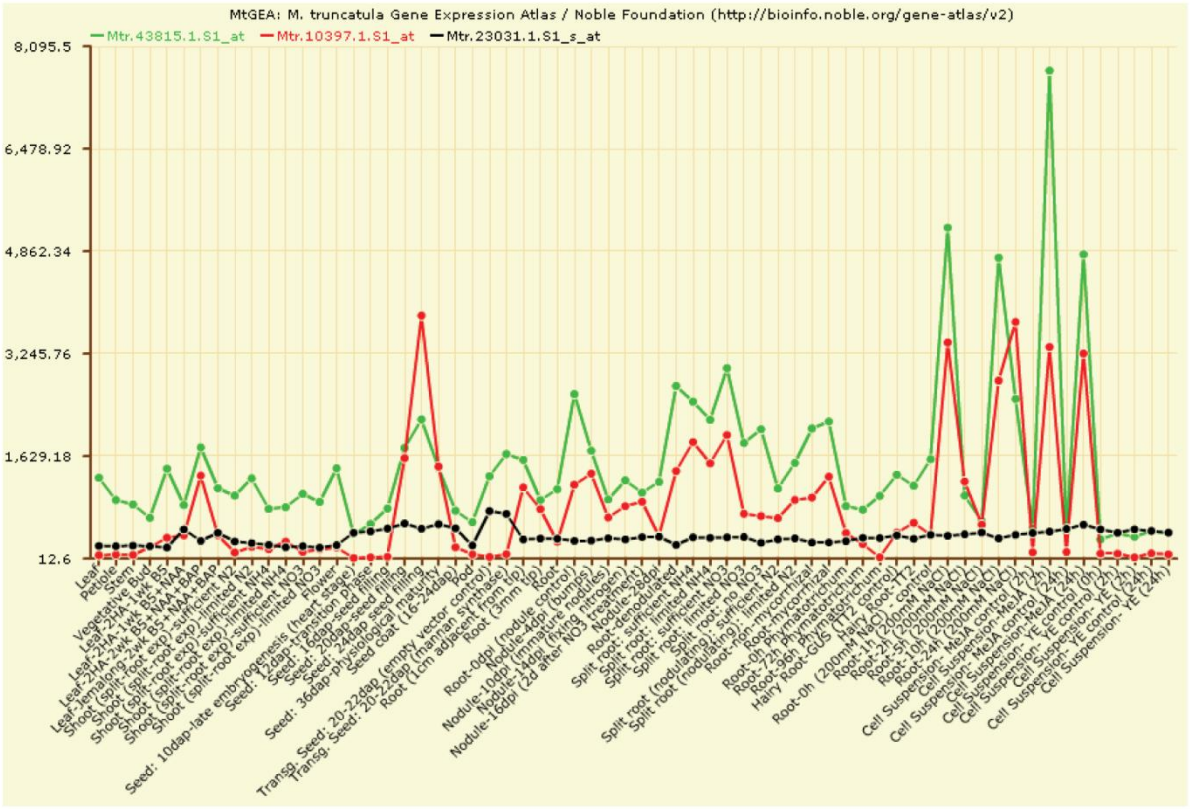


Figure 8. Co-expression of *M. truncatula MKB1* and *HMGR1*. Screenshot of the co-expression analysis of *M. truncatula MKB1* (green), *HMGR1* (red) and *HRD1* (black) with the MtGEA software.

To experimentally support this hypothesis we first checked whether MKB1 could target yeast HMGR and thereby complement a yeast strain devoid of functional HRD1. Complementation was carried out in a yeast *hrd1-1* strain and scored by determining the ability of transformed strains to grow on lovastatin containing medium. Strikingly, the lovastatin tolerance of the *hrd1-1* strain was returned to wildtype levels after transformation with *MKB1* but not with the ligase-dead *MKB1mut* version, demonstrating that the E3 ligase activity of MKB1 is required to overcome HRD1 deficiency (Figure 9B). To verify that the *hrd1-1* complementation resulted from the re-established HMGR degradation, we carried out cycloheximide (CHX) chase assays and monitored the levels of the 6myc-tagged Hmg2p protein in the transformed yeast strains. In the wildtype *HRD1* yeast background, the tagged Hmg2p was nearly absent after a 4 h CHX chase, whereas in the *hrd1-1* strain, it could still be detected. Basal Hmg2p levels were also significantly higher in the *hrd1-1* strain (Figure 9C). In agreement with the lovastatin growth complementation assay, expression of FL wildtype *MKB1*, but not the ligase-dead *MKB1mut*, restored the Hmg2p levels to those of the *HRD1* strain, both before and after CHX chase (Figure 9C). Together, these results indicate that, despite the extensive

sequence divergence, MKB1 can functionally complement the yeast HRD1 and, thereby, target the yeast HMGR for Ub-mediated degradation.

To experimentally support this hypothesis we first checked whether MKB1 could target yeast HMGR and thereby complement a yeast strain devoid of functional HRD1. Complementation was carried out in a yeast *hrd1-1* strain and scored by determining the ability of transformed strains to grow on lovastatin containing medium. Strikingly, the lovastatin tolerance of the *hrd1-1* strain was returned to wildtype levels after transformation with *MKB1* but not with the ligase-dead *MKB1mut* version, demonstrating that the E3 ligase activity of MKB1 is required to overcome HRD1 deficiency (Figure 9B). To verify that the *hrd1-1* complementation resulted from the re-established HMGR degradation, we carried out cycloheximide (CHX) chase assays and monitored the levels of the 6myc-tagged Hmg2p protein in the transformed yeast strains. In the wildtype *HRD1* yeast background, the tagged Hmg2p was nearly absent after a 4 h CHX chase, whereas in the *hrd1-1* strain, it could still be detected. Basal Hmg2p levels were also significantly higher in the *hrd1-1* strain (Figure 9C). In agreement with the lovastatin growth complementation assay, expression of FL wildtype *MKB1*, but not the ligase-dead *MKB1mut*, restored the Hmg2p levels to those of the *HRD1* strain, both before and after CHX chase (Figure 9C). Together, these results indicate that, despite the extensive sequence divergence, MKB1 can functionally complement the yeast HRD1 and, thereby, target the yeast HMGR for Ub-mediated degradation.

To assess whether MKB1 also targets plant HMGR proteins, we produced a firefly luciferase (fLUC)-tagged version of the *M. truncatula* HMGR1 protein in tobacco (*Nicotiana tabacum*) protoplasts combined or not with MKB1. This analysis showed that cotransfection with MKB1 decreased the HMGR1-fLUC signal almost by 2-fold, whereas it did not reduce the activity of free fLUC (Figure 9D), indicating that MKB1 could target *M. truncatula* HMGR proteins for Ub-mediated degradation.

2.8. Altered *MKB1* expression alters expression of triterpene saponin but not of sterol biosynthesis genes

Having established the role of MKB1 in the regulation of HMGR levels and, hence, in the control of the flux toward the triterpene saponins, we assessed whether loss- or gain-of-function of MKB1 in the transgenic $MKB1^{KD}$ and $MKB1^{OE}$ lines, respectively, altered the steady-state levels of the gene transcripts corresponding to enzymes involved in the triterpene

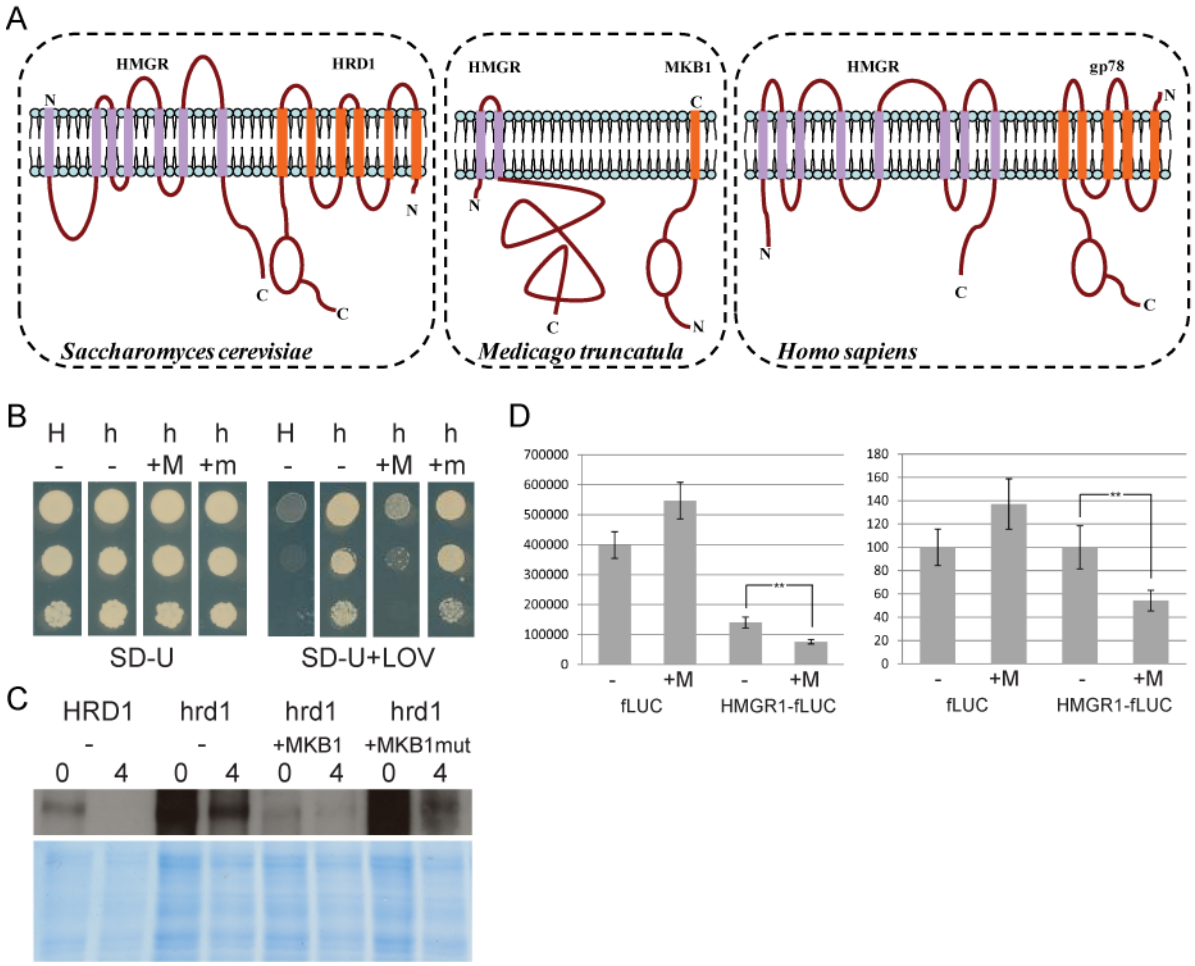


Figure 9. MKB1 targets HMGR. (A) Schematic overview of the topology of HMGR enzymes and ERAD E3 Ub ligases from yeast, *M. truncatula* and human. (B) Complementation of the yeast *hrd1-1* mutant. HRD1 (H) or hrd1-1 (h) Yeast cells were transformed with MKB1 (M) or MKB1mut (m), spotted in 10- or 100-fold dilutions on SD medium lacking URA supplemented (SD-U+LOV) or not (SD-U) with lovastatin and grown for 2 days at 30ºC. The destination vector pAG426GPD was used as a control (-). (C) Degradation of yeast Hmg2p. Immunoblot analysis for 6myc-tagged Hmg2p protein in HRD1 or hrd1-1 yeast cells transformed with MKB1 or MKB1mut, 0 h and 4 h after CHX treatment. The destination vector pAG426GPD was used as a control (-). (D) Degradation of *M. truncatula* HMGR1-fLUC in tobacco protoplasts. Left and right panels indicate the absolute fLUC values obtained with each construct in each transfection (numbers in the Y-axis) and the percentage stability relative to the fLUC value measured in the absence of MKB1 (set at 100%), respectively. Error bars indicate the standard error (n=24). Statistical significance was determined by Student's *t*-test (** P<0.01).

and other (secondary) metabolic pathways. First, the steady-state levels of the gene corresponding to the MKB1 target, HMGR, for which five isoforms are present in the *M. truncatula* genome (Kevei et al., 2007), were measured. Although some fluctuations –but generally less than 1.5-fold- were observed, expression of none of the five *HMGR* genes, nor of the *SQS*, *SQE1*, and *SQE2* genes, was markedly altered in the MKB1^{KD} lines (Figure 10),

indicating that expression of the genes encoding the enzymes catalyzing triterpene synthesis up to the oxidosqualene precursor was not affected by silencing of the *MKB1* gene. Similarly, steady-state levels of the putative *M. truncatula* homologs of genes corresponding to sterol biosynthesis enzymes from *Arabidopsis*, such as cycloartenol synthase (CAS), sterol-C24-methyltransferase (C24MT and CVP1), obtusifoliol 14α-demethylase (CYP51G1), sterol C-14 reductase (FACKEL), sterol C-8,7 isomerase (HYDRA1), and CYP710A15 (Schrick et al., 2004; Morikawa et al., 2006) were not altered either, indicating that alteration in the expression of *MKB1* did not affect the transcriptional regulation of sterol biosynthesis (Figure 11).

In contrast, the expression of almost all of the hitherto known triterpene saponin biosynthetic genes was strongly downregulated in the MKB1^{KD} lines, pointing to a possible specific feedback mechanism caused by the ectopic accumulation of incompletely glycosylated saponins. *BAS* and *UGT73F3* transcript levels were reduced 2- to 4-fold, whereas expression of *UGT73K1* was reduced 5- to 10-fold (Figure 10). Also, the expression of the *Mt175-CYP93E2* gene, which is the *M. truncatula* gene with the highest homology to the soybean CytP450 gene encoding the β-amyrin and sophoradiol 24-hydroxylase (Shibuya et al., 2006), was reduced 5- to 50-fold in the MKB1^{KD} lines (Figure 10). The only presumed saponin biosynthesis gene of which the expression was not affected by silencing of *MKB1*, was *UGT71G1* (Figure 10). However, in contrast to the other two glycosyltransferases, UGT71G1 also recognized isoflavones and flavonols as substrates and glycosylated these compounds with higher efficiencies than triterpenes (Achnine et al., 2005), suggesting it might not correspond to a saponin-specific biosynthesis gene. The expression of any of the tested metabolic pathway genes did not differ in MKB1^{OE} lines (Figure 10; Figure 11), which is in accordance with the lack of any observable (metabolic) phenotype in the MKB1^{OE} lines.

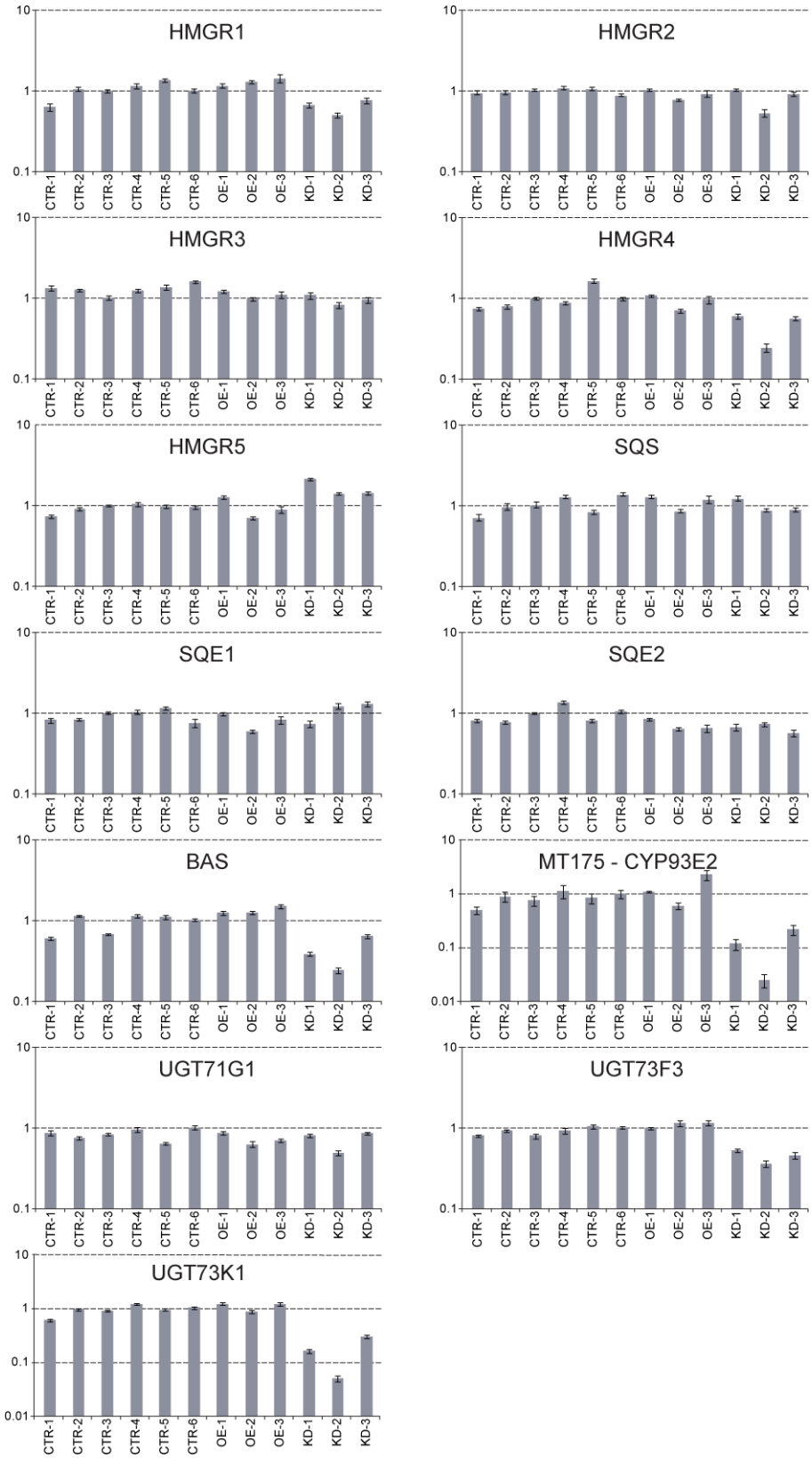


Figure 10. qRT-PCR analysis of saponin biosynthetic genes in MKB1^{KD} and MKB1^{OE} lines. CTR, control lines; OE, MKB1^{OE} lines; KD, MKB1^{KD} lines.

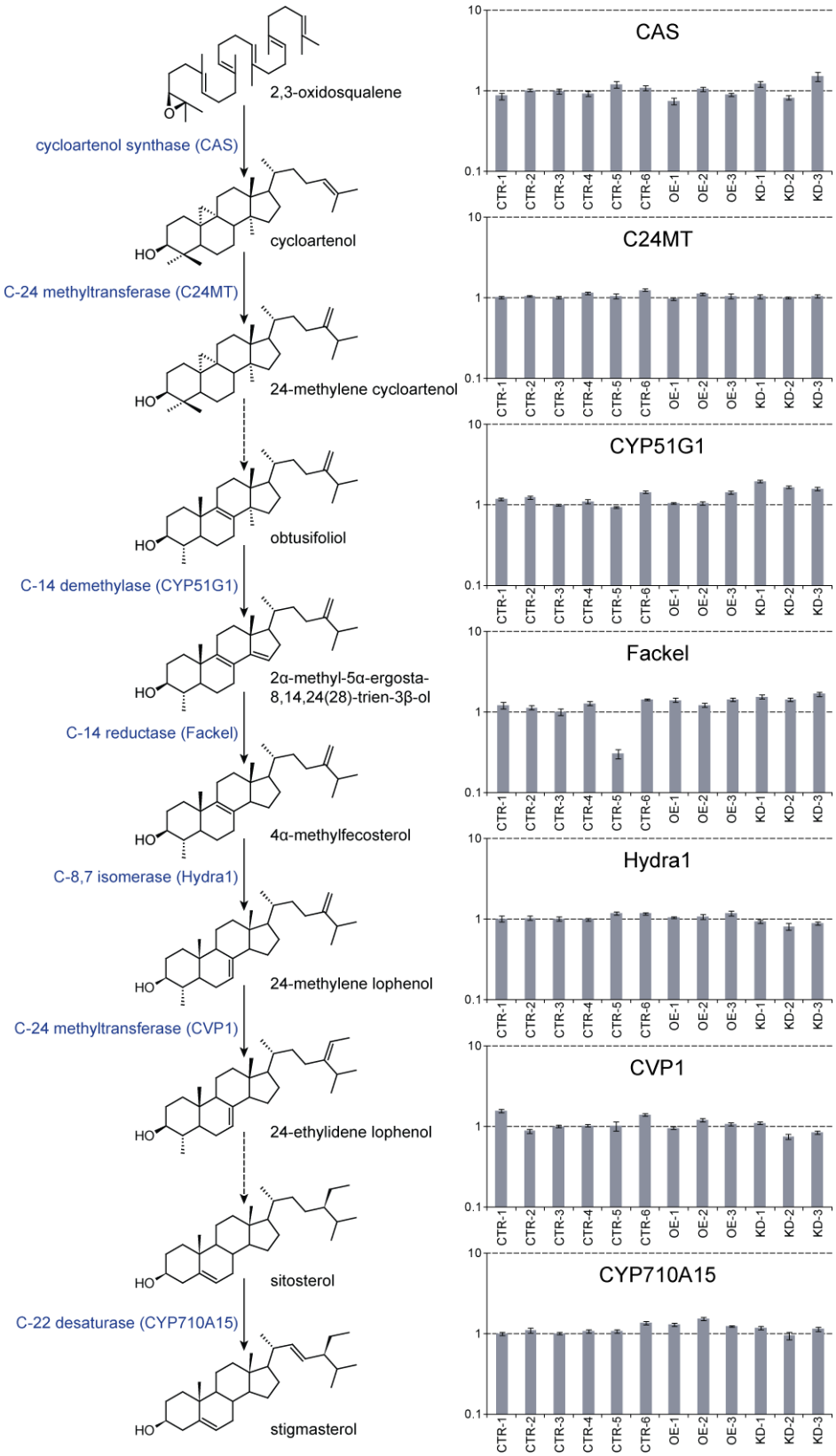


Figure 11. qRT-PCR analysis of sterol biosynthetic genes in MKB1^{KD} and MKB1^{OE} lines. A schematic overview of the sterol biosynthesis pathway and the qRT-PCR patterns are shown left and right, respectively. CTR, control lines; OE, MKB1^{OE} lines; KD, MKB1^{KD} lines.

3. Discussion

3.1. Importance of the MVA pathway in legume triterpene saponin biosynthesis

In this study, a cDNA-AFLP-based transcriptome analysis of a time series covering the early response of *M. truncatula* cells to MeJA elicitation was carried out, with the aim to unravel the signaling cascades that steer triterpene saponin biosynthesis. These plant natural products share a common biogenic origin with the ubiquitous sterols since they are both derived from oxidosqualene and depend on the cytosolic pathway that involves MVA as a key intermediate for isopentenyl diphosphate supply. In yeast and mammals, the activity of HMGR, encoding the rate-limiting enzyme in MVA biosynthesis, is regulated from the transcriptional to the posttranslational level. Many of the regulatory mechanisms involved have been characterized in these organisms, but, in plants, little is known to date. Mostly, regulation at the transcriptional level has been reported, for instance, in response to several environmental and developmental factors or to perturbations of the metabolic flux through the sterol pathway (Nieto et al., 2009). Accordingly, in MeJA-elicited *M. truncatula* cells, the MVA pathway is transcriptionally upregulated, as reflected by the co-activation of different isoforms of the HMGR gene with the known saponin-specific biosynthesis genes (Figure 1, chapter 2). Hence, transcriptional upregulation of the MVA pathway may indeed account for or, at least, contribute to the MeJA-mediated increase in saponin accumulation in legume species.

Also in plants, mechanisms operating at the posttranslational level appear to control the HMGR activity, as suggested by the observation that some environmental or developmental factors or perturbations trigger an increase in HMGR activity but no changes in the HMGR transcript levels (Nieto et al., 2009). However, to our knowledge, a possible posttranslational regulatory mechanism of plant HMGR has been reported only once, in a study of plant SNF1-related protein kinases (SnRKs) that phosphorylate and, thereby, inactivate *Arabidopsis* HMGR (Dale et al., 1995). Here, we showed that MeJA-modulated (transcriptional) elicitation of saponin biosynthesis in *M. truncatula* was co-regulated with the transcriptional induction of a RING E3 Ub ligase, *MKB1*, that controls the stability of HMGR through a mechanism analogous to the ERAD system directing the sterol-regulated destruction of the HMGR enzyme in yeast and mammals, but hitherto undetected in plants.

3.2.*M. truncatula* uses an ERAD system different from the one that regulates HMGR stability in yeast and mammals

Remarkably, MKB1 does not belong to the HRD-type RING E3 Ub ligases known to be involved in the ERAD system that controls HMGR stability in both yeast and mammals, but instead to the RMA-type of RING E3 Ub ligases. In mammalian cells, RMA1 is involved in the ERAD system that targets aberrant or misfolded proteins for degradation by the 26S proteasome. Thus far, no evidence for an involvement of mammalian RMA1 in ERAD surveyed HMGR degradation has been reported, in contrast to the mammalian HRD-type E3 ligase gp78, which, like its yeast counterpart HRD1, seems to be the sole determinant in the HMGR control system. Furthermore, no RMA-type ERAD ligase equivalents exist in yeast or fungi, suggesting that sterol-mediated feedback control of HMGR stability by HRD-type ligases is broadly conserved and strictly an HRD-type ERAD matter. Nonetheless, we show that MKB1 can complement the yeast *hrd1-1* mutant, implying that RMA-type ERAD ligases can take over some of the functions of HRD proteins in the absence of the latter, at least in yeast, and likely use the same molecular machinery as HRD1.

Furthermore, some common mechanistic principles between the yeast/mammalian HRD-type and the *M. truncatula* RMA-type ERAD ligases in the control of HMGR stability were apparent. In mammalian cells, coexpression of the Insig proteins is needed to restore the resistance of overproduced HMGR to sterol-induced ubiquitination and degradation suggesting that ectopic reductase saturates the endogenous ERAD machinery. Accordingly, RNAi against *Insig* or *Gp78* abrogates sterol-mediated HMGR ubiquitination and degradation (Sever et al., 2003), whereas RNAi against the mammalian Hrd1, not involved in HMGR control, does not affect HMGR ubiquitination (Jo and Debose-Boyd, 2010). Similarly, in *M. truncatula*, RNAi against *MKB1*, the RMA-protein shown to target at least one isoform of the *M. truncatula* HMGR proteins can apparently provoke an uncontrolled flux through the saponin pathway, leading to the accumulation of incompletely glycosylated saponins. The increased and/or ectopic accumulation of partially glycosylated saponins is harmful as illustrated by the dramatic phenotype of $MKB1^{KD}$ roots, which is analogous to that of an oat mutant deficient in avenacin A1 glycosyltransferases and accumulating incompletely glycosylated avenacin A1 (Mylona et al., 2008). Furthermore, the possible toxicity of ectopically accumulated partially glycosylated saponins is supported by structure-activity relationship studies that indicated that the monodesmoside derivatives of medicagenic acid

possessed a higher hemolytic activity, as well as a higher anti-fungal and allelopathic activity than the related bi- and tridesmosides (Oleszek, 1990; Voutquenne et al., 2002; Tava and Avato, 2006). Hemolytic activity is generally attributed to membrane bursts caused by the interaction between the saponins and the sterols of the erythrocyte membrane. Hence, co-induction of *MKB1* by MeJA treatment might be a safety break to prevent saturation of the ERAD system and to restrict the flux to the saponin pathway. This hypothesis is further supported by the marked co-regulation of *M. truncatula HMGR1* and *MKB1* observed in the MtGEA.

3.3.Conservation of ERAD modifier activities and substrate sequences

Plant SnRKs phosphorylate and inactivate HMGR, for instance *Arabidopsis* HMG1 at a site (Ser-577) equivalent to that at which the mammalian HMGR is phosphorylated by AMPK, a major metabolic regulator in mammals (Dale et al., 1995). Hence, the activity of the modifier and the primary sequence of the physiological substrate might be conserved between plants and mammals, although the signaling molecule that activates plant SnRKs is not AMP (Dale et al., 1995). For the ERAD control of HMGR stability, such a conservation of the primary sequence of the substrate is seemingly improbable. Point mutations within the HMGR membrane domain, which is also the sterol-sensing domain, at amino acid residues that are conserved between human, hamster, *Xenopus*, zebrafish, sea urchin, and, to some extent, also *Drosophila*, disrupt Insig binding and abolish sterol-accelerated degradation of the HMGR enzyme (Lee et al., 2007). In mammals, the HMGR transmembrane domain is also the site where ubiquitin is conjugated (Doolman et al., 2004). In contrast to the high conservation of the catalytic C-terminal domain, the N-terminal, membrane-spanning domain of plant HMGR proteins is highly divergent and little primary sequence conservation can be observed. For instance, the amino acid residues important for sterol-sensing in animal HMGR proteins are not conserved. Furthermore, the N-terminal domain is truncated and lacks the complex membrane-spanning architecture, i.e. plant HMGR proteins only possess two membrane-spanning domains, whereas in yeast and human, they contain seven and eight transmembrane domains, respectively (Figure 5; Figure 9A) (Learned and Fink, 1989). Hence, recognition of the plant HMGR proteins by RMA-type ligases, such as MKB1, might have evolved alternative recognition and ubiquitination sites. Alternatively, the ERAD-regulated degradation of plant HMGR might depend on a ‘distributed degron’, requiring structural features distributed over the entire transmembrane domain and not visible in primary

sequence motifs, as found for yeast Hmg2p (Gardner and Hampton, 1999; Hampton, 2002b). Although plant HMGR proteins have a truncated membrane domain, such structural features might have been conserved. In support of the latter hypothesis, despite the structural resemblance, the primary sequence resemblance between the membrane domains of yeast and mammalian HMGR proteins is only limited (<20% identity over 340 amino acids; (DeBose-Boyd, 2008)).

The discovery of MKB1 and its role in the regulation of HMGR stability generates several other intriguing questions. Is a role for the RMA-type E3 Ub ligases in the ERAD control of HMGR stability conserved in plants? Only very recently, the first putative substrates for two of the *Arabidopsis* RMA ligases, AtRMA1 (At4g03510) and AtRMA2 (At4g28270), have been found, namely aquaporin and the auxin receptor auxin-binding protein 1, respectively (Lee et al., 2009; Son et al., 2010), but the regulatory mechanisms involved have not been elucidated yet. Assuming a conserved role for RMA proteins in the ERAD control of plant HMGR stability, would it have evolved de novo and independently of the HRD system, or conversely, would it cooperate with plant HRD proteins in targeting HMGR for degradation? Like mammals, plants seem to possess both ERAD systems, in contrast to fungi with only the HRD system. For instance, besides at least three RMA-type proteins (Lee et al., 2009), BLAST searches indicate that *Arabidopsis* encodes two HRD-like proteins (At1g65040 and At3g16090; http://www.arabidopsis.org/). Similarly, at least two putative HRD proteins (Medtr4g106990 and Medtr5g007790; http://www.medicago.org/genome/IMGAG/) can be identified in the *M. truncatula* genome. More *M. truncatula* HRD homologs might exist, because the genome sequencing is still incomplete. The gene corresponding to MKB1 is still absent in the current version, but one RMA-type homolog is present (Medtr5g019250; http://www.medicago.org/genome/IMGAG/). Yet, several TCs with a 100% match to MKB1 and at least five homologs could be identified in the DFCI Medicago Gene Index (Figure 6B; http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago), confirming that also *M. truncatula* possesses several HRD- and RMA-type ligases. Conversely, in view of the conserved coexistence of the two ERAD systems in higher eukaryotes, could mammalian RMA proteins play a hitherto unnoticed role in the regulation of HMGR stability, and consequently the RMA-type ERAD have originated early in evolution with the capacity to target HMGR? Cooperation between the mammalian HRD-type gp78 ligase and RMA1 has been reported in the ERAD of mutant forms of the cystic fibrosis transmembrane conductance regulator (Morito et al., 2008). Would similar (metabolite) signals determine the

recognition of HMGR by the ERAD system in plants and mammals? In addition to sterols in mammals, both yeast and mammalian cells appear to use FPP-derived non-sterols, such as geranylgeraniol or geranylgeranyl pyrophosphate, to regulate HMGR stability/degradation (Sever et al., 2003; Garza et al., 2009). Actually, the primary sequence differences between the membrane domains of yeast and mammalian HMGR proteins have been postulated to account for the observation that HMGR degradation is triggered by non-sterol isoprenoids in yeast, and not by sterols, as in mammals (Hampton, 2002b; DeBose-Boyd, 2008; Garza et al., 2009). In analogy, the divergent sequence of plant HMGR proteins might have allowed the generation of a plant-specific gateway to the ERAD-mediated control of HMGR stability. Perhaps both the general primary and the more plant-specific secondary metabolic pathways, such as sterol and saponin biosynthesis, respectively, in *M. truncatula*, can generate specific intermediates that can activate the ERAD control of HMGR stability and regulate the flux through their respective pathways. Further biochemical and protein-protein interaction studies will address at least some of these questions in the near future.

4. Methods

4.1.*M. truncatula* suspension cell culture maintenance and elicitation

M. truncatula cell cultures (kindly provided by Richard Dixon, Ardmore, USA), were maintained as described (Suzuki et al., 2002). For elicitation, 7 days after inoculation of 75 mL of a 14-day-old suspension culture into 175 mL fresh medium, cells were treated with 100 μM MeJA (dissolved in EtOH) or an equivalent amount of EtOH as a control. Samples were harvested, vacuum filtrated, and frozen at -80°C. For transcript profiling, samples were taken 0, 0.5, 1, 2, 4, 8, 12, and 24 h after elicitor or mock treatments.

4.2.Transcript profiling

Total RNA from *M. truncatula* cells was prepared with TRIZol (Invitrogen, Carlsbad, CA) and reverse transcribed to double-stranded cDNA as described (Vuylsteke et al., 2007). After appropriate sample preparation, cDNA-AFLP based transcript profiling was done with all 128 possible *Bst*YI+1/*Mse*I+2 primer combinations (Vuylsteke et al., 2007). Gel images were analyzed with the AFLP-QUANTARPRO software (Keygene, Wageningen, The Netherlands), allowing accurate quantification of band intensities. The intensity of all individual bands was determined and the obtained raw expression data were corrected for lane variations (due to PCR or loading differences) by dividing the raw data by a correction

factor. The correction factor was calculated by dividing the sum of the expression levels of all fragments within one lane by the highest sum of all lanes within a primer combination. Subsequently, the standard deviation (SD) and the average were calculated for each individual band. Individual gene expression profiles were variance normalized by subtracting the calculated average from each individual data point, after which the obtained value was divided by the SD. A coefficient of variation (CV) was obtained by dividing the SD by the calculated average. Gene tags displaying expression values with a CV ≥0.5 were considered as differentially expressed. Based on this cut-off value, together with visual inspection of the cDNA-AFLP gels, differentially expressed gene tags were selected for further analysis. Cluster analysis, sequencing, and BLAST analysis were done as described (Rischer et al., 2006).

For quantitative Real Time PCR (qRT-PCR), total RNA was extracted with the RNeasy mini kit (Qiagen), and cDNA prepared with SuperScript™ II Reverse Transcriptase (Invitrogen). Primers were designed with Beacon Designer version 4.0 (Premier Biosoft International, Palo Alto, CA, USA). qRT-PCR was carried out with a Lightcycler 480 (Roche) and SYBR Green QPCR Master Mix (Stratagene). For reference genes, 40S ribosomal protein S8 (40S) (TC160725 of the MTGI from TIGR) and translation elongation factor 1α (ELF1α) (TC148782 of the MTGI from TIGR) were used. Reactions were done in triplicate and for the relative quantification with multiple reference genes qBase was used (Hellemans et al., 2007).

4.3. Generation of DNA constructs

For hpRNAi, the 471-bp *MKB1* cDNA-AFLP fragment was PCR-amplified and by Gateway™ recombination cloned into the binary vector pK7GWIWG2D(II) (Karimi et al., 2002). The resulting expression clone was transformed into the *Agrobacterium rhizogenes* strain LBA 9402/12 for generation of hairy roots.

To identify the FL-ORF of *MKB1*, the cDNA-AFLP tag sequence was used for a BLASTN search against the *Medicago truncatula* Gene Index database (http://compbio.dfci.harvard.edu/tgi/). The *MKB1* FL-ORF consensus sequence (TC149901; Genbank accession JF714982) and the *M. truncatula* HMGR1 sequence (Genbank accession EU302813) (Kevei et al., 2007) were PCR-amplified and by Gateway™ recombination cloned into the entry vector pDONR221. To obtain entry clones with and without stop codon,

Gateway primers were designed according to Underwood et al. (2006). The *MKB1* entry vector was used as a template to amplify truncated versions of the *MKB1* sequence as well as to create point mutations with the GeneTailor Site-Directed Mutagenesis system (Invitrogen). All entry constructs were sequence-verified. For stable *MKB1* overexpression experiments, Gateway recombination was carried out with the pK7WG2D binary vector (Karimi et al., 2002), and the resulting clone transformed to *A. rhizogenes*. For transient *MKB1* overexpression in plant protoplasts and for MKB1 localization experiments, Gateway recombination was carried out with the p2GW7 and pK7WGF2 vectors, respectively (Karimi et al., 2002). For recombinant protein production, the *MKB1* sequences were recombined in pDEST15 and pDEST17 expression vectors, and the resulting clones transformed to *E. coli* BL21 (DE3) cells. For the yeast complementation assays, the pAG426GPD vector (Alberti et al., 2007) was used as the destination vector for the *MKB1* gene. For the degradation assays in plant protoplasts, the *M. truncatula* HMGR1 ORF was fused at its C-terminus with the firefly luciferase ORF by a fusion PCR and Gateway recombined in the p2GW7 vector.

4.4.Generation and cultivation of transgenic *M. truncatula* hairy roots

A. rhizogenes-mediated transformation of *M. truncatula* (ecotype Jemalong J5) hairy roots was done according to Boisson-Dernier et al. (2001) with modifications. After scarification with sulfuric acid for 5 minutes, seeds were surface sterilized with 12% sodium hypochlorite for 2 minutes and washed with sterile water. Subsequently, seeds were treated with 1 μM 6-benzylaminopurine for 3 hours and thereafter allowed to germinate on wet, sterile 3 mm Whatman paper at room temperature in the dark. After 2 days of germination, the seedlings were wounded by cutting approximately 2 mm from the root tip of the radicle. The wounded seedlings were infected with *A. rhizogenes* grown on solid YEB medium, and placed on slanted agar plates containing Murashige and Skoog (MS) medium (pH 5.8) supplemented with vitamins (Duchefa). The plates were sealed with micropore tape, placed vertically, and cocultivation was allowed for 10 days under a 16 h/8 h light/dark regime at 22°C. After 10 days, the plants were transferred to new plates, containing 100 mg/L cefotaxime to prevent *Agrobacterium* growth and incubated under identical conditions. Ten days later, hairy roots were excised from the plants and transferred to liquid MS medium (pH 5.8) supplemented with vitamins, 1% sucrose (w/v), and 100 mg/L cefotaxime. The hairy roots were grown for 7 days in the dark at 24°C and shaking at 300 rpm. Subsequently, the roots were transferred to horizontal Petri dishes containing solid MS medium (pH 5.8) supplemented with 1%

sucrose (w/v) and 100 mg/L cefotaxime, grown in the dark at 24°C, and subcultured every 3 weeks.

4.5.Particle bombardment of onion epidermis cells

The constructs for localization were transformed into onion epidermis cells by microparticle bombardment with a PDS-1000/He Biolistic® Particle Delivery System (Bio-Rad Laboratories, Hercules, CA, USA). To this end, 1 mg of 1.6 μm Gold Microcarriers (Bio-Rad Laboratories) were coated with 5 μg plasmid DNA according to the manufacturer's instructions. The coated particles were bombarded into onion epidermis slices of approximately 2.5×2.5 cm, placed on solid MS medium (pH 5.8) supplemented with 1% (w/v) sucrose, with 1100 psi rupture discs and a vacuum of 0.1 bar. Subsequently, the onion slices were stored in the dark for 24 hours at room temperature and analyzed by confocal microscopy with the FV10 ASW Olympus Confocal with a water immersion 63× objective.

4.6.Phenotypic analysis of transgenic *M. truncatula* hairy roots

Samples for SEM were prepared as described (Van Damme et al., 2006). Briefly, after the first fixation step in 4% paraformaldehyde, 1% glutaraldehyde in 2 mM sodium phosphate buffer, the root samples were fixed in 1% osmium tetroxide solution (Fluka) for 2 hours, and subsequently subjected to a dehydratation series to 100% ethanol. Next, the root samples were critical-point dried and sputter-coated with gold particles before they were examined with the SEM microscope (JEOL JSM-5600 LV) under an acceleration voltage of 10 kV.

For TEM, root tips from 12-day-old *M. truncatula* hairy roots grown in liquid medium were briefly immersed in 20% (w/v) BSA and frozen immediately in a high-pressure freezer (EM Pact, Leica Microsystems, Vienna, Austria). Freeze substitution was carried out using a Leica EM AFS in dry acetone containing 0.1% uranyl acetate, 1% (w/v) OsO_4 and 0.2% glutaraldehyde over a 4-day period as follows: -90°C for 26 hours, 2°C per hour increase for 15 hours, -60°C for 8 hours, 2°C per hour increase for 15 hours, and -30°C for 8 hours. Samples were then slowly warmed up to 4°C, infiltrated stepwise over 3 days at 4°C in Spurr's resin and embedded in capsules. The polymerization was performed at 70°C for 16 hours. Ultrathin sections were made with an ultramicrotome (Leica EM UC6) and post-stained in a Leica EM AC20 for 40 min in uranyl acetate at 20 °C and for 10 min in lead stain at 20 °C. Grids were viewed with a JEM 1010 transmission electron microscope (JEOL, Tokyo, Japan) operating at 80 kV.

4.7.Saponin profiling

M. truncatula hairy roots were grown for 21 days in liquid medium. The hairy roots were harvested and the medium collected from five biological repeats of three independent transgenic lines per transgene construct. The hairy roots were rinsed with purified water under vacuum filtration and ground in liquid nitrogen. Of the ground material, 400 mg was extracted with 1 ml methanol at room temperature for 10 minutes, followed by centrifugation for 10 minutes at 20,800×g. Under vacuum, 500 µl of the supernatant was evaporated to dryness and the residue was dissolved in 600 µl water/cyclohexane (2/1, v/v). Samples were centrifuged (10 minutes at 20,800×g) and 200 µl of the aqueous phase was retained for analysis.

To remove salts from the samples of the medium, 1 mL of medium was brought on a 100 mg Extract-Clean^{TM} SPE column (Mandel, Guelph, Ontario, Canada) preconditioned with 1 mL 100% MeOH and 1 mL water acidified with 0.1% (v/v) acetic acid. After washing with 1 mL acidified water, samples were eluted in 1 mL methanol. The methanol eluent was evaporated to dryness under vacuum and the residue dissolved in 200 µl water for analysis.

For reversed phase LC, an Acquity UPLC^{TM} BEH C18 column (150 x 2.1 mm, 1.7 µm; Waters, Milford, MA) was serially coupled to an Acquity UPLCTM BEH C18 column (100 x 2.1 mm, 1.7 µm) and mounted on an ultrahigh performance LC system consisting of a Accela pump (Thermo Electron Corporation, Bremen, Germany) and Accela autosampler (Thermo Electron Corporation). The Accela LC system was hyphenated to a LTQ FT Ultra (Thermo Electron Corporation) via an electrospray ionization source. The following gradient was run with water:acetonitrile (99:1, v/v) acidified with 0.1% (v/v) acetic acid (solvent A) and acetonitrile:water (99:1, v/v) acidified with 0.1% (v/v) acetic acid (solvent B): time 0 min, 5% B; 30 min, 55% B; 35 min, 100% B. The loop size, flow, and column temperature were 25 µl, 300 µl/min, and 80°C, respectively. Full loop injection was applied. Negative ionization was obtained with the following parameter values: capillary temperature 150°C, sheath gas 25 (arbitrary units), auxiliary gas 3 (arbitrary units), and spray voltage 4.5 kV. Full FT-MS spectra between *m/z* 120–1400 were recorded at a resolution of 100,000. For identification, full MS spectra were interchanged with a dependent MS^2 scan event in which the most abundant ion in the previous full MS scan was fragmented, and two dependent MS^3 scan events of the two most abundant daughter ions. The collision energy was set at 35%.

Elucidation of the MS^n spectra was according to Pollier et al. (2011) for the saponins and Morreel et al. (2006; 2010) for the flavonoids and (neo-)lignans.

The resulting chromatograms were integrated and aligned with the XCMS package (Smith et al., 2006) in R version 2.6.1. with the following parameter values: xcmsSet(fwhm=8, max=300, snthresh=5, mzdiff=0.5), group(bw=8, max=300), retcor(method=loess, family=symmetric). A second grouping was done with the same parameter values. Due to in-source fragmentation, multiple *m*/*z* peaks for each compound were often observed. The number of compounds was estimated with "peak groups" consisting of *m*/*z* peaks with the same retention time (window, x s) that were correlated (Pearson; threshold, 0.85) across all control samples.

4.8.Phylogenetic analysis

The protein sequences were aligned with ClustalW and the resulting alignments were manually adjusted (Dataset 1). The phylogenetic tree was generated in MEGA 4.0.1 software (Tamura et al., 2007), by the Neighbor-Joining method, and bootstrapping was done with 10,000 replicates. The evolutionary distances were computed with the Poisson correction method, and all positions containing gaps and missing data were eliminated from the data set (complete deletion option).

```
>At-Rma1/AT4g03510
DCNICLDSVQEPVVTLCGHLFCWPCIHKWLDVQSFSTSCPVCKSKVSHSTLVPLYGRGKVPKRPVGPYMVGEMVATRLFGGPRMRRRIMQADKSLGRIFFFFMCCVVLCLLLF
>At-Rma2/AT4g27470
DCNICLDTAHDPVVTLCGHLFCWPCIYKWLHVQL-SSVCPVCKSNITITSLVPLYGRGSIPRRPAPTGMFGDLVYTRIFGSQRM----MQREKSLNRVSIFFLCCIILCLLLF
>At-Rma3/AT4g28270
DCNICLDQVRDPVVTLCGHLFCWPCIHKWTYASNNSRQCPVCKSDVSEATLVPIYGRGSVPSRPTG-YVVCEMVYRRLFG-VRSRRRAMQAEESLSRVYLFLLCFMFMCLFLF
>Ca-Rma1H1
DCNICLDCVHEPVITLCGHLYCWPCIYKWIYFQSVSSECPVCKAEVSEKTLIPLYGRGKIPQRPPSLSMIGEVAYARIFGSPRMRRQLSQADRSLGRICFFLFCCFVTCLILF
>Ce-RNF5
ECNICLDAAKDAVVSLCGHLFCWPCLSQWLDTRPNNQVCPVCKSAIDGNKVVPIYGRGKVPPRPKGPGIFPLSFVASFFGSAAPGSRMAQEEEYLSNIFKYIGFFMLFWLLFV
>Hs_RNF5/RMA1
ECNICLETAREAVVSVCGHLYCWPCLHQWLETRPERQECPVCKAGISREKVVPLYGRGKTPPRPQGPGAFPFGFFTTVFNDLGQG---HPASSWQDSLFLFLAIFFFFWLLSI
>MKB1
DCNICLESANDPVVTLCGHLYCWPCIYKWLNVQS-SSVCPVCKAVISHTSLVPLYGRGNIPHRPPPTGLFGEMVSTRMFGSPRMRRQEMQIDKSLNRVSIFLLCCFILCLLLF
>TC144757
DCNICLDFAHEPVVTLCGHLYCWSCIYKWLFVQS-ASLCPVCKDGISHTKMVPLYGRGSIPPRPPAA-LFPQFVFG--FGSSRWRRQEMLANKSLNRISFFLFLCFLLCFILF
>TC156053
ECNICFDLAQDPVITLCGHLFCWPCLYRWLHHHSHSQECPVCKALVQEEKLVPLYGRGKIPRRPSGNGLMNIHFHG--FQSQATGQVQRQEDNCLKNLLMLIGFLVLLTVIFF
>TC150776
DCNICLDFANEPVVTLCGHLYCWPCIYKWLHVQSDDSLCPVCKENISHTTMVPLYGRGSVPPRPTAAGMFEEMIFSGVVGSSRLRRQELQTDKFLNRISNFLFCCFLLCLIVF
>TC152655
DCNICLECVQDPVVTLCGHLYCWPCIYKWLNFHAENQ-CPVCKSEISKSSLVPLYGRGNIPPRPLGPQVFGEMIYARIFGNPRIRRHLMRADKSLGRICFFLLCCTVLCLLLF
>TC154271
DCNICLDCVQDPVVTFCGHLYCWPCIYKWLDIQSGISSCPVCKSELSQSSLVPLYGRGKVPRRPTGPSIFGEMIYGRMFGNPRMRRHLMQVDKSLNRISFFLFCSLVLCVLLF
>AT1g19310
ECNICLDLAQDPIVTLCGHLFCWPCLYKWLHLHSQSKDCPVCKAVIEEDRLVPLYGRGKVPNRPSGPGLFNLHFHG--FPYQRHTGRQGQQDHHLR-ILLLIVFVVVVFSLFL
>AT2g23780
ECNICFELAQDPIVTLCGHLFCWPCLYRWLHHHSHSQECPVCKAVVQDDKLVPLYGRGKIPNRPTGPGLFNFQFHG--FHERPMARGGNQSDAFLKNILFFVGICVVIFLIW-
```

Dataset 1. The amino acid alignments used to generate the phylogenetic tree presented in Figure 6.

4.9.Ubiquitination assay

Recombinant GST-MKB1 fusion proteins (wildtype, truncated, or mutated) were purified from transformed *E. coli* cells, pretreated for 2 hours with isopropyl-β-D-1-thiogalactopyranoside (IPTG), with Glutathione Sepharose^{TM} 4B resin columns (GE Healthcare) according to the manufacturer's instructions. A protein refolding step to assure the full ion Zn charge of the GST-MKB1 fusion protein was included by incubation with refolding buffer (20 mM Hepes, pH 7.4, 0.02 mM ZnCl, 1.5 mM $MgCl_2$, 150 mM KCl, 0.2 mM EDTA, 20% glycerol, 0.05% Triton-X100) for 1 hour at 4°C.

Ubiquitination reactions were done in a total volume of 30 µl using 15 µl of the refolded GST-MKB1 in a glutathione resin. The reaction contained 300 ng of GST-MKB1 fusion protein as E3 ligase, 250 ng of the Ub-activating Enzyme (UBE1) from rabbit (BostonBiochem), 400 ng of human recombinant UbcH5a protein (BostonBiochem), and 2 µg of His6-Ub from human (BostonBiochem) in ubiquitination buffer (50 mM Hepes, pH 7.4, 2 mM ATP, 5 mM $MgCl_2$, 2 mM DTT, 0.02 mM $ZnCl_2$). The ubiquitination reactions were incubated for 1 hour at 30°C and stopped by adding 2X Laemmli buffer. Samples were resolved on 8% SDS-PAGE followed by protein immunoblot analysis with α-penta/tetra His antibody (Qiagen) and α-GST (GE Healthcare) antibodies.

4.10. Yeast complementation and protein degradation assays

The *S. cerevisiae* strains RHY400 (*a, ade2-101, his3Δ200, lys2-801, met2, hmg1::LYS2, hmg2::HIS3, ura3-52::6MYC-HMG2*) expressing 6myc-Hmg2p and its *hrd1-1* mutant RHY401 were used for the complementation and protein degradation assays (Hampton et al., 1996). Transformations were carried out with the high-efficiency lithium acetate/single-stranded carrier DNA/polyethylene glycol method. The transformed yeast strains were selected on minimal medium (2.67% minimal SD Base with 0.077% -Ura DO supplement; Clontech) supplemented with 30 mg/l adenine and methionine.

For the *hrd1-1* mutant phenotype complementation assays, minimal medium supplemented with 100 µg/ml lovastatin was used. A stock solution of 25 mg/ml lovastatin was prepared by the hydrolysis of a 100 mg/ml solution in 95% ethanol with 1 N NaOH at 55°C for 40 min, followed by addition of 1 M Tris-HCl (pH 8.0) and adjustment of pH to 8.0 with 1 N HCl.

The protein degradation assay was done by a CHX chase of log phase cells at 0 h and 4 h, and subsequent immunoblotting to follow the degradation of 6myc-Hmg2p (Hampton and Rine, 1994). Whole cell protein extracts were prepared from yeast cells by washing them with minimal medium containing 0.1% NaN_3 followed by resuspension in 100 µl of SUTE buffer (8 M urea, 1% SDS, 10 mM Tris base, 10 mM EDTA, pH adjusted to 7.5) containing Complete protease inhibitors (Roche) at pH 6.8. The cells were lysed by vortexing at high speed with acid-washed 0.5 mm glass beads. The lysate was boiled for 10 min at 65°C after addition of 100 µl USB buffer (8 M urea, 4% SDS, 0.125 M Tris-HCl, pH 6.8, 10% β-mercaptoethanol, pH adjusted to 6.8). Ten µl of the clear liquid lysate was loaded on SDS-PAGE gels for protein separation, followed by immunoblotting with the 9E10 monoclonal anti-myc antibody.

4.11. HMGR degradation assays in plant protoplasts

Protoplast preparation from tobacco Bright Yellow-2 cells, automated transfection, lysis, and firefly luciferase assays were carried out as described (De Sutter et al., 2005).

4.12. Accession numbers

Sequence data from this article can be found in the EMBL/Genbank data libraries under accession number JF714982.

Acknowledgements

We thank Wilson Ardiles-Diaz for sequencing the cDNA-AFLP tags, Kristiina Himanen, Jan Geerinck, and Lander Ingelbrecht for excellent technical advice and assistance, Martine De Cock for help in preparing the manuscript, and Richard Dixon (Samuel Roberts Noble Foundation, Ardmore, USA) for generously providing the *M. truncatula* cell line. This work was supported by the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" project SBO040093). Tessa Moses is indebted to the VIB International PhD Fellowship Program for a predoctoral fellowship.

References

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

Alberti, S., Gitler, A.D., and Lindquist, S. (2007). A suite of Gateway® cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. Yeast **24**, 913-919.

Augustin, J.M., Kuzina, V., Andersen, S.B., and Bak, S. (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry **72**, 435-457.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y., and Udvardi, M.K. (2008). A gene expression atlas of the model legume *Medicago truncatula*. Plant J **55**, 504-513.

Boisson-Dernier, A., Chabaud, M., Garcia, F., Bécard, G., Rosenberg, C., and Barker, D.G. (2001). *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. Mol Plant Microbe Interact **14**, 695-700.

Browse, J. (2009). Jasmonate passes muster: a receptor and targets for the defense hormone. Annu Rev Plant Biol **60**, 183-205.

Chappell, J. (2002). The genetics and molecular genetics of terpene and sterol origami. Curr Opin Plant Biol **5**, 151-157.

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J.M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F.M., Ponce, M.R., Micol, J.L., and Solano, R. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. Nature **448**, 666-671.

Dale, S., Arró, M., Becerra, B., Morrice, N.G., Boronat, A., Hardie, D.G., and Ferrer, A. (1995). Bacterial expression of the catalytic domain of 3-hydroxy-3-methylglutaryl-CoA reductase (isoform HMGR1) from *Arabidopsis thaliana*, and its inactivation by phosphorylation at Ser577 by *Brassica oleracea* 3-hydroxy-3-methylglutaryl-CoA reductase kinase. Eur J Biochem **233**, 506-513.

DeBose-Boyd, R.A. (2008). Feedback regulation of cholesterol synthesis: sterol-accelerated ubiquitination and degradation of HMG CoA reductase. Cell Res **18**, 609-621.

De Sutter, V., Vanderhaeghen, R., Tilleman, S., Lammertyn, F., Vanhoutte, I., Karimi, M., Inzé, D., Goossens, A., and Hilson, P. (2005). Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J **44**, 1065-1076.

Dixon, R.A., and Sumner, L.W. (2003). Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol **131**, 878-885.

Doolman, R., Leichner, G.S., Avner, R., and Roitelman, J. (2004). Ubiquitin is conjugated by membrane ubiquitin ligase to three sites, including the N terminus, in transmembrane region of mammalian 3-hydroxy-3-methylglutaryl coenzyme A reductase: implications for sterol-regulated enzyme degradation. J Biol Chem **279**, 38184-38193.

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.M., Gimenez-Ibanez, S., Geerinck, J., Eeckhout, D., Schweizer, F., Godoy, M., Franco-Zorrilla, J.M., Pauwels, L., Witters, E., Puga, M.I., Paz-Ares, J., Goossens, A., Reymond, P., De Jaeger, G., and Solano, R. (2011). The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of JA responses. Plant Cell **23**, 701-715.

Fonseca, S., Chico, J.M., and Solano, R. (2009a). The jasmonate pathway: the ligand, the receptor and the core signalling module. Curr Opin Plant Biol **12**, 539-547.

Fonseca, S., Chini, A., Hamberg, M., Adie, B., Porzel, A., Kramell, R., Miersch, O., Wasternack, C., and Solano, R. (2009b). (+)-7-*iso*-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. Nat Chem Biol **5**, 344-350.

Gardner, R.G., and Hampton, R.Y. (1999). A 'distributed degron' allows regulated entry into the ER degradation pathway. EMBO J **18**, 5994-6004.

Garza, R.M., Tran, P.N., and Hampton, R.Y. (2009). Geranylgeranyl pyrophosphate is a potent regulator of HRD-dependent 3-Hydroxy-3-methylglutaryl-CoA reductase degradation in yeast. J Biol Chem **284**, 35368-35380.

Hampton, R.Y. (2002a). ER-associated degradation in protein quality control and cellular regulation. Curr Opin Cell Biol **14**, 476-482.

Hampton, R.Y. (2002b). Proteolysis and sterol regulation. Annu Rev Cell Dev Biol **18**, 345-378.

Hampton, R.Y., and Rine, J. (1994). Regulated degradation of HMG-CoA reductase, an integral membrane-protein of the endoplasmic reticulum, in yeast. J Cell Biol **125**, 299-312.

Hampton, R.Y., Gardner, R.G., and Rine, J. (1996). Role of 26S proteasome and HRD genes in the degradation of 3-hydroxy-3-methylglutaryl-CoA reductase, an integral endoplasmic reticulum membrane protein. Mol Biol Cell **7**, 2029-2044.

He, J., Benedito, V.A., Wang, M.Y., Murray, J.D., Zhao, P.X., Tang, Y.H., and Udvardi, M.K. (2009). The *Medicago truncatula* gene expression atlas web server. BMC Bioinformatics **10**, 441.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol **8**, R19.1-R19.14.

Hirsch, C., Gauss, R., Horn, S.C., Neuber, O., and Sommer, T. (2009). The ubiquitylation machinery of the endoplasmic reticulum. Nature **458**, 453-460.

Jo, Y., and Debose-Boyd, R.A. (2010). Control of cholesterol synthesis through regulated ER-associated degradation of HMG CoA reductase. Crit Rev Biochem Mol Biol **45**, 185-198.

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY^{TM} vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci **7**, 193-195.

Kevei, Z., Lougnon, G., Mergaert, P., Horváth, G.V., Kereszt, A., Jayaraman, D., Zaman, N., Marcel, F., Regulski, K., Kiss, G.B., Kondorosi, A., Endre, G., Kondorosi, E., and Ané, J.M. (2007). 3-Hydroxy-3-methylglutaryl coenzyme A reductase1 interacts with NORK and is crucial for nodulation in *Medicago truncatula*. Plant Cell **19**, 3974-3989.

Kosarev, P., Mayer, K.F.X., and Hardtke, C.S. (2002) Evaluation and classification of RING-finger domains encoded by the *Arabidopsis* genome. Genome Biol **3**, 0016.1-0016.12.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J Mol Biol **157**, 105-132.

Learned, R.M., and Fink, G.R. (1989). 3-Hydroxy-3-methylglutaryl-coenzyme A reductase from *Arabidopsis thaliana* is structurally distinct from the yeast and animal enzymes. Proc Natl Acad Sci U S A **86**, 2779-2783.

Lee, H.K., Cho, S.K., Son, O., Xu, Z., Hwang, I., and Kim, W.T. (2009). Drought stress-induced Rma1H1, a RING membrane-anchor E3 ubiquitin ligase homolog, regulates aquaporin levels via ubiquitination in transgenic Arabidopsis plants. Plant Cell **21**, 622-641.

Lee, P.C.W., Nguyen, A.D., and DeBose-Boyd, R.A. (2007) Mutations within the membrane domain of HMG-CoA reductase confer resistance to sterol-accelerated degradation. J. Lipid Res **48**, 318-327.

Matsuda, N., Suzuki, T., Tanaka, K., and Nakano, A. (2001). Rma1, a novel type of RING finger protein conserved from Arabidopsis to human, is a membrane-bound ubiquitin ligase. J Cell Sci **114**, 1949-1957.

Morikawa, T., Mizutani, M., Aoki, N., Watanabe, B., Saga, H., Saito, S., Oikawa, A., Suzuki, H., Sakurai, N., Shibata, D., Wadano, A., Sakata, K., and Ohta, D. (2006). Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in *Arabidopsis* and tomato. Plant Cell **18**, 1008-1022.

Morito, D., Hirao, K., Oda, Y., Hosokawa, N., Tokunaga, F., Cyr, D.M., Tanaka, K., Iwai, K., and Nagata, K. (2008). Gp78 cooperates with RMA1 in endoplasmic reticulum-associated degradation of CFTRΔF508. Mol Biol Cell **19**, 1328-1336.

Morreel, K., Goeminne, G., Storme, V., Sterck, L., Ralph, J., Coppieters, W., Breyne, P., Steenackers, M., Georges, M., Messens, E., and Boerjan, W. (2006). Genetical metabolomics of flavonoid biosynthesis in *Populus*: a case study. Plant J **47**, 224-237.

Morreel, K., Kim, H., Lu, F.C., Dima, O., Akiyama, T., Vanholme, R., Niculaes, C., Goeminne, G., Inzé, D., Messens, E., Ralph, J., and Boerjan, W. (2010) Mass spectrometry-based fragmentation as an identification tool in lignomics. Anal Chem **82**, 8095-8105.

Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., Hill, L., Qi, X., Bakht, S., Melton, R., and Osbourn, A. (2008). Sad3 and sad4 are required for saponin biosynthesis and root development in oat. Plant Cell **20**, 201-212.

Naoumkina, M., Farag, M.A., Sumner, L.W., Tang, Y., Liu, C.J., and Dixon, R.A. (2007). Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc Natl Acad Sci U S A **104**, 17909-17915.

Naoumkina, M.A., He, X., and Dixon, R.A. (2008). Elicitor-induced transcription factors for metabolic reprogramming of secondary metabolism in *Medicago truncatula*. BMC Plant Biol **8**, 132.

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

Nieto, B., Forés, O., Arró, M., and Ferrer, A. (2009). Arabidopsis 3-hydroxy-3-methylglutaryl-CoA reductase is regulated at the post-translational level in response to alterations of the sphingolipid and the sterol biosynthetic pathways. Phytochemistry **70**, 53-59.

Oleszek, W. (1990). Structural specificity of alfalfa (*Medicago sativa*) saponin hemolysis and its impact on 2 hemolysis-based quantification methods. J Sci Food Agr **53**, 477-485.

Oleszek, W., Price K.R., Colquhoun, I.J., Jurzysta, M., Ploszynski, M., and Fenwick, G.R. (1990). Isolation and identification of alfalfa (*Medicago sativa* L.) root saponins: their activity in relation to a fungal bioassay. J Agric Food Chem **38**, 1810-1817.

Pauwels, L., Fernández Barbero, G., Geerinck, J., Tilleman, S., Grunewald, W., Cuéllar Pérez, A., Chico, J.M., Vanden Bossche, R., Sewell, J., Gil, E., García-Casado, G., Witters, E., Inzé, D., Long, J.A., De Jaeger, G., Solano, R., and Goossens, A. (2010). NINJA connects the co-repressor TOPLESS to jasmonate signaling. Nature **464**, 788-791.

Pauwels, L., Inzé, D., and Goossens, A. (2009). Jasmonate-inducible gene: what does it mean? Trends Plant Sci **14**, 87-91.

Phillips, D.R., Rasbery, J.M., Bartel, B., and Matsuda, S.P.T. (2006). Biosynthetic diversity in plant triterpene cyclization. Curr Opin Plant Biol **9**, 305-314.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. J Nat Prod, **in press**.

Rischer, H., Orešič, M., Seppänen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M.C.E., Inzé, D., Oksman-Caldentey, K.M., and Goossens, A. (2006). Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc Natl Acad Sci U S A **103**, 5614-5619.

Schrick, K., Fujioka, S., Takatsuto, S., Stierhof, Y.D., Stransky, H., Yoshida, S., and Jürgens, G. (2004). A link between sterol biosynthesis, the cell wall, and cellulose in *Arabidopsis*. Plant J **38**, 227-243 (Err. Plant J **38**: 562).

Sever, N., Song, B.L., Yabe, D., Goldstein, J.L., Brown, M.S., and DeBose-Boyd, R.A. (2003). Insig-dependent ubiquitination and degradation of mammalian 3-hydroxy-3-methylglutaryl-CoA reductase stimulated by sterols and geranylgeraniol. J Biol Chem **278**, 52479-52490.

Sheard, L.B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T.R., Kobayashi, Y., Hsu, F.F., Sharon, M., Browse, J., He, S.Y., Rizo, J., Howe, G.A., and Zheng, N. (2010). Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. Nature **468**, 400-405.

Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T., and Ebizuka, Y. (2006). Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. FEBS J **273**, 948-959.

Smith, C.A., Want E.J., O'Maille G., Abagyan R., and Siuzdak G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem **78**, 779-787.

Son, O., Cho, S.K., Kim, S.J., and Kim, W.T. (2010). *In vitro* and *in vivo* interaction of AtRma2 E3 ubiquitin ligase and auxin binding protein 1. Biochem Biophys Res Commun **393**, 492-497.

Sparg, S.G., Light, M.E., and van Staden, J. (2004). Biological activities and distribution of plant saponins. J Ethnopharmacol **94**, 219-243.

Sun, H.X., Xie, Y., and Ye, Y.P. (2009). Advances in saponin-based adjuvants. Vaccine **27**, 1787-1796.

Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P.T., and Dixon, R.A. (2002). A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. Plant J **32**, 1033-1048.

Suzuki, H., Reddy, M.S.S., Naoumkina, M., Aziz, N., May, G.D., Huhman, D.V., Sumner, L.W., Blount, J.W., Mendes, P., and Dixon, R.A. (2005). Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. Planta **220**, 696-707.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24**, 1596-1599.

Tava, A., and Avato, P. (2006) Chemical and biological activity of triterpene saponins from *Medicago* species. Nat Prod Commun **1**, 1159-1180.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S.Y., Howe, G.A., and Browse, J. (2007). JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. Nature **448**, 661-665.

Underwood, B.A., Vanderhaeghen, R., Whitford, R., Town, C.D., and Hilson, P. (2006). Simultaneous high-throughput recombinational cloning of open reading frames in closed and open configurations. Plant Biotechnol J **4**, 317-324.

Van Damme, D., Coutuer, S., De Rycke, R., Bouget, F.Y., Inzé, D., and Geelen, D. (2006). Somatic cytokinesis and pollen maturation in Arabidopsis depend on TPLATE, which has domains similar to coat proteins. Plant Cell **18**, 3502-3518.

Voutquenne, L., Lavaud, C., Massiot, G., and Le Men-Olivier, L. (2002). Structure-activity relationships of haemolytic saponins. Pharm Biol **40**, 253-262.

Vuylsteke, M., Peleman, J.D., and van Eijk, M.J. (2007). AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat Protoc **2**, 1399 1413.

Wang, Z., Dai, L., Jiang, Z., Peng, W., Zhang, L., Wang, G., and Xie, D. (2005). *GmCOI1*, a soybean F-box protein gene, shows ability to mediate jasmonate-regulated plant defense and fertility in *Arabidopsis*. Mol Plant-Microbe Interact **18**, 1285-1295.

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

Zhao, J., Davis, L.C., and Verpoorte, R. (2005). Elicitor signal transduction leading to production of plant secondary metabolites. Biotechnol Adv **23**, 283-333.

Chapter 7

Perspectives

1. The Combinatorial Biosynthesis in Plants (COMBIPLAN) project

This PhD project was conducted in the framework of the COMBIPLAN SBO project ("Strategisch Basisonderzoek" project SBO040093), with the aim to demonstrate the potential and technical feasibility of a platform allowing semi-rational combinatorial engineering of the biosynthesis of existing and novel plant secondary metabolites in plant cell tissue cultures, with the secondary metabolite class of the triterpene saponins as the target. The COMBIPLAN project was a collaboration between three research groups, and was coordinated by Prof. Alain Goossens, from the VIB department of Plant Systems Biology, Ghent University. The collaborating partners consisted of the research groups of Prof. Danny Geelen of the department Plant Production, Ghent University, and of Prof. Luc Pieters of the department of Pharmaceutical Sciences, University of Antwerp.

1.1. Contribution to the COMBIPLAN project

The first major deliverable of this thesis for the COMBIPLAN project was the gene discovery program for combinatorial biosynthesis in plants. The creation of conditions of differential saponin production in the target plants by elicitor treatment allowed to pinpoint relevant genes through a genome-wide transcriptome analysis. This transcript profiling was performed with cDNA-AFLP, and led to the visualization of the expression of more than 70,000 genes, of which around 250 were prioritized for further analysis based on their putative annotation and transcriptional co-regulation with known saponin biosynthesis genes (Chapter 2). To obtain the FL-clones of the selected genes, a medium-throughput cDNA-library screening strategy to convert the cDNA-AFLP tag sequences to the corresponding FL-clones was developed and applied (Chapter 3). This screening led to a library of 140 FL-clones, encoding enzymes potentially involved in saponin biosynthesis (geneset 1) and transcription factors that potentially regulate the saponin biosynthesis (geneset 2).

The second major aim of this PhD project was delivering proof of concept of combinatorial biosynthesis of triterpene saponins in plants. To this end, more than 50 gene constructs (from geneset 2) leading to ectopic expression of candidate saponin biosynthesis genes were introduced in *M. truncatula* hairy roots. A method for metabolite profiling of triterpene saponins in *M. truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry was designed (Chapter 4), and employed to screen for novel molecules, which were detected in three different transgenic *M. truncatula* hairy root

lines. Structural elucidation of the novel molecules indicated that two of the transgenic lines produced novel triterpene saponins, whereas the third line produced a set of novel, non-saponin compounds. Hence, proof of concept of combinatorial biosynthesis in plants was obtained.

1.2.Are the new compounds truly novel?

Questions may arise about the true novelty of the compounds. Indeed, they have been shown to be new to both the *M. truncatula* hairy roots and the plants of which the genes were derived, but, since their exact chemical structure is not know, it is still possible they are not new to nature. Nuclear magnetic resonance (NMR) spectroscopy is the most commonly used method to determine the exact chemical structure of new compounds. Especially for complicated molecules such as triterpene saponins, two-dimensional (2D)-NMR may be needed to fully elucidate the structures. 2D-NMR reveils nuclei that are spin-coupled to each other, and since spin-coupled nuclei are usually separated by one to three bonds only, the connections within the compound are revealed. Homonuclear ^{1}H-^{1}H correlation spectroscopy (COSY) reveils spin-coupled protons, whereas heteronuclear ^{1}H-^{13}C correlation spectroscopy reveals either one-bond carbon-hydrogen interactions (such as heteronculear single quantum coherence (HSQC) or heteronculear multiple quantum coherence (HMQC) techniques) or multiple-bond carbon-hydrogen correlations (e.g. heteronuclear multiple bond correlation (HMBC) spectroscopy). By combining the correlations reveiled by the different 2D-NMR experiments, the chemical structure of a compound can be determined. However, to perform an NMR experiment, a significant amount (>1 mg) of purified (>90%) material is required. Since the new compounds in the ML385 lines are highly abundant, this may prove to be relatively easy, however, because of the low abundance of the new saponins in the PG0408 and BF0567 lines, to obtain enough pure compound may become quite a challenge. The concentration of saponins in root tissue of *M. truncatula* was shown to be 5.92 mg/g (Huhman et al., 2005), and the new saponins detected in the PG0408 lines only occur at 0.15 to 0.66% of the highest abundant compound, soyasaponin I, which accounts for 32% of the total saponins in the *M. truncatula* roots (Huhman et al., 2005). Thus, to obtain 1 mg of pure compound, 80-340 g of dry roots would be required. After metabolite extraction of the roots, purification of the low-abundant compounds could be performed via preparative column chromatography.

It is not unthinkable, however, that after purification it turns out that the required amount of 1 mg of pure compound for NMR analysis is not obtained for some of the lower abundant compounds. When lesser amounts of compound are available, the various NMR techniques may not be able to reveal all correlated nuclei. Nonetheless, provided that similar compounds as the unknown compound are available, performing NMR on lower amounts of the unknown compound may be still useful. By comparing the NMR spectrum that is obtained with low amounts of compound with the NMR spectrum of a similar compound, it may still be possible to identify the changes between the "standard" and the unknown compound. For instance, in the case of the PG0408 lines, hydroxylation on another position of the dammarenediol backbone will only cause the shift of a few of the peaks obtained by NMR. Thus, comparison of the NMR spectrum of the commercially available ginsenoside aglycones with the low resolution NMR spectrum of the unknown compound may still lead to structural elucidation of the unknown compound.

2. Initiated follow-up research

In this PhD project, a library of 140 FL-clones was generated comprising genes that encode enzymes potentially involved in saponin biosynthesis (geneset 1) and transcription factors that potentially regulate the saponin biosynthesis (geneset 2). During my thesis period, three other PhD projects that use these genesets as a basis have been started within this department.

2.1.Biosynthesis of novel plant derived drugs in engineered yeast

The goal of the PhD project of Tessa Moses is to create a combinatorial biosynthesis platform in yeast that allows the production of novel biologically active plant derived molecules. After engineering the yeast strains to produce the basic saponin backbone (β-amyrin, lupeol, dammarenediol, …), the strains will be further engineered with multiple candidate saponin biosynthesis genes from geneset 1, in new-to-nature combinations, in order to create yeast cells that produce structurally novel saponins or sapogenin precursors. This will not only allow for the biosynthesis of new compounds, but will also generate the opportunity to functionally characterize the corresponding enzymes. Thus far, yeast strains capable of accumulating high levels of several triterpene backbones have been generated. Furthermore, super-transformation of these yeast strains with some of the genes encoding CytP450 enzymes of geneset 1 led to the identification of a novel functional *M. lanceolata* CytP450 enzyme that hydroxylates the β-amyrin backbone.

2.2. Identification and functional characterization of jasmonate-modulated master switches that steer the biosynthesis of plant pharmaceuticals

In her PhD project, Azra Gholami will try to functionally characterize some of the candidate regulators of geneset 2. In first instance, candidate regulators of *M. truncatula* will be overexpressed or silenced in *M. truncatula* hairy root lines, and the resulting transgenic lines will be phenotypically characterized. This will include metabolite profiling of triterpene saponins with the method developed within my PhD (Chapter 4) and targeted and genome-wide transcript profiling. For a selection of lead genes an in-depth functional analysis is envisaged using various approaches, such as biochemical analysis, mutant and/or chemical genetic screens, yeast screens, bimolecular fluorescence complementation (BiFC) analysis, in situ expression analysis, which should allow the elucidation of their mode of action and the mapping of their position in the JA-signalling machineries. Finally, the utility of these genes as tools for metabolic engineering towards increased production of plant-derived pharmaceuticals will be assessed.

2.3. Study of protein complexes involved in the elicitation of secondary metabolism in *M. truncatula*

In his PhD project, Nathan De Geyter will set up a Tandem Affinity Purification (TAP) platform (Van Leene et al., 2008) in *M. truncatula* hairy root cultures. As baits, 11 proteins of the candidate regulators of geneset 2 will be used, and the proteins that are co-purified with the baits will be identified. As such, Nathan will try to identify new components of the protein complexes involved in the jasmonate signal transduction cascade that leads to the elicitation of secondary metabolism in *M. truncatula*. When the protein complexes are identified, their role in elicitation of secondary metabolism in *M. truncatula* will be investigated in overexpression or silencing lines of the corresponding genes. To this end, targeted or genome-wide transcript profiling could be performed and the effect on metabolite accumulation could be assessed using the method developed in this thesis (Chapter 4).

3. Potential of the technology platform and the generated data

Besides the initiated PhD projects, the generated data and the technology platform itself may be of economic and scientific importance in various research fields.

3.1.Drug discovery

In this PhD project, proof of concept of combinatorial biosynthesis in plants was obtained with the secondary metabolite class of the triterpene saponins as target (Chapter 5). Triterpene saponins are a large and structurally diverse group of plant natural products that possess various pharmaceutical properties (Chapter 1), and are thus of considerate economical importance. For instance, recently a license fee of $450 million to develop and commercialize a chemical derivative of oleanolic acid, bardoxolone methyl, for the treatment of chronic kidney disease has been paid by Abbott to Reata Pharmaceuticals (http://www.abbott.com/global/url/pressReleases). Also ginsenosides, the dammarane-type saponins restricted to the genus *Panax*, possess a variety of pharmaceutical properties. Cultivation of ginseng is a slow process and productivity from ginseng tissue and cell cultures has been low. In this PhD project, several (novel) ginsenosides were produced in a plant not belonging to the genus *Panax* (Chapter 5), suggesting an opportunity for alternative production methods for ginsenosides. Hence, the application of combinatorial biosynthesis to increase the structural variation of triterpene saponins may result in the discovery of novel compounds with a significant pharmaceutical (and thus economical) impact.

One of the major challenges in the development of plant-derived drugs is obtaining sufficient amounts of the compounds. In a lot of cases, the complex bioactive molecules are available at very low concentrations *in planta*. Furthermore, plant growth can be very slow, or the molecules are found in plant species that are rare or threatened with extinction. As such, the establishment of combinatorial biosynthesis programs *in planta* may be compromised, at least in some plant species. Also here, the library of 140 FL genes may offer opportunities. Overexpression or silencing of regulatory genes may lead to a higher accumulation of the desired compounds *in planta*. For instance, silencing of the RING E3 ubiquitin ligase MAKABISH1 that regulates the triterpene saponin biosynthesis in *M. truncatula* leads to an increased flux through the mevalonate pathway, with an increased accumulation of monoglycosylated saponins as a consequence (Chapter 6). Furthermore, also reconstitution of the biosynthesis pathway in a heterologous host could overcome low production amounts *in planta*, or the need to sacrifice rare or endangered species. The genes present in our FL library may be lead candidates for the reconstitution of the saponin biosynthesis in a suitable host, and thus for the development of a more robust microbial production system.

3.2.Gene discovery for combinatorial biosynthesis

To exploit the full potential of plant natural products, knowledge of the genes involved in the biosynthesis of the metabolites is essential, but until now most, if not all, of the biosynthesis pathways leading to the production of plant natural products remain incompletely understood. With the commercialization of several next-generation sequencing techniques, the last five years witnessed a revolution in functional genomics. This tremendous progress now brings the whole genome or transcriptome of any plant species within reach and will undoubtedly greatly accelerate gene discovery for secondary metabolite pathways. Simultaneously, the enormous amount of data generated by such next generation sequencing platforms demands for efficient strategies to identify and select the relevant biosynthetic genes from the pools of thousands of sequenced genes, or at least for efficient ways to exploit them.

Next to the use of elicitors (Chapter 1 and 2), overexpression or silencing of regulatory genes may allow pinpointing the desired subsets of the total gene pools as candidate genes responsible for the biosynthesis of the secondary metabolites. For instance, silencing of the the RING E3 ubiquitin ligase MAKABISH1 in *M. truncatula* hairy roots led to a strong downregulation of practically all of the hitherto known triterpene saponin biosynthetic genes, whereas no effects were observed on genes of the related sterol biosynthesis (Chapter 6). Thus, comparison of the transcriptome of MAKABISH1 silencing lines with the transcriptome of control hairy root lines may allow to specifically pinpoint the hitherto unknown genes involved in saponin biosynthesis.

References

Huhman, D.V., Berhow, M.A., and Sumner, L.W. (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. J Agric Food Chem **53**, 1914-1920.

Van Leene, J., Witters, E., Inzé, D., and De Jaeger, G. (2008). Boosting tandem affinity purification of plant protein complexes. Trends Plant Sci **13**, 517-520.

Summary

Combinatorial biosynthesis comprises a series of methods that establish novel enzyme-substrate combinations in vivo, which in turn lead to the biosynthesis of new, natural product-derived compounds that can be used in drug discovery programs. Because of the more complex genetic makeup of plants and the lack of tools to manipulate them, combinatorial biosynthesis in plants is still in its infancy. (Chapter 1). The main goal of this PhD research was to establish a combinatorial biosynthesis platform in plants with the secondary metabolite class of triterpene saponins as the target.

Since several of the steps involved in the saponin biosynthesis were uncharacterized at the molecular level, a functional genomics approach to identify candidate genes involved in the saponin biosynthesis of five different plants was applied. The creation of conditions of differential saponin production in the target plants by elicitor treatment allowed to pinpoint relevant genes through a genome-wide transcriptome analysis. This transcript profiling was performed with cDNA-AFLP, and led to the visualization of the expression of more than 70,000 genes, of which around 250 were prioritized for further analysis based on their putative annotation and transcriptional co-regulation with known saponin biosynthesis genes (Chapter 2). To obtain the FL-clones of the selected genes, a medium-throughput cDNA-library screening strategy to convert the cDNA-AFLP tag sequences to the corresponding FL-clones was developed and applied (Chapter 3). This screening has led to a library of 140 FL-clones, encoding enzymes potentially involved in saponin biosynthesis and transcription factors that potentially regulate the saponin biosynthesis. To deliver proof of concept of combinatorial biosynthesis of triterpene saponins in plants, more than 50 gene constructs leading to heterologous expression of candidate saponin biosynthesis genes were introduced in *M. truncatula* hairy roots. A method for metabolite profiling of triterpene saponins in *M. truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry was designed (Chapter 4), and employed to screen for novel molecules, which were detected in three different transgenic *M. truncatula* hairy root lines. Structural elucidation of the novel molecules indicated that two of the transgenic lines produced novel triterpene saponins, whereas the third line produced a set of novel, non-saponin compounds. Hence, proof of concept of combinatorial biosynthesis in plants was obtained (Chapter 5). Furthermore, overexpression and silencing of a set of candidate genes involved in the

regulation of saponin biosynthesis has led to the identification of MKB1, a RING E3 ubiquitin ligase that regulates triterpene saponin biosynthesis in *Medicago truncatula* by targeting 3-hydroxy-3-methylglutaryl-CoA reductase, the enzyme catalyzing the rate-limiting step in the mevalonate pathway (Chapter 6).

Samenvatting

Combinatoriële biosynthese omvat een aanal technieken die nieuwe enzym-substraat combinaties genereren *in vivo*, welke resulteren in de biosynthese van nieuwe molecules die gebruikt kunnen worden in drug discovery programma's. Hoewel combinatoriële biosynthese technieken in micro-organismen regelmatig worden toegepast, is het gebruik ervan in planten gelimiteerd omwille van de complexere genetische structuur van planten en het ontbreken van technieken om planten te manipuleren (Hoofdstuk 1). Het hoofddoel van dit doctoraatsonderzoek was dan ook om een combinatoriële biosynthese platform in planten te ontwikkelen, met de triterpeen saponines, een klasse van secundaire metabolieten in planten, als doel.

Daar tot op heden nog steeds verschillende stappen in de biosynthese van triterpeen saponines ongekend zijn, werd er in de eerste stap van dit doctoraatsonderzoek in vijf verschillende planten gezocht naar genen die mogelijks betrokken zijn in de biosynthese van triterpeen saponines. Het behandelen van de verschillende planten met methyl jasmijnzuur resulteerde in een verhoogde saponine accumulatie, als gevolg van de activatie van de saponine biosynthese genen. Het onderzoeken van de genexpressie van meer dan 70,000 genen van de vijf verschillende planten met behulp van de cDNA-AFLP techniek stelde ons in staat om een set van een 250-tal genen te priorizeren voor verdere analyse. Een gen werd enkel beschouwd als kandidaat-gen als de expressie ervan overeen kwam met de expressie van de gekende saponine biosynthese genen (co-regulatie) en als het gen ook nog eens een geannoteerde functie had die nodig was voor de ongekende biosynthese stappen (hoofdstuk 2). Om de volledige coderende sequentie van de genen te bekomen werd er een strategie ontwikkeld voor het efficiënt screenen van cDNA bibliotheken naar cDNA clones die overeenkomen met de geselecteerde gentags (hoofdstuk 3). Dit resulteerde in een collectie van clones van een 140-tal genen, welke coderen voor enzymes die mogelijk betrokken zijn bij de saponine biosynthese of de regulatie ervan. Om te bewijzen dat combinatoriële biosynthese in planten mogelijk is werden meer dan 50 kandidaat-biosynthese genen tot overexpressie gebracht in *M. truncatula* hairy roots. Een methode voor het profileren van de door de hairy roots geproduceerde saponines werd ontwikkeld (hoofdstuk 4) en toegpast voor het opsporen van nieuwe metabolieten in de transgene hairy roots, welke gedetecteerd werden in drie verschillende transgene lijnen. Het gedeeltelijk

ophelderen van de structuur van de nieuwe componenten toonde aan dat twee van de lijnen nieuwe triterpeen saponines produceerden, terwijl er in de derde lijn accumulatie van nieuwe, niet-saponine componenten werd waargenomen. Het bewijs van combinatoriële biosynthese in planten werd hiermee geleverd (hoofdstuk 5). Bovendien werd er ook een set van kandidaat genen betrokken bij de regulatie van de saponine biosynthese onderzocht. Dit leidde tot de identificatie van MKB1, een RING E3 ubiquitin ligase dat de triterpeen saponine biosynthese reguleert in *Medicago truncatula* door het enzyme 3-hydroxy-3-methylglutaryl-CoA reductase van de mevalonaat biosynthese te labelen voor degradatie (hoofdstuk 6).

Acknowledgements

Finally my PhD is over! What initially seemed a straightforward project of four years has finally become an endeavor of just over six years. The results are what they are, but they would never have been there without the help and support of many others. Many thanks to…

…Alain, for allowing me to start my PhD in your group. Your guiding, enthusiasm, ideas, and especially your unabating working spirit (your e-mails start arriving around 8 am, continue throughout the day and do not stop arriving before 11 pm) were very stimulating for me throughout my PhD. Also many thanks for the input for my thesis and all the reading and very fast correcting. I sincerely hope I can stay a while longer in your group.

…Kris Morreel, for introducing me into the fascinating world of metabolite profiling in the last two years of my PhD. Every time I had a problem or question (on almost daily basis), you were willing to help and explain. The coffee-addiction I already had before working with you, but you infected me with your passion for metabolites. How to separate and identify them, how do they split, how to purify them... If one day I have only half of your knowledge in this field, I will proudly call myself an expert… Thank you very much, Kris!

…Peter Marhavý, for the hours spent in the confocal room picturing my hairy roots, onion peels, BY2s, and *Nicotiana* leaves that were supposed to have GFP expression... Whenever I had samples, you were always there to help… Thanks!

…All people involved in the COMBIPLAN-project: Miguel, Lander, Seppe and students Sandy, Ursula, and Inge; Prof. Danny Geelen, Ellen Van Twembeke, Ellen Lambert and Faizal from the department of plant production (Ghent University); and Prof. Luc Pieters, Sandra, Kenn, Mart and Taposh from the Department of Pharmaceutical Sciences (University of Antwerp). Without your results and input, this thesis would seem rather thin.

…All past and present metabollers, for the nice and stimulating environment and all the help. Thank you Robin, Freya, Emmy, Rebecca, Valerie, Lies, Sofie, Laurens, Jan, Amparo, Astrid, Nathan, Azra, Michele, Michiel, Gino, Antonio, Miriam, Hasnain, Inês, and Janine. It was a pleasure to work with you all! Also thanks to Anelia and Katrijn, whom I could help with their cDNA-AFLP and library screenings.

…Prof. Dirk Inzé, head of the department, and the ‘PSB-support’ team, Jackie and Kristof for all the autoclaving, Nino for the primers and robots, Wilson for the sequencing, Karel for the pictures, Martine for all the corrections, Nico for watering and pruning our plants, Raf and Hendrik to solve the computer-trouble.

…My parents and family for their support, my scouts-friends and the Indian gang, for the adventurous camps, fun and crazy parties!

And last and most importantly… Tessa. My dearest, the day you started working in the lab, November 30, 2007, you must have thought that I truly was a complete idiot. Drinking too much and falling on my face… Later, Alain asked you to write a review together with that fool… At first, this was going more or less fine, but soon, that fool became much more interested in the co-author than the review. Now, years later, that review is still not published… but you became my life and now that fool is yours forever. Thank you Tess, for all your support and patience, for all your help and understanding,… Thank you for being mine!

Appendix: Additional publications

Besides the articles described in the main body of this PhD thesis, which represent the main focus of my PhD research, I was also involved in several other projects, some of which led to publications. The abstracts of the published or submitted articles are added in this appendix. My contribution to each of the papers is indicated underneath the abstracts.

Determination of saponins in *Maesa lanceolata* by LC-UV: development and validation

Theunis, M.H.B.L., Foubert, K., Pollier, J., González-Guzmán, M., Goossens, A., Vlietinck A.J., Pieters, L.A.C., and Apers, S.

Triterpene saponins are a class of plant natural products with a wide range of bioactivities, which makes them an interesting research subject. The small tree *Maesa lanceolata*, growing in African countries, is used in traditional medicine against various diseases. In previous work a triterpenoid saponin mixture was isolated from the leaves of *M. lanceolata* and the compounds were identified as closely related oleanane type triterpenes. The compounds showed virucidal, haemolytic, molluscicidal and antiangiogenic activity. Here we report the development of an extraction and quantification method to analyse saponin compounds in roots and leaves of *M. lanceolata*. After a purification step using C_{18} solid phase extraction (SPE) cartridges, the samples were analysed on a LC-UV/MS system. The identification of the peaks from the different saponins was confirmed based on the retention time and mass spectrum. The quantification was performed using the UV signals. The standard oleanolic acid curve was linear over a concentration range of 2.8-140.0 μg/mL. The recovery from the leaves was 94.5%. The precision of the method with respect to time and concentration was acceptable, with relative standard deviation (RSD%) values of 4.9 and 4.3, respectively.

Published in: Phytochemistry (IF: 3.104), **2007**, volume 68, pages 2825-2830.

Contribution: Cultivation and harvesting of the *M. lanceolata* plants.

Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants

Maes, L., Van Nieuwerburgh, F.C.W., Zhang, Y., Reed, D.W., Pollier, J., Vande Casteele, S.R.F., Inzé, D., Covello, P.S., Deforce, D.L.D., and Goossens, A.

• Biosynthesis of the sesquiterpene lactone and potent antimalarial drug artemisinin occurs in glandular trichomes of *Artemisia annua* plants and is subjected to a strict network of developmental and other regulatory cues.
• The effects of three hormones, jasmonate, gibberellin and cytokinin, were studied at the structural and molecular levels in two different *A. annua* chemotypes by microscopic analysis of gland development, and by targeted metabolite and transcript profiling. Furthermore, a genome-wide cDNA-amplified fragment length polymorphism (AFLP)-based transcriptome profiling was carried out of jasmonate-elicited leaves at different developmental stages.
• Although cytokinin and gibberellin positively affected at least one aspect of gland formation, these two hormones did not stimulate artemisinin biosynthesis. Only jasmonate simultaneously promoted gland formation and coordinated transcriptional activation of biosynthetic gene expression, which ultimately led to increased sesquiterpenoid accumulation with chemotype-dependent effects on the distinct pathway branches. Transcriptome profiling revealed a trichome-specific fatty acyl- coenzyme A reductase, trichome-specific fatty acyl-CoA reductase 1 (TFAR1), the expression of which correlates with trichome development and sesquiterpenoid biosynthesis.
• TFAR1 is potentially involved in cuticular wax formation during glandular trichome expansion in leaves and flowers of *A. annua* plants. Analysis of phytohormone-modulated transcriptional regulons provides clues to dissect the concerted regulation of metabolism and development of plant trichomes.

Published in: New Phytologist (IF: 6.033), **2011**, volume 189, pages 176-189.

Contribution: qRT-PCR analysis of artemisinin biosynthetic genes (Figure 4).

Characterization of a RABE GTPase expressed during morphogenesis of the unicellular green alga *Micrasterias denticulata* (Zygnemophyceaea, Streptophyta)

Vannerum, K., De Rycke, R., Pollier, J., Goossens, A., Inzé, D., and Vyverman, W.

Rab GTPases are central regulators of cell shape in land plants by coordinating vesicle trafficking during morphogenesis. To date, relatively little is known about the role of these ubiquitous signaling proteins during cell growth of microalgae, in particular of the related charophyte algae. This paper identifies the first charophyte Rab GTPase, MdRABE1, in *Micrasterias denticulata* Bréb., a convenient model organism for studying morphogenesis. Its expression correlated with the onset of morphogenesis and structural analysis indicated that it belongs to the RABE subclass. Confocal fluorescence and immuno-electron microscopy of transiently GFP-MdRABE1 overexpressing interphase cells demonstrated that the GFP MdRABE1 protein was localized to the endoplasmic reticulum, dictyosomes, exocytotic vesicles, the cell margin, the membranes of cell organelles and in the isthmus zone around the nucleus. Although overexpression phenotyping of both N- and C-terminal GFP-fusions failed to indicate further functional evidence of the MdRABE1 protein due to mortality of those transgenic cells, its expression profile, bioinformatics, and its intracellular localization suggest a role in vesicle trafficking during morphogenesis.

Submitted to: Journal of Phycology (IF: 2.270).

Contribution: Screening for the RABE GTPase in a *M. denticulata* cDNA library via colony hybridization.

Transcriptional analysis of cell growth and morphogenesis in *Micrasterias* (Streptophyta) provides insights into the evolution of expansins in the green algal progenitors of land plants

Vannerum, K. Huysman, M.J.J., De Rycke, R. Vuylsteke, M., Leliaert, F., Pollier, J., Lütz-Meindl, U., Gillard, J., De Veylder, L., Goossens, A., Inzé, D., and Vyverman, W.

Background

Streptophyte green algae share several features of cell growth and cell wall formation with their relatives, the embryophytic land plants. The multilobed cell wall of *Micrasterias denticulata*, which rebuilds symmetrically after cell division and consists of pectin and cellulose, makes this unicellular streptophyte alga an interesting model system to study the evolution of molecular controls on cell shape and cell wall formation within the green plant lineage.

Results

By genome-wide transcript expression profiling of synchronously growing cells, 107 genes were identified of which the expression correlated with the growth phase. Four transcripts showed high similarity to expansins, genes that had not been examined previously in green algae. Phylogenetic analysis demonstrated that they are most closely related to the plant EXPANSIN A family, although their domain organization is very divergent. A GFP-tagged version of the expansin-resembling protein MdEXP2 localized to the cell wall and in Golgi-derived vesicles. Overexpression phenotypes ranged from lobe elongation to loss of growth polarity and planarity. These results indicate that MdEXP2 can alter cell wall structure and thus may exert a function related to land plant expansins during cell morphogenesis.

Conclusions

Our results demonstrate the potential of *M. denticulata* as a unicellular model system, extended by the discovery of cell growth mechanisms similar to those in land plants. This study adds to growing body of evidence that the evolutionary origins of many cell wall components and regulating genes in embryophytes lie in an era prior to the colonization of land.

Submitted to: BMC Plant Biology (IF: 3.77).

Contribution: Help with screening for the FL cDNAs in a *M. denticulata* cDNA library via RACE PCR.

Appendix: Curriculum vitae

PERSONAL DETAILS

Name: Jacob Pollier
Date of Birth: July 21, 1980, Eeklo, Belgium.
Nationality: Belgian.
Marital status: Married to Tessa Moses.
Voskenslaan 453, 9000 Gent, Belgium
Tel: + 32 (0)498.85.02.58
E-mail:japol@psb.vib-ugent.be
jacobpollier@hotmail.com

EDUCATION

2005-2011 PhD in Sciences: Biotechnology
Institute: Ghent University, Belgium
Thesis: Combinatorial biochemistry of triterpene saponins in plants.
Promotor: Prof. Dr. Alain Goossens.

2003-2004 Master of Science in Molecular Biotechnology (with Distinction)
Institute: Ghent University, Belgium
Thesis: Post-transcriptional gene silencing in plants.
Promotor: Prof. Dr. Ann Depicker.

2001-2003 Industrieel Ingenieur in Chemie (optie Biochemie) (with Distinction)
(Master of Science in Industrial Sciences: Chemistry)
Institute: BME-CTL; Hogeschool Gent, Belgium.
Thesis: Integration of DNA-markers in *Lolium perenne* selection schedules.
Promotor: Prof. Dr. Isabel Roldán-Ruiz.

1998-2001 Gegradueerde in Chemie (optie biochemie) (with Distinction)
(Bachelor of Science: Chemistry)
Institute: BME-CTL; Hogeschool Gent, Belgium.
Thesis: Isolation and characterization of autofluorescent proteins from *Anthozoa* species.
Performed in Devgen.

RESEARCH PROJECTS

2010-2011	HAWKEYE	Industrial collaboration (SoluCel)
2010	PHOENIX	Industrial collaboration (Bayer CropScience)
2005-2009	COMBIPLAN	Combinatorial Biochemistry of Triterpene Saponins in Plants.

PUBLICATIONS

Pollier, J., González-Guzmán, M., Ardiles-Diaz, W., and Goossens, A. **An integrated PCR colony hybridization approach to screen cDNA libraries for full-length coding sequences corresponding to incomplete cDNAs.** Submitted to *PLoS ONE*.

Pollier, J., Moses, T., and Goossens, A. **Combinatorial biosynthesis in plants: A (p)review on its potential and future exploitation.** Submitted to *Natural Product Reports*.

Vannerum, K., De Rycke, R., Pollier, J., Goossens, A., Inzé, D., and Vyverman, W. **Characterization of a RABE GTPase expressed during morphogenesis of the unicellular green alga *Micrasterias denticulata* (Zygnemophyceaea, Streptophyta).** Submitted to *Journal of Phycology*.

Vannerum, K. Huysman, M.J.J., De Rycke, R. Vuylsteke, M., Leliaert, F., Pollier, J., Lütz-Meindl, U., Gillard, J., De Veylder, L., Goossens, A., Inzé, D., and Vyverman, W. **Transcriptional analysis of cell growth and morphogenesis in *Micrasterias* (Streptophyta) provides insights into the evolution of expansins in the green algal progenitors of land plants.** Submitted to *BMC Plant Biology*.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). **Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry.** Accepted by *Journal of Natural Products*.

Maes, L., Van Nieuwerburgh, F.C.W., Zhang, Y., Reed, D.W., Pollier, J., Vande Casteele, S.R.F., Inzé, D., Covello, P.S., Deforce, D.L.D., and Goossens, A. (2011). **Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants.** *New Phytologist* 189: 176-189.

Theunis, M.H.B.L., Foubert, K., Pollier, J., González-Guzmán, M., Goossens, A., Vlietinck A.J., Pieters, L.A.C., and Apers, S. (2007). **Determination of saponins in *Maesa lanceolata* by LC-UV: Development and validation.** *Phytochemistry* 68: 2825-2830.

OTHER SCIENTIFIC COMMUNICATIONS

Pollier, J., Moses, T., González-Guzmán, M., Ingelbrecht, L., Lambert, E., Theunis, M. Foubert, K., Apers, S., Pieters, L., Geelen, D., Inzé, D., and Goossens, A. (2009). **Identification of structural and regulatory genes involved in the biosynthesis of triterpene saponins in medicinal plants.** Oral presentation at TERPNET 2009 – 9th international meeting: Biosynthesis and Function of Isoprenoids in Plants, Microorganisms and Parasites. Tokyo (Japan), May 25-29, 2009.

Appendix: Supplementary data

Supplementary Table S1. LC ESI FT-ICR MS analysis of MKB1^{KD} roots.

Change	and tentative identification of the observed compounds	$[M-H]^-$	Formula (δ ppm)	FT-ICR MS^n: *m/z* (% base peak)
Up	?	675.23439	$C_{26}H_{43}O_{20}$ (-1.6)	Low abundant compound, no MS^n info available
Up	pentose derivate	425.16616	$C_{17}H_{29}O_{12}$ (-0.68)	**MS^2 [425.17]**: 383 (100), 365 (9), 293 (17) **MS^3 [425.17→383]**: 251 (100)
Up	?	741.22494	$C_{33}H_{41}O_{19}$ (0.25)	**MS^2 [741.22]**: 427 (46), 313 (100)
Up	?	473.14549	$C_{24}H_{25}O_{10}$ (0.36)	**MS^2 [473.15]**: 413 (1), 311 (3), 269 (100) **MS^3 [473.15→269]**: 254 (85), 253 (33), 237 (100), 227 (50), 148 (6), 145 (60) **MS^3 [473.15→311]**: 268 (100)
Up	?	559.20408	$C_{25}H_{35}O_{14}$ (1.52)	**MS^2 [559.20]**: 407 (100) **MS^3 [559.20→407]**: 365 (100)
Up	?	165.01928	$C_8H_5O_4$ (-0.32)	Low abundant compound, no MS^n info available
Up	?	670.16455	$C_{28}H_{32}O_{18}N$ (3.08)	Low abundant compound, no MS^n info available
Up	naringenin + hexose + 42 Da	475.16168	$C_{24}H_{27}O_{10}$ (1.49)	**MS^2 [475.16]**: 433 (43), 415 (10), 271 (100) **MS^3 [475.16→271]**: 239 (12), 165 (100), 137 (7) **MS^3 [475.16→433]**: 271 (100)
Up	?	1119.52638	$C_{53}H_{83}O_{25}$ (3.12)	Low abundant compound, no MS^n info available
Up	?	597.21986	$C_{28}H_{37}O_{14}$ (1.64)	**MS^2 [597.22]**: 193 (100)
Up	naringenin + hexose + malonic acid	519.15162	$C_{25}H_{27}O_{12}$ (1.58)	**MS^2 [519.15]**: 475 (100) **MS^3 [519.15→475]**: 433 (18), 415 (5), 271 (100) **MS^4 [519.15→475→271]**: 239 (12), 165 (100), 147 (6), 137 (12)
Up	?	313.07184	$C_{17}H_{13}O_6$ (0.25)	**MS^2 [313.07]**: 298 (100), 151 (33) **MS^3 [313.07→298]**: 283 (100), 270 (45), 226 (33) **MS^3 [313.07→151]**: 107 (100)
Up	?	606.18215	$C_{41}H_{24}O_3N_3$ (-0.27)	**MS^2 [606.18]**: 562 (100)
Up	divanilloyl hexose + 280 Da	759.21685	$C_{36}H_{39}O_{18}$ (3.507)	**MS^2 [759.22]**: 479 (100), 311 (6), 297 (4) **MS^3 [759.22→479]**: 311 (44), 297 (100)
Up	Afrormosin + malonyl hexose	1091.27251	$C_{52}H_{51}O_{26}$ (4.678)	**MS^2 [1091.27]**: 1047 (100), 799 (73), 500 (60)
Up	Afrormosin (297) + ???	1113.2565	$C_{54}H_{49}O_{26}$ (4.263)	**MS^2 [1113.26]**: 907 (21), 865 (100), 864 (12), 863 (9), 821 (38), 820 (8), 617 (62), 567 (20), 523 (8), 379 (3) **MS^3 [1113.26→617]**: 297 (100)
Up	?	585.38103	$C_{35}H_{53}O_7$ (2.311)	Low abundant compound, no MS^n info available
Up	dHex-Hex-Hex-hederagenin	941.51536	$C_{48}H_{77}O_{18}$ (4.058)	**MS^2 [941.51]**: 795 (100)$[M\text{-dHex-H}]^-$, 777 (12)$[M\text{-dHex-H}_2O\text{-H}]^-$, 633 (13)$[M\text{-dHex-Hex-H}]^-$, 615 (2)$[M\text{-dHex-Hex-H}_2O\text{-H}]^-$ **MS^3 [941.51→795]**: 633 (100)$[M\text{-dHex-Hex-H}]^-$, 615 (5)$[M\text{-dHex-Hex-H}_2O\text{-H}]^-$, 471 (20)$[Agly\text{-H}]^-$ **MS^3 [941.51→633]**: 471 (100)$[Agly\text{-H}]^-$, 393 (3)
Up	Hex-Bayogenin	649.39774	$C_{36}H_{57}O_{10}$ (3.108)	**MS^2 [649.40]**: 487 (100)$[Agly\text{-H}]^-$, 469 (15), 441 (28), 409 (36)
Up	Hex-Medicagenic acid	663.37701	$C_{36}H_{55}O_{11}$ (3.051)	**MS^2 [663.37]**: 645 (5)$[M\text{-H}_2O\text{-H}]^-$, 501 (13)$[M\text{-Hex-H}]^-$, 439 (100)$[Agly\text{-}CO_2\text{-}H_2O\text{-H}]^-$ **MS^3 [663.37→439]**: 393 (83), 391 (100)

Up	Malonyl-Hex-malonyl-Hex-Bayogenin	983.45391	$C_{48}H_{71}O_{21}$ (4.654)	**(MS)**: 983 (21)[M-H]$^-$, 895 (21)[M-CO_2-CO_2-H]$^-$, 734 (11)[M-CO_2-CO_2-Hex-H]$^-$, 691 (100)[M-malonyl-Hex-CO_2-H]$^-$, 649 (18)[M-malonyl-Hex-malonyl-H]$^-$, 631 (62)[M-malonyl-Hex-malonyl-H_2O-H]$^-$, 487 (39)[Agly-H]$^-$
Up	Malonyl-Hex-malonyl-Hex-Medicagenic acid	997.4345	$C_{48}H_{69}O_{22}$ (5.918)	**MS^2 [997.43]**: 953 (51)[M-CO_2-H]$^-$, 909 (100)[M-CO_2-CO_2-H]$^-$, 733 (6), 705 (24)[M-malonyl-Hex-CO_2-H]$^-$, 691 (26) **MS^3 [997.43→909]**: 891 (4)[M-CO_2-CO_2-H_2O-H]$^-$, 865 (6)[M-CO_2-CO_2-CO_2-H]$^-$, 849 (3)[M-CO_2-CO_2-malonyl-H_2O-H]$^-$, 703 (4)[M-CO_2-CO_2-CO_2-Hex-H]$^-$, 661 (6)[M-CO_2-CO_2-malonyl-Hex-H]$^-$, 643 (15)[M-CO_2-CO_2-malonyl-Hex-H_2O-H]$^-$, 439 (100)[Agly-H]$^-$
Up	Hex-Hederagenin	633.40242	$C_{36}H_{57}O_9$ (2.547)	**MS^2 [633.40]**: 615 (33)[M-H_2O-H]$^-$, 513 (15), 471 (100)[Agly-H]$^-$, 453 (24), 405 (24), 393 (39)
Up	Hex-Aglycone A	647.38233	$C_{36}H_{55}O_{10}$ (3.489)	**MS^2 [647.38]**: 485 (100)[Agly-H]$^-$
Up	Malonyl-Hex-malonyl-Hex-Hederagenin	967.45877	$C_{48}H_{71}O_{20}$ (4.498)	**MS^2 [879.47]**: 861 (2)[M-CO_2-CO_2-H_2O-H]$^-$, 819 (6)[M-malonyl-CO_2-H_2O-H]$^-$, 777 (5)[M-malonyl-malonyl-H_2O-H]$^-$, 717 (27)[M-Hex-CO_2-CO_2-H]$^-$, 675 (100)[M-Hex-malonyl-CO_2-H]$^-$, 657 (3)[M-Hex-malonyl-CO_2-H_2O-H]$^-$, 615 (72)[M-Hex-malonyl-malonyl-H_2O-H]$^-$
Up	?	981.43761	$C_{48}H_{69}O_{21}$ (2.545)	Low abundant compound, no MS^n info available
Down	?	863.24545	$C_{36}H_{47}O_{24}$ (-0.96)	Low abundant compound, no MS^n info available
Down	kaempferol + pentose + deoxyhexose + 2 hexoses	887.24717	$C_{38}H_{47}O_{24}$ (1.01)	**MS^2 [887.25]**: 725 (100)[M-Hex-H]$^-$ **MS^3 [887.25→725]**: 593 (68)[M-Hex-Pen-H]$^-$, 575 (100)[M-Hex-Pen-H_2O-H]$^-$, 327 (11), 285 (49)[Kaempferol-H]$^-$
Down	?	782.20754	$C_{41}H_{36}O_{15}N_1$ (-1.92)	**MS^2 [782.21]:** 764 (10)[M-H_2O-H]-, 739 (6), 461 (100) **MS^3 [782.21→461]**: 299 (9), 281 (100) **MS^4 [782.21→461→281]**: 266 (100)
Down	?	595.16713	$C_{27}H_{31}O_{15}$ (0.48)	**MS^2 [595.17]:** 505 (44), 475 (100), 385 (69), 355 (51)
Down	?	773.22731	$C_{32}H_{41}O_{20}N_2$ (1.93)	**MS^2 [773.23]:** 730 (13), 705 (100) **MS^3 [773.23→705]**: 429 (100)
Down	?	435.12997	$C_{21}H_{23}O_{10}$ (1.4)	**MS^2 [435.13]:** 417 (13)[M-H_2O-H]$^-$, 345 (31), 315 (100) **MS^3 [435.13→315]**: 297 (100), 271 (23), 243 (20), 209 (91), 191 (71), 190 (25), 163 (11), 151 (13) **MS^3 [435.13→345]**: 327 (100), 299 (9), 283 (18), 239 (55), 221 (52), 219 (36), 181 (10), 163 (6)
Down	?	995.34649	$C_{39}H_{63}O_{29}$ (0.44)	**MS^2 [995.35]:** 935 (100) **MS^3 [995.35→935]**: 875 (100)
Down	naringenin + hexose	433.11434	$C_{21}H_{21}O_{10}$ (0.74)	**MS^2 [433.11]:** 327 (79), 271 (49)[M-Hex-H]$^-$, 253 (100), 243 (40) **MS^3 [433.11→253]**: 225 (84), 211 (12), 209 (53), 197 (15), 181 (10), 135 (100), 133 (15), 117 (17) **MS^3 [435.13→327]**: 281 (6), 221 (6), 219 (13), 207 (14), 191 (64), 179 (12), 167 (25), 165 (100), 164 (9), 163 (15), 151 (5), 137 (53)
Down	?	328.12025	$C_{18}H_{18}O_5N$ (3.67)	**MS^2 [328.12]:** 241 (100), 223 (3) **MS^3 [328.12→241]**: 223 (100), 207 (14), 205 (6), 195 (23), 177 (10)
Down	naringenin + hexose + ?	637.17872	$C_{29}H_{33}O_{16}$ (2.31)	**MS^2 [637.18]:** 595 (5), 531 (23), 475 (13)[M-Hex-H]-, 253 (100) **MS^3 [637.18→253]**: 235 (17), 225 (100), 209 (39), 197 (15), 180 (9), 135 (78), 133 (11) **MS^3 [637.18→531]**: 369 (15), 351 (100), 333 (14), 321 (76), 263 (27), 219 (11), 207 (14), 179 (5)
Down	carboxybenzoic acid (Phthalic acid or isomer)	165.01929	$C_8H_5O_4$ (-0.26)	**MS^2 [165.02]:** 121 (100)[M-CO_2-H]$^-$ **MS^3 [165.02→121]**: 77 (100)[M-CO_2-CO_2-H]$^-$
Down	hydroxymethoxybenzoic acid + hydroxybenzoic acid + 276 Da	581.15228	$C_{26}H_{29}O_{15}$ (1.87)	**MS^2 [581.15]:** 413 (100), 167 (2) **MS^3 [581.15→413]**: 137 (100) **MS^3 [581.15→167]**: 152 (72), 123 (100), 108 (10)

Down	naringenin + hexose + malonic acid	519.1152	$C_{24}H_{23}O_{13}$ (1.5)	**MS^2 [519.12]:** 475 (100)[M-CO_2-H]$^-$ **MS^3 [519.12→475]**: 433 (7)[M-malonyl-H]$^-$, 369 (45), 271 (43)[M-malonyl-Hex-H]$^-$, 253 (100)[M-malonyl-Hex-H_2O-H]-, 243 (43), 165 (5) **MS^4 [519.12→475→253]**: 225 (100), 211 (5), 209 (39), 208 (4), 197 (4), 185 (9), 181 (3), 135 (59), 133 (16)
Down	?	695.18495	$C_{31}H_{35}O_{18}$ (2.97)	**MS^2 [695.18]:** 607 (100), 519 (9), 413 (11)
Down	pentose + hexose derivate	367.16127	$C_{15}H_{27}O_{10}$ (0.82)	**MS^2 [367.16]:** 235 (100)[M-Pen-H]$^-$, 161 (15) **MS^3 [367.16→235]**: 161 (100), 101 (22) **MS^3 [367.16→161]**: 101 (100)
Down	?	521.13069	$C_{24}H_{25}O_{13}$ (1.2)	**MS^2 [521.13]:** 503 (24), 477 (90), 459 (59), 435 (100) **MS^3 [521.13→435]**: 417 (12), 345 (32), 315 (100) **MS^3 [521.13→477]**: 459 (10), 387 (15), 357 (100) **MS^4 [521.13→435→315]**: 297 (100), 271 (20), 243 (24), 209 (87), 191 (65), 190 (27), 163 (10), 151 (12), 123 (3)
Down	?	993.25443	$C_{44}H_{49}O_{26}$ (2.69)	Low abundant compound, no MS^n info available
Down	malonic acid + pentose + hexose derivate	453.16205	$C_{18}H_{29}O_{13}$ (1.51)	**MS^2 [453.16]:** 409 (100)[M-CO_2-H]$^-$ **MS^3 [453.16→409]**: 367 (100)[M-malonyl-H]$^-$, 349 (11), 235 (5)[M-malonyl-Pen-H]$^-$ **MS^4 [453.16→409→367]**: 235 (100), 161 (13)
Down	?	629.21054	$C_{28}H_{37}O_{16}$ (2.91)	Low abundant compound, no MS^n info available
Down	?	879.31497	$C_{40}H_{55}O_{14}N_4S_2$ (-1.37)	Low abundant compound, no MS^n info available
Down	?	699.23788	?	Low abundant compound, no MS^n info available
Down	hexose derivate + hydroxymethoxybenzoic acid + hydroxybenzoic acid + 276 Da	891.22321	$C_{40}H_{43}O_{23}$ (3.53)	**MS^2 [891.22]:** 753 (47), 729 (10)[M-Hex-H]$^-$, 581 (100), 477 (44), 413 (96) **MS^3 [891.22→581]**: 413 (100) **MS^3 [891.22→413]**: 137 (100)
Down	?	597.12626	$C_{29}H_{25}O_{14}$ (2.16)	**MS^2 [597.13]:** 553 (71)[M-CO_2-H]$^-$, 491 (81), 287 (38), 269 (100), 259 (57)
Down	?	305.11427	$C_{15}H_{17}O_5N_2$ (0.25)	**MS^2 [305.11]:** 287 (6)[M-H_2O-H]$^-$, 277 (33), 233 (6), 229 (16), 203 (100), 189 (61), 135 (6) **MS^3 [305.11→203]**: 185 (83), 175 (100)
Down	?	271.06125	$C_{15}H_{11}O_5$ (0.38)	**MS^2 [271.06]:** 256 (100)[M-$^{\bullet}CH_3$-H]$^-$, 243 (49)[M-CO-H]$^-$, 227 (12)[M-CO_2-H]$^-$, 212 (14)[M-$^{\bullet}CH_3$-CO_2-H]$^-$, 148 (3), 123 (4), 108 (2) **MS^3 [271.06→256]**: 241 (7)[M-2$^{\bullet}CH_3$-H]$^-$, 228 (67)[M-$^{\bullet}CH_3$-CO-H]$^-$, 227 (100), 212 (10)[M-$^{\bullet}CH_3$-CO_2-H]$^-$, 108 (8) **MS^3 [271.06→243]**: 228 (100)[M-$^{\bullet}CH_3$-CO-H]$^-$, 225 (70)[M-CO-H_2O-H]$^-$
Down	?	783.3852		Low abundant compound, no MS^n info available
Down	?	599.14158	$C_{29}H_{27}O_{14}$ (1.59)	**MS^2 [599.14]:** 493 (100), 463 (52), 449 (32), 273 (40), 269 (44), 259 (31)
Down	?	695.25791	$C_{33}H_{43}O_{16}$ (3.24)	**MS^2 [695.26]:** 673 (2), 665 (100), 653 (2), 635 (4) **MS^3 [695.26→665]**: 647 (12), 623 (100), 605 (59), 587 (4), 533 (6), 329 (40) **MS^4 [695.26→665→623]**: 605 (56), 459 (10), 329 (100), 315 (29), 311 (12) **MS^4 [695.26→665→605]**: 587 (77), 575 (12), 329 (100), 315 (25)
Down	Similar Rt 9,7	515.17784	$C_{23}H_{31}O_{13}$ (1.6)	**MS^2 [515.18]:** 413 (100) **MS^3 [515.18→413]**: 137 (100)
Down	?	863.28541	$C_{37}H_{51}O_{23}$ (3.18)	**MS^2 [863.29]:** 845 (48), 725 (49), 449 (24), 431 (100), 387 (72)
Down	naringenin + rhamnose +	581.13109	$C_{29}H_{25}O_{13}$ (1.76)	**MS^2 [581.13]:** 537 (100)[M-CO_2-H]$^-$, 475 (27), 431 (4), 415 (5), 271 (19), 253 (44), 243 (15)

	pentose?			**MS^3 [581.13→537]**: 431 (34), 271 (34), 253 (100), 243 (42), 165 (4) **MS^3 [581.13→253]**: 225 (100), 211 (6), 209 (43), 135 (68)
Down	?	428.18085	$C_{22}H_{26}O_6N_3$ (-4.34)	Low abundant compound, no MS^n info available
Down	?	975.28121	$C_{45}H_{51}O_{24}$ (3.73)	Low abundant compound, no MS^n info available
Down	pinoresinol + hexose + pentose	651.2314	$C_{31}H_{39}O_{15}$ (3)	**MS^2 [651.23]:** 633 (85), 621 (17), 487 (10), 357 (100), 343 (10), 339 (37), 327 (13) **MS^3 [581.13→357]**: 339 (63), 151 (100) **MS^3 [581.13→633]**: 339 (100)
Down	lariciresinol + hexose + pentose + malonic acid?	739.2483	$C_{34}H_{43}O_{18}$ (3.8)	**MS^2 [739.25]:** 695 (100) **MS^3 [739.25→695]**: 653 (93), 635 (100), 563 (4), 359 (8), 329 (33) **MS^3 [739.25→437]**: 305 (100), 161 (20)
Down	malonic acid + pentose derivate?	523.2041	$C_{22}H_{35}O_{14}$ (1.66)	**MS^2 [523.20]:** 479 (100) **MS^3 [523.20→479]**: 437 (100), 419 (8), 305 (6), 297 (4)
Down	lariciresinol + hexose + malonic acid?	563.21454	$C_{28}H_{35}O_{12}$ (2.02)	**MS^2 [563.21]:** 503 (6), 359 (80), 329 (100) **MS^3 [563.21→329]**: 314 (17), 299 (56), 284 (11), 193 (26), 192 (21), 178 (100), 175 (11), 161 (12), 160 (15)
Down	dihydroxyflavone or hydroxyflavonol derivate?	583.14668	$C_{29}H_{27}O_{13}$ (1.66)	**MS^2 [583.15]:** 477 (100), 459 (5), 449 (8), 433 (31), 431 (14), 297 (35), 271 (26), 255 (10), 253 (45), 243 (24) **MS^3 [583.15→253]**: 225 (100)
Down	?	1027.33379	$C_{46}H_{59}O_{26}$ (3.68)	**MS^2 [1027.33]:** 613 (25), 595 (29), 431 (100), 413 (7) **MS^3 [1027.33→413]**: 137 (100)
Down	?	503.1204	$C_{24}H_{23}O_{12}$ (1.79)	Low abundant compound, no MS^n info available
Down	pinoresinol + hexose + pentose + malonic acid	737.23297	$C_{34}H_{41}O_{18}$ (4.25)	**MS^2 [737.23]:** 693 (100) **MS^3 [737.23→693]**: 675 (37), 663 (8), 651 (100), 633 (71), 615 (5), 357 (58), 339 (9), 327 (6) **MS^4 [737.236→693→651]**: 633 (100), 621 (26), 519 (8), 487 (14), 357 (87), 343 (26), 339 (36), 327 (16), 313 (5), **MS^4 [737.236→693→357]**: 342 (11), 327 (10), 311 (15), 151 (100), 136 (32)
Down	hydroxybenzoic acid + pentose + hexose derivative	613.21538	$C_{28}H_{37}O_{15}$ (2.59)	**MS^2 [613.22]:** 431 (100) **MS^3 [613.22→431]**: 299 (38), 137 (100)
Down	?	923.28627	$C_{42}H_{51}O_{23}$ (3.86)	Low abundant compound, no MS^n info available
Down	?	919.40594	$C_{39}H_{67}O_{24}$ (3.4)	**MS^2 [919.41]:** 901 (100), 857 (27), 757 (22), 743 (83), 725 (11), 669 (47), 651 (20) **MS^3 [919.41→901]**: 857 (51), 743 (100), 669 (21), 651 (30) **MS^3 [919.41→743]**: 669 (100), 651 (75)
Down	?	785.2327	$C_{38}H_{41}O_{18}$ (3.86)	**MS^2 [785.23]:** 767 (100), 755 (18), 413 (8) **MS^3 [785.23→767]**: 413 (100), 353 (98)
Down	?	997.24566	$C_{50}H_{45}O_{22}$ (4.88)	**MS^2 [997.24]:** 743 (19), 725 (5), 683 (4), 623 (13), 581 (100) **MS^3 [997.24→581]**: 537 (100), 475 (27), 271 (19) **MS^3 [997.24→743]**: 581 (100)
Down	?	1083.45469	$C_{54}H_{71}O_{21}N_2$ (-0.73)	Low abundant compound, no MS^n info available
Down	malonic acid derivative	1015.29832	$C_{55}H_{51}O_{19}$ (-4.61)	**MS^2 [1015.29]:** 971 (100) **MS^3 [1015.29→971]**: 929 (48), 911 (74), 621 (38), 579 (100), 557 (97)
Down	?	993.34472	$C_{50}H_{57}O_{21}$ (4.97)	**MS^2 [993.34]:** 975 (100)
Down	?	1035.35403	$C_{34}H_{67}O_{35}$ (6.95)	Low abundant compound, no MS^n info available

Down	?	461.20364	$C_{21}H_{33}O_{11}$ (1.745)	Low abundant compound, no MS^n info available
Down	?	1051.33501	$C_{48}H_{59}O_{26}$ (4.761)	**MS^2 [1051.34]:** 991 (9), 973 (12), 949 (100), 907 (57), 889 (34), 725 (27) **MS^3 [1051.34→949]**: 889 (48), 767 (78), 725 (100), 707 (34) **MS^3 [1051.34→907]**: 725 (100), 707 (27)
Down	pentose + malonic acid derivative	539.23538	$C_{23}H_{39}O_{14}$ (1.58)	**MS^2 [539.24]:** 495 (100)[M-CO_2-H]$^-$ **MS^3 [539.24→495]**: 453 (100)[M-malonyl-H]$^-$ **MS^4 [539.24→495→453]**: 321 (100)[M-malonyl-Pen-H]$^-$
Down	G(8-O-4)FA + 2 G?	781.29551	$C_{37}H_{49}O_{18}$ (3.932)	**MS^2 [781.30]:** 739 (16), 585 (100), 389 (70), 371 (23), 341 (26) **MS^3 [781.30→389]**: 371 (100), 341 (59), 193 (55)
Down	malonic acid + pentose derivate?	507.2092	$C_{22}H_{35}O_{13}$ (1.746)	**MS^2 [507.21]:** 463 (100)[M-CO_2-H]$^-$ **MS^3 [507.21→463]**: 421 (100)[M-malonyl-H]$^-$, 403 (10), 331 (3), 289 (4)[M-malonyl-Pen-H]$^-$ **MS^4 [507.21→463→421]**: 289 (100), 161 (35), 159 (5) **MS^5 [507.21→463→421→289]**: 161 (100), 159 (11)
Down	malonic acid + pentose derivate?	507.20895	$C_{22}H_{35}O_{13}$ (1.25)	**MS^2 [507.21]:** 463 (100)[M-CO_2-H]$^-$ **MS^3 [507.21→463]**: 421 (100)[M-malonyl-H]$^-$
Down	dHex-Hex-HexA-Aglycone D	973.50386	$C_{48}H_{78}O_{20}$ (2.56)	**MS^2 [973,50]**: 955 (100)[M-H_2O-H]$^-$, 929 (10)[M-CO_2-H]$^-$, 911 (53)[M-H_2O-CO_2-H]$^-$, 827 (3)[M-dHex-H]$^-$, 809 (7)[M-dHex-H_2O-H]$^-$, 783 (6)[M-dHex-CO_2-H]$^-$, 765 (19)[M-dHex-H_2O-CO_2-H]$^-$, 665 (2)[M-dHex-Hex-H]$^-$, 647 (26)[M-dHex-Hex-H_2O-H]$^-$, 629 (14)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 603 (7)[M-dHex-Hex-H_2O-CO_2-H]$^-$, 557 (13)[M-dHex-Hex-H2O-$C_3H_6O_3$-H]$^-$, 489 (11)[Agly-H]$^-$ **MS^3 [973.50→955]**: 911 (100)[M-H_2O-CO_2-H]$^-$, 765 (21)[M-dHex-H_2O-CO_2-H]$^-$, 679 (9), 629 (49)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 557 (55)[M-dHex-Hex-H2O-$C_3H_6O_3$-H]$^-$, 489 (7)[Agly-H]$^-$
Down	?	397.14851	$C_{19}H_{25}O_9$ (-4.77)	**MS^2 [397.15]**: 371 (100), 289 (7) **MS^3 [397.15→371]**: 289 (100) **MS^3 [397.15→289]**: 161 (100), 159 (9), 143 (4), 113 (12), 101 (22)
Down	?	961.4853	$C_{43}H_{77}O_{23}$ (-0.85)	Low abundant compound, no MS^n info available
Down	?	397.14853	$C_{19}H_{25}O_9$ (-4.72)	**MS^2 [397.15]**: 353 (100) **MS^3 [397.15→353]**: 311 (100), 183 (30)
Down	?	773.32414	$C_{36}H_{53}O_{18}$ (0.52)	Low abundant compound, no MS^n info available
Down	?	1165.57122	$C_{55}H_{89}O_{26}$ (5.546)	Low abundant compound, no MS^n info available
Down	Hex-Hex-Hex-Bayogenin	973.50407	$C_{48}H_{77}O_{20}$ (2.776)	**MS^2 [973.50]**: 955 (10)[M-H_2O-H]$^-$,811 (100)[M-Hex-H]$^-$, 649 (20)[M-Hex-Hex-H]$^-$, 631 (7)[M-Hex-Hex-H_2O-H]$^-$, 487 (8) [Agly-H]$^-$ **MS^3 [973.50→811]**: 649 (100)[M-Hex-Hex-H]$^-$, 631 (42)[M-Hex-Hex-H_2O-H]$^-$, 487 (86) [Agly-H]$^-$ **MS^3 [973.50→649]**: 487 (100) [Agly-H]$^-$ 409 (1), 393 (1), 391 (1)
Down	Hex-Hex-Hex-Medicagenic Acid	987.48516	$C_{48}H_{75}O_{21}$ (4.584)	**MS^2 [987.49]**: 825 (100)[M-Hex-H]$^-$, 781 (5)[M-Hex-CO_2-H]$^-$, 711 (3), 663 (4)[M-Hex-Hex-H]$^-$, 601 (4)[M-Hex-Hex-CO_2-H_2O-H]$^-$, 439 (53)[Agly-CO_2-H_2O-H]$^-$ **MS^3 [987.49→825]**: 439 (100)[Agly-CO_2-H_2O-H]$^-$
Down	?	756.35992	?	**MS^2 [756.36]**: 738 (79), 712 (100), 694 (18), 641 (72), 640 (36), 597 (19), 596 (26) **MS^3 [756.36→712]**: 694 (53), 597 (83), 596 (100) **MS^3 [756.36→738]**: 720 (12), 694 (45), 640 (100), 623 (52), 622 (31), 596 (33)
Down	?	1121.52513	$C_{49}H_{85}O_{28}$ (1.645)	Low abundant compound, no MS^n info available
Down	?	1149.57585	$C_{55}H_{89}O_{25}$ (5.227)	**MS^2 [1149.58]**: 1069 (100), 1025 (62), 941 (7), 879 (30), 761 (27), 743 (22), 671 (26)
Down	Hex-Hex-Bayogenin	811.45241	$C_{42}H_{68}O_{15}$ (4.763)	**MS^2 [811.45]**: 649 (100)[M-Hex-Hex-H]$^-$, 631 (8)[M-Hex-Hex-H_2O-H]$^-$, 487 (62) [Agly-H]$^-$

				MS^3 [811.45→649]: 487 (100) $[Agly-H]^-$ **MS^3 [811.45→487]**: 409 (100), 403 (21), 393 (38), 391 (74)
Down	?	840.32397	$C_{46}H_{50}O_{14}N$ (0.347)	**MS^2 [840.32]**: 711 (100), 649 (18) **MS^3 [840.32→711]**: 667 (100), 610 (48), 562 (23) **MS^3 [840.32→649]**: 619 (89), 605 (57), 575 (36), 564 (100), 432 (18), 415 (26)
Down	Malonyl-Hex-HexA-Bayogenin	911.43371	$C_{45}H_{67}O_{19}$ (6.041)	**MS^2 [867.44]**: 825 (34)[M-malonyl-H]$^-$, 807 (25)[M-malonyl-H_2O-H]$^-$, 705 (9)[M-CO_2-Hex-H]$^-$, 645 (77)[M-malonyl-Hex-H_2O-H]$^-$, 601 (100)[M-malonyl-Hex-H_2O-CO_2-H]$^-$ **MS^3 [867.44→601]**: 583 (3), 529 (100), 487 (7)$[Agly-H]^-$
Down	?	711.28055	$C_{24}H_{47}O_{20}N_4$ (2.302)	**MS^2 [711.28]**: 693 (23), 681 (12), 667 (100), 610 (47), 592 (14), 562 (21), 375 (7), 358 (4) **MS^3 [711.28→667]**: 649 (74), 637 (100), 619 (29), 562 (21), 416 (15), 375 (14)
Down	Hex-HexA-Aglycone A	823.41492	$C_{42}H_{63}O_{16}$ (3.352)	**MS^2 [823.41]**: 661 (100)[M-Hex-H]$^-$, 617 (33)[M-Hex-CO_2-H]$^-$, 599 (52)[M-Hex-H_2O-CO_2-H]$^-$, 485 (15)$[Agly-H]^-$ **MS^3 [823.41→661]**: 585 (5), 485 (100)$[Agly-H]^-$, 467 (34) **MS^3 [823.41→599]**: 527 (100), 485 (3)$[Agly-H]^-$, 467 (2)
Down	?	591.20735	$C_{29}H_{35}O_{13}$ (-1.631)	**MS^2 [591.21]**: 477 (64), 403 (100), 385 (20), 331 (21), 259 (33), 247 (24), 187 (73) **MS^3 [591.21→403]**: 385 (17), 331 (100) **MS^3 [591.21→187]**: 143 (100), 121 (62)
Down	?	1073.52242	$C_{52}H_{81}O_{23}$ (4.665)	Low abundant compound, no MS^n info available
Down	?	563.21203	$C_{28}H_{35}O_{12}$ (-2.432)	Low abundant compound, no MS^n info available
Down	dHex-Hex-HexA-soyasapogenol A	957.51003	$C_{48}H_{77}O_{19}$ (3.735)	**MS^2 [957.51]**: 939 (100)[M-H_2O-H]$^-$, 895 (50)[M-H_2O-CO_2-H]$^-$, 811 (8)[M-dHex-H]$^-$, 767 (3)[M-dHex-CO_2-H]$^-$, 749 (29)[M-dHex-H_2O-CO_2-H]$^-$, 649 (1)[M-dHex-Hex-H]$^-$, 631 (29)[M-dHex-Hex-H_2O-H]$^-$, 613 (15)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 541 (23)[M- dHex-Hex-H_2O-$C_3H_6O_3$-H]$^-$, 473 (8)$[Agly-H]^-$ **MS^3 [957.51→939]**: 895 (100)[M-H_2O-CO_2-H]$^-$, 749 (14)[M-dHex-H_2O-CO_2-H]$^-$, 613 (61)[M-dHex-Hex-H_2O-H_2O-H]$^-$, 541 (62)[M- dHex-Hex-H_2O-$C_3H_6O_3$-H]$^-$ **MS^3 [957.51→895]**: 749 (100)[M-dHex-H_2O-CO_2-H]$^-$, 587 (5)[M-dHex-Hex-H_2O-CO_2-H]$^-$, 473 (13)$[Agly-H]^-$
Down	?	641.35192	$C_{24}H_{53}O_{17}N_2$ (-3.916)	**MS^2 [641.35]**: 623 (5), 597 (100), 480 (6), 327 (3) **MS^3 [641.35→597]**: 579 (30), 553 (24), 535 (14), 480 (48), 327 (100), 309 (19)
Down	?	835.34199	$C_{41}H_{55}O_{18}$ (3.115)	**MS^2 [835.34]**: 793 (22), 639 (100), 389 (73), 371 (15), 341 (28) **MS^3 [835.34→639]**: 597 (31), 463 (36), 445 (15), 403 (15), 193 (100)
Down	Malonyl-Hex-HexA-Hederagenin	851.44613	$C_{45}H_{67}O_{18}$ (3.136)	**MS^2 [851.45]**: 809 (29)[M-malonyl-H]$^-$, 791 (27)[M-malonyl-H_2O-H]$^-$, 689 (82)[M-Hex-CO_2-H]$^-$, 629 (84)[M-malonyl-H_2O-H]$^-$, 585 (100)[M-malonyl-CO_2-H_2O-H]$^-$, 471 (4)$[Agly-H]^-$ **MS^3 [851.45→585]**: 513 (100), 471 (13)$[Agly-H]^-$ **MS^3 [851.45→629]**: 629 (100)[M-malonyl-H_2O-H]$^-$, 513 (54), 471 (78)[Agly-H]$^-$, 453 (45), 405 (64), 393 (74), 387 (22)
Down	Vanillic acid + unknown	613.25165	$C_{29}H_{41}O_{14}$ (2.399)	**MS^2 [613.25]**: 445 (100)[M-Vanillic acid-H]$^-$ **MS^3 [613.25→445]**: 403 (100)
Down	Malonyl-Hex-HexA-Aglycone A	865.42579	$C_{44}H_{65}O_{17}$ (3.543)	**MS^2 [865.43]**: 823 (3)[M-malonyl-H]$^-$, 821 (4)[M-CO_2-CO_2-H]$^-$, 805 (28)[M-malonyl-H_2O-H]$^-$, 761 (4)[M-malonyl-CO_2-H_2O-H]$^-$, 703 (8)[M-CO_2-Hex-H]$^-$, 643 (100)[M-malonyl-Hex-H_2O-H]$^-$, 599 (35)[M-malonyl-Hex-CO_2-H_2O-H]$^-$, 485 (1)$[Agly-H]^-$ **MS^3 [865.43→643]**: 509 (100), 485 (64)$[Agly-H]^-$, 467 (41), 421 (41)
Down	?	1135.5037	$C_{49}H_{83}O_{29}$ (1.013)	Low abundant compound, no MS^n info available

Down	Vanillic acid + unknown	613.25278	$C_{29}H_{41}O_{14}$ (4.241)	**MS^2 [613.25]**: 445 (100)[M-Vanillic acid-H]$^-$ **MS^3 [613.25→445]**: 403 (100)
Down	Hex-Pen-hederagenin	765.44543	$C_{41}H_{65}O_{13}$ (3.088)	**MS^2 [765.45]**: 603 (100)[M-Hex-H]$^-$, 585 (7)[M-Hex-H_2O-H]$^-$, 471 (6)[Agly-H]$^-$ **MS^3 [765.45→603]**: 471 (100)[Agly-H]$^-$, 423 (14)
Down	Ferulic acid? + unknown	639.26765	$C_{31}H_{43}O_{14}$ (2.848)	**MS^2 [639.27]**: 463 (21), 445 (22), 403 (11), 193 (100) **MS^3 [639.27→193]**: 178 (97), 149 (100), 134 (60)
Down	?	809.43631	$C_{42}H_{65}O_{15}$ (4.22)	Low abundant compound, no MS^n info available
Down	?	712.37007	$C_{39}H_{54}O_{11}N$ (-0.231)	**MS^2 [712.37]**: 668 (100), 650 (11), 593 (8) **MS^3 [712.37→668]**: 650 (24), 624 (100), 593 (20), 549 (14), 527 (13), 439 (33)
Down	Hex-dHex-Malonyl-Medicagenic acid	895.43821	$C_{45}H_{68}O_{18}$ (5.496)	**MS^2 [851.45]**: 809 (49)[M-malonyl-H]$^-$, 791 (8)[M-malonyl-H_2O-H]$^-$, 689 (37)[M-Hex-CO_2-H]$^-$, 629 (48)[M-malonyl-Hex-H_2O-H]$^-$, 585 (100)[M-malonyl-Hex-H_2O-CO_2-H]$^-$, 439 (9)[Agly-CO_2-H_2O-H]$^-$
Down	?	817.33168	$C_{41}H_{53}O_{17}$ (3.495)	**MS^2 [817.33]**: 799 (100) **MS^3 [817.33→799]**: 353 (100)

HexA, uronic acid, such as glucuronic acid or galacturonic acid; Hex, hexose, such as glucose or galactose; dHex, 6-deoxyhexose, such as rhamnose or fucose; Pen, pentose, such as arabinose or xylose.