Metabolic engineering for production of triterpenoid saponin building blocks in plants and yeast

Tessa Moses

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

Academic year: 2011-2012
Promoter: Prof. Alain Goossens
Co-promoter: Prof. Johan Thevelein
This work was conducted in the VIB Department of Plant Systems Biology, Ghent University and the VIB Department of Molecular Microbiology, Katholieke Universiteit Leuven.

This work was supported by the Flanders Institute for Biotechnology (VIB) International PhD Fellowship Program of 2007.
Board of Examiners

Prof. Ann Depicker (Chair)
VIB Department of Plant Systems Biology, Department of Plant Biotechnology and Bioinformatics, Faculty of Sciences, Ghent University

Prof. Alain Goossens (Promoter)
VIB Department of Plant Systems Biology, Department of Plant Biotechnology and Bioinformatics, Faculty of Sciences, Ghent University

Prof. Johan Thevelein (Co-promoter)
VIB Department of Molecular Microbiology, Laboratory of Molecular Cell Biology, Katholieke Universiteit Leuven

Prof. Bartel Vanholme*
VIB Department of Plant Systems Biology, Department of Plant Biotechnology and Bioinformatics, Faculty of Sciences, Ghent University

Dr. Sofie De Maeseneire*
Department of Biochemical and Microbial Technology, Faculty of Bioscience Engineering, Ghent University

Prof. Anne Osbourn
Department of Metabolic Biology, John Innes Centre

Prof. Danny Geelen*
Department of Plant Production, Faculty of Bioscience Engineering, Ghent University

Prof. Jan Van Bocxlaer
Department of Bio-analysis, Faculty of Pharmaceutical Sciences, Ghent University

Prof. Luc Pieters*
Department of Pharmaceutical Sciences, Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, University of Antwerp

Prof. Nico Callewaert*
VIB Department for Molecular Biomedical Research, Department of Biochemistry and Microbiology, Faculty of Sciences, Ghent University

* Members of the Reading Committee
# Table of contents

Commonly used abbreviations and gene names 001  
Abstract 005  
Chapter 1  Scope of Research 007  
Chapter 2  Bioengineering of plant terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro* 011  
Chapter 3  An introduction to Gas Chromatography – Mass Spectrometry (GC-MS) and its application for triterpenoid sapogenin analysis 059  
A novel β-amyrin hydroxylase from *Bupleurum falcatum*, its utilization for combinatorial biosynthesis in *Saccharomyces cerevisiae* and a novel approach for extraction of triterpene sapogenins from yeast cells 073  
Chapter 5  Unravelling triterpenoid saponin biosynthesis in *Maesa lanceolata* using *Saccharomyces cerevisiae* 105  
Chapter 6  Combinatorial biosynthesis of triterpene sapogenins in *Saccharomyces cerevisiae* 131  
Chapter 7  The protein quality control system manages plant defense compound synthesis 155  
Chapter 8  Outlook 193  
Summary 205  
Addendum 1  List of yeast strains generated in this study and the sapogenin backbones or sapogenins they produce. 207  
Addendum 2  Cloning strategy and PCR fragments amplified for generation of strain *TM1*. 213  
Additional publications and patents 215  
Acknowledgements 219  
Curriculum vitae 223
Commonly used abbreviations and gene names

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>αCD</td>
<td>α-cyclodextrin</td>
</tr>
<tr>
<td>ACT</td>
<td>Artemisinin based combination therapy</td>
</tr>
<tr>
<td>ADS</td>
<td>Amorpha-4,11-diene synthase</td>
</tr>
<tr>
<td>βCD</td>
<td>β-cyclodextrin</td>
</tr>
<tr>
<td>bAS</td>
<td>β-amyrin synthase</td>
</tr>
<tr>
<td>BF567</td>
<td><em>Bupleurum falcatum</em> putative β-amyrin 21-hydroxylase</td>
</tr>
<tr>
<td>CD</td>
<td>Cyclodextrin</td>
</tr>
<tr>
<td>cDNA-AFLP</td>
<td>Complementary DNA - amplified fragment length polymorphism</td>
</tr>
<tr>
<td>CDS</td>
<td>Coding sequence</td>
</tr>
<tr>
<td>CECF</td>
<td>Crude extract cell-free system</td>
</tr>
<tr>
<td>CPR</td>
<td>Cytochrome P450 reductase</td>
</tr>
<tr>
<td>CTR</td>
<td>Control</td>
</tr>
<tr>
<td>CytP450</td>
<td>Cytochrome P450 monooxygenase</td>
</tr>
<tr>
<td>CYP716A12</td>
<td><em>Medicago truncatula</em> β-amyrin 28-oxidase</td>
</tr>
<tr>
<td>CYP88D6</td>
<td><em>Glycyrrhiza uralensis</em> β-amyrin 11-oxidase</td>
</tr>
<tr>
<td>CYP93E2</td>
<td><em>Medicago truncatula</em> β-amyrin 24-hydroxylase</td>
</tr>
<tr>
<td>CYPED</td>
<td>Cytochrome P450 engineering database</td>
</tr>
<tr>
<td>DER3</td>
<td>HMGR degradation 1</td>
</tr>
<tr>
<td>DMAPP</td>
<td>Dimethylallyl pyrophosphate</td>
</tr>
<tr>
<td>DO</td>
<td>Drop out supplement</td>
</tr>
<tr>
<td>DOXP</td>
<td>1-deoxy-D-xylulose 5-phosphate</td>
</tr>
<tr>
<td>DXS</td>
<td>Deoxyxylulose 5-phosphate synthase</td>
</tr>
<tr>
<td>EI</td>
<td>Electron ionization</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAD</td>
<td>Endoplasmic reticulum associated degradation</td>
</tr>
<tr>
<td>ERG7</td>
<td>Lanosterol synthase</td>
</tr>
<tr>
<td>ERG9</td>
<td>Squalene synthase</td>
</tr>
<tr>
<td>EtOH</td>
<td>Ethanol</td>
</tr>
</tbody>
</table>
FL-ORF  Full length - open reading frame
fLUC    Firefly luciferase
FPP     Farnesyl pyrophosphate
FPPS    Farnesyl pyrophosphate synthase
FT-ICR-MS Fourier transform - ion cyclotron resonance - mass spectrometry
γCD     γ-cyclodextrin
GC-MS   Gas chromatography – mass spectrometry
GGPP    Geranylgeranyl pyrophosphate
GMA     3-O-Glc-medicagenic acid
GPP     Geranyl pyrophosphate
GRAS    Generally recognized as safe
gTME    global transcription machinery engineering
HMG-CoA 3-hydroxy-3-methylglutaryl-CoA
HMGR    3-hydroxy-3-methylglutaryl-CoA reductase
HRD1    HMGR degradation 1
IPP     Isopentenyl pyrophosphate
IPTG    Isopropyl-β-D-1-thiogalactopyranoside
IVC     In vitro compartmentalization
JA      Jasmonic acid
KD      Knock down
LC-ESI  Liquid chromatography - electrospray ionization
MβCD    Methyl β-cyclodextrin
MCS     Multiple cloning site
MeJA    Methyl jasmonate
MEP     2-C-methyl-D-erythritol 4-phosphate
MKB1    Makibishi 1, RMA-like E3 ubiquitin ligase
ML257   *Maesa lanceolata* β-amyrin 28-oxidase
ML593   *Maesa lanceolata* putative β-amyrin 21-hydroxylase
MtGEA   *Medicago truncatula* gene expression atlas
MVA     Mevalonic acid
NMR     Nuclear magnetic resonance
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OE</td>
<td>Over expression</td>
</tr>
<tr>
<td>OSC</td>
<td>Oxidosqualene cyclase</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>RMA1</td>
<td>Ring-finger protein with membrane anchor 1</td>
</tr>
<tr>
<td>RMBβCD</td>
<td>Random methyl β-cyclodextrin</td>
</tr>
<tr>
<td>SB</td>
<td>Synthetic biology</td>
</tr>
<tr>
<td>SD</td>
<td>Synthetic defined medium</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning electron microscopy</td>
</tr>
<tr>
<td>SEP</td>
<td>Synthetic enzymatic pathway</td>
</tr>
<tr>
<td>SQE</td>
<td>Squalene epoxidase</td>
</tr>
<tr>
<td>SQS</td>
<td>Squalene synthase</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission electron microscopy</td>
</tr>
<tr>
<td>TIA</td>
<td>Terpenoid indole alkaloid</td>
</tr>
<tr>
<td>TS</td>
<td>Taxa-4(5),11(12)-diene synthase</td>
</tr>
<tr>
<td>TTCED</td>
<td>Triterpene cyclase engineering database</td>
</tr>
<tr>
<td>UGT</td>
<td>UDP-dependent glycosyltransferase</td>
</tr>
<tr>
<td>UGT73F3</td>
<td>UDP-dependent glycosyltransferase</td>
</tr>
<tr>
<td>UGT73K1</td>
<td>UDP-dependent glycosyltransferase</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
Plants are a rich source of natural products with diverse structures and biological activities beneficial to humans. This has led to their application as flavors and fragrances, colorants and therapeutics. Terpenoids or isoprenoids constitute the largest class of plant natural products with over 25,000 characterized structures. This dissertation focuses on the building blocks of triterpenoid saponins that structurally consist of 30 carbon atoms arranged in four or five ring structures with several attached oxygens and sugar moieties. The triterpenoids have been valued in traditional herbal medicine for their healing power and general health promoting activities. Despite their potential therapeutic applications, triterpenoids have not been widely employed in modern medicine, owing to the usually very low amounts in which they are produced in their natural source, the plants themselves. Engineering plants and cell cultures for the enhanced production of triterpenoids has been limited. Therefore, first and foremost a heterologous microbial Saccharomyces cerevisiae based yeast platform was established for the production of high amounts of the triterpene saponin building blocks, resulting in 36.2 mg/L β-amyrin and 46.3 mg/L lupeol producing yeast strains. Besides the establishment of a production host, the yeast strains also facilitated the identification and characterization of novel genes encoding enzymes catalyzing saponin biosynthesis from the medicinal plants Bupleurum falcatum and Maesa lanceolata, and the establishment of a combinatorial biosynthetic platform in yeast to generate novel triterpene sapogenins. Furthermore, to increase the productivity and simplify the purification pipeline of triterpene sapogenins, a methyl β-cyclodextrin based culturing approach was established to promote their release from yeast cells into the growth medium. Finally, by down regulating the expression of a Medicago truncatula saponin regulatory gene that targets specific isoforms of 3-hydroxy-3-methylglutaryl-CoA reductase, the rate controlling enzyme of the mevalonate pathway, monoglycosylated saponin building blocks could be accumulated in transgenic hairy roots, thereby serving as a plant based production system. In conclusion, two in vivo production systems were established in the model legume M. truncatula and the
conventional yeast *S. cerevisiae*, for the synthesis of glycosylated and unglycosylated saponin building blocks, respectively.
Scope of Research
From time immemorial plant-derived natural products have played a significant role in curing diseases and promoting general human health. The production of these high-value, biologically active natural products is hindered by the unfeasibility of total chemical synthesis due to their structural complexity whereas, the plants themselves most often cannot provide a sustainable supply. An alternative route of obtaining these compounds is by engineering plant cultures or heterologous microbial hosts like *Escherichia coli* and *Saccharomyces cerevisiae*. The growing knowledge of plant secondary metabolism, together with the advances in engineering tools, facilitates translating plant-based biosynthetic routes into such microorganisms for the production of natural products.

The main objective of this PhD was to address the lack of a robust microbial production system for triterpenoid saponins and hence, to develop an *S. cerevisiae* based production platform for saponin building blocks (or sapogenins). Such a yeast based system could serve several purposes, which include (a) the characterization of novel genes involved in triterpene saponin biosynthesis, (b) the engineering of a production host for the targeted synthesis of a desired sapogenin and (c) the generation of new-to-nature sapogenins using combinatorial biosynthesis to widen the spectrum of lead molecules with biological activities for pharmaceutical applications.

Saponins and its building blocks have been primarily valued as therapeutics owing to their adjuvant, anti-angiogenic, anti-inflammatory, anti-cancer, antiviral, antibacterial and general health promoting activities in both traditional and modern medicine. The amphipathic nature of saponins also render them biologically active as surfactants that can enhance the penetration of macromolecules, like proteins, through cell membranes. Additionally, the surface active properties of saponins has been exploited in beverages, soaps and shampoos to produce stable foams. Saponins have also found applications in animal feed as anti-parasitics and fecal odor reductants. Next to these applications, the plants themselves produce saponins as defense compounds to overcome pathogen and pest attacks, thereby also highlighting their pesticidal activities. The multiple applications of saponins and its building blocks render them industrially valuable molecules that can
be harvested only in limited amounts from the plants. Therefore, the development of a yeast based production platform allows creating a source for the sustainable supply of saponin and saponin building blocks in the future.
Bioengineering of plant terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*
Publication status:

Tessa Moses, Jacob Pollier, Johan M. Thevelein and Alain Goossens.

Bioengineering of plant terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro.

Manuscript submitted for publication

Author contribution:

Writing of the manuscript.
Abstract

Terpenoids, also called isoprenoids, constitute a large and diverse class of natural products that serve many functions in nature. More than half of the over 40,000 discovered terpenoid structures are synthesized by plants, where they can function as primary metabolites involved in growth and development, or as secondary metabolites involved in the interaction of the plant with its environment. Several plant terpenoids are economically important molecules that are used as pharmaceuticals, flavors and fragrances, colorants, pesticides and elastomers. Two of the major challenges encountered for the commercialization of plant-derived (terpenoid) compounds is their low production or accumulation in planta and the continuous demand of the (pharmaceutical) industry for novel molecules with new or superior biological activities. In this review, we highlight several methods to enhance and diversify the production of plant terpenoids, with a foresight into triterpenoid engineering. Higher yields or a cheaper production of valuable terpenoids may be obtained by metabolic engineering of plants or by heterologous production of the compounds in other plants or microbes. Alternatively, novel compounds can be generated by applying combinatorial biosynthesis or directed enzyme evolution approaches. Ultimately, synthetic biology may lead to the production of high amounts of plant terpenoids in in vitro systems or custom-designed artificial biological systems.
Introduction

Plants synthesize and accumulate a wide range of small molecules or natural products that are involved in fundamental physiological and ecological processes. Some of these natural products have therapeutic potential that has been exploited by humans for thousands of years in the form of traditional herbal medicine. In recent years, with our growing understanding of their biosynthesis, regulation and functioning, plant-derived natural products have emerged as high-value therapeutics, flavors and fragrances, colorants and health-promoting agents. Based on their structure and biosynthetic origin, plant natural products can be classified in different groups, such as the terpenoids, alkaloids and phenolic compounds (Croteau et al., 2000). This review focuses on the terpenoids, of which over 25,000 compounds have been characterized from plants, making them the largest class of plant natural products. The terpenoids or isoprenoids comprise structurally diverse compounds that are associated with primary as well as secondary metabolism. Gibberellin, abscisic acid and brassinosteroid phytohormones, phytosterols and carotenoid pigments are primary metabolic terpenoids involved in basic functions such as regulation of plant growth and development, photosynthesis, membrane permeability and fluidity (Bohlmann & Keeling, 2008; Vranová et al., 2012). However, the majority of the plant terpenoids are secondary metabolites that play a crucial role in the interaction of the plant with its environment, for instance by serving as pollinator attractants, herbivore repellents, anti-feedants, toxins or antibiotics (Gershenzon & Dudareva, 2007).

The structural variety and inherent biological activities of many plant terpenoids have rendered them widely applicable. With an annual production of $10^7$ tons, natural rubber is undoubtedly the most abundant terpenoid produced, which serves as biological material in the non-food industry. Its unique properties in terms of resilience, abrasion and impact resistance, make natural rubber the preferred material for the production of heavy-duty tires and vibration dampers. Furthermore, latex products for medical applications, such as surgical gloves, are often based on natural rubber (van Beilen & Poirier, 2007). Other examples of plant terpenoids with significant economical value include menthol, a monoterpenoid extracted from peppermint and used in the flavor and fragrance industry;
abietic acid, a diterpenoid isolated from conifer rosin that is used in lacquers, varnishes and soap; and the anti-malarial and anti-cancer drugs artemisinin and taxol, respectively (Bohlmann & Keeling, 2008).

A major hurdle in the commercialization of plant terpenoids is that they often accumulate in very low concentrations in planta, thereby hindering their purification in large amounts from the natural source. For instance, the diterpenoid taxol is harvested from the inner bark of the yew tree where it corresponds to only ~0.001-0.05% of the dry weight. Taxol can also be produced semi-synthetically from taxol precursors, baccatin III and 10-deacetylbaccatin III, which can reach concentrations of ~0.3% of the dry weight in certain Taxus species. As such, the production cost of taxol is high and the supply from natural sources cannot fulfill the growing demand for the compound (Expósito et al., 2009).

When extraction of a natural product from its natural source is not sufficient, several alternative approaches can be explored, including (a) plant breeding and genetic engineering to generate cultivars or transgenics accumulating higher levels of the desired compounds, (b) developing scalable plant cell or root cultures, and (c) engineering microbial hosts to produce the compound (Kirby & Keasling, 2009). Commercially viable production systems have already been established for some terpenoids, which is reflected in the emergence of companies like Phyton Biotech (http://www.phytonbiotech.com/), a global provider of chemotherapeutics including paclitaxel extracted from Taxus cell cultures and Amyris (http://www.amyris.com/) that uses a synthetic biology platform for the production of artemisinin in yeast.

The pharmaceutical industry is in constant search for novel molecules, primarily due to the discovery of new drug targets, the emergence of new diseases and, in the case of infectious diseases, the growing resistance of microbes against the currently marketed antibiotics (Pollier et al., 2011; Winter & Tang, 2012). Additionally, the business model of pharmaceutical companies is under threat, as leading blockbuster drugs will soon lose patent protection and become available for market competition, which often leads to lower market prices, thereby rendering the production of the drug non-profitable to the
original developer. As traditional pharmacological screening of medicinal plants is time-consuming and expensive, and the output of combinatorial chemistry libraries is low in terms of new drugs, alternative approaches to generate new molecules or scaffolds are required (Koehn & Carter, 2005; Welsch et al., 2010; Carter, 2011). Combinatorial biosynthesis accelerates the process of natural evolution and multiplies the natural diversity by generating novel enzyme-substrate combinations that result in the production of novel compounds (Kirschning et al., 2007; Pollier et al., 2011). By integrating emerging technologies such as synthetic biology and directed enzyme evolution, combinatorial biosynthesis can be rationally applied to custom design new compounds.

In this review, we summarize the current status of terpenoid engineering in plants and microbial hosts, thereby highlighting the latest approaches available for enhancing compound production and increasing structural diversity. We discuss the booming trend of synthetic biology in the perspective of terpenoid production and give a futuristic view into the engineering of triterpenoids, the least engineered class of terpenoids with pharmaceutical potential.

**Classification and biosynthesis of plant terpenoids**

All terpenes are derived from the repetitive fusion of isoprene \((C_5H_8)\) units and have the molecular formula \((C_5H_8)_n\), where ‘n’ stands for the number of linked isoprene units. These isoprene units are linked most commonly in a ‘head-to-tail’ fashion, but products with ‘head-to-head’ and ‘head-to-middle’ fusions also occur (Fig. 1). The terpenes resulting from the fusion of isoprene units undergo various modifications, like carbon rearrangements or oxidations, to result in the formation of a plethora of different terpenoid structures.
Bioengineering plant terpenoids

The terpenoids are classified by the number of isoprene units they comprise (Croteau et al., 2000). The C₅H₈ ‘hemiterpenes’ are the smallest terpenoids consisting of a single isoprene unit. The volatile isoprene is a hemiterpene, and oxygenated derivatives, such as prenol, angelic acid and isovaleric acid, are hemiterpenoids found in plants. The C₁₀H₁₆ terpenes consist of two isoprene units and are called ‘monoterpenes’. They form the base unit from which terpene nomenclature is derived and consist of α-, mono- and bicyclic compounds. The monoterpenes comprise volatiles, such as like linalool, menthol, thymol and camphor, that are used as flavors and fragrances (Schwab et al., 2008). The C₁₅H₂₄ ‘sesquiterpenes’ are α-, mono-, bi- or tricyclic compounds derived from the fusion of three isoprene units. Sesquiterpenes can also acquire a lactone ring resulting in the formation of sesquiterpene lactones, such as artemisinin. The ‘diterpenes’ are C₂₀H₃₂ compounds derived from the fusion of four isoprene units. The plant hormone gibberellin and the anti-cancer agent taxol belong to this class of terpenoids (Croteau et al., 2000). Although rare in plants, the ‘sesterterpenes’ with the molecular formula C₂₅H₄₀ are derived from five isoprene units. To date, they have been identified in only two medicinal plant genera, Salvia (Moghaddam et al., 2010) and Leucosceptrum (Choudhary et al., 2004). The ‘triterpenes’ are composed of six isoprene units and are formed by the ‘head-to-head’ joining of two C₁₅ chains, in which the isoprene units are joined ‘head-to-tail’. The resulting C₃₀H₄₈ compounds constitute a large class of molecules, including phytosterols, brassinosteroids and saponins. The ‘tetraterpenes’ consist of the well known carotenoids with the molecular formula C₄₀H₆₄ and are composed of eight isoprene units.
They serve as light absorbing accessory pigments in plants and have been exploited as natural colorants and food additives for their beneficial effects on human health (Rao & Rao, 2007). The ‘polyterpenes’ contain more than eight isoprene units. The tail of the electron carrier ubiquinone or coenzyme Q$_{10}$ and rubber are polyterpenes. Finally, the ‘meroterpenes’ include members from the terpenoid indole alkaloids (TIAs) and cytokinins that have mixed origins, but are partially derived from terpenes (Croteau et al., 2000). In bacteria, a relatively new family of terpenes made up of seven isoprene units has been encountered, denominated sesquarterpenes (Sato et al., 2011), but such compounds have not been detected in plants yet.

Despite their enormous structural diversity, terpenoids share a common biosynthetic origin and follow similar synthesis routes. The biosynthesis begins with the generation of isopentenyl pyrophosphate (IPP), the principal precursor. This is followed by consecutive condensation of IPP units to form prenylated pyrophosphates, the immediate precursors of the different terpenoid classes. Specific terpenoid synthases (often cyclases) (Chen et al., 2011) modify these precursors to terpenoid skeletons, which are subsequently decorated by various enzymatic modifications to generate the structural and functional diversity of terpenoids.

In higher plants, the biosynthesis of the central intermediate IPP occurs through two distinct pathways, the mevalonate (MVA)/3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase pathway and the 2-C-methyl-D-erythritol 4-phosphate (MEP)/1-deoxy-D-xylulose 5-phosphate (DOXP)/non-mevalonate pathway (Rodríguez-Concepción & Boronat, 2002). A clear compartmentalization for the generation of IPP and the synthesis of terpenoids exists in plants (Lange et al., 2000). The IPP generated from the MVA pathway in the cytosol, peroxisome and endoplasmic reticulum (ER) is used for the biosynthesis of sesqui- and triterpenes (Sapir-Mir et al., 2008; Clastre et al., 2011), whereas the plastidic MEP pathway synthesizes IPP for the generation of hemi-, mono-, di- and tetraterpenes. Additionally, parts of the isoprenoid pathway have also been detected in the mitochondria (Rodríguez-Concepción & Boronat, 2002) (Fig. 2).
Figure 2. Terpenoid biosynthesis in plants. Two distinct pathways for the synthesis of the universal precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) exist in plants: the cytoplasm, peroxisome and endoplasmic reticulum (ER) localized mevalonate (MVA) pathway (purple) and the plastid localized methyl erythritol phosphate (MEP) pathway (blue). The prenyltransferases (orange) generate the immediate precursors to the different terpene classes (green). Dashed arrows indicate multiple reactions. Dotted grey boxes indicate subcellular localization of the pathway. Grey arrows indicate metabolites that are transported between subcellular compartments. AACT, acetoacetyl-CoA thiolase; CMK, 4-diphosphocytidyl-methylerythritol kinase; CMS, 4-diphosphocytidyl-methylerythritol synthase; DXR, deoxyxylulose 5-phosphate reductoisomerase; DXS, deoxyxylulose 5-phosphate synthase; FPPS, farnesyl pyrophosphate; FPPS, FPP synthase; GGPPS, geranylgeranyl pyrophosphate; GGPPS, GGPP synthase; GPP, geranyl pyrophosphate; GPPS, GPP synthase; HDR, hydroxymethylbutenyl 4-diphosphate reductase; HDS, hydroxymethylbutenyl 4-diphosphate synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; IDI, isopentenyl diphosphate isomerase; MDS, methylerythritol 2,4-cyclodiphosphate synthase; MVK, mevalonate kinase; PDC, pyruvate dehydrogenase complex; PMD, 5-diphosphatevalonate decarboxylase; PMK, 5-phosphatevalonate kinase; PSY, phytoene synthase; SQS, squalene synthase.
The MVA pathway begins with a two-step condensation of three acetyl-CoA molecules to generate HMG-CoA, which is reduced to MVA that is sequentially phosphorylated in an ATP-dependent manner, and finally decarboxylated to generate IPP (Fig. 2). In the plastidic MEP pathway, pyruvate and glyceraldehyde-3-phosphate generate the first C₅ intermediate DOXP, which is rearranged and reduced to form MEP that is subsequently converted to IPP (Fig. 2). The generated IPP is first isomerized to its allylic isomer dimethylallyl pyrophosphate (DMAPP) at the site of IPP synthesis and in the mitochondria. The allylic double bond in DMAPP allows its condensation with IPP to yield the C₁₀ intermediate geranyl pyrophosphate (GPP), the precursor for monoterpenoids. Subsequent addition of one or two IPP units yields C₁₅ farnesyl pyrophosphate (FPP), the precursor for sesquiterpenoids, and C₂₀ geranylgeranyl pyrophosphate (GGPP), the diterpenoid precursor, respectively. FPP is the only terpene synthesized in the mitochondria and is the precursor for the synthesis of the ubiquinone tail (Rodríguez-Concepción & Boronat, 2002). The tri- and tetraterpenoid precursors are generated by the fusion of two FPP and GGPP units, respectively. The condensation reactions are catalyzed by specific prenyltransferases and are named according to the product they generate. The synthesis of terpenoid skeletons from the prenylated precursors and their subsequent modifications are vast and highly specific and are reviewed elsewhere (Croteau et al., 2000; Degenhardt et al., 2009; Augustin et al., 2011; Hamberger et al., 2011).

**Regulation of terpenoid biosynthesis in plants**

The biosynthesis of terpenoids is tightly controlled in plants, since they serve many functions in plant growth, development and response to biotic and abiotic environmental factors (Tholl, 2006; Nagegowda, 2010; Vranová et al., 2012). Terpenoid synthesis occurs within specific tissues or at specific plant developmental stages (Nagegowda, 2010). For instance, many plant species have glandular trichomes, specialized structures for the synthesis of secreted terpenoid natural products (Markus Lange & Turner, 2012). *Artemisia annua* plants synthesize and store artemisinin in specialized glandular
Bioengineering plant terpenoids

trichomes on the leaves, stems and inflorescences to prevent autotoxic effects of the compound (Duke et al., 1994; Maes et al., 2011). In Arabidopsis thaliana specific floral tissues emit volatile terpenoids at particular times of the flowering period, presumably to attract pollinators and to defend the flowers against microbial pathogens and herbivores (Tholl et al., 2005; Tholl, 2006). Such specific terpenoid synthesis is mainly regulated at the transcriptional level (Nagegowda, 2010). For example, the two terpenoid synthase genes responsible for the production of the complex mixture of volatile sesquiterpenoids in A. thaliana flowers are expressed exclusively in specific flower tissues (Tholl et al., 2005), and all specialized genes responsible for artemisinin biosynthesis in A. annua are highly expressed in the glandular trichomes (Maes et al., 2011).

Besides this spatiotemporal regulation, induced terpenoid biosynthesis is often observed in response to herbivore feeding, pathogen attack or various abiotic stresses (Nagegowda, 2010; Vranová et al., 2012). For instance, when plants are damaged by herbivorous insects, volatile terpenoids that attract carnivorous enemies of the herbivores are produced de novo, thereby providing indirect defense to the plants (Paré & Tumlinson, 1999; Pichersky & Gershenzon, 2002; Degenhardt et al., 2003). Likewise, when white pine weevils attack Sitka spruce (Picea sitchensis), the plants respond by forming traumatic resin ducts and by increasing the expression of terpene synthase genes, leading to the accumulation of oleoresin terpenoids that give direct protection against the weevils (McKay et al., 2003; Miller et al., 2005). The increased accumulation or release of terpenoids in response to various (a)biotic stresses is often mediated by an increased transcriptional activity of the specific terpenoid biosynthetic genes (Tholl, 2006; Nagegowda, 2010). This transcriptional response is controlled by a complex signaling cascade in which jasmonate hormones (JAs) play a crucial role. Hence, treatment of plants with JAs often causes transcriptional and metabolic changes comparable to pathogen or herbivore attack. Similar to the white pine weevil attack, the expression of terpenoid synthase genes strongly increases and traumatic resin ducts accumulate when Sitka spruce plants are treated with methyl jasmonate (MeJA) (Miller et al., 2005).

The concerted transcriptional activation of entire secondary metabolic pathways by JAs is conserved across the plant kingdom. However, downstream of the conserved JA
perception and initial signaling cascade, species-specific transcriptional machineries that regulate the transcriptional activity of the specific biosynthetic genes exist (Pauwels et al., 2009; Pauwels & Goossens, 2011; De Geyter et al., 2012). Several transcription factors regulated by the JA signaling cascade that activate the transcription of terpenoid biosynthetic genes have already been characterized (De Geyter et al., 2012). The JA-inducible cotton (Gossypium arboreum) WRKY1 transcription factor activates the promoter of the CAD1-A gene, which encodes the sesquiterpene (+)-δ-cadinene synthase, that catalyzes the first committed step of gossypol biosynthesis (Xu et al., 2004). Similarly, the AP2/ERF transcription factors AaERF1 and AaERF2 are JA-responsive and induce the transcription of genes encoding amorpha-4,11-diene synthase (ADS) and CYP71AV1, 2 key enzymes of the artemisinin biosynthetic pathway (Yu et al., 2012).

Jasmonates are important, but definitely not the only regulators of terpenoid metabolism in plants. For instance, when white pine weevils attack Sitka spruce, there is a higher accumulation of traumatic resin ducts and a more complex terpenoid emission profile, as compared to MeJA-treated plants (Miller et al., 2005). Indeed, the plant’s defense is often regulated by a complex cross-talk between various signaling cascades in which next to JAs, abscisic acid, salicylic acid and ethylene play important roles (De Vos et al., 2005). As such, together with specific attacker-induced signals, plants can generate defense responses that are adapted to the different attackers they have to face (De Geyter et al., 2012).

Many terpenoids also play an essential role in plant growth and development as phytohormones, accessory pigments and building blocks of cellular components. Hence, also their biosynthesis is strictly regulated. For instance, the high expression of gibberellin biosynthetic genes in plant tissues is correlated with a strong accumulation of active gibberellins, and several transcription factors involved in the developmental regulation of gibberellin biosynthesis have been characterized (Hedden & Thomas, 2012).

Besides the transcriptional, developmental and spatiotemporal modulation of terpenoid biosynthetic genes, post-translational regulatory mechanisms also exist in terpenoid
Bioengineering plant terpenoids

biosynthesis. For instance, in addition to the transcriptional control of a key gibberellin biosynthetic enzyme, the gibberellin levels in developing *A. thaliana* embryos are maintained at low levels by the oxidative inactivation of the hormone by the enzyme gibberellin 2-oxidase, the transcription of which is stimulated by the MADS domain transcription factor AGL15 (Wang *et al.*, 2004). The activity of HMGR, the enzyme that catalyzes the key regulatory step of the MVA pathway, is controlled at the protein level via dephosphorylation of the enzyme through the action of protein phosphatase 2A (Leivar *et al.*, 2011).

Bioengineering of terpenoids in planta

Because of their strict regulation, most terpenoids are produced in very low amounts in their natural sources. The low yield makes extraction expensive, which is eventually reflected in their market value. Taxol, for instance, costs $300,000/kg (Expósito *et al.*, 2009). Consequently, there is a wide gap between demand and supply of terpenoids, which hampers their widespread application. The classical approach to ensure a constant or improved yield is by selecting and propagating high-producing cultivars or by producing (transgenic) plant (cell) cultures (Georgiev *et al.*, 2009; Lim & Bowles, 2012; Wilson & Roberts, 2012). Additionally, the use of elicitors (Zhao *et al.*, 2005; Khosroushahi *et al.*, 2006; Lambert *et al.*, 2011), optimization of culturing conditions (Cusidó *et al.*, 2002) and addition of precursors (Lee-Parsons & Royce, 2006) has been implemented to increase productivity. Our growing understanding of terpenoid biosynthesis, regulation and accumulation, together with the development of functional genomics and systems biology toolkits, has enabled metabolic engineering of whole plants and plant cell cultures to enhance productivity and also alter terpenoid distribution *in planta* (Aharoni *et al.*, 2006; Roberts, 2007; Dudareva & Pichersky, 2008).

Since terpenoid biosynthesis is strictly regulated and often controlled by specific transcription factors, one way of increasing productivity is by modulating the expression of such or other regulatory factors (Broun, 2004; De Geyter *et al.*, 2012). For example,
overexpression of the transcription factors AaERF1 and AaERF2 leads to an increased accumulation of artemisinin in transgenic A. annua plants (Yu et al., 2012). However, in spite of the identification of transcription factors that steer the biosynthesis of terpenoids, overexpression of a single transcription factor often does not lead to a higher production of the compounds. For instance, overexpression of ORCA3, an AP2/ERF transcription factor that controls the expression of several TIA biosynthetic genes, is not sufficient to induce TIA production in C. roseus cell cultures, indicating that only a part of TIA biosynthesis is under the control of this transcription factor (van der Fits & Memelink, 2000). Hence, further elucidation of the complex signaling cascades that lead to an increased accumulation of terpenoids is mandatory for large-scale metabolic engineering of terpenoid production by using transcription factors.

A second way to increase productivity is by specifically overexpressing the rate-limiting enzymes in the pathway. Overexpression of genes encoding such enzymes, like HMGR, deoxyxylulose 5-phosphate synthase (DXS) and prenyltransferases, has been used to elevate terpenoid levels in plant tissue cultures (Degenhardt et al., 2003). For instance, expression of a C. roseus HMGR and the native ADS gene in A. annua plants enhanced artemisinin production by over 7.5-fold to 1.73 mg/g of dry weight (Alam & Abdin, 2011). Enhanced terpenoid production has also been observed upon alteration of the subcellular localization of enzymes, presumably resulting from the uncoupling of biosynthesis and regulation (Bouwmeester, 2006; Kumar et al., 2012). For instance, expression of an avian farnesyl pyrophosphate synthase (FPPS) gene and A. annua ADS in plastids of tobacco (Nicotiana tabacum) resulted in a 1000-fold increase in amorpha-4,11-diene production than when the same enzymes were targeted to the cytosol (Wu et al., 2006). Similarly, eight times more artemisinin accumulated in transgenic N. tabacum plants expressing truncated yeast HMGR and A. annua ADS, CYP71AV1, cytochrome P450 monooxygenase (CytP450) reductase (CPR) and artemisinic aldehyde reductase (DBR2) when the A. annua ADS was targeted to the mitochondria rather than to the cytosol (Farhi et al., 2011).

Besides enhancing terpenoid production, in planta engineering has also been used as a tool to modulate the terpenoid composition of plants. By employing heterologous gene
expression and native gene knock down strategies, staple crops like rice, maize, potato, carrot, tomato and canola have been engineered for β-carotene accumulation to increase their nutritional value (Farré et al., 2011). Similarly, potato tubers have been engineered to accumulate high levels of the sesquiterpene flavor volatile α-copaene (Morris et al., 2011). The different strategies and tools available for in planta engineering and production of biomolecules with respect to their cost-effectiveness have been summarized elsewhere (Wilson & Roberts, 2012).

Contrary to sesqui-, tetra- and meroterpenoids, in planta triterpenoid (particularly saponin) engineering has been hampered by the lack of knowledge about regulatory mechanisms controlling gene expression and the tissue- or organ-specific nature of its accumulation (Sawai & Saito, 2011). Hence, a challenge for future saponin research will be to identify the transcription or other regulatory factors that steer their biosynthesis. Only a single study reports on an attempt to engineering triterpene synthesis in tobacco by the heterologous expression of an avian FPPS and a yeast squalene synthase (SQS) gene targeted to the cytoplasm or plastid. No differences in squalene accumulation caused by specific targeting of the enzymes have been observed but comparison of transgenic tobacco plants expressing the same genes driven by a constitutive viral promoter or a trichome-specific promoter revealed higher squalene accumulation, accompanied by negative effects on plant growth and physiology, in case the enzymes were directed to the trichomes (Wu et al., 2012). Nonetheless, this study underscored the potential of engineering triterpenoids in planta by relocation of the biosynthetic pathway and enhancement of the precursor flux, and encourages future research on this terpenoid class.

Heterologous biosynthesis and engineering of plant terpenoids in microbial hosts

Compared to plant production systems, microorganisms are attractive alternatives as heterologous hosts because of their rapid doubling time, robustness under process
conditions, ease of scalability, simplicity of product purification due to the absence of competing contaminants and cost-effectiveness resulting from the conversion of inexpensive feedstock to valuable compounds (Zhang et al., 2011). The choice of a suitable host (or “chassis”) is critical and should be based on multiple factors, including the chemical nature and complexity of the product to be synthesized, the genetic amenability of the host, the intrinsic availability of precursors for product biosynthesis, the codon usage bias of the host, the need for post-translational modifications and the feasibility to metabolically engineer the host to boost productivity (Keasling, 2010). Microbial synthesis of any plant natural product can be achieved by ‘precursor-mediated product synthesis’, where an existing host pathway is altered to incorporate a heterologous pathway, or by ‘de novo synthesis’, where new-to-host biosynthetic routes are imported, thereby avoiding feedback regulation (Chang & Keasling, 2006). After establishing heterologous synthesis, it is usually imperative to metabolically engineer the host to optimize production yield and rate (Chemler & Koffas, 2008).

The colloquial hosts *Escherichia coli* and *Saccharomyces cerevisiae* have been employed for both precursor-mediated and de novo synthesis of mono-, di-, sesqui-, tri- and tetraterpenes (Misawa, 2011). The prokaryotic *E. coli* has an inherent MEP pathway and the eukaryotic *S. cerevisiae* has the MVA pathway to produce IPP and its isomer DMAPP. Theoretically, terpenoid biosynthesis can be incorporated into these hosts by expressing the corresponding genes, but low yields may be obtained because of the limited intracellular IPP pool (Carter et al., 2003). The rate-limiting nature of the reactions catalyzed by *dxs* and *dxr* in *E. coli* (Rodríguez-Concepción & Boronat, 2002) and *HMGR* in *S. cerevisiae* (Gardner & Hampton, 1999) controls the IPP production. The IPP and subsequent precursor levels have been supplemented by metabolic engineering of (1) the MVA pathway in *E. coli* (Campos et al., 2001), (2) the MEP pathway and prenyltransferases in *E. coli* (Kajiwara et al., 1997), (3) the MVA pathway by a feedback-regulation deficient *HMGR* in *S. cerevisiae* (Ro et al., 2006), (4) the MVA pathway by decreasing downstream enzymes to accumulate precursors in *S. cerevisiae* (Paradise et al., 2008), (5) the global transcription factor regulating sterol biosynthesis in *S. cerevisiae*
(Davies et al., 2005) and (6) protein scaffolds for the MVA pathway in *S. cerevisiae* (Dueber et al., 2009) (Fig. 3).

**Figure 3.** Strategies employed to enhance the production of IPP and terpenoids thereof, in *E. coli* and *S. cerevisiae*. (a) Expression of the *S. cerevisiae* MVA pathway in *E. coli*. (b) Expression of rate-limiting MEP enzymes in *E. coli*. (c) Expression of a truncated copy of HMGR in *S. cerevisiae*. (d) Downregulation of endogenous sterol biosynthesis to accumulate terpenoid precursors in *S. cerevisiae*. (e) Expression of a mutant version (*upc2-1*) of the global transcription factor (*UPC2*) regulating sterol biosynthesis in *S. cerevisiae* upregulates expression of native sterol biosynthesis. (f) Protein scaffolding to prevent rate-limitation in *S. cerevisiae* by spatially organizing rate-limiting sterol biosynthetic enzymes in a modulated ratio.
Alongside targeted engineering, global approaches have been applied to improve the terpenoid pathway flux in microbial hosts. A ‘chromosomal promoter engineering’ strategy was used to express some of the endogenous MEP genes from a strong bacteriophage T5 promoter in an *E. coli* strain harboring \( \beta \)-carotene biosynthetic genes, resulting in the enhanced production of \( \beta \)-carotene compared to the parental strain (Yuan *et al.*, 2006). Similarly, a ‘global transcription machinery engineering’ (gTME) on the *rpoD* gene encoding \( \sigma^{70} \), the primary sigma factor, resulted in an increased lycopene production in *E. coli* (Alper & Stephanopoulos, 2007). In this case, a single gTME perturbation outperformed the traditional engineering strategies. Furthermore, the order of gene assembly in a synthetic zeaxanthin operon in *E. coli* influenced the expression of each gene based on its distance from the promoter (Nishizaki *et al.*, 2007).

Once precursor synthesis has been optimized, another major hurdle to overcome is to achieve functional expression of the pathway genes downstream of the precursor, particularly CytP450s. Plant CytP450s are ER membrane localized enzymes with a prerequisite for a CPR partner for efficient functioning (Podust & Sherman, 2012). The CPRs are also ER membrane localized electron donor proteins that mediate electron transfer from a cofactor, like NADPH, to the CytP450. In this regard, *S. cerevisiae*, with its native CytP450s and CPR, has an advantage over *E. coli* for the expression of complex terpenoid pathways (Hamann & Möller, 2007). Nevertheless, plant CytP450s supplemented with a plant CPR have been successfully expressed in both *E. coli* and *S. cerevisiae* (Arsenault *et al.*, 2008).

Two pioneering examples of microbial metabolic engineering for the production of terpenoids are artemisinin (Fig. 4a) and taxol (Fig. 4b). Artemisinin-based combination therapy (ACT) has been recommended by the World Health Organisation (WHO) as the first line therapy for the treatment of malaria. Artemisinin (Fig. 5a) kills the majority of the parasites at the beginning of infection and the slowly released partner drugs kill the remaining parasites, thus preventing the emergence of artemisinin-resistant *Plasmodium falciparum*. The global procurement of ACT increased by 14.5% in 2010 as compared to 2009 (WHO, 2011), reflecting the constant demand for higher amounts of the drug, which is still extracted from a plant source and hence is prone to fluctuations in yearly
harvest and market price. Microbial production is considered as a valid alternative to stabilize the price of artemisinin and to meet its clinical demand. Furthermore, semi-synthesis of microbially produced intermediates is possible. In precursor engineered *S. cerevisiae*, production of 153 mg/l amorpha-4,11-diene (Ro *et al.*, 2006) and 2.5 g/l artemisinic acid under optimized fermentation conditions has been achieved (Lenihan *et al.*, 2008). The amount of artemisinic acid produced by engineered *S. cerevisiae* in 4-5 day fermentations is comparable to that obtained from several month-old *A. annua* plants. Additionally, *S. cerevisiae* secretes artemisinic acid into the growth medium, facilitating purification and economizing the production pipeline (Zeng *et al.*, 2008). In an isoprenoid-optimized *E. coli* strain, under carbon- and nitrogen-depleted fermentative conditions, 27.4 g/l of amorpha-4,11-diene has been produced (Tsuruta *et al.*, 2009). The highest titer of amorpha-4,11-diene observed thus far, 41 g/l, was produced in *S. cerevisiae* CEN.PK2 by upregulating the MVA pathway and downregulating ergosterol synthesis (Westfall *et al.*, 2012). Hence, these two microbial-based cell factories are promising alternatives to meet the market demand for artemisinin.

![Chemical structures](image)

**Figure 4** Chemical structures of (a) artemisinin, (b) taxol and (c) bisabolane.

Production of the diterpenoid taxol (paclitaxel) has also been extensively engineered in *E. coli* and *S. cerevisiae*. Taking advantage of the characterization of a few biosynthetic genes (Fig. 5b) (Heinig & Jennewein, 2009), taxol production employing metabolically engineered *E. coli* (Liu & Khosla, 2010) and *S. cerevisiae* (Engels *et al.*, 2008) was attempted. In the first endeavor, eight biosynthetic genes were expressed in *S. cerevisiae*,
resulting in the production of 1 mg/l taxadiene and 25 µg/l taxadien-5α-ol. The low yields and failure to detect taxadien-5α-yl acetate or taxadien-5α,10β-diol monoacetate indicated a pathway restriction (Dejong et al., 2006). Expression of a codon-optimized taxadiene synthase (TS) in a precursor-optimized S. cerevisiae, resulted in the accumulation of 8.7 mg/l taxadiene and 33.1 mg/l geranylgeraniol, suggesting room for further improvement (Engels et al., 2008). Simultaneously, E. coli was engineered using a ‘multivariate-modular pathway engineering’ approach resulting in the production of 300 mg/l taxadiene (Ajikumar et al., 2010). Here, the taxol biosynthetic pathway was first subdivided into modules and the expression of genes in each module was varied simultaneously to obtain an optimal expression balance or maximal taxadiene output. In a controlled fed-batch fermentation, this strain produced 1 g/l taxadiene, leaving room for further improvement by optimizing fermentation conditions. Additionally, overexpression of a codon-optimized translational fusion chimeric enzyme of taxadiene 5α-hydroxylase and the CPR partner from T. cuspidata in this E. coli strain resulted in the conversion of 98% of the taxadiene to 58 mg/l taxadien-5α-ol and the byproduct 5(12)-oxa-3(11)-cyclotaxane. However, taxadiene production in this strain was much lower than that of the parental strain, emphasizing the need for further engineering to boost productivity (Ajikumar et al., 2010).

A more recent example of microbial biosynthesis of plant terpenoids is bisabolane (Fig. 4c), an alternative to D2 diesel fuel. Through a combination of enzyme screening and metabolic engineering of FPP-overproducing E. coli and S. cerevisiae, strains producing over 900 mg/l of the sesquiterpene bisabolene were engineered. The biosynthesized precursor bisabolene is then chemically hydrogenated to the advanced biofuel bisabolane (Peralta-Yahya et al., 2011).

S. cerevisiae has also been employed for the expression of triterpenoid saponin biosynthetic genes. Although triterpenoids constitute a diverse class of compounds, their biosynthesis is generally catalyzed by three sets of enzymes, the oxidosqualene cyclases (OSCs), CytP450s and UDP-dependent glycosyltransferases (UGTs). Through its native ergosterol biosynthesis, S. cerevisiae produces oxidosqualene, the precursor of saponins. In engineered strains optimized to accumulate oxidosqualene, different OSCs and
CytP450s have been expressed, mainly for their functional characterization (Augustin et al., 2011) (Fig. 6).

**Figure 5.** Biosynthetic pathway for, (a) Artemisinin and (b) Taxol. Dashed arrows indicate multiple steps. Highlighted enzymes (red) and compounds (blue) have been expressed and detected, respectively in *E. coli* and/or *S. cerevisiae*. ADS, amorpha-4,11-diene synthase; *ALDH1*, aldehyde dehydrogenase 1; *CPR*, cytochrome P450 reductase; *CYP71AV1*, amorpha-4,11-diene monooxygenase; *BAPT*, C-13-phenylpropanoyl-CoA transferase; *DBAT*, 10-deacetylbaccatin III-10-O-acetyltransferase; *DBR2*, artemisicinic aldehyde Δ11(13) double-bond reductase; *DBTNBT*, 3'-N-debenzoyl-2'-deoxytaxol N-benzoyl transferase; *DMAPP*, dimethylallyl pyrophosphate; *FPP*, farnesyl pyrophosphate; *FPPS*, FPP synthase; *GGPP*, geranylgeranyl pyrophosphate; *GGPPS*, GGPP synthase; *IPP*, isopentenyl pyrophosphate; *IDI*, isopentenyl isomerase; *TAT*, taxadiene-5α-O-acetyltransferase; *TBT*, taxane-2α-O-benzoyltransferase; *THY5α*, taxadiene-5α-hydroxylase; *THY10β*, taxane-10β-hydroxylase; *THY13α*, taxane-13α-hydroxylase; *TS*, taxadiene synthase.
However, engineering efforts have been limited to the production of β-amyrin only. Through a conventional pathway engineering approach, a 50% improvement in the production of β-amyrin with a final titer of 6 mg/l was demonstrated in a S. cerevisiae strain expressing a β-amyrin synthase (bAS) from A. annua (Kirby et al., 2008). Subsequent to a genotype-to-phenotype linking study, a 500% improvement in β-amyrin production was achieved by overexpressing three native genes in a S. cerevisiae strain expressing a Pisum sativum bAS, resulting in a final titer of 3.93 mg/l (Madsen et al., 2011). The β-amyrin levels produced by the parent strains in the above reports reflect the cyclization efficiency of the enzymes employed. Therefore, by employing a more efficient bAS (or any other saponin biosynthetic gene), followed by targeted and/or global engineering, it should be possible to further enhance β-amyrin (or triterpenoid) levels.

**Combinatorial biosynthesis of terpenoids**

Plants are a rich source of structurally diverse biomolecules with therapeutic potential. Many of these therapeutics, particularly those used in the treatment of infectious diseases, are susceptible to development of acquired resistance. For instance, resistance against artemisinin easily develops when used as the sole treatment (WHO, 2011). Similarly, many anti-bacterials are rapidly losing their therapeutic value because of the emergence of single and multi-drug resistant bacterial strains (Levy & Marshall, 2004). Likewise, many cancers are prone to acquire resistance to commonly used anti-cancer drugs (Gottesman, 2002). Additionally, there are still many clinical targets for which effective drugs have not been developed, such as influenza, polio and many types of cancers. Hence, despite the chemical diversity of biologically active compounds from plants and other organisms, there is a great demand for novel molecules. One way of fulfilling this demand is by generating ‘natural product analogs’, in which the majority of the biological functionality of the starting molecule is retained.
Figure 6. A simplified scheme of triterpenoid saponin biosynthesis as expressed in *S. cerevisiae*. Dashed arrows indicate multiple steps. Highlighted enzymes (red) and compounds (blue) were expressed and detected, respectively. αAS, α-amyrin synthase; As, *Avena strigosa*; bAS, β-amyrin synthase; CAS, cycloartenol synthase; Cr, *Catharanthus roseus*; DDS, dammarenediol synthase; Gm, *Glycine max*; Gu, *Glycyrrhiza uralensis*; LAS, lanosterol synthase; LUP, lupeol synthase; Mt, *Medicago truncatula*; Pg, *Panax ginseng*; SHC, squalene-hopane cyclase; SQE, squalene epoxidase; Vv, *Vitis vinifera*. 
Natural product scaffolds exhibit a broad diversity in chemical space (Grabowski et al., 2008). Hence, the generation of ‘lead-like’ natural products (or analogs), inspired by existing privileged scaffolds in the bioactive chemical space, is encouraging because of the predictive knowledge about their absorption, distribution, metabolism, excretion and toxicity (ADMET) properties (Norinder & Bergström, 2006; Hong, 2011).

Combinatorial biosynthesis-based reconstitution of pathways is a useful tool to generate known and novel natural products, which can be further modified by semi-synthesis. In its simplest form, combinatorial biosynthesis is the process of generating different but structurally related molecules, in a short span of time, through the assembly of genes from different organisms in a single host (Kirschning et al., 2007) (Fig. 7a). Plants possess an immense potential for combinatorial biosynthesis (Pollier et al., 2011). For instance, expression of a bacterial halogenase in C. roseus resulted in the generation of novel chlorinated TIA's in the plant (Runguphan et al., 2010). However, to date there have been no reports on the combinatorial biosynthesis of any other terpenoid (or any secondary metabolite) in plants. Nonetheless, the existing chemical diversity together with our growing understanding of their biosynthesis, renders terpenoids appealing compounds for the combinatorial generation of novel analogs. For instance, the screening of a synthetic triterpenoid combinatorial library derived from betulinic and ursolic acid, led to the identification of compounds with an enhanced anti-malarial activity compared to the parent compounds (Pathak et al., 2002).

Combinatorial biosynthesis of triterpenoid saponins holds great potential, as they exhibit a plethora of biological activities. For instance, the soluble Quillaja saponin fraction QS-21 is a potent vaccine adjuvant (Sun et al., 2009) and the semi-synthetic derivative of oleanolic acid, bardoxolone methyl, has been subject of clinical trials for the treatment of chronic kidney disease. The synthesis of bardoxolone methyl occurs through chemical modifications of the three active portions of oleanolic acid. These modifications render the derivative biologically more potent than the starter molecule (Sporn et al., 2011). The enzymatic addition of extra functionalities on the triterpene backbone could increase the number of sites that can be accessed for further synthetic modifications (Pollier & Goossens, 2012). These reactive moieties could be rationally engineered on the triterpene
backbones to generate unnatural compounds using combinatorial biosynthesis in *S. cerevisiae*, to ensure the efficient expression of CytP450s.

A major drawback of generating novel molecules *in planta* lays in the complexity of plant metabolite extracts and the complications of purifying a compound of interest from a large pool of different molecules, including compounds with similar structure and physicochemical properties. Therefore, combinatorial biosynthesis of plant secondary metabolites has also been performed in microorganisms, which lack the production of compounds similar to the target compound (Julsing *et al.*, 2006) (Fig. 7b). Novel carotenoid structures with an enhanced antioxidative activity have been generated in *E. coli* by the combinatorial expression of bacterial and plant genes (Sandmann, 2002). To date, this is the only report on the combinatorial biosynthesis of plant-derived terpenoids in microbes. A probable obstacle to the wider utilization of combinatorial biosynthesis could be the missing information about many pathway genes encoding the enzymes that catalyze the biosynthetic reactions. Ways to solve this bottleneck include programs for gene discovery in medicinal plants by means of the next-generation sequencing platforms (Fox *et al.*, 2009) or for directed evolution of enzymes towards a particular (novel) function (Kwon *et al.*, 2012).

**Enzyme engineering or directed evolution of terpenoid biosynthetic enzymes**

Small molecule drugs, considered relevant as lead molecules, often have a high degree of chemical complexity with multiple functional groups and defined stereochemistry (Nannemann *et al.*, 2011). In their natural source these small molecules are most often synthesized by enzymes that have a high regio- and stereoselectivity, a high catalysis rate and a relaxed substrate specificity. Nonetheless, natural enzymes often cannot meet the requirements of industrial chemists in terms of substrate tolerance, efficiency, process tolerance and economical viability.
Figure 7. Strategies to generate novel triterpenoid saponins. (a) Combinatorial biosynthesis in the model legume *Medicago truncatula* which produces 3-Glc-28-Glc-medicagenic acid. The overexpression of *CYP88D6*, a cytochrome P450 monooxygenase from *Glycyrrhiza uralensis* roots which produces glycyrrhizin, in *M. truncatula* could lead to the formation of a combinatorial product 3-Glc-28-Glc-11-oxo-medicagenic acid along with the naturally occurring saponins. (b) Combinatorial biosynthesis of saponins in a sterol reduced *S. cerevisiae* by the heterologous expression of saponin biosynthetic genes from *M. truncatula* and *G. uralensis*. (c) The process of directed enzyme evolution involves mutagenesis and selection for desired enzyme properties. Here, the evolution of a multifunctional enzyme to increase its reaction specificity is depicted. Glc, glucose; GlcUA, glucuronic acid. *MtBAS*, *M. truncatula* ß-amyrin synthase, *MtCytP450s*, *M. truncatula* cytochrome P450 monooxygenases; UGT, UDP-glucosyltransferase.
Hence, enzymes have been engineered by directed evolution to improve one or more of their properties under defined conditions (Dalby, 2011). In the last few years, directed enzyme evolution has progressed tremendously, and it is now feasible to engineer enzymes to accept unnatural substrates and catalyze regio- and stereospecific reactions with an efficiency comparable to that of the natural enzymes (Goldsmith & Tawfik, 2012). Directed evolution mimics the process of natural evolution and employs a set of methodologies to enhance or modify the function of a progenitor enzyme to accept an unnatural substrate or to catalyze a new biosynthetic reaction (Dalby, 2011; Nannemann et al., 2011) (Fig. 7c).

The promiscuous nature of proteins gives them an inherent ability to generate novel or altered functions with a small number of amino acid substitutions (Aharoni et al., 2005), and computational methods like computational methodologies like catalytic active site prediction (CLASP, (Chakraborty et al., 2011)) and directed evolution using CLASP: an automated flow (DECAAF, (Chakraborty, 2012)) utilize virtual screening for spatial, electrostatic and scaffold matching to identify target progenitor proteins. Enzymes catalyzing branch-point reactions in multi-branched pathways, in which a substrate is converted to multiple products, have a high evolvability. Additionally, evolvable enzymes exhibit multiple mutational residues and are ‘locally-specific’ as they recognize a common motif on structurally diverse substrates (Umeno et al., 2005).

In terpenoid biosynthesis, carotenoid biosynthetic genes have been evolved to generate novel natural product analogs. Random mutagenesis of carotenoid desaturases and subsequent expression in E. coli resulted in their acceptance of an unnatural C\textsubscript{35} carotenoid backbone to generate colonies with a spectrum of colors (Umeno & Arnold, 2003). Following a site-saturation mutagenesis to generate a library of all possible mutations at given sites, the specificity of a carotenoid synthase was altered to generate unnatural C\textsubscript{45} and C\textsubscript{50} backbones in E. coli (Umeno & Arnold, 2004). Directed evolution also served as a tool to define the product specificity of a sesquiterpene synthase expressed in E. coli. A γ-humulene synthase from Abies grandis that cyclizes FPP to 52 different sesquiterpenes was evolved by site-saturation mutagenesis of its promiscuous active site, to generate seven independent synthases, each producing one or a few
products derived from a predominant reaction pathway. Four of the seven generated synthases were new-to-nature, emphasizing the possibility of exploring novel structures from highly promiscuous natural enzymes (Yoshikuni et al., 2006).

Oxidosqualene, the immediate precursor of triterpenoid biosynthesis is a versatile molecule that is cyclized into multiple products by different OSCs. Several of these OSCs are multifunctional in nature and generate multiple products in a single reaction (Phillips et al., 2006), highlighting the promiscuity, thus evolvability of the enzymes. Through directed evolution, the major cyclization product of a multifunctional OSC could be redefined to a specific or novel product. This evolution approach could also be extended to downstream enzymes, in particular the CytP450s. Triterpene saponin backbones constitute of 30 carbons, of which approximately 20 are accessible for CytP450-mediated modifications, as deduced from known saponins (Dinda et al., 2010). Additionally, diverse functional groups are observed at the modifiable carbons pointing to the existence of specific CytP450s that catalyze these highly specific reactions. For instance, the C-11 position of many triterpene backbones can be oxidized with either an α- or β-hydroxyl group. The CytP450 specifically catalyzing the α-hydroxylation at C-11 (CYP88D6) has already been characterized (Seki et al., 2008). This implies that CytP450s that attack the C-11 position have naturally evolved to target the same substrate, but catalyze distinct reactions. To date, only seven CytP450 families involved in triterpene modifications have been identified (Fig. 6). Through directed evolution it should be possible to (a) broaden their substrate acceptance to divergent backbones, (b) target specific carbon positions and (c) specify the functional group to be added to the triterpene skeleton. Understanding the sequence-to-structure-to-function relationships is key for directed evolution studies. Integrated databases of triterpene cyclases (TTCED, Racolta et al. (2012)) and CytP450s (CYPED, Sirim et al. (2009)) facilitate the identification of functionally relevant and selectivity-determining amino acid residues, within members of a protein family by extensive sequence analysis.

Next to the generation of novel molecules, directed evolution of enzymes can also be applied to enhance enzyme efficiency or to abolish feedback regulation on enzymes. For instance, adaptive evolution was employed to engineer two enzymes, HMGR and γ-
humulene synthase, by minimizing their unfavorable in vivo properties for optimal functioning in *E. coli* and thereby improving the product flux (Yoshikuni *et al.*, 2008). In another report, the expression of a *Solanum lycopersicum* 4-coumarate:CoA ligase, which had been directly evolved to remove allosteric feedback inhibition by naringenin, resulted in an enhanced flux through the phenylpropanoid pathway in *S. cerevisiae* (Alberstein *et al.*, 2012). A similar approach could be applied to triterpenoid biosynthetic genes to increase the exogenous production of desired compounds.

**Engineering of terpenoid production via in vivo synthetic biology**

Synthetic biology (SB) is the redesigning of complex natural living systems in a rational and systematic way to simplified, predictable and controllable modules that can be modeled and manipulated to generate industrially scalable systems with a defined purpose. A simplified SB framework constitutes of computational designing and modeling, followed by experimental synthesis and analysis of a microbial host (Fig. 8a). SB aims at making metabolic engineering uniform for any goal and eliminates multiple cycles of trial and error for the generation of robust producers in a short time and cost effective manner (McArthur & Fong, 2010). To complement metabolic engineering, SB offers a plethora of tools through the generation of minimal hosts, standard biological parts, regulatory elements, vectors, assembly methods and in silico computer-aided design tools.

The first and main requirement for the production of any natural product is the availability of a robust host. SB facilitates the generation of ‘minimal hosts’ that contain only genes essential for their growth to synthesize macromolecules from simple and inexpensive feedstock. Minimal hosts of *E. coli* have been generated with about 15% genome reduction by the deletion of non-essential genes (Pósfai *et al.*, 2006). Similarly, for *S. cerevisiae*, the synthetic yeast genome project Sc2.0 aims at designing fully synthetic minimized hosts without transposable elements and telomeric sequences, with relocated tRNAs and with site-specific recombination sites incorporated in the genome.
Two partially synthetic *S. cerevisiae* chromosomes synIXR and semi-synVIL with genome reductions of 20.3% and 15.7%, respectively, were generated and successfully incorporated into the yeast (Dymond *et al.*, 2011). Likewise, the *Streptomyces avermitilis* linear chromosome was reduced to 81.46% of the wild-type chromosome by stepwise deletion of a region of >1.4 Mb, including genes coding for the synthesis of all endogenous secondary metabolites. This minimized strain produced 10-30 µg/ml of amorpha-4,11-diene upon expression of a synthetic codon-optimized *A. annua ADS* (Komatsu *et al.*, 2010). Additionally, the feasibility of generating completely artificial synthetic hosts with a desired set of genes has been demonstrated by the cloning of a chemically synthesized and assembled *Mycoplasma genitalium* genome in *S. cerevisiae* (Gibson *et al.*, 2008).

Most often metabolic engineering focuses on maximizing the production of a final compound with less attention to the behavior of intermediates. This approach makes engineering time-consuming, laborious and unextendable to other products. SB on the other hand begins with the deconvolution of metabolic pathways to independent parts that are optimized for host-specific expression, and are subsequently incorporated in a rational bottom-up approach to build elegant production modules. The repositories of functional parts (promoters, ribosomal binding sites, protein domains, terminators, etc.) generated within SB initiatives facilitate the assembly of metabolic pathways (Boyle & Silver, 2012). Two depositories with codon-optimized parts for pathway engineering in *E. coli* (The Registry of Standard Biological Parts, partsregistry.org/Main_Page) and terpene engineering in *S. cerevisiae* (Serber *et al.*, 2012) have been described. SB also promotes variable expression of related biosynthetic genes to avoid metabolic bottlenecks. Robust synthetic promoter libraries with defined promoter strengths enable modular gene expression in bacteria and yeast (Hammer *et al.*, 2006; Nevoigt *et al.*, 2006; Hartner *et al.*, 2008; Blazeck *et al.*, 2011). Tunable intergenic regions that generate mRNA secondary structures and RNase recognition sites have also been employed for differential stabilization of segments of mRNA encoding multiple enzymes in the form of operons (Pfleger *et al.*, 2006). Alternatively, synthetic protein scaffolds that are particularly efficient in overcoming rate-limiting steps, have been generated to increase
flux through metabolic pathways by tethering enzymes together (Dueber et al., 2009). Several of these SB approaches resulted in an enhanced amorpha-4,11-diene production by over a million-fold in *E. coli* (Keasling, 2012).

Natural product biosynthesis typically involves multigene pathways, thus implementing the necessity for simultaneous expression of multiple genes in a microbial chassis in an SB program. Both *in vitro* and *in vivo* methods facilitate multigene assembly in *E. coli* and *S. cerevisiae* (Ellis et al., 2011; Wang et al., 2012), some of which have been employed to assemble terpenoid biosynthetic pathways. For instance, using the λ-Red recombinase-based recombineering technique eight astaxanthin synthetic genes were successfully integrated in *E. coli* (Lemuth et al., 2011). In *S. cerevisiae*, the DNA assembler method, based on *in vivo* homologous recombination, was used to integrate an eight-gene zeaxanthin pathway (Shao et al., 2009). In parallel, viral mechanisms, such as internal ribosome entry site (IRES) and 2A oligopeptide sequences have been adapted for polycistronic gene expression (de Felipe, 2002). IRES are \( \geq 500 \) bp elements which allow initiation of translation in the middle of the mRNA in a 5’ cap-independent manner (Hellen & Sarnow, 2001). When this sequence is inserted between protein coding elements, the genes are translated independently, facilitating the expression of multiple genes from a single promoter. However, the size of the IRES and the lower yield from the downstream elements make them less attractive than the 2A peptides that are self-cleavable 19-33 amino acid peptides possessing a ribosome skipping mechanism. When the 2A peptide is inserted between concatenated genes, the ribosome skips at the glycine-proline residue of the 2A without forming a peptide bond and continues to translate the next gene, resulting in discrete proteins that are produced at similar levels (de Felipe, 2004).

Computational tools are an indispensable part of SB and make use of mathematical models for *in silico* predictions and analysis of alternative design strategies that guide biologists in choosing optimal translatable approaches (Marchisio & Stelling, 2009). Model-guided designing limits system variability, increases predictability and decreases the time spent on combinatorial assembly, validation and optimization. Optimization algorithms like Optknock (Burgard et al., 2003), OptGene (Patil et al., 2005), OptORF...
Bioengineering plant terpenoids

(Kim & Reed, 2010), OptFlux (Rocha et al., 2010) and OptForce (Ranganathan et al., 2010) have been developed for identifying gene deletion and overexpression strategies. Genome-scale metabolic network reconstructions like flux balance analysis (FBA, Gianchandani et al. (2010)), minimization of metabolic adjustments (MOMA, Segrè et al. (2002)), dynamic flux balance analysis (DFBA, Mahadevan et al. (2002)) and integrated DFBA (idFBA, Lee et al. (2008)) have been developed to effectively calculate steady-state flux distribution and enables in silico characterization of growth yield, culture conditions and robustness upon gene manipulations. In silico driven metabolic engineering based on OptGene and MOMA has been applied in yeast to significantly increase cubebol (sesquiterpenoid) production (Asadollahi et al., 2009). Additionally, computational networks like the biochemical network integrated computational explorer (BNICE) algorithm have been employed to design novel metabolic routes by starting from a specific substrate and reaction rules to result in a defined product (Hatzimanikatis et al., 2005).

Lastly, it is essential to combine metabolomics with transcriptomic approaches for the engineering of secondary metabolism, as metabolic profiling permits the characterization of novel metabolic networks and the identification of metabolic bottlenecks, and plays a pivotal role in model-based engineering (Nguyen et al., 2012). For instance, by applying targeted metabolomics, the bottlenecks in the synthesis of coenzyme Q₁₀ in E. coli were determined by looking for accumulating pathway intermediates. Subsequently, the pathway limitations were overcome by the overexpression of the corresponding genes, resulting in three-fold more production (Cluis et al., 2011).

**In vitro synthetic biology: an evolving tool**

Contrary to cell-based SB, in which the cell’s growth and survival objectives might interfere with the engineering objective, *i.e.* the production of a compound, cell-free in vitro SB provides a platform where all available resources are concentrated on the user-defined objective, which eventually might result in improved production systems (Harris
& Jewett, 2012). A cell-free environment is highly flexible and devoid of genetic regulation or transport barriers, facilitating substrate addition and product purification. *In vitro* SB systems can comprise ‘synthetic enzymatic pathways’ (SEPs), in which purified enzymes are combined in an aqueous environment to convert a substrate to a product through a series of reactions. Alternatively, ‘crude extract cell-free’ (CECF) systems, in which cells are grown, harvested and lysed to obtain a crude extract, can be utilized for the conversion of a substrate to a product (Hodgman & Jewett, 2012) (Fig. 8b).

The choice between SEP and CECF is influenced by time, cost and the need for cellular reinforcement to support the desired network. A CECF approach, for instance, is more suited for a reaction requiring a constant supply of energy such as protein synthesis (Carlson et al., 2012), however, unlike SEP, CECF reactions can exhibit undesirable activities due to the crude nature of the cellular extract.

The multireaction nature of biochemical networks, low protein concentrations, enhanced substrate diffusion, low enzyme proximity and low reaction rates due to unbalanced enzyme activity still hamper the efficiency of cell-free SB. *In vitro* compartmentalization (IVC) is one way of achieving proximity of reaction components. In IVC, genes are coupled to a substrate and encapsulated in water-in-oil emulsions together with transcription and translation machineries to facilitate enzyme synthesis and consequent product formation (Fig. 8c). Novel enzymes have been uncovered by linking product formation to genes in a confined micro-environment through IVC (Rothe et al., 2006). IVC is also being employed as a screening approach for the directed evolution of enzymes (Arnold & Volkov, 1999; Forster & Church, 2007). Besides enclosing reaction components in a defined environment through IVC, metabolic channeling has been employed as an alternative to reduce substrate diffusion lengths. Protein scaffolding (Fig. 8d), surface tethering of enzymes (Fig. 8e), covalently linked enzyme aggregates (Fig. 8f) and foam dispersion of enzymes with liposomes using surfactants (Fig. 8g) have been employed to facilitate the spatial organization of pathway components (Hodgman & Jewett, 2012).
Figure 8. Synthetic biology platforms. (a) *In vivo* synthetic biology where standard biological parts are mathematically modeled to obtain optimal expression networks which are expressed in microbial hosts like *E. coli*. (b) *In vitro* synthetic biology consists of synthetic enzymatic pathways where purified enzymes are combined with reaction components in an aqueous environment to convert a substrate to a product through a series of reactions, and crude-extract cell-free systems where resources from the cell convert an externally provided substrate to a product. (c) *In vitro* compartmentalization using water-in-oil emulsions. The encapsulated water phase consists of a substrate coupled to a gene which is transcribed and translated *in vitro* to generate an enzyme that can convert the substrate to the product. Metabolic channeling allows bringing enzymes in close proximity with their substrate by (d) protein scaffolding, (e) tethering enzymes to a surface, (f) covalently linking related enzymes into aggregates and (g) foam dispersion techniques in which the enzymes are encapsulated using surfactants. a,b,c,d,e,f,g,h, native enzymes; b*,d*,g*, synthetically modified enzymes; B,C,D,E,F,G,H, intermediates; P, product; S, substrate.
Current applications of \textit{in vitro} SB are limited to proteins, nucleic acids and small molecule ligands. Nonetheless, these tools can undoubtedly be extended to natural product (or terpenoid) engineering in the future. For instance, IVC could be employed as a tool for the directed evolution of CytP450s. A potential hurdle is the membranous nature of CytP450s, which prevents their solubilization in the aqueous reaction environment, but which may be overcome by the utilization of nanodisc membranes (Denisov & Sligar, 2011). A great advantage of using \textit{in vitro} SB in terpenoid engineering would be the simplicity and ease of catalyzing precise regio- and stereospecific reactions with a high efficiency in a relatively pure form, which may overcome the drawbacks of chemical synthesis, metabolic engineering and product purification.

\textbf{Conclusions and perspectives}

Terpenoids, the largest class of structurally and biologically diverse molecules, play a pivotal role in the plant’s growth and survival. Since time immemorial they have also been exploited by mankind for their overall health promoting and therapeutic potential. Presently, mono-, di- and tetraterpenoids are valued as natural colorants, flavors and fragrances. Although many plant terpenoids have exploitable properties, they are not utilized excessively due to their limiting availability. Like all secondary metabolites, the biosynthesis of terpenoids is strictly regulated in plants, which is often reflected by a very low accumulation \textit{in planta}. Numerous efforts to enhance the production of valuable terpenoids in endogenous plants through the application of elicitors, the external supply of precursors and the optimization of culture conditions, concurrent with the selection for high-producing cultivars, have resulted in limited success. Metabolic engineering of plants to produce desired endogenous and heterologous terpenoids has thus far also been uneconomical due \textit{e.g.} to the high costs associated with the cultivation of plant cell cultures, the time and effort required to obtain competitive or sustainable yields, the challenges in purifying compounds from complex mixtures and the failure or limited success of the engineering itself.
Microorganisms, particularly *E. coli* and *S. cerevisiae*, have emerged as industrially effective biological factories for the production of terpenoids, not only because of their scalability, but primarily due to their amenability to metabolic engineering for enhanced productivity. The advances in engineering strategies further support the potential of heterologous microbial production of secondary metabolites in general, by providing an ever-growing toolkit for the rational engineering of organisms to produce profitable compounds in a short time span. The combinatorial biosynthesis of novel compounds, the evolution of enzymes to carry out defined functions and SB approaches are being increasingly used for tetra- and sesquiterpene engineering, which in turn is generating a knowledge base for the future engineering of other terpenoid classes that have hardly been touched so far. For instance, the triterpene saponins with their grandeur of chemical diversity and biological activity, constitute one of the least engineered terpenoid classes with regard to heterologous production in microbes. The vast structural diversity and distribution of triterpene saponins in plants, highlights the capacity for combinatorial synthesis of novel molecules using existing natural enzyme resources. The spectrum of saponins seen in nature is generated predominantly from a single squalene precursor through specific enzyme-driven chemical reactions, underscoring the promiscuity of the involved enzymes and therefore their amenability toward engineering programs. Hence, through the combined application of directed evolution strategies and synthetic biology toolboxes, triterpene saponin engineering could become the epitome in the near future.

**Acknowledgements**

We thank Annick Bleys for help in preparing the manuscript. This work was supported by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. Tessa Moses is indebted to the VIB International PhD Fellowship Program for a predoctoral fellowship. Jacob Pollier is a postdoctoral fellow of the Research Foundation-Flanders.
References


Bioengineering plant terpenoids


Kumar S, Hahn FM, Baidoo E, Kahlon TS, Wood DF, McMahan CM, Cornish K, Keasling JD, Daniell H, Whalen MC. 2012. Remodeling the isoprenoid pathway in


Bioengineering plant terpenoids


Bioengineering plant terpenoids


An introduction to Gas Chromatography – Mass Spectrometry (GC-MS) and its application for triterpenoid sapogenin analysis
Author contribution:

Writing of the manuscript.
Abstract

Gas chromatography – mass spectrometry has been widely applied for the separation, detection and identification of a wide range of non-polar compounds, including sterols. The saponin building blocks or the sapogenins, akin sterols are extremely hydrophobic in nature. Based on sterol analysis using GC-MS, methods for the separation and identification of sapogenins using electron impact ionization have been established. Here, we give a short summary on the general principles of GC-MS and expand on the application of this technique for triterpenoid sapogenin analysis. In addition, the principles of sample preparation, gas chromatographic separation and interpretation of derived mass spectra for both tetra- and pentacyclic triterpenoid sapogenins is summarized.
Introduction to GC-MS

Gas chromatography – mass spectrometry (GC-MS) is a synergistic combination of two analytical techniques, where the gas chromatograph separates the components of a mixture in time and the mass spectrometer provides information to aid the structural identification of the components. The basic operating principle of the GC-MS involves the volatilization of the sample mixture in a heated inlet port (the injector), separation of the mixture components in a column, and the detection of each component by a detector (Fig. 1).

Figure 1. Functional diagram of Gas chromatography – mass spectrometry (GC-MS) instrumentation.

The carrier gas (mobile phase) is an integral part of the technique and allows the transfer of the sample from the injector, through the column into the detector. Gases like hydrogen and helium are the most commonly used carrier gases. The column and its
packing material correspond to the stationary phase. The separation of components in a mixture is determined by the distribution of each component between the mobile and the stationary phase. A component that sticks less to the stationary phase elutes quickly from the column and will therefore have a short retention time on the column.

The GC column contains a thin layer of a non-volatile chemical that is either, coated onto the wall of the column (capillary column), or coated on an inert solid that is added to the column (packed column). The components of an injected sample are carried onto the column by the carrier gas and selectively retained on the stationary phase. The most commonly used general purpose stationary phases which are less prone to column bleed are the highly non-polar dimethylsiloxanes (DB-1 or equivalent) and 5% phenyl/95% dimethylsiloxane (DB-5 or equivalent). During a GC run, the temperature of the oven where the GC column is placed is increased gradually (ramped) at 4-20°C/min so that the higher boiling and more strongly retained components are successively released from the column. Therefore, GC is limited to compounds that are volatile or can be made volatile and are sufficiently stable to flow through the column. Sample derivatization is frequently employed to increase the volatility and stability of acids, amines, amides, sterols and many other compound classes for GC analysis.

The chromatogram produced during a GC run is a graphical representation of the signal generated from a compound eluting from the GC column into the detector. The X-axis represents the retention time ($t_R$) of the compound on the column and the Y-axis shows the intensity or abundance of the signal. Each peak represents an individual compound that was separated from a sample mixture. If the GC conditions (oven temperature ramp, column type, etc.) are the same, a given compound will always elute from the column at nearly the same $t_R$. Therefore, by knowing the $t_R$ of a compound, some assumptions about its identity can be made. However, compounds with similar properties often have similar $t_R$ and more information is usually required for the identification of a compound in a sample containing unknown components.

The MS allows the identification of components that elute from the GC column. An interface in the GC-MS allows transporting the effluent from the GC to the MS without
condensing nor decomposing the components during the transfer. The MS measures the mass-to-charge ratio (m/z) of gas phase ions and the abundance of each ionic species. The measurement is calibrated against ions of known m/z. However, the charge of ions in a GC-MS is almost always 1 and therefore the m/z corresponds to the mass of the ion in Daltons (Da) or atomic mass units.

The MS separates the ions in a low pressure environment or vacuum by the interaction of a magnetic and electrical field on the charged particles. The electron impact ionization (EI) is the most commonly used ionization method for GC-MS, although chemical ionization (CI) is also employed. During EI, electrons from a hot wire or ribbon (the filament) are accelerated by 70V (70eV) before entering the ion source through a small aperture. The effluent from the GC also enters this partially enclosed ion source. When electrons pass near neutral molecules they impart sufficient energy to remove the outer shell of electrons from these molecules, thereby producing additional free electrons and positive molecular ions. The electron-electron interactions cause molecules to lose both the incoming electron and a bound electron. Owing to the high energy imparted by EI, and the absence of other compounds as is the case in vacuum, the molecular ions break up or ‘fragment’ into other ions, radicals and neutral atoms (Fig. 2).

\[ M + e^- \rightarrow M^{++} + 2e^- \]

\[ F^+ + N^- \]

**Figure 2.** Electron impact ionization (EI) of a compound (M). e\(^-\), electron; F\(^+\), fragment ion; N\(^-\), neutral atom.

The mass spectrum is a graphical representation of the ions observed by the MS over a range of m/z values, where the X-axis is the m/z scale and the Y-axis corresponds to the intensity scale. The mass spectrum produced by a compound is essentially the same every
time and therefore, a fingerprint for the molecule, that can be used for the identification of compounds in sample mixtures. When all the components of a mixture eluting from the GC column are analyzed on the MS, the total ion current (TIC) corresponds to a full MS scan, and both the intensity and the m/z of a component are important for interpreting the mass spectrum. Full MS scans are typically used for untargeted component analysis and consists of a compilation of mass spectral information from all the components in the mixture. However, most mass spectra usually have several peaks that correspond to background and not to the specific component. These background peaks occur due to the presence of contamination in the ion source, the transfer line, the GC column, and due to column bleed.

The MS can also be used for recording the ion current at selected masses for targeted component analysis, by scanning for characteristic m/z corresponding to a compound. This is referred to as selected ion monitoring (SIM), where the MS does not spend time scanning the entire mass range, but rapidly changes between m/z values for which characteristic ions are expected. The SIM method allows quantitative analysis at parts per billion (ppb) level and is advantageous because of its high specificity and sensitivity.

**Sample preparation for GC-MS analysis of triterpenoid sapogenins**

Triterpenoid sapogenins are composed of 30 carbon atoms that are arranged in 4 or 5 ring structures with or without a side chain to form tetra- and pentacyclic triterpenoids, respectively. For the analysis of triterpenoid sapogenins various chromatography based analytical techniques, like thin layer chromatography (TLC, (Oleszek, 2002)), liquid chromatography (LC, (Burnouf-Radosevich and Delfel, 1984)), capillary electrophoresis (CE, (Unger, 2009)) and gas chromatography (GC) with flame ionization (Janicsák et al., 2003) or mass spectrometry (MS, (Burnouf-Radosevich et al., 1985)) detection are described. Among these, GC-MS is the most powerful technique for the separation, quantification and structural determination of triterpenoid sapogenins, including position isomers with very similar structures, such as α-amyrin (Fig. 3a) and β-amyrin (Fig. 3b).
Some of the sapogenin backbones, like α-amyrin, β-amyrin, lupeol, etc., are sufficiently volatile at high temperatures and can be detected by GC-MS without compound derivatization. However, sapogenins derived from these backbones have high molecular weights and are apolar, which considerably lowers their volatility. Therefore, detection of underivatized sapogenins by GC-MS is laborious.

**Figure 3.** Chemical structure of (a) α-amyrin, (b) β-amyrin.

Derivatization of sapogenins by silylation, acetylation or methylation prior to GC-MS enhances their volatility and hence, detectability. Silylation is the most prevalent derivatization method applied to tetra- and pentacyclic triterpenoid sapogenins. Silylation produces silyl derivatives by replacing active hydrogens through nucleophilic attack with a trimethylsilyl (TMSi) group. The better the leaving group, the better the silylation reaction. Silylation reagents readily react with water. Therefore, samples and solvents used for derivatization should be water-free. Pyridine is the most commonly employed solvent, which is also an acid scavenger that can positively influence the reaction. The ease of reactivity of a functional group with an active hydrogen for silylation follows the order; hydroxyl of alcohol (-OH) > hydroxyl of phenol (-OH) > carboxyl (-COOH) > amine (-NH2) > amide (-CONH2). The silylation reaction is driven by a good leaving group which has low basicity and therefore the ability to stabilize a negative charge in the transitional state, and has negligible or no π bonding between the leaving group and
silicon atom. There are multiple silylating reagents that can be used in combination with pyridine for silylation reactions. We utilized the N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) silylating reagent, which contains TFA as the strong leaving group, and reacts completely with the sample in a short time.

GC separation of derivatized triterpenoid sapogenins

The retention of sapogenins and sapogenin backbones in a GC column is influenced by the number and type of functional groups they possess. We utilized a VF-5ms capillary column, which is equivalent to the DB-5 column, for the separation of sapogenins. The retention time of sapogenins on this column generally increased with increasing molecular weight of the derivatized triterpenoid. Among the pentacyclic triterpenoids, the ursane-type sapogenins α-amyrin and uvaol, were retained longer on the column than their respective oleanane-type isomers β-amyrin and erythrodiol (Fig. 4). This might be a result of the increasing planarity of the molecule due to the shifting of the methyl (-CH₃) group from the axial conformation on C-20 in β-amyrin to the equatorial conformation on C-19 in α-amyrin, thereby increasing its retention on the column.

Interpretation of EI fragmentation pattern of triterpenoid sapogenins

In general the triterpenoid sapogenins can be divided into three categories based on their EI fragmentation pattern, (a) the pentacyclic triterpenoids containing a C-12 – C-13 double bond, like α- or β-amyrin and their derivatives, (b) the pentacyclic triterpenoids without a double bond, like lupeol and its derivatives, and (c) the tetracyclic triterpenoids with a side chain, like dammarenediol and its derivatives. In this section, the interpretation of the EI fragmentation pattern of these main categories of sapogenins derivatized with TMS will be illustrated with examples.
The pentacyclic triterpenoids with a C-12 – C-13 double bond undergo a characteristic retro Diels-Alder (rDA) cleavage of the C-ring to form fragments consisting of the ABC*-rings (dienophile) and the C*DE-rings (diene), where C* represents only a portion of the C-ring (Fig. 5a). The charge is retained on the diene which undergoes further fragmentation mostly generating an ion with the m/z 203, as seen in the mass spectrum of β-amyrin (Fig. 5b) and oleanolic acid (Fig. 5c). Similarly, the dienophile looses the silylated hydroxyl (TMSiOH) on C-3 to form the fragment ion with m/z 189. Together with the abundant intensities of the fragment ions generated by the rDA fragmentation, also signals corresponding to m/z of the trimethylsilylated molecular (parent, M+) ion, loss of methyl group (M+-CH$_3$, 15 Da), loss of trimethylsilylated hydroxyl group (M+-TMSiOH, 90 Da), loss of trimethylsilylated carboxyl group (M+-TMSiCOOH, 118 Da), combinations of different above mentioned losses, and the m/z 73 corresponding to TMSiH can be observed (Fig. 5b-c). In addition, the loss of hydroxyl- and carboxyl groups on ring-carbons can also be distinguished from loss of hydroxyl groups from methyl-carbons (Burnouf-Radosevich et al., 1985).
In the pentacyclic triterpenoids, like lupeol, that do not have a double bond the EI fragmentation is initiated by a C-14 – C-27 cleavage and consequent methyl elimination (Fig. 6a). The fragment ions so generated are not observed as they readily decompose to fragment ions with lower m/z. The fragment ions with m/z 279 and 189 are formed, like in lupeol (Fig. 6b), as a result of two competitive pathways, which can be used to detect the presence of substituents on A, B, C, D or E rings of the pentacyclic triterpenoids (de Carvalho et al., 2010). Also the m/z corresponding to the M+ and M+−CH₃ ions can be detected in the EI spectrum.
GC-MS for triterpenoid sapogenin analysis

Figure 6. (a) EI pattern of trimethylsilylated lupeol, a pentacyclic triterpenoid without double bond. (b) Mass spectra of a lupeol standard. The X-axis of the mass spectra correspond to m/z values.

The tetracyclic triterpenoids, like dammarenediol have characteristic side chains attached to the D ring. The m/z corresponding to the trimethylsilylated parent M⁺ ion, is often hard to detect. However, the m/z of M⁺-CH₃ and M⁺-TMSiOH fragment ions, corresponding to the loss of 15 Da and 90 Da, respectively can be readily observed. The characteristic cleavage on either side of the silyloxy group at C-20 gives a m/z of 505 and 199, the latter of which represents the most abundant m/z in the EI fragmentation pattern (Fig. 7) (Spencer, 1981).
Figure 7. (a) EI pattern of trimethylsilylated dammarenediol, a tetracyclic triterpenoid. (b) Mass spectra of dammarenediol. The X-axis of the mass spectra correspond to m/z values.

References


A novel β-amyrin hydroxylase from *Bupleurum falcatum*, its utilization for combinatorial biosynthesis in *Saccharomyces cerevisiae* and a novel approach for extraction of triterpene sapogenins from yeast cells
Publication status:

**Tessa Moses**, Jacob Pollier, Lorena Almagro Romero, Dieter Buyst, María A. Pedreño, Jose Martins, Johan M. Thevelein and Alain Goossens.

*Bupleurum falcatum* β-amyrin hydroxylase and a novel approach for extraction of triterpene sapogenins from *Saccharomyces cerevisiae*.

Manuscript in preparation

Author contribution:

Generation of yeast strains, culturing and metabolite extractions, GC-MS profiling, data analysis and writing of the manuscript.
Abstract

Saikosaponins are triterpenoid saponins abundantly present in the roots of the genus *Bupleurum*, the species of which are widely used in Asian traditional medicine. The saikosapogenins are the aglycosylated forms of the saponin and constitute of oleanane- and ursane-type backbones with multiple oxygenations mainly in the D and E rings of the backbone. In this study we have identified a cytochrome P450 (BF567) from *Bupleurum falcatum* that is involved in the hydroxylation of triterpenes during the biosynthesis of saikosapogenins. We engineered a yeast strain that endogenously produces 36.2 mg/L β-amyrin and used it for the ectopic expression of BF567. Additionally, by varying the ratio of the cytochrome P450 reductase to BF567 we demonstrate improvements in the endogenous β-amyrin hydroxylation efficiency of BF567 and thereby the relevance of this ratio for the production of triterpene sapogenins. We also establish a novel approach for the extraction of hydrophobic triterpene sapogenins from yeast cells into the medium, using methylated cyclodextrin to increase productivity and facilitate downstream processing. Finally, we present proof of concept for combinatorial triterpene sapogenin biosynthesis in *S. cerevisiae* and apply it for the characterization of BF567.
Introduction

Triterpenoid saponins are a class of secondary metabolites exhibiting a wide range of structural diversity and biological activities among many plant species (Suzuki et al., 2002; Sparg et al., 2004). Saponins can be structurally defined as, glycosides of “sapogenins” composed of 30 carbon cyclic isoprenoids arranged in 4 or 5 ring structures, that are decorated by additional functional groups (hydroxy-, carboxy-, oxo-, etc). Saponins are synthesized by multiple glycosylations of the sapogenin building blocks, which in turn are produced by multiple cytochrome P450 monooxygenase (CytP450) or oxido-reductase mediated modifications of basic sapogenin backbones like β-amyrin (oleanane-type), α-amyrin (ursane-type), lupeol and dammarenediol (Fig. 1a). These diverse backbones are generated by specific oxidosqualene cyclase (OSC) mediated cyclization of 2,3-oxidosqualene, which is also an intermediate in the synthesis of membrane sterols in plants, mammals and yeast. Sapogenins are valuable as biologically active compounds and can also serve as important starter molecules for further synthetic modifications. For instance, the naturally occurring triterpenoid sapogenin oleanolic acid and its derivatives possess several promising pharmacological activities, such as hepato-protective effects, anti-inflammatory, antioxidant and anticancer activities (Pollier and Goossens, 2012), of which bardoxolone methyl, a semisynthetic derivative of oleanolic acid, is currently in a late-stage clinical development for the treatment of advanced chronic kidney disease in type 2 diabetes mellitus patients.

The genus *Bupleurum* constitutes of perennial herbs and forms an integral part of Asian traditional medicine in which it is used either alone, or in combination with other ingredients for the treatment of common colds, fever and inflammatory disorders in the form of over the counter herbal teas. *Bupleurum* species are also enlisted in the Chinese and Japanese Pharmacopoeias and the WHO monographs on selected medicinal plants. Saikosaponins constitute the largest class of secondary metabolites in *Bupleurum* and account for ~7% of the total dry weight of roots. More than 120 closely related glycosylated oleanane- and ursane-type saikosaponins have been identified from this genus that can be distinguished only by the position and number of double bonds in rings.
C and D and oxygenation patterns on C-16, C-23, C-28 and C-30 (Fig. 1b) (Ashour and Wink, 2011). The presence of oxygenations at various positions on saikosapogenins suggest the presence of specific enzymes, generally CytP450s, capable of catalyzing these modifications on the β-amyrin and/or α-amyrin backbone in the genus Bupleurum. However, to date not a single CytP450 or oxido-reductase involved in triterpene sapogenin biosynthesis has been identified from Bupleurum species.

In general the triterpene sapogenin biosynthesis is not well elucidated and only a handful of CytP450s have been characterized from different plants, which include a C-11 oxidase (CYP88D6 (Seki et al., 2008)), C-24 hydroxylases (CYP93E1 (Shibuya et al., 2006), CYP93E2 (Fukushima et al., 2011) and CYP93E3 (Seki et al., 2008)), C-28 oxidases (CYP716A12 (Carelli et al., 2011), CYP716A15, CYP716A17 (Fukushima et al., 2011) and CYP716AL1 (Huang et al., 2012)) and C-30 hydroxylases (CYP72A154 and CYP72A63 (Seki et al., 2011)) that modify the β-amyrin backbone (indicated by asterisk, Fig. 1b). Additionally, a C-6 (CYP716A53v2 (Han et al., 2012)) and C-12 (CYP716A47 (Han et al., 2011)) hydroxylase that modify the dammarenediol-II backbone have also been identified. To characterize these CytP450s, yeast strains either producing β-amyrin or externally fed with the substrate have been employed for ectopic expression of the CytP450 and identification of modified products. From these studies it is naturally evident that yeast cells can be employed, not only for the characterization of novel enzymes, but also as a heterologous host for the production of triterpenoid sapogenins. From the perspective of a production host, in spite of several OSCs having been expressed in S. cerevisiae, little effort has been made towards engineering the host to enhance production of triterpenes. For instance, Kirby and coworkers could produce 6 mg/L of β-amyrin in S. cerevisiae by expressing a β-amyrin synthase (bAS) from Artemisia annua and manipulating two enzymes, 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) and lanosterol synthase (ERG7), of the ergosterol biosynthetic pathway. In this strain they observed a 12-fold increase in squalene levels indicating scope for further improvements in β-amyrin production (Kirby et al., 2008).
Figure 1. (a) A simplified overview of saponin biosynthesis in plants depicting the different branches of enzymes involved. DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate; GPP, geranyl pyrophosphate; IPP, isopentenyl pyrophosphate. (b) Chemical structure of oleanane-type sapogenin backbone. Asterisk (*) indicate the carbon positions for which a CytP450 has already been characterized. (c) Schematic representation of the modified yeast ergosterol biosynthetic pathway. Multiple steps are indicated by dashed lines.

Here we describe the identification and characterization of a novel plant cytochrome P450 (BF567), from *B. falcatum* involved in the biosynthesis of saikosaponins. BF567 was identified through a genome wide cDNA-AFLP based transcript profiling and
ectopically expressed in a β-amyrrin producing *S. cerevisiae* strain with modified sterol biosynthesis (Fig. 1c), where it was identified as a novel hydroxylase from *B. falcatum*. Further, using *BF567* we established a combinatorial pipeline for the production of triterpene sapogenins. Additionally, we propose a novel approach for the extraction of triterpene sapogenins from yeast cells into the medium to facilitate downstream purification by utilizing cyclodextrins.

**Results**

**Transcript profiling of MeJA-treated *B. falcatum* roots reveals *BF567*.**

To identify new saponin biosynthetic genes, we performed a genome-wide cDNA-AFLP based transcript profiling on the roots of hydroponically grown *B. falcatum* plants. Using the complete set of 128 *Bst*YI+1/*Mse*I+2 primer combinations, the expression of a total of 18,800 transcript tags was monitored over time. In total, 1,771 MeJA-responsive transcript tags were isolated (hereafter referred to as BF tags). Direct sequencing of the reamplified BF tags gave good quality sequences for 1217 (68.7%) of the fragments. To the remaining 554 tags (31.3%), no unique sequence could be attributed unambiguously, indicating that they might not represent unique gene tags and hence were not considered for further analysis. A BLAST search with the nucleotide sequences of the 1217 unique cDNA-AFLP tags led to the annotation of 776 (63.7%) of the BF tags.

Average linkage hierarchical clustering analysis of the expression profiles of the 776 annotated BF tags showed that, upon MeJA treatment the selected genes are either transcriptionally activated (Clusters I-V, Fig. 2a) or transcriptionally repressed (Cluster VI, Fig. 2a). The activated gene tags can be divided into five subclusters, based on the timing of the MeJA response. A first cluster consists of genes with a response immediately after (≤ 1 h) MeJA treatment (cluster II). In a second and third cluster (Clusters III and IV, respectively), genes reach maximum expression between 2-4 h and 4-8 h after MeJA treatment, respectively. In a fourth cluster (Cluster I), genes are
activated 24 h after MeJA treatment, and in a fifth cluster (Cluster V), genes are activated within 2 h after the MeJA treatment, and their expression remains high thereafter. In the latter group, tags corresponding to genes encoding enzymes that catalyze early steps in the triterpene saponin biosynthesis, including squalene synthase (SQS) and bAS can be found. These tags displayed an almost identical expression pattern, suggesting a tight co-regulation, and reached maximum levels of expression 8-24 h post-elicitation (Fig. 2b). The gene tag BF567 is tightly co-regulated with these genes (Fig. 2b), and shows homology to the M. truncatula gene encoding the cytochrome P450 enzyme CYP716A12, characterized to oxidize β-amyrin in a sequential three-step oxidation on C-28 to yield oleanolic acid through erythrodiol (Carelli et al., 2011; Fukushima et al., 2011). The full-length open reading frame (BF567) corresponding to the gene tag BF567 was picked up from a B. falcatum Uncut Nanoquantity cDNA library (Pollier et al., 2011b). Additionally, a phylogenetic analysis (Fig. 2c) confirmed that BF567 groups together with the CYP716 family of CytP450s involved in triterpene sapogenin biosynthesis.

\textit{Saccharomyces cerevisiae} strain \textsc{tm1} with modified sterol biosynthesis.

The ergosterol biosynthetic pathway of a S288c BY4742 strain of \textit{S. cerevisiae} was modified as described (Kirby et al., 2008) with adaptations, to facilitate accumulation of 2,3-oxidosqualene (Fig. 1c), a common intermediate of sterol biosynthesis between plants and yeast. First, the \textit{S. cerevisiae} lanosterol synthase (ERG7) responsible for the cyclization of 2,3-oxidosqualene to lanosterol was made conditionally down-regulatable, by replacing the native ERG7 promoter with a methionine-repressible MET3 promoter to generate strain \textsc{tm1}. The amount of ergosterol produced by \textsc{tm1} in the presence of different concentrations of methionine was quantified using the sterol quantification method (Arthington-Skaggs et al., 1999) and a 60% reduction in ergosterol accumulation was observed with 1.5 mM methionine, when compared to wild type cells. This concentration of methionine was thus used for all following experiments. An overview of all yeast strains derived from \textsc{tm1} is given in Table 1.
Further, a truncated feedback uncoupled copy of isoform 1 of the rate limiting enzyme 3-hydroxy-3-methylglutaryl-CoA reductase (tHMG1) was generated and cloned into the multiple cloning site (MCS) 1 of the high copy number plasmid pESC-URA behind the
galactose inducible \textit{GAL10} promoter to generate \textit{pESC-URA[GAL10/tHMG1]}, which was transformed into \textit{TM1} to generate a control strain \textit{TM5}.

\textbf{β-amyrin producing strains \textit{TM2} and \textit{TM3}.}

Two β-amyrin producing strains \textit{TM2} and \textit{TM3} were generated from \textit{TM1} by expressing \textit{bAS} of \textit{M. truncatula} (GenBank accession number AJ430607; (Suzuki et al., 2002)) or \textit{G. glabra} (GenBank accession number AB037203; (Hayashi et al., 2001)), respectively from the plasmid \textit{pESC-URA[GAL10/tHMG1; GAL1/bAS]}. The \textit{in vivo} activity of the \textit{bAS} in \textit{TM2} and \textit{TM3} was compared by quantifying levels of β-amyrin produced by 72 h cultures, using GC-MS. The strains were grown in galactose medium for 24 h, to accumulate \textit{bASp} and increase flux through ergosterol pathway, prior to repression of \textit{ERG7} with 1.5 mM methionine for 48 h, to accumulate 2,3-oxidosqualene and simultaneously convert it to β-amyrin. Analysis of cell pellets showed the presence of a single peak at 27.2 min with an electron ionization (EI) pattern corresponding to a β-amyrin standard (Fig. 3d), in the GC chromatograms of \textit{TM2} (Fig. 3a) and \textit{TM3} (Fig. 3b), but not the control strain \textit{TM5} (Fig. 3c). Further, we observed higher accumulation of β-amyrin in strain \textit{TM3} (36.2 mg/L) compared to \textit{TM2} (19 mg/L) and hence continued with \textit{TM3} for further experiments.

\textbf{In vivo activity of BF567 in strain \textit{TM7}.}

Strain \textit{TM7} was generated by super-transforming strain \textit{TM3} with plasmids \textit{pAG423[GAL1/BF567]} and \textit{pAG425[GAL1/AtATR1]}, expressing BF567 and the \textit{A. thaliana} CytP450 reductase (CPR), \textit{AtATR1} (At4g24520), respectively from the galactose inducible \textit{GAL1} promoter. In parallel, a control strain \textit{TM26} harboring only \textit{pAG425[GAL1/AtATR1]} in \textit{TM3} was also generated. Cell pellets analyzed by GC-MS showed the presence of a unique new peak eluting at 31.8 min in \textit{TM7} (Fig. 4a), but not
TM26 (Fig. 4c). The EI pattern of this peak corresponded to a hydroxylated derivative of \( \beta \)-amyrin, with the alcohol function on either the D or E ring of \( \beta \)-amyrin (Fig. 4d).

**Figure 3.** GC chromatograms showing \( \beta \)-amyrin at 27.2 min, where the X- and Y-axis correspond to time in minutes and total ion current, respectively. (a) Extraction from cells of strain TM2, expressing tHMG1 and MtbAS. (b) Extraction from cells of strain TM3, expressing tHMG1 and GgbAS. (c) Extraction from cells of control strain TM5, expressing tHMG1. (d) A \( \beta \)-amyrin standard. Right panel shows the mass spectra extracted from the indicated (*) peak, where X-axis corresponds to m/z values.
Since BF567 was tentatively annotated as a homolog of the M. truncatula CYP716A12 (GenBank accession number FN995113; (Carelli et al., 2011)), we compared the GC elution time and EI pattern of the new peak in TM7 with a standard of erythrodiol (28-hydroxy-β-amyrin). We also generated strain TM10 by transforming plasmid pAG423[GAL1/CYP716A12] along with pAG425[GAL1/AtATR1] in TM3, and compared its GC-MS profile with TM7, TM26 and an erythrodiol standard. A peak corresponding to the elution time and EI of standard erythrodiol (Fig. 4e) was observed at 32.5 min in the GC chromatogram of TM10 (Fig. 4b) but not TM7 and TM26, indicating that BF567 hydroxylates β-amyrin at a position different than CYP716A12. Therefore, we looked at the oleanane-type triterpenoid sapogenins found in Bupleurum (Ashour and Wink, 2011) and narrowed down the possible hydroxylation positions of BF567 on rings D and E of β-amyrin to C-16β, C-21β and C-29 (Fig. 4d), which are positions for which a CytP450 has not been characterized to date.

Effect of CPR:CytP450 ratio on in vivo activity of BF567.

The endoplasmic reticulum (ER) localized CPRs are flavoproteins, containing both a redox cofactor flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), that serve as electron donor proteins for several ER oxygenases, including CytP450s. Therefore, optimal interaction between CPR and CytP450 is essential to allow the reducing equivalents from NADPH to pass from the CPR to the CytP450 (Reed and Backes, 2012). In an attempt to increase the efficiency of hydroxylation of β-amyrin by BF567, we determined the effect of the ratio of CPR to CytP450, by varying the expression level of AtATR1 while keeping the expression of BF567 constant.
Figure 4. GC chromatograms corresponding to, (a) Extraction from cells of strain TM7, expressing tHMG1, GgbAS, AtATR1 and BF567. (b) Extraction from cells of strain TM10, expressing tHMG1, GgbAS, AtATR1 and CYP716A12. Right panels show the mass spectra extracted from the indicated (*) peaks. (c) Extraction from cells of control strain TM26, expressing tHMG1, GgbAS and AtATR1. (d) Mass spectra extracted from the peak indicated (*) at 31.8 min of strain TM7. The possible hydroxylation positions on β-amyrin are indicated by arrow heads. (e) Mass spectra of an erythrodiol standard. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively. The X-axis of the mass spectra correspond to m/z values.
The CPR:CytP450 ratios between 1:5 and 1:30 have been reported to be ideal for the efficient functioning of yeast and mammalian CytP450s (Reed and Backes, 2012), therefore we expressed the AtATR1 from either an integrated (pAG305, 1 copy per cell), low-copy number (pAG415, 3-5 copies per cell), or high-copy number (pAG425, 10-40 copies per cell) vector, in combination with BF567 always expressed from the high-copy number plasmid (pAG423, 10-40 copies per cell). Thus, we generated two strains TM8 and TM9, overexpressing pAG423[GAL1/BF567] along with pAG305[GAL1/AtATR1] or pAG415[GAL1/AtATR1], respectively and compared the amount of hydroxylated β-amyrin produced by these strains with that of TM7. In accordance with this assumption, strain TM9, accumulated higher levels of hydroxylated β-amyrin compared to TM8 and TM7, with the lowest accumulation in the strain expressing the integrated copy of CPR (Fig. 5a). Therefore, we expressed AtATR1 from the low-copy number vector pAG415, for further experiments.

Cyclodextrin facilitates production and release of triterpene sapogenins from yeast cells.

Cyclodextrins (CDs) are cyclic oligosaccharides composed of α-D-glucopyranoside units, that can be represented by a toroidal topology and are extensively employed for solubilizing and stabilizing pharmaceuticals (Stella and He, 2008). Owing to their toroidal arrangement, the interior of the CD is considerably more hydrophobic than the exterior aqueous environment enabling them to host hydrophobic moieties, like cholesterol (Christian et al., 1997). Therefore, considering the hydrophobic nature of sapogenins and their common biosynthetic origin with sterols, sapogenins could also be sequestered by CD and transported from within the yeast cells to the external growth medium.

First the concentration of methyl β-cyclodextrin (MβCD) to be applied to the yeast culture, to facilitate the extraction of ergosterol and β-amyrin from yeast cells to the spent medium was determined. We compared the GC chromatograms of extracts from cell
Bupleurum falcatum CytP450 and extraction of sapogenins from yeast

pellet and spent medium of strain TM3 treated with 5 mM or 25 mM MβCD and compared it with an untreated control. Both ergosterol and β-amyrin were detected in the cell pellet of all the samples, but only in the spent medium of MβCD treated samples, suggesting the ability of MβCD to sequester not only yeast sterols but also triterpene sapogenin backbones into the medium (Fig. 5b). Since both 5 mM and 25 mM MβCD could extract the hydrophobic compounds from the yeast cells, we decided to continue further experiments with the lower concentration of MβCD.

Using 5 mM MβCD we examined the amount of β-amyrin extracted from the yeast cells by culturing strain TM3 in the presence and absence of the CD and quantified the amount of β-amyrin in the cell pellet and spent medium by GC-MS. GC chromatograms confirmed the presence of β-amyrin in both the cell pellet (20.4 mg/L) and spent medium (37.3 mg/L) upon MβCD treatment, with higher concentrations of β-amyrin quantified from the extracts of spent medium (Fig. 5c). Additionally, the total concentration of β-amyrin was found to be 1.6-fold higher in cultures treated with MβCD, when compared to untreated controls.

Next, we determined if MβCD facilitates the extraction of β-amyrin from strain TM3 in a dose dependent manner. For this we applied MβCD at different times during culturing and generated a total of 7 culturing conditions which included, an untreated control (C, Fig. 5d) and 6 treated samples, with MβCD added to a concentration of 5 mM each time. To samples I1, I2 and I3 (Fig. 5d) MβCD was added on Day 1 after inoculation into galactose medium. To samples I2 and I3 an additional dose of MβCD was added on Day 2 together with addition of methionine, and additionally to sample I3 a third dose of MβCD was added on Day 3. Further, to samples R1, R2, and AR1 (Fig. 5d) MβCD was added on Day 2 only, Day 2 and Day 3, and Day 3 only, respectively. Extractions were performed on the spent medium of all the samples on Day 4 and quantified for β-amyrin using GC-MS. We observed a direct corelation between the amount of β-amyrin quantified from the spent medium and the number of times cyclodextrin was added to the sample, thereby suggesting the dose dependent nature of this extraction (Fig. 5d).
Figure 5. (a) Effect of CPR:CytP450 ratio on the *in vivo* activity of BF567 in strains TM8, TM9 and TM7, where AtATR1 is expressed from vectors having integrative (INT), centromeric (CEN) and 2 micron (2µ) origin of replication. (b) GC chromatograms showing ergosterol and β-amyrin in spent medium of strain TM3 treated with 5 mM or 25 mM MβCD and untreated control. The X- and Y-axis correspond to time in minutes and total ion current, respectively. (c) Quantification of β-amyrin from the cells and spent medium of strain TM3, where 100% corresponds to 36.2 mg/L of β-amyrin. (d) Dose dependent extraction of β-amyrin quantified for strain TM3, where cyclodextrin was added on Day 1 (I1), Day 1 and 2 (I2), Day 1, 2 and 3 (I3), Day 2 (R1), Day 2 and 3 (R2), Day 3 (AR) and untreated control (C). (e) Relative amounts of hydroxylated β-amyrin quantified from the cells and spent medium of strain TM9. (f) Relative amounts of hydroxylated β-amyrin quantified from spent medium of strain TM9 treated with different variants of CD.
For the following experiments with strain TM9, we employed condition R2 to avoid the excessive release and hence loss of β-amyrin, the precursor for BF567, from the cells into the medium. Surprisingly, the hydroxylated β-amyrin eluting at 31.8 min was only observed in the GC chromatograms of the spent medium and not cell pellets of TM9 upon MβCD treatment (Fig. 5e), suggesting the complete release of the hydroxylated product from the yeast cells into the medium.

We also determined the specificity of the type of CD used, on the sequestering of hydroxylated β-amyrin from the cells into the spent medium of strain TM9. The most abundant variants of CD are α, β and γCD which have 6, 7 and 8 glucose units, respectively. Therefore, we applied αCD, βCD, γCD, Random MβCD (RMβCD) or MβCD to a final concentration of 5 mM as in condition R2 and analyzed the spent medium on Day 4, for quantification of the amount of hydroxylated β-amyrin released into the medium. Sequestering was only observed with the βCD and its methylated versions and the highest amount of hydroxylated β-amyrin was detected upon RMβCD and MβCD treatment, suggesting a strong specificity of the methylated forms of CD over the unmethylated forms, for sequestering of triterpenoid sapogenins from yeast cells (Fig. 5f).

**Combinatorial biosynthesis using BF567 and CYP716A12.**

Combinatorial biosynthesis also known as combinatorial biochemistry involves the combination of genes from different organisms, in a heterologous host to produce bioactive compounds by establishing novel enzyme-substrate combinations in vivo, which in turn could lead to the biosynthesis of novel natural products (Pollier et al., 2011a). Although BF567 is tentatively annotated as a homolog of CYP716A12, the GC elution time and EI pattern of the β-amyrin hydroxylation product of BF567 is different from erythrodiol (Fig. 4). We reasoned that if the two enzymes hydroxylate β-amyrin at two different carbon positions, it should be possible to combine the enzymes in the yeast strain TM3 and produce a combinatorial compound not produced by either of the
enzymes alone. Therefore, we generated strain TM30 from TM3 by overexpressing the plasmids pAG415[GAL1/AtATR1] and pAG423[GAL1/BF567-T2A-CYP716A12], where BF567 and CYP716A12 are stitched together into a self-processing polypeptide via the 2A oligopeptide (de Felipe et al., 2006) which is expressed from a single galactose inducible GAL1 promoter. The spent medium of strain TM30 cultured in the presence of MβCD was analyzed by GC-MS and compared to the GC chromatograms of spent medium from strains TM9, and TM17 overexpressing pAG423[GAL1/CYP716A12] and pAG415[GAL1/AtATR1] in TM3.

We observed a unique peak at 40.5 min in strain TM30 (Fig. 6a) but not TM9 (Fig. 6b) and TM17 (Fig. 6c), strongly supporting the fact that BF567 and CYP716A12 catalyze hydroxylations of two different carbons on β-amyrin. Additionally, the EI pattern of this peak suggested the presence of carboxyl and alcohol functions on β-amyrin, indicating the C-28 carboxylation by CYP716A12 and the C-16, C-21 or C-29 hydroxylation by BF567. Considering the close proximity of C-16 and C-28 on the β-amyrin molecule (Fig. 1b) and the tentative annotation of BF567 as a homolog of CYP716A12 we reasoned that BF567 might hydroxylate C-16 of β-amyrin. Therefore, we compared the GC chromatogram and EI pattern of an echinocystic acid standard (Fig. 6d) with that of the new peak at 40.5 min in TM30 (Fig. 6a). However, the GC elution time and fragmentation of the new peak at 40.5 min did not match that of echinocystic acid (3β,16α-dihydroxyolean-28-oic acid) ruling out the possibility of a C-16 α-hydroxylation by BF567. Due to the absence of authentic 3β,16β-dihydroxyolean-28-oic acid, 3β,21β-dihydroxyolean-28-oic acid and 3β,29α-dihydroxyolean-28-oic acid standards we could not confirm the identity of the combinatorial compound.
**Figure 6.** GC chromatograms corresponding to, (a) Extraction from spent medium of strain TM30, expressing tHMG1, GgbAS, AtATR1, CYP716A12 and BF567. (b) Extraction from spent medium of strain TM9, expressing tHMG1, GgbAS, AtATR1 and BF567. (c) Extraction from spent medium of strain TM17, expressing tHMG1, GgbAS, AtATR1 and CYP716A12. (d) An echinocystic acid standard. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively. (e) Mass spectra extracted from the peak indicated (*) at 40.5 min. Parts of the structure in blue depict the possible unknown hydroxylation position. (f) Mass spectra extracted from echinocystic acid standard. The X-axis of the mass spectra correspond to m/z values. (g) Oxidation of β-amyrrin by BF567 and CYP716A12.
Table 1. List of yeast strains generated and used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>S288c BY4742</td>
<td>MATα; his3Δ1; leu2Δ0; ura3Δ0; lys2Δ0</td>
</tr>
<tr>
<td>TM1</td>
<td>S288c BY4742; P_{erg7}::P_{MET3}-ERG7</td>
</tr>
<tr>
<td>TM2</td>
<td>TM1; pESC-URA[GAL10/tHMG1; GAL1/MtbAS] (19.03 mg/L β-amyrin)</td>
</tr>
<tr>
<td>TM3</td>
<td>TM1; pESC-URA[GAL10/tHMG1; GAL1/ GgbAS] (36.2 mg/L β-amyrin)</td>
</tr>
<tr>
<td>TM5</td>
<td>TM1; pESC-URA[GAL10/tHMG1]</td>
</tr>
<tr>
<td>TM7</td>
<td>TM3; pAG423[GAL1/BF567], pAG425[GAL1/AtATR1]</td>
</tr>
<tr>
<td>TM8</td>
<td>TM3; pAG423[GAL1/BF567], pAG305[GAL1/AtATR1]</td>
</tr>
<tr>
<td>TM9</td>
<td>TM3; pAG423[GAL1/BF567], pAG415[GAL1/AtATR1]</td>
</tr>
<tr>
<td>TM10</td>
<td>TM3; pAG423[GAL1/CYP716A12], pAG425[GAL1/AtATR1]</td>
</tr>
<tr>
<td>TM17</td>
<td>TM3; pAG423[GAL1/CYP716A12], pAG415[GAL1/AtATR1]</td>
</tr>
<tr>
<td>TM26</td>
<td>TM3; pAG423, pAG425[GAL1/AtATR1]</td>
</tr>
<tr>
<td>TM30</td>
<td>TM3; pAG423[GAL1/BF567-T2A-CYP716A12], pAG415[GAL1/AtATR1]</td>
</tr>
</tbody>
</table>

Discussion

Saikosaponins constitute the most abundant secondary metabolites produced by the genus *Bupleurum*, an integral plant in Asian traditional medicine. The biological activity of the genus *Bupleurum* is generally attributed to the saikosaponins, which are glycosylated oleanane- and ursane-type sapogenins. The saikosapogenins are very closely related in their structure and contain oxygenations on C-11, C-16, C-21, C-23, C-28, C-29 or C-30 (Fujioka et al., 2003; Ashour and Wink, 2011; Nakahara et al., 2011). However, not a single CytP450 involved in the oxygenation of any of these carbon positions has been described so far from *Bupleurum*. We identified BF567, a CytP450 that was tightly co-regulated with known triterpene saponin biosynthetic genes, through a cDNA-AFLP based transcript profiling of MeJA-treated *B. falcatum* roots. The BF567 was tentatively annotated as a homolog of the *M. truncatula* CYP716A12 and thought to oxygenate C-28 of the oleanane- and ursane-type sapogenin backbones. In this study, using a combinatorial approach in *S. cerevisiae* we revealed that BF567 is not a C-28 oxygenase
but, a novel cytochrome P450 that encodes a β-amyrin 16β/21β/29α-hydroxylase involved in the synthesis of saikosaponins in *Bupleurum*.

For the purpose of this study we generated a highly efficient β-amyrin producing *S. cerevisiae* strain TM3 by expressing an already characterized *bAS* from *G. glabra*. By employing the same strategy as (Kirby et al., 2008), but utilizing an efficient *bAS* we could generate a yeast strain capable of accumulating 36.2 mg/L β-amyrin in 72 h from synthetic medium. This is a considerable (6-fold) improvement over the 6 mg/L of β-amyrin produced by expressing an *A. annua bAS* (Kirby et al., 2008). Our lesser efficient strain TM2 expressing a *M. truncatula bAS* was also able to produce 3.2-fold more β-amyrin (19 mg/L) than (Kirby et al., 2008). However, we have not looked at the levels of the other intermediates in sterol biosynthesis of yeast, like squalene and ergosterol among others, that can provide vital clues for pathway engineering to further improve the amounts of β-amyrin produced by *S. cerevisiae*, which might be relevant to achieve our ultimate goal of creating an industrially compatible triterpene sapogenin producing yeast platform.

We subsequently utilized strain TM3 for the ectopic expression of BF567 and identified a hydroxylated product of β-amyrin, that was produced in considerably low amounts in the strain TM7 that was used for this analysis. Therefore, we tried to improve the efficiency of this reaction by manipulating the relative expression level of the redox partner *AtATRI* we employed, the isoform 1 of cytochrome P450 reductase from *A. thaliana*, whilst keeping the expression of BF567 constant. In this study we have shown for the first time in heterologous triterpene sapogenin (or plant secondary metabolite) biosynthesis that, the ratio of CPR to CytP450 plays a key role in the efficient functioning of CytP450s, and should be considered while engineering cytochrome P450 mediated reactions in non-native hosts. On gene expression level we have demonstrated the generally accepted ratio of CPR:CytP450 between 1:5 and 1:30 to also hold true for the expression of plant CytP450s in heterologous systems. Nonetheless, these ratios are yet to be confirmed on protein expression and enzyme activity levels.
We also addressed the concern of the terpene research community, which has reverberated around the fact that triterpene production may not be amenable to engineering efforts, like volatile sesquiterpenes and monoterpenes that can diffuse out of cells (Kirby et al., 2008), thereby enhancing productivity by disposal of toxic heterologous compounds from the cells internal environment. Here, we have described a novel approach for transporting triterpenes into the culture medium from within the yeast cells by the use of methylated cyclodextrins. We have proven using our engineered triterpene yeast strains that it is possible to release hydrophobic compounds like β-amyrin and triterpene sapogenins into the spent medium to, a) relieve the cells from exogenous compounds, which in turn increases production of the compound and b) facilitate downstream processing for product identification and purification. Although the approach has been demonstrated for triterpene sapogenins, it can be extended to the extraction of any hydrophobic compound from living cells without damaging the integrity of the cell, and theoretically can be extended to continuous culture systems for the production of valuable hydrophobic compounds, like triterpene sapogenins.

Lastly, in this study we have generated a combinatorial triterpene sapogenin biosynthesis product, using our new CytP450 from *B. falcatum*, BF567 and CYP716A12 from *M. truncatula*, in our yeast strain expressing a *G. glabra* β-amyrin synthase and a *A. thaliana* cytochrome P450 reductase. Although we did not characterize the exact hydroxylation position of BF567 using nuclear magnetic resonance (NMR), we decided to combine BF567 with CYP716A12 in an attempt to prove the different reaction specificities of the two enzymes. We observed a combinatorial product in strain TM30, the identity of which could not be confirmed with the use of an echinocystic acid standard, that harbors an α-hydroxyl-group on C-16 and a β-carboxyl-group on C-28 of β-amyrin. However, we could exclude a 16α-hydroxylation by BF567. We looked into literature for the other possible hydroxylation positions in *Bupleurum* species and narrowed the possibilities to a 16β, 21β or 29α (Ashour and Wink, 2011). We still speculate the possibility of a 16β hydroxylation based on the EI pattern of the combinatorial product and its longer retention time on the column. In case of a 16β hydroxylation, its close proximity with the 28β carboxyl group could stearically hinder the complete trimethylsilyl derivatization of...
the compound, resulting in only the 3β and 28β positions being derivatized. As a result, the completely trimethylsilylated echinocystic acid would have a lower vaporizing temperature than the dimethylsilylated 3β,16β-dihydroxyolean-28-oic acid, which would be reflected by a longer column retention or higher elution time of the latter. Additionally, the EI pattern could be interpreted to a biderivatized compound further supporting this possibility. However, this does not exclude the possibility of hydroxylation at C-21 or C-29 and therefore an NMR on the hydroxylated β-amyrin is essential to truly characterize BF567. Nonetheless, a combinatorial approach originally developed to generate novel bioactive products, could be exploited for the partial characterization of a novel plant β-amyrin 16β/21β/29α-hydroxylase.

Materials and Methods

Chemicals. β-amyrin, Erythrodiol and Echinocystic acid were purchased from Extrasynthese. Hexane, Pyridine and N-Methyl-N-(trimethylsilyl)trifluoroacetamide were purchased from Sigma-Aldrich. Methyl-β-cyclodextrin was purchased from CAVASOL. α-cyclodextrin, β-cyclodextrin, γ-cyclodextrin and random methyl-β-cyclodextrin (degree of substitution, DS~12) were purchased from Cyclolab.

Cultivation and elicitation of B. falcatum plants. 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 16 h/8 h light/dark regime, at 21°C. The pH was monitored daily and 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 methyl jasmonate (MeJA) (dissolved in ethanol (EtOH)) or an equivalent amount of EtOH as a control, by adding the EtOH or MeJA solution directly to the hydroponics medium. For transcript profiling, roots were harvested 0, 0.5, 1, 2, 4, 8 and 24 h after treatment, frozen in liquid nitrogen, and stored at -70°C. For each sample, 3 individual plants were pooled.
Transcript profiling. After appropriate sample preparation, cDNA-AFLP based transcript profiling with all 128 possible BrsYI+1/MseI+2 primer combinations was performed as described (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 was 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 and the average was 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 standard deviation. A coefficient of variation (CV) was obtained by dividing the standard deviation 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 in publicly available protein databases using a translated nucleotide query (blastx) or nucleotide database using a nucleotide query (blastn) was performed as described (Rischer et al., 2006).

Phylogenetic analysis. The protein sequences were retrieved from GenBank and aligned with ClustalW. The phylogenetic tree was generated in MEGA 4.0.1 software (Tamura et al., 2007), by the Neighbor-Joining method, and bootstrapping was performed with 1000 replicates. The evolutionary distances were computed using the Poisson correction method, and all positions containing gaps and missing data were eliminated from the data set (complete deletion option).

Full length cloning of BF567. Using the primers P11 + P12, the full-length coding sequence of BF567 was screened for in a B. falcatum Uncut Nanoquantity cDNA library (custom-made by Invitrogen, Carlsbad, CA, USA) as reported (Pollier et al., 2011b). The
full-length open reading frame (FL-ORF) of BF567 was PCR amplified for Gateway\textsuperscript{TM} cloning using the primer pair P21 + P22 (Table 2).

**Generation of strain TMI.** The native ERG7 promoter of a wild type S288c BY4742 strain was replaced with a MET3 promoter using a PCR based gene targeting approach (see addendum 2) as described previously for ERG9 in S. cerevisiae CEN.PK 113-7D (Asadollahi et al., 2008). First, plasmid pIP007 was created by amplifying the MET3 promoter from the genomic DNA of BY4742 using primers P1 + P2 (Table 2) followed by ligation of the digested PCR fragment into the SpeI and SacII sites of plasmid pUG6, which harbors the KanMX selection cassette. Next, two overlapping PCR fragments a and b, were amplified from pIP007, using the primer pairs P3 + P4 and P5 + P6 (Table 2), respectively. Further, fragment c, corresponding to the genomic region 500 bp upstream of ERG7 promoter and fragment d, corresponding to 500 bp of the start of ERG7 coding sequence (CDS) were amplified from the genome of BY4742 using the primer pairs P7 + P8 and P9 + P10 (Table 2), respectively. The gel purified PCR fragments were then combined in pairs, a+c and b+d to generate two fused overlapping fragments e and f, using the primer pairs P7 + P4 and P5 + P10, respectively. A S. cerevisiae BY4742 wild type strain was transformed with the gel purified fragments e and f, using the lithium acetate mediated transformation method, and transformants with the ERG7 promoter replaced by the MET3 promoter, were selected on YPD medium (Yeast extract Peptone Dextrose, Clontech) supplemented with 200 \(\mu\)g/mL G-418 Disulfate (Geneticin, Duchefa). An overview of all the yeast strains generated from TMI is given in Table 1.

**Generation of plasmid vectors.** The tHMG1 gene (coding amino acids 532 to 1054 of HMG1p (Polakowski et al., 1998)) was amplified from the chromosomal HMG1 gene using primers containing NotI and PacI restriction sites P13 + P14, and inserted into the MCS1 of pESC-URA (Agilent Technologies) to generate pESC-URA[GAL10/tHMG1]. The bAS was inserted into the MCS2 of the generated plasmid to create pESC-URA[GAL10/tHMG1; GAL1/bAS]. The bAS of M. truncatula and G. glabra were amplified using XhoI and KpnI containing primers, P15 + P16 and P17 + P18 (Table 2),
respectively. All constructs were first inserted into Gateway™ vector pDONR221 to facilitate sequence verification prior to insertion into their respective expression vectors.

The FL-ORF of *A. thaliana* CPR (*AtATR1, At4g24520*), *B. falcatum* BF567 and *M. truncatula* CYP716A12 were PCR amplified using primers P19 + P20, P21 + P22, and P23 + P24 (Table 2), respectively and cloned into the entry vector pDONR221 by Gateway™ recombination. The *AtATR1* was further recombined into the integrating, low-copy number and high-copy number yeast expression vectors pAG305GAL-ccdB (Addgene plasmid 14137), pAG415GAL-ccdB (Addgene plasmid 14145) and pAG425GAL-ccdB (Addgene plasmid 14153), respectively and all having the galactose inducible GAL1 promoter and LEU2 auxotrophic marker. The entry clones of BF567 and CYP716A12 were recombined into the high-copy number expression vector pAG423GAL-ccdB (Addgene plasmid 14149) with the GAL1 promoter and HIS3 auxotrophic marker.

The self-processing polyprotein of BF567 and CYP716A12, was generated by amplifying the FL-ORF of BF567 without a stop codon and having a 3’- overhang of the partial T2A sequence using the primer pair P21 + P25. The FL-ORF of CYP716A12 was amplified with a 5’- overhang of the partial T2A sequence using the primers P26 + P24 such that there was an overlap of 7 bp between the two amplified sequences. Since the primers P25 and P26 contain an Uracil each, the BF567 and CYP716A12 were PCR amplified using the Pfu Turbo Cx polymerase (Stratagene). The purified gel fragments were used for Uracil-Specific Excision or USER™ Cloning (New England Biolabs) to generate two fragments with complementary sticky ends which were ligated *in vitro* using the T4 DNA ligase (Invitrogen). The ligated DNA product was once again gel purified and used as template for amplification with the primers P21 + P24. This amplicon was Gateway™ recombined into pDONR221, sequence verified and further recombined into pAG423GAL-ccdB to generate pAG423[GAL1/ BF567-T2A-CYP716A12].

**Culturing of yeast cells for metabolite analysis.** Precultures were grown at 30°C, 250 rpm for 18-20 h in 5 ml synthetic defined (SD) medium containing glucose (Duchefa, Clontech) and appropriate drop out (DO) supplements (Duchefa, Clontech).
The precultures were washed twice with water to remove glucose prior to inoculating 15 ml SD Gal/Raf medium containing galactose and raffinose as carbon source (Duchefa, Clontech) and appropriate DO supplements to a starting optical density of 0.25. The cultures were incubated as before for 24 h before ERG7 repression by adding 2.25 ml of 10 mM methionine to a final concentration of 1.5 mM. The cultures were incubated further for 48 h.

MβCD treatment was carried out by adding 300 µl or 1500 µl of, a 250 mM MβCD stock to a final concentration of 5 mM or 25 mM, respectively. To determine the specificity of CD variants, αCD, βCD, γCD, RMβCD and MβCD treatment was carried out by adding 750 µl of a 100 mM stock to a final concentration of 5 mM.

**Metabolite extraction and GC-MS analysis.** An organic extract of the yeast cells or spent medium was prepared for identification and quantification using GC-MS. For extraction from yeast cells, a 1 ml culture was pelleted by centrifugation at 10,600 x g for 1 min, resuspended in an ethanolic potassium hydroxide solution (250 µl 40% KOH, 250 µl 50% ethanol), boiled at 95°C for 10 min, and allowed to cool to room temperature. For extraction from the spent medium, only the supernatant post centrifugation of a 1 ml culture was used. On both the lyzed cell pellet and spent medium, a first extraction was performed with 500 µl hexane by vortexing at high speed for 1 min. The phases were separated by centrifugation at 10,600 x g for 1 min, and the organic phase was transferred to a fresh tube. The extraction was repeated two more times on the aqueous phase with 500 µl hexane each and the organic phases pooled (~1400 µl) and vaporized to dryness. A trimethylsilyl derivatization was performed on the dry material, by adding 20 µl pyridine and 100 µl N-methyl-N-(trimethylsilyl)trifluoroacetamide directly to the tube, vortexing and incubating at room temperature for 15 min. The derivatized material was transferred to GC vials and subjected to GC-MS (GC model 6890, MS model 5973, Agilent). A 1 µl aliquot was injected (splitless mode) into a VF-5ms capillary column (Varian CP9013, Agilent) and operated at a constant helium flow of 1 ml/min. The injector temperature was set to 280°C and the oven temperature was held at 80°C for 1 min post injection, ramped to 280°C at 20°C/min, held at 280°C for 45 min, ramped to 320°C at 20°C/min, held at 320°C for 1 min, and finally cooled down to 80°C at 50°C/min at the end of the
run. The MS transfer line was set to 250°C, the MS ion source to 230°C and the quadrupole to 150°C, throughout. For identification of metabolites a full mass spectra was generated by scanning the m/z range of 60-800 with a solvent delay of 7.8 min. For quantification of β-amyrin a select ion mode was operated detecting the ions 498, 483, 393, 279, 218, 203 and 189. The areas of the peaks were calculated using the default settings of the AMDIS software (version 2.6, NIST, USA).

Table 2. Primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generation of strain TM1</strong></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>GGactagtCCCTGGTGATAAGGTAGGGGTCACAGG</td>
</tr>
<tr>
<td>P2</td>
<td>TCCccgcgGGAATACCACCCGTCAGGACAGCATG</td>
</tr>
<tr>
<td>P3</td>
<td>GATCCGGGGAATTGCACTAGCAGTGGACGAGCAGACC</td>
</tr>
<tr>
<td>P4</td>
<td>CCATGAGTGACGACTGAATCCCG</td>
</tr>
<tr>
<td>P5</td>
<td>CTATCGATTTGTATGGAAGGCC</td>
</tr>
<tr>
<td>P6</td>
<td>CGATGTGTGTCAGAATAAAATCTCTGATGTAAATTATATTTTATCTGATAT TATATAC</td>
</tr>
<tr>
<td>P7</td>
<td>TAATACCCCTTGGAGAATGTCTTC</td>
</tr>
<tr>
<td>P8</td>
<td>CATGGCAATTTCGGGATCGCGGAAAAACATCGACTTTATCGAG</td>
</tr>
<tr>
<td>P9</td>
<td>ATGACAGAATTATATTCGACACAAATCG</td>
</tr>
<tr>
<td>P10</td>
<td>CCTTCCCATTTTATAAAGTTTATAGTGC</td>
</tr>
<tr>
<td><strong>Generation of plasmid vectors</strong></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>CCTCCCTATACATCTCCATCC</td>
</tr>
<tr>
<td>P12</td>
<td>TAGGGTCTCTCTCTCTCCTT</td>
</tr>
<tr>
<td>P13</td>
<td>TGGAAGccgcgcgTTTCTAGTACTATGGAACATTGGAAGGAAACTG</td>
</tr>
<tr>
<td>P14</td>
<td>GCTGCTGCTCTTTTTCTCAGTATAGTGGAAGGAAACTG</td>
</tr>
<tr>
<td>P15</td>
<td>GCAGGCTCATATATATGCTAGTGAACACGTGACATGTCGCTGGTGTGTT</td>
</tr>
<tr>
<td>P16</td>
<td>GCTGCTGCTCATATATCTAGTGAACACGTGACATGTCGCTGGTGTGTT</td>
</tr>
<tr>
<td>P17</td>
<td>GCAGGCTCATATATCTAGTGAACACGTGACATGTCGCTGGTGTGTT</td>
</tr>
<tr>
<td>P18</td>
<td>GCTGCTGCTCATATATCTAGTGAACACGTGACATGTCGCTGGTGTGTT</td>
</tr>
<tr>
<td>P19</td>
<td>GGGGACAAGTTTGTGTAACAAAAAGCAGGCTTAATGACTTCTTGTTGATGC</td>
</tr>
<tr>
<td>P20</td>
<td>GGGGACAAGTTTGTGTAACAAAAAGCAGGCTTAATGACTTCTTGTTGATGC</td>
</tr>
<tr>
<td>P21</td>
<td>GGGGACAAGTTTGTGTAACAAAAAGCAGGCTTAATGACTTCTTGTTGATGC</td>
</tr>
<tr>
<td>P22</td>
<td>GGGGACAAGTTTGTGTAACAAAAAGCAGGCTTAATGACTTCTTGTTGATGC</td>
</tr>
<tr>
<td>P23</td>
<td>GGGGACAAGTTTGTGTAACAAAAAGCAGGCTTAATGACTTCTTGTTGATGC</td>
</tr>
<tr>
<td>P24</td>
<td>GGGGACCACTTTGTACAAGGAAAGGCTGGTACAGGCTATCCCAATGGAGGCTATATATTCTATCTC TCCC</td>
</tr>
<tr>
<td>P25</td>
<td>GGGGACCACTTTGTACAAGGAAAGGCTGGTACAGGCTATCCCAATGGAGGCTATATATTCTATCTC TCCC</td>
</tr>
<tr>
<td>P26</td>
<td>ACCGCAUGTTAGAGACACTTCCCTTCTGCCCCTAGTAGGATTTGTGGGGAT</td>
</tr>
<tr>
<td>P27</td>
<td>ATCGGGUGACGTCAGGGAGAATCTGCGGCCCAATGGAGCCTAAATATCTATC</td>
</tr>
</tbody>
</table>

The sequences in lower case represent the restriction recognition site used for restriction enzyme mediated cloning. The underlined sequence corresponds to T2A partial sequences.
Acknowledgements

We thank Wilson Ardiles-Diaz for sequencing the cDNA-AFLP tags, Miguel González-Guzmán and Lander Ingelbrecht for excellent technical advice and assistance, and Prof. Jan Van Bocxlaer (Universiteit Gent) for the fruitful discussions on GC-MS profiling. 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. Lorena Almagro was supported by a short-term scientific mission grant from the COST action FA1006-Plant Engine.

References


Unravelling triterpenoid saponin biosynthesis in *Maesa lanceolata* using *Saccharomyces cerevisiae*
Publication status:

**Tessa Moses**, Jacob Pollier, Lorena Almagro, María A. Pedreño, Danny Geelen, Johan M. Thevelein and Alain Goossens.

**Unravelling triterpenoid saponin biosynthesis in *Maesa lanceolata* using *Saccharomyces cerevisiae***

Manuscript in preparation

Author contribution:

Generation of yeast strains, culturing and metabolite extractions, GC-MS profiling, data analysis and writing of the manuscript.
Abstract

Maesasaponins are oleanane-type saponins with diverse biological activities produced by the indigenous African plant *Maesa lanceolata*. The maesasapogenins harbor modifications on positions C-16, C-21, C-22 and C-28 of the oleanane backbone, the latter being involved in hemiacetal and ester bridge formation. In this study we have characterized three enzymes, a β-aminin synthase which is the first committed enzyme of the oleanane-type triterpene saponin biosynthesis, a C-28 oxidase involved in bridge formations and a putative C-21 hydroxylase, involved in the biosynthesis of maesasaponins.
Introduction

Triterpenoid saponins are a large class of diverse plant secondary metabolites produced by many plant species (Sparg et al., 2004; Vincken et al., 2007). They belong to the class of isoprenoid natural products and are composed of 6 isoprene units resulting in a total of 30 carbon atoms, that are arranged in 4 or 5 ring structures to form triterpene sapogenin backbones like β-amyrin, α-amyrin, lupeol and dammarenediol. These diverse sapogenin backbones are derived from a common precursor 2,3-oxidosqualene, an intermediate of membrane sterol biosynthesis in eukaryotes, via specific plant oxidosqualene cyclase (OSC) mediated cyclizations. The triterpenoid saponin structural diversity is further aggravated by multiple cytochrome P450 monoxygenase (CytP450) mediated modifications of the sapogenin backbone, which generates “sapogenins” containing reactive functional groups (hydroxyl-, carboxyl-, oxo-, etc). The subsequent UDP-glycosyltransferase mediated glycosylations of these functional groups results in the formation of biologically active saponin moieties (Osbourn et al., 2011).

Maesa lanceolata, a member of the Myrsinaceae family, is a shrub or small tree indigenous to Africa. African traditional healers use extracts and/or parts of the plant for the treatment of a wide range of diseases including infectious hepatitis, bacillary dysentery, impetigo, ozena, dermatoses and neuropathies. Methanol extracts of M. lanceolata leaves are rich in maesasaponins and have been shown to possess virucidal, molluscicidal, fungistatic and antimutagenic activities (Sindambiwe et al., 1998). The maesasaponins identified so far are derived from an oleanane skeleton via modifications of the β-amyrin backbone, resulting in a characteristic C-13,28 hemiacetal or ester bridge and oxidations on C-16, C-21 and C-22 (Fig. 1) (Foubert et al., 2010; Manguro et al., 2011). The hemiacetal or ester bridge between C-13 and C-28 is thought to occur through the reaction between a C-13 hydroxyl and C-28 aldehyde or carboxyl group, respectively (Vincken et al., 2007). The presence of these diverse oleanane maesasaponins suggests the presence of a β-amyrin specific OSC (or β-amyrin synthase) along with specific CytP450s catalyzing oxygenations at C-16, C-21, C-22 and C-28 in M. lanceolata. However, to date not a single triterpene saponin biosynthetic gene has been identified from Maesa.
In the past, several plant β-amyrin synthases have been functionally characterized in yeast, including from *Artemisia annua* (Kirby et al., 2008), *Medicago truncatula* (Suzuki et al., 2002), *Glycyrrhiza glabra* (Hayashi et al., 2001) and *Saponaria vaccaria* (Meesapyodsuk et al., 2007). Additionally, cytochrome P450s catalyzing the C-28 oxidation of triterpene backbones have been identified from *M. truncatula* (Corelli et al., 2011), *Vitis vinifera* (Fukushima et al., 2011) and *Catharanthus roseus* (Huang et al., 2012). Furthermore, we identified a C-16β/21β/29α hydroxylase from *Bupleurum falcatum* (see chapter 4). These CytP450s have also been functionally characterized using yeast cells.

The current study describes the identification of putative maesasaponin biosynthetic genes through cDNA-AFLP based transcript profiling, the application of degenerate PCR for gene cloning and the functional characterization of a β-amyrin synthase and two CytP450s involved in the oxidation of C-28 (ML257) and most probably, hydroxylation of C-21β (ML593) of the oleanane backbone using an engineered *Saccharomyces cerevisiae* strain.
Results

Transcript profiling of MeJA treated *M. lanceolata* shoots reveals ML257 and ML593.

To identify new saponin biosynthetic genes, we performed a transcript profiling on methyl jasmonate (MeJA) treated *M. lanceolata* axenic shoot cultures. Using the complete set of 128 BstYI+1/MseI+2 primer combinations, a genome-wide cDNA-AFLP transcript profiling analysis (Vuylsteke et al., 2007) was carried out to monitor the expression of a total of 13,558 transcript tags over time. In total, 733 MeJA-responsive transcript tags were isolated (hereafter referred to as ML tags). Direct sequencing of the reamplified ML tags gave good quality sequences for 545 (74.4%) of the fragments. To the remaining 188 tags (25.6%), no unique sequence could be attributed unambiguously, indicating that they might not represent unique gene tags and hence, these were not considered for further analysis. A BLAST search with the nucleotide sequences of the 545 unique cDNA-AFLP tags led to the annotation of 312 (57.2%) of the ML tags.

Average linkage hierarchical clustering analysis of the expression profiles of the ML tags showed that, upon MeJA treatment the genes are either transcriptionally activated (Clusters I-V, Fig. 2a) or transcriptionally repressed (Cluster VI, Fig. 2a). The activated gene tags can be divided into 5 subclusters, based on their MeJA response time. A first cluster consists of genes with a response immediately after (≤ 2 h) MeJA treatment (cluster V). In a second cluster (Cluster IV), genes reach maximum expression 4-8 h after MeJA treatment. In a third cluster (Cluster II), genes are activated 2 h after MeJA treatment and their expression remains high thereafter. In a fourth and fifth cluster (Clusters III and I, respectively), genes are activated 24 h after MeJA treatment. In cluster III a gene tag that reached maximum levels of expression 24-48 h post-elicitation and corresponding to squalene epoxidase (*SQE*) can be found (Fig. 2b).

The gene tags ML257 and ML593, corresponding to CytP450s are tightly co-regulated with this gene (Fig. 2b). The gene tag ML257 shows homology to the *M. truncatula CYP716A12* that was shown to oxidize β-amyrin in a sequential three-step oxidation on
Unravelling *Maesa lanceolata* saponin biosynthesis

Figure 2. Transcript profiling of jasmonate elicited *M. lanceolata* plants. (a) General view on the average linkage hierarchical cluster and tree of all 545 unique ML tags. (b) Subcluster of the *M. lanceolata* transcriptome, comprising all tags corresponding to genes reported to be involved in terpene biosynthesis, or with high sequence similarity to such genes, and all gene tags corresponding to CytP450s. Treatments and time points (in h) are indicated on top. Blue and yellow boxes reflect transcriptional activation and repression relative to the average expression level, respectively. Gray boxes correspond to missing time points. (c) Phylogenetic relationship between CytP450 proteins involved in triterpene saponin biosynthesis. Arrowheads indicate the CytP450s functionally defined in this study. Scale bar indicates the number of amino acid substitutions per site.

C-28 to yield oleanolic acid through erythrodiol (Carelli et al., 2011; Fukushima et al., 2011), and the gene tag ML593 shows homology to the *A. thaliana* steroid 22α-hydroxylase gene encoding a CytP450 enzyme that catalyzes the oxidation of sterols on the C-22 position (Fujita et al., 2006). The full-length open reading frame of *ML257* and *ML593* corresponding to the respective gene tags was picked up from a *M. lanceolata* Uncut Nanoquantity cDNA library (Pollier et al., 2011). Phylogenetic analysis of all the CytP450s characterized to be involved in triterpene saponin biosynthesis (Fig. 2c) showed that *ML257* groups together with the *CYP716* family of CytP450s and *ML593*
formed an isolated branch. The ML257 encodes a polypeptide of 485 amino acids showing 76% identity with CYP716AL1 from Catharanthus roseus (AEX07773), 73% identify with CYP716A15 from Vitis vinifera (BAJ84106) and 69% identity with CYP716A12 from Medicago truncatula (ABC59076).

**Cloning and in vivo characterization of *M. lanceolata* β-amyrin synthase.**

Since no cDNA-AFLP tag corresponding to β-amyrin synthase (bAS) was encountered in the cDNA-AFLP transcript profiling analysis, we amplified a gene tag corresponding to bAS from a *M. lanceolata* cDNA template using degenerate oligonucleotide primers. The obtained sequence allowed us to generate a gene tag that was used to screen for the full-length open reading frame of *M. lanceolata* β-amyrin synthase (MlbAS) in a cDNA library (Pollier et al., 2011). The MlbAS encodes a polypeptide of 760 amino acids showing 75% identity with GgbAS from Glycyrrhiza glabra (BAA89815) and 74% identity with MtbAS from Medicago truncatula (AES86318).

Using our yeast strain TM1 (Table 1) with the modified ergosterol biosynthesis (see chapter 4), we generated a strain TM4 harboring the plasmid pESC-URA[GAL10/tHMG1; GAL1/MlbAS]. We cultured TM4 and the control strain TM5 (Table 1) in galactose medium for 24 h, prior to repression of ERG7 with 1.5 mM methionine for 48 h. The cell pellets of 72 h cultures were used for metabolite extraction and analyzed by GC-MS. The GC chromatogram of TM4 showed a single peak at 27.2 min (Fig. 3a) with electron ionization (EI) pattern corresponding to a β-amyrin standard (Fig. 3c). This peak was absent in the control strain TM5 (Fig. 3b). However, the β-amyrin peak in TM4 could only be confirmed on the mass spectrometer (MS) with select ion monitoring, indicating the low concentration of β-amyrin produced by the strain, which in turn suggests the poor 2,3-oxidosqualene cyclization efficiency of MlbAS, to generate β-amyrin. Therefore, we decided to continue further experiments with the high β-amyrin producing strain TM3 (Table 1, chapter 4).
Figure 3. GC chromatograms corresponding to, (a) Extraction from cells of strain TM4, expressing \textit{tHMG1} and \textit{MlbAS}. (b) Extraction from cells of control strain TM5, expressing \textit{tHMG1}. (c) A \(\beta\)-amyrin standard. Right panels show the mass spectra extracted from the indicated (*) peak, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively.

\textit{In vivo} characterization of \textit{ML257} as a C-28 oxidase.

We generated strain \textit{TM20} by super-transforming the \(\beta\)-amyrin producing strain \textit{TM3} (Table 1) with the plasmids pAG415[\textit{GAL1/AtATR1}] and pAG423[\textit{GAL1/ML257}], where the cytochrome P450 reductase (CPR, \textit{AtATR1}) was expressed from a low-copy number plasmid (pAG415) and the CytP450 from a high-copy number plasmid (pAG423), to maintain a CPR:CytP450 ratio between 1:5 and 1:30 (see chapter 4). In parallel, we generated a control strain \textit{TM27} not expressing any CytP450, but harboring the plasmids
Unravelling *Maesa lanceolata* saponin biosynthesis

pAG415[\(\text{GAL1/AtATRI}\)] and pAG423GAL-ccdB in the strain TM3. The strains were cultured in the presence of methyl-\(\beta\)-cyclodextrin (MbCD), to facilitate extraction of hydrophobic metabolites from the yeast cells (see chapter 4), and the spent medium was analyzed for the presence of \(\beta\)-amyrin and hydroxylated \(\beta\)-amyrin by GC-MS.

Strain TM20 expressing the CytP450 ML257, annotated as a homolog of the *M. truncatula* CYP716A12, showed three new peaks eluting at 32.5 min, 35.4 min and 36.6 min in the extracts from the spent medium (Fig. 4a). The CYP716A12 (GenBank accession number FN995113; (Carelli et al., 2011)) is a C-28 oxidase that catalyzes the conversion of \(\beta\)-amyrin to oleanolic acid in a three-step reaction via the intermediates erythrodiol and most likely an oleanolic aldehyde (Carelli et al., 2011). Therefore, we compared the GC chromatograms of TM20 with that obtained from the spent medium of strain TM17 (Table 1) expressing the *M. truncatula* CYP716A12 and the control strain TM27 (Fig. 4c). Alongside we also determined the elution time and EI pattern of erythrodiol (Fig. 4d) and oleanolic acid (Fig. 4e) standards. We observed the same three peaks eluting at 32.5 min, 35.4 min and 36.6 min in TM17 (Fig. 4b), as we observed in TM20 and could confirm the identity of these peaks as erythrodiol (32.5 min) and oleanolic acid (35.4 min). Although the EI pattern of the peak eluting at 36.6 min could be interpreted as oleanolic aldehyde in both strains TM20 and TM17, we could not confirm its elution time due to the lack of an authentic standard.

Nonetheless, our data confirms ML257 to be a C-28 oxidase that catalyzes the three-step oxidation of \(\beta\)-amyrin to oleanolic acid. Additionally, we observed differences in the efficiency of the three step oxidation of \(\beta\)-amyrin to oleanolic acid between strains TM20 and TM17. The *M. lanceolata* ML257 in strain TM20 exhibited a lower efficiency to produce oleanolic acid than CYP716A12. When comparing the amounts of \(\beta\)-amyrin, erythrodiol, oleanolic aldehyde and oleanolic acid between the two strains (Fig. 5), TM20 was found to accumulate 3-fold more erythrodiol than TM17, while the relative amounts of \(\beta\)-amyrin and oleanolic aldehyde remained comparable between the two strains. Contrary to the substrate and the intermediates, the amount of oleanolic acid was 2-fold lesser in strain TM20 compared to TM17. The high efficiency of CYP716A12 to oxidize \(\beta\)-amyrin to oleanolic acid has been reported earlier for both *in vitro* (Carelli et al., 2011)
and in vivo (Fukushima et al., 2011) activity of the CytP450, where either no or very low amounts of erythrodiol was detected, respectively.

**Figure 4.** GC chromatograms (first column) corresponding to, (a) Extraction from spent medium of strain TM20, expressing tHMG1, GgbAS, AtATRI and ML257. (b) Extraction from spent medium of strain TM17, expressing tHMG1, GgbAS, AtATRI and CYP716A12. (c) Extraction from spent medium of control strain TM27, expressing tHMG1, GgbAS and AtATRI. (d) An erythrodiol standard. (e) An oleanolic acid standard. Columns two to four show the mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively.
Figure 5. Relative amounts of β-amyrin, erythrodiol, oleanolic aldehyde and oleanolic acid produced by the strains TM20 and TM17, expressing ML257 and CYP716A12, respectively.

In vivo characterization of ML593.

To characterize the second putative CytP450, ML593 from our transcript profiling, we generated strain TM21 from the β-amyrin producing strain TM3 (Table 1), by super-transforming with the plasmids pAG415[GAL1/AtATR1] and pAG423[GAL1/ML593]. The strains TM21 and TM27 (Table 1) were cultured in the presence of MβCD and the spent medium analyzed by GC-MS. We observed a new peak eluting at 31.8 min corresponding to a hydroxylated β-amyrin in strain TM21 (Fig. 6a), but not in the control strain TM27 (Fig. 6c). The EI pattern of this peak (Fig. 6a) corresponded to a hydroxylation on the D or E ring of the oleanane structure and was similar to that observed with strain TM9 (Table 1) expressing BF567 (Fig. 6b). The strong similarity between the elution time and EI pattern of the hydroxylated β-amyrin in strain TM21 and TM9 further supports this assumption. We also observed that strain TM21 produced 8-fold more hydroxy β-amyrin than strain TM9, highlighting the better efficiency of ML593 for hydroxylating β-amyrin as compared to BF567.
Figure 6. GC chromatograms corresponding to, (a) Extraction from spent medium of strain TM21, expressing tHMG1, GgbAS, AtATRI and ML593. Arrow heads indicate the positions that could be hydroxylated by ML593 and are common with predicted positions of BF567. (b) Extraction from spent medium of strain TM9, expressing tHMG1, GgbAS, AtATRI and BF567. (c) Extraction from spent medium of control strain TM27, expressing tHMG1, GgbAS and AtATRI. Right panel shows mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively.

To confirm the identity of ML593 as a functional homolog of BF567, we generated strain TM31, similar to strain TM30 (Table 1) (see chapter 4). Strain TM31 was created by supertransforming TM3 with the plasmids pAG415[GAL1/AtATRI] and pAG423[GAL1/ML593-T2A-CYP716A12], where ML593 and CYP716A12 form a self-processing polyprotein stitched together by the 2A oligopeptide (de Felipe et al., 2006).
and is expressed from a single \textit{GAL1} promoter. The strains \textit{TM31}, \textit{TM17} and \textit{TM21} (Table 1) were cultured in the presence of MβCD and the spent medium was analyzed using GC-MS. Similar to strain \textit{TM30} (Fig. 7d), we observed a new peak eluting at 40.5 min in strain \textit{TM31} (Fig. 7a), but not the strains \textit{TM17} (Fig. 7b) and \textit{TM21} (Fig. 7c) expressing only the CytP450 \textit{CYP716A12} and \textit{ML593}, respectively. From the EI pattern of this peak we could confirm its identity as being the same as in strain \textit{TM30}. \textit{Maesa} saponins with C-16α and/or C-21β hydroxylations on the rings D and E of β-amyrin have been reported so far. Similarly, the saikosaponins from \textit{Bupleurum} have C-16β, C-21β and/or C-29 hydroxylations on the D and E rings of β-amyrin. The oxidation positions known from literature together with the lack of homology, of the peak at 40.5 min in strains \textit{TM30} and \textit{TM31}, with echinocystic acid suggests a β-hydroxylation on C-21 by \textit{BF567} and \textit{ML593}, the only remaining position commonly hydroxylated between \textit{Bupleurum} and \textit{Maesa} species.

\begin{table}[h]
\centering
\caption{List of yeast strains generated and used in this study.}
\begin{tabular}{ll}
\hline
Name & Construct \\
\hline
\textit{S288c BY4742} & \textit{MATa; his3Δ1; leu2Δ0; ura3Δ0; lys2Δ0} \\
\textit{TM1} & \textit{S288c BY4742; P}_{\textit{erg7}}::P_{\textit{MET3}}-\textit{ERG7} \\
\textit{TM3} & \textit{TM1; pESC-URA[GAL10/tHMG1; GAL1/GgbAS]} (36.2 mg/L \textit{β}-amyrin) \\
\textit{TM4} & \textit{TM1; pESC-URA[GAL10/tHMG1; GAL1/MlbAS]} \\
\textit{TM5} & \textit{TM1; pESC-URA[GAL10/tHMG1]} \\
\textit{TM9} & \textit{TM3; pAG423[GAL1/BF567, pAG415[GAL1/AtATR1]} \\
\textit{TM17} & \textit{TM3; pAG423[GAL1/CYP716A12, pAG415[GAL1/AtATR1]} \\
\textit{TM20} & \textit{TM3; pAG423[GAL1/ML257, pAG415[GAL1/AtATR1]} \\
\textit{TM21} & \textit{TM3; pAG423[GAL1/ML593, pAG415[GAL1/AtATR1]} \\
\textit{TM27} & \textit{TM3; pAG423, pAG415[GAL1/AtATR1]} \\
\textit{TM30} & \textit{TM3; pAG423[GAL1/BF567-T2A-CYP716A12, pAG415[GAL1/AtATR1]} \\
\textit{TM31} & \textit{TM3; pAG423[GAL1/ML593-T2A-CYP716A12, pAG415[GAL1/AtATR1]} \\
\hline
\end{tabular}
\end{table}
Figure 7. GC chromatograms corresponding to, (a) Extraction from spent medium of strain TM31, expressing tHMG1, GgbAS, AtATR1, CYP716A12 and ML593. (b) Extraction from spent medium of strain TM17, expressing tHMG1, GgbAS, AtATR1 and CYP716A12. (c) Extraction from spent medium of strain TM21, expressing tHMG1, GgbAS, AtATR1 and ML593. (d) Extraction from spent medium of strain TM30, expressing tHMG1, GgbAS, AtATR1, CYP716A12 and BF567. (e) An echinocystic acid standard. Right panel shows the mass spectra of indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively. Parts of structure highlighted in blue indicate probable hydroxylation positions.
Discussion

The maesasaponins constitute a suite of oleanane derived triterpenoid saponins distinguishable by their characteristic C-13,28 hemiacetal or ester bridges, which result from the reaction between a C-13 hydroxyl and C-28 aldehyde or carboxyl group, respectively. Additional oxidations also occur on C-16, C-21 and C-22 of the oleanane backbone. However, to date not a single enzyme involved in either, the formation of or, modification of the oleanane backbone had been elucidated from *M. lanceolata*. Here, we identified a β-amyrin synthase, the first committed enzyme for the production of oleanane-type triterpene saponins and two CytP450 monooxygenases (*ML257* and *ML593*), involved in the biosynthesis of maesasaponins.

Through a degenerate PCR approach we cloned a β-amyrin synthase from *M. lanceolata* and functionally characterized it using a sterol modified *S. cerevisiae* strain. However, only very low amounts of β-amyrin was produced, which we tentatively attribute to a bad ‘exogenous functionality’ of the *MlbAS* gene. For instance, yeast expressing the β-amyrin synthase from *Artemisia annua* has been reported to produce 6 mg/L of β-amyrin (Kirby et al., 2008), while we have reported a 36 mg/L and a 19 mg/L β-amyrin producing strain by expressing the *bAS* from *Glycyrrhiza glabra* or *Medicago truncatula*, respectively, even though all these yeast strains harbored the same modifications in their ergosterol biosynthesis (see chapter 4). This variability in the exogenous activity of *bAS* in yeast, might not necessarily be reflective of their *in planta* enzyme efficiency. The inconsistencies in enzyme activity can be attributed to factors including the lack of upstream elements for the optimal expression of plant genes in yeast, unoptimized codon usage, protein misfolding, protein instability, etc.

Therefore, for the characterization of the CytP450s we employed a high β-amyrin producing yeast strain. We identified *ML257* as a homolog of the *M. truncatula* C-28 oxidase (*CYP716A12*) and could confirm its function in the oxidation of β-amyrin to oleanolic acid via the intermediates erythrodiol and oleanolic aldehyde. Although the role of the intermediate, oleanolic aldehyde, in this reaction is yet to be generally accepted, we detected its presence in our yeast strains expressing either the *ML257* or the
Unravelling *Maesa lanceolata* saponin biosynthesis

CYP716A12. Additionally, in our yeast strain the ML257 favored the C-28 hydroxylation of β-amyrin to erythrodiol over its subsequent oxidation to oleanolic acid. The relative higher amounts of erythrodiol in ML257 strain, could be reflective of the presence of non-bridged C-28 hydroxylated saponins in *M. lanceolata* (Manguro et al., 2011), a branch of saponins never detected in *M. truncatula* from which the CYP716A12 originated. The further oxidation of erythrodiol to oleanolic aldehyde and oleanolic acid are essential for their involvement in the hemiacetal and ester bridge formation with C-13, in *M. lanceolata*.

In our genome-wide transcript profiling we also identified a second CytP450 (ML593) that was annotated as a homolog of the 22α-hydroxylase of sterols (CYP90B1), but tightly co-regulated with other triterpene saponin biosynthetic genes. Using our yeast strain we could functionally characterize this gene to catalyze the same hydroxylation of β-amyrin (or triterpene sapogenins) as BF567 (characterized in chapter 4) and was found to be its functional homolog. Although the BF567 and ML593 belong to different CytP450 families (CYP716A and CYP90B, respectively), in our yeast strain they demonstrate the same catalytic activity. To the best of our knowledge this is a first report demonstrating the functional homology of CytP450s belonging to two different families, involved in triterpene saponin biosynthesis.

The hydroxylations at C-16 and C-21 in maesasaponins occur as α and β-hydroxy groups, respectively. The lack of El homology between the combinatorial product in strain TM31 and an echinocystic acid standard suggests, that ML593 does not catalyze a 16α-hydroxylation on the triterpene backbone and most likely is involved in the hydroxylation of C-21. This can only be verified by NMR of the hydroxylated β-amyrin structure produced by strain TM21 or TM31 due to the absence of authentic 3β,21β-dihydroxyolean-12-ene and 3β,21β-dihydroxyolean-28-oic acid standards.

Hence, we have in this study identified and characterized three genes involved in the biosynthesis of maesasaponins. These include a β-amyrin synthase which is the first committed enzyme involved in the cyclization of the precursor 2,3-oxidosqualene to the oleanane backbone, a putative C-21β hydroxylase and a C-28 oxidase catalyzing a) the
hydroxylation of the oleanane backbone to 28-hydroxy sapogenins involved in the production of non-bridged maesasaponins, b) the C-28 aldehyde formation to generate the hemiacetal-bridged maesasaponins and c) the carboxylation of the oleanane backbone to generate the ester-bridged maesasaponins.

Materials and Methods

Chemicals. β-amyrin, Erythrodiol, Oleanolic acid and Echinocystic acid were purchased from Extrasynthese. Hexane, Pyridine and N-Methyl-N-(trimethylsilyl)trifluoroacetamide were purchased from Sigma-Aldrich. Methyl-β-cyclodextrin was purchased from CAVASOL.

Cultivation and elicitation of M. lanceolata plants. M. lanceolata axenic shoot cultures were generated and maintained as described (Faizal et al., 2011). For elicitation, each pot of shoot culture was sprayed with 2 ml deionized water containing 0.05% (v/v) Tween-20 in combination with 500 µM MeJA (10 µl of 100 mM stock dissolved in ethanol) or an equivalent amount of ethanol as control. For transcript profiling, samples were collected 0, 0.5, 1, 2, 4, 8, 24 and 48 h after elicitor or mock treatments. For each sample, 3 different plants were pooled.

Transcript profiling. The cDNA-AFLP based transcript profiling was performed with all 128 possible BstYI+1/MseI+2 primer combinations as described (Vuylsteke et al., 2007). For accurate quantification of band intensities the gel images were analyzed with the AFLP-QUANTARPRO software (Keygene, Wageningen, The Netherlands). After determining the intensity of all individual bands, the raw expression data was corrected for lane variations by dividing with a correction factor, which was calculated by dividing the sum of the expression levels of all fragments within one lane with the highest sum of all lanes within a primer combination. Next, the standard deviation and average was calculated for each individual band. Individual gene expression profiles were variance normalized by subtracting the calculated average from each individual data point, and the
obtained value was divided by the calculated standard deviation. A coefficient of variation (CV) was obtained by dividing the standard deviation by the calculated average. Gene tags displaying expression values with a CV ≥ 0.5 were considered as differentially expressed. Based on this cut-off and visual inspection of the cDNA-AFLP gels, differentially expressed gene tags were selected for further analysis. Cluster analysis, sequencing, and BLAST analysis in publicly available protein databases using a translated nucleotide query (blastx) or nucleotide database using a nucleotide query (blastn) was performed as described (Rischer et al., 2006).

**Phylogenetic analysis.** The protein sequences were retrieved from GenBank and aligned with ClustalW. The phylogenetic tree was generated in MEGA 4.0.1 software (Tamura et al., 2007), by the Neighbor-Joining method and bootstrapping was performed with 1000 replicates. The evolutionary distances were computed using the Poisson correction method, and all positions containing gaps and missing data were eliminated from the data set (complete deletion option).

**Full length open reading frame (FL-ORF) cloning of ML257, ML593 and MlbAS.** The FL-ORF of ML257 and ML593 were screened for in a M. lanceolata Uncut Nanoquantity cDNA library (custom-made by Invitrogen, Carlsbad, CA, USA) as reported (Pollier et al., 2011), using the primer pairs P27 + P28 and P29 + P30 (Table 2), respectively.

Degenerate oligonucleotide primers were designed based on highly conserved amino acid regions of known plant bAS genes and used to amplify a fragment of the MlbAS gene. First a PCR was performed on M. lanceolata cDNA using the primers P31 and P32, corresponding to the amino acid motifs DGGWGLH and LKAARHLP, respectively. The amplicon was purified using the NucleoSpin Extract II kit (Macherey-Nagel) according to the manufacturer’s instructions and used as template for a nested PCR with the primers P33 and P34, corresponding to the amino acid motifs FLPMHPAKMW and EQAGAPEWA, respectively. The resulting PCR fragment was purified once again and cloned into the pGEM®-T easy vector (Promega) for sequencing. The obtained sequence was used to design the primers P35 and P36, which were used to generate a DNA probe
for screening the FL-ORF of *MlbAS* in the *M. lanceolata* cDNA library as reported (Pollier et al., 2011).

The FL-ORFs of *MlbAS*, *ML257* and *ML593* were PCR amplified for Gateway™ cloning in pDONR221 using the primer pairs P37 + P38, P39 + P40 and P41 + P42, respectively (Table 2).

**Generation of plasmid vectors.** The *MlbAS* was PCR amplified using *XhoI* and *KpnI* containing primers, P37 + P38 (Table 2) and cloned into Gateway™ vector pDONR221 for sequence verification prior to insertion into the MCS2 of the plasmid pESC-URA[GA10/tHMG1] (described in chapter 4) to create the expression vector pESC-URA[GA10/tHMG1; GA1/MlbAS].

The *A. thaliana* CPR (AtATR1, At4g24520) and *CYP716A12* were cloned (described in chapter 4) into the low-copy number yeast expression vector pAG415GAL-ccdB (Addgene plasmid 14145) and high-copy number expression vector pAG423GAL-ccdB (Addgene plasmid 14149), respectively. The entry clones of *ML257* and *ML593* were Gateway™ recombined into the high-copy number expression vector pAG423GAL-ccdB behind the galactose inducible *GA1* promoter and having the *HIS3* auxotrophic marker.

The self-processing polyprotein of *BF567* with *CYP716A12* was generated (as described in chapter 4). Similarly, a self-processing polyprotein of *ML593* with *CYP716A12* was generated by amplifying the FL-ORF of *ML593* without a stop codon and having a 3’-overhang of the partial T2A sequence using the primer pair P41 + P43. The FL-ORF of *CYP716A12* was amplified with a 5’-overhang of the partial T2A sequence using the primers P26 + P24 such that there was an overlap of 7 bp between the two amplified sequences. Since the primers P43 and P26 contain an Uracil each, the *ML593* and *CYP716A12* were PCR amplified using the Pfu Turbo Cx polymerase (Stratagene). The purified gel fragments were used for Uracil-Specific Excision or USER™ Cloning (New England Biolabs) to generate two fragments with complementary sticky ends which were ligated *in vitro* using the T4 DNA ligase (Invitrogen). The ligated DNA product was once again gel purified and used as template for amplification with the primers P41 + P24.
This amplicon was Gateway™ recombined into pDONR221, sequence verified and further recombined into pAG423GAL-ccdB to generate pAG423[GAL1/ML593-T2A-CYP716A12].

Table 2. Primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P24</td>
<td>GGGGACCACTTTTGTAACAGAGGCTGGTAATAGCCTTTGTGGATGAAAGGCG</td>
</tr>
<tr>
<td>P26</td>
<td>ATGCGGGUGGAGCTTCAGGAGGCAATCCCTGGCCCAATGGAGGCTAATTCTATC</td>
</tr>
<tr>
<td>P27</td>
<td>TGGACCAAGCCTCTTACACATT</td>
</tr>
<tr>
<td>P28</td>
<td>GAAGAAGTGGGTCGGAAAGAC</td>
</tr>
<tr>
<td>P29</td>
<td>CTTCTGCAATCTAAATCCGAAC</td>
</tr>
<tr>
<td>P30</td>
<td>AGCAGAAGATTCGCTTTGCTA</td>
</tr>
<tr>
<td>P31</td>
<td>GAYGGNGGNTGAGGGNYTNCA</td>
</tr>
<tr>
<td>P32</td>
<td>NARYTNNGCNGCYCRTGARNGG</td>
</tr>
<tr>
<td>P33</td>
<td>TTYTNCNATGCAAYCCNGCNAARATGTGG</td>
</tr>
<tr>
<td>P34</td>
<td>TCYTGNGCNCNGCGNGYTCCCANGC</td>
</tr>
<tr>
<td>P35</td>
<td>ATGTGGGTGTTACTGCGATTG</td>
</tr>
<tr>
<td>P36</td>
<td>TAGCAAAATCTTTGCCGACATAGGA</td>
</tr>
<tr>
<td>P37</td>
<td>GGGGACTTTTGTAACAGAGGCTGGGCTGggtaccCTAAGACGGCAATTTGA</td>
</tr>
<tr>
<td>P38</td>
<td>GGGGACCACTTTTGTAACAGAGGCTGGTAATAGCCTTTGTGGATGAAAGGCG</td>
</tr>
<tr>
<td>P39</td>
<td>GGGGACCACTTTTGTAACAGAGGCTGGTAATAGCCTTTGTGGATGAAAGGCG</td>
</tr>
<tr>
<td>P40</td>
<td>GGGGACCACTTTTGTAACAGAGGCTGGTAATAGCCTTTGTGGATGAAAGGCG</td>
</tr>
<tr>
<td>P41</td>
<td>GGGGACCACTTTTGTAACAGAGGCTGGTAATAGCCTTTGTGGATGAAAGGCG</td>
</tr>
<tr>
<td>P42</td>
<td>GGGGACCACTTTTGTAACAGAGGCTGGTAATAGCCTTTGTGGATGAAAGGCG</td>
</tr>
<tr>
<td>P43</td>
<td>ACAGAGGCTGGTAATAGCCTTTGTGGATGAAAGGCG</td>
</tr>
</tbody>
</table>

The sequences in lower case represent the restriction recognition site used for restriction enzyme mediated cloning. The underlined sequence corresponds to T2A partial sequences.

**Culturing of yeast cells for metabolite analysis.** Precultures were grown at 30°C, 250 rpm for 18-20 h in 5 ml synthetic defined (SD) medium containing glucose (Duchefa, Clontech) and appropriate drop out (DO) supplements (Duchefa, Clontech). Cultures were induced by inoculating to a starting optical density of 0.25 in 15 ml SD Gal/Raf medium containing galactose and raffinose as carbon source (Duchefa, Clontech) and appropriate DO supplements, post washing of the preculture in water. The cultures
were incubated at 30°C, 250 rpm for 24 h before addition of 2.25 ml of 10 mM methionine to a final concentration of 1.5 mM and subsequently incubated further for 48 h. MβCD treatment was carried out by adding 300 µl of a 250 mM MβCD stock to a final concentration of 5 mM, twice during culturing, at the time of methionine repression and 24 h after methionine addition.

**Metabolite extraction and GC-MS analysis.** An organic extract of the yeast cells or spent medium was prepared (as described in chapter 4) for identification and quantification of metabolites using GC-MS. In brief, first a hexane extraction was performed on alkali-lyzed cell pellets or spent medium, which was concentrated by evaporation of the hexane and trimethylsilyl derivatized before GC-MS analysis. For identification of metabolites a full mass spectra was generated by scanning the m/z range of 60-800 with a solvent delay of 7.8 min. For select ion monitoring of β-amyrin the ions 498, 218 and 203 were screened. The areas of the peaks were calculated using the default settings of the AMDIS software (version 2.6, NIST, USA).

**Acknowledgements**

We thank Wilson Ardiles-Diaz for sequencing the cDNA-AFLP tags, Miguel González-Guzmán and Lander Ingelbrecht for excellent technical advice and assistance, Prof. Danny Geelen (Universiteit Gent) for the *M. lanceolata* axenic shoot cultures and Prof. Jan Van Bocxlaer (Universiteit Gent) for the fruitful discussions on GC-MS profiling. This work was supported by the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" project SBO040093) and the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. Tessa Moses is indebted to the VIB International PhD Fellowship Program for a predoctoral fellowship. Lorena Almagro was supported by a short-term scientific mission grant from the COST action FA1006- Plant Engine.
References


Rischer, H., Orešič, M., Seppänen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M.C., Inzé, D., Oksman-Caldentey, K.M.,
Combinatorial biosynthesis of triterpene sapogenins in *Saccharomyces cerevisiae*
Publication status:

**Tessa Moses**, Jacob Pollier, Johan M. Thevelein and Alain Goossens.

**Combinatorial biosynthesis of triterpene sapogenins in Saccharomyces cerevisiae.**

Manuscript in preparation

Author contribution:

Generation of yeast strains, culturing and metabolite extractions, GC-MS profiling, data analysis and writing of the manuscript.
Abstract

Combinatorial biosynthesis is the process of establishing novel enzyme-substrate combinations *in vivo*, leading to the synthesis of new natural product derived compounds that can be used in drug discovery programs. Plants are a rich source of bioactive natural products and continue to possess huge potential for drug discovery. However, these valuable natural products cannot always be isolated from the plant source in commercially feasible quantities. Additionally, the biosynthetic pathways leading to the synthesis of these secondary metabolites in plants are not completely elucidated making the translation of pathways to heterologous hosts complicated. Therefore, we utilized a combinatorial approach for the synthesis of triterpenoid sapogenins in *Saccharomyces cerevisiae* by coexpression of known saponin biosynthetic enzymes. Using this approach we were able to synthesize both known as well as hitherto undescribed and possibly novel triterpene scaffolds.
Introduction

The secondary metabolites of plants are structurally diverse natural products with significant therapeutic applications. For instance, the anti-malarial drug artemisinin, a sesquiterpene, and the anti-cancer drug taxol, a diterpene were originally isolated from the shrub *Artemisia annua* and the Pacific yew tree (*Taxus cuspidata*), respectively. Together with plant cell cultures, the plants themselves remain the only source of these drugs to date, rendering them susceptible to fluctuations in yield and market price. Although total chemical synthesis of both of these molecules has been established, their intricate structural complexity makes the process expensive and fairly complicated. Nonetheless, there are continuous efforts being made towards engineering microbes for the production of plant-based drugs to stabilize prices and to meet the ever increasing market demand. The fast growing, genetically amenable hosts *Saccharomyces cerevisiae* and *Escherichia coli* continue being the first choice for the microbial production of plant natural products (Ajikumar et al., 2010; Westfall et al., 2012). Additionally, the recent advances in metabolic engineering facilitate directed synthesis of high levels of desired products in heterologous hosts.

Combinatorial biosynthesis is an extension of the metabolic engineering process, where genes catalyzing individual metabolic steps from different organisms are combined in a single host to create novel branches in metabolic pathways to synthesize products that were not accessible previously (Kirschning et al., 2007; Pollier et al., 2011). Synthesis of novel scaffolds by combinatorial biosynthesis generates structural diversification and broadens the spectrum of lead molecules for therapeutic applications. Structural novelty can be achieved either by rational redesigning of biosynthetic pathways, or by the random mixing of biosynthetic genes from related pathways. However, the combinatorial approach can also be applied for the biosynthesis of natural products, of which complete biosynthetic pathways are not yet elucidated (Julsing et al., 2006; Chemler and Koffas, 2008) or even to enhance the productivity of certain desired products. For instance, in *S. cerevisiae* the production of taxadiene, the precursor of taxol, was enhanced by the combination of bacterial and plant genes (Engels et al., 2008).
The triterpenoid branch of plant secondary metabolites comprises the saponins that display a plethora of biological activities and are widely distributed among many plant species in nature (Sparg et al., 2004). The triterpenoid saponins are glycosides of “sapogenins”, that are valuable not only as biologically active compounds but also serve as starting molecules for semi-synthetic modifications to enhance structural diversity. For instance, the naturally occurring sapogenins, oleanolic acid (Pollier and Goossens, 2012) and betulinic acid (Fulda, 2008) are widespread in the plant kingdom and exhibit various biological activities that have been accessed for future clinical development. Chemical derivatives of these molecules have been generated not only for structural diversification, but also to reduce cytotoxicity, enhance efficacy and increase solubility of the sapogenins. One such semi-synthetic derivative of oleanolic acid, bardoxolone methyl is currently in a late-stage clinical development for the treatment of advanced chronic kidney disease in type 2 diabetes mellitus patients (Pollier and Goossens, 2012).

In the current study we established a combinatorial platform in *S. cerevisiae* for the production of triterpene sapogenins by coexpressing saponin biosynthetic genes from different (medicinal) plants. Through this untargeted biosynthesis approach we generated both known sapogenin structures as well as novel scaffolds that might harbor plausible biological activities.

**Results**

**Oxidation of β-amyrin by a putative C-21β hydroxylase and a C-11 oxidase.**

To generate a combinatorial scaffold using β-amyrin we expressed *ML593* (see chapter 5) and *CYP88D6* in strain *TM3*. The CytP450 *CYP88D6* (GenBank accession number AB433179; (Seki et al., 2008)) catalyzes a two-step oxidation of β-amyrin to 11-oxo-β-amyrin through a 11-hydroxy-β-amyrin intermediate. First, we confirmed the activity of *CYP88D6* in our β-amyrin producing strain by generating strain *TM18* (Table 1) and culturing it in the presence of MβCD along with the control strain *TM27*. GC
Combinatorial biosynthesis in *Saccharomyces cerevisiae*

chromatograms of the spent medium of strain TM18 showed the presence of four unique peaks (Fig. 1c) that were absent in the control strain TM27 (Fig. 1d). Two of these peaks eluting at 25.4 min and 37.5 min corresponded to 11-hydroxy β-amyrin and 11-oxo β-amyrin (Seki et al., 2008), respectively. The two remaining peaks eluting at 24.4 min and 26.6 min could not be assigned an identity despite their clear EI pattern. The highest mass observed in the mass spectra extracted from these peaks (Fig. 1e) was lower than that of trimethylsilylated β-amyrin (M+=498), indicating their possibly non-triterpenoid origin. These additional peaks observed in our yeast strain were not reported when the CYP88D6 was expressed in a wild type yeast strain expressing a β-amyrin synthase from *Lotus japonicus* (Seki et al., 2008).

Further, we generated strain TM32 by supertransforming strain TM3 with the plasmids pAG415[GAL1/AtATRI] and pAG423[GAL1/ML593-T2A-CYP88D6], where ML593 and CYP88D6 are stitched together with the 2A oligopeptide resulting in the generation of a self-processing polypeptide (de Felipe et al., 2006). The spent medium of strain TM32 (Fig. 1a) cultured in the presence of MβCD was compared with that of strains TM21 (Fig. 1b), TM18 (Fig. 1c) and TM27 (Fig. 1d). Strain TM32 showed the presence of two unique peaks eluting at 32.1 min and 44.7 min that could correspond to 11α,21β(?)-dihydroxy β-amyrin and 11-oxo-21β(?)-hydroxy β-amyrin based on their EI pattern, respectively.

Oxidation of lupeol and lupanediol by C-28 oxidases.

The *Maesa lanceolata* ML257 like the *Medicago truncatula* CYP716A12 is a C-28 oxidase of β-amyrin (see chapter 5). We determined the substrate specificity of ML257 for the oxidation of other triterpene sapogenin backbones, like lupeol. First, we generated a lupeol producing strain TM6 from the engineered strain TM1 (Table 1) by expressing the *Arabidopsis thaliana* isoform 1 of lupeol synthase (AtLUP1) (GenBank accession number U49919; (Herrera et al., 1998)) from the plasmid vector pESC-URA[GAL10/tHMG1; GAL1/AtLUP1]. We cultured TM6 and the control strain TM5
Combinatorial biosynthesis in Saccharomyces cerevisiae

Figure 1. GC chromatograms corresponding to, (a) Extraction from spent medium of strain TM32, expressing tHMG1, GgbAS, AtATR1, CYP88D6 and ML593. (b) Extraction from spent medium of strain TM21, expressing tHMG1, GgbAS, AtATR1 and ML593. (c) Extraction from spent medium of strain TM18, expressing tHMG1, GgbAS, AtATR1 and CYP88D6. (d) Extraction from spent medium of control strain TM27, expressing tHMG1, GgbAS and AtATR1. (e) Mass spectra extracted from indicated (*) peaks of strain TM18 and TM32. Right panels show the mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively. (f) Oxidation of β-amin by ML593 and CYP88D6.
(Table 1) and quantified lupeol accumulation from the cell pellets using GC-MS. A peak eluting at 28.9 min and corresponding to 46.3 mg/L of a lupeol standard (Fig. 2c) was observed in the GC chromatograms of TM6 (Fig. 2a), but not the control strain TM5 (Fig. 2b).

**Figure 2.** GC chromatograms corresponding to, (a) Extraction from cells of strain TM6, expressing \( tHMG1 \) and AtLUP1. (b) Extraction from cells of control strain TM5, expressing \( tHMG1 \). (c) A lupeol standard. Right panels show mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively. (d) Cyclization of 2,3-oxidosqualene to lupane and lupanediol. (e) Quantification of lupeol from the cells and spent medium of strain TM6, where 100% corresponds to 46.3 mg/L of lupeol.
Additionally, in this strain we also observed a new peak eluting at 40.2 min (Fig. 2a) the EI pattern of which corresponded to 3β,20-dihydroxylupane (lupanediol), resulting from the addition of water to the lupanyl cation (Fig. 2d) in the course of its cyclization by \textit{AtLUP1} (Kushiro et al., 2006). Notably, the relative amount of lupanediol eluting at 40.2 min (corresponding to the area under the curve in the GC chromatogram), was higher than lupeol eluting at 28.9 min.

The \textit{AtLUP1} has been reported to cyclize oxidosqualene to multiple triterpene alcohols and oxacyclic triterpenoids with the major products being lupeol and lupanediol together with multiple minor products, including β-amyrin, germanicol, ψ-taraxasterol (Segura et al., 2000), and 17,24-epoxybaccharane, 20,24-epoxydammarane (Shan et al., 2005). However, in our strain we only detected lupeol and lupanediol and none of the minor products that have been reported earlier.

When the strains \textit{TM6} and \textit{TM5} were cultured in the presence of MβCD, the same two peaks corresponding to lupeol and lupanediol were observed in the extracts of the spent medium of \textit{TM6} but not \textit{TM5} and hence, to facilitate metabolite extractions we utilized MβCD in all subsequent experiments. A strikingly higher concentration of lupeol was quantified from the extracts of spent medium (164.2 mg/L) compared to cell pellet (41.7 mg/L) on MβCD treatment and the total lupeol concentration was 4.4-fold higher in treated cultures compared to non-treated controls (Fig. 2e). This pattern was also observed for the production and extraction of lupanediol from \textit{TM6} in the presence of MβCD.

Next, we supertransformed strain \textit{TM6} with the plasmids pAG423[\textit{GAL1/ML257}], pAG423[\textit{GAL1/CYP716A12}] or pAG423GAL-ccdB in combination with pAG415[\textit{GAL1/AtATR1}] to generate strains \textit{TM23}, \textit{TM22} and \textit{TM28}, respectively. These strains were cultured in the presence of MβCD and the spent medium analyzed by GC-MS for the presence of lupeol, lupanediol and their hydroxylation products. We observed new peaks eluting at 35.2 min corresponding to betulin (Fig. 3d) and at 46.8 min most likely corresponding to the C-28 hydroxylation of lupanediol in strains \textit{TM23} (Fig. 3a) and \textit{TM22} (Fig. 3b), but not the control strain \textit{TM28} (Fig. 3c). Strain \textit{TM22} also showed
the presence of an extra new peak at 54.7 min, the EI pattern of which could not be interpreted due to the very low abundance of the peak in the chromatogram, but which could correspond to the carboxylation of C-28 of lupeol. However, we did not detect betulinic acid in either TM23 or TM22 expressing ML257 or CYP716A12, respectively, contrary to (Fukushima et al., 2011) where both betulin and betulinic acid were detected in an yeast strain expressing a G. uralensis lupeol synthase in combination with the M. truncatula CYP716A12.

Figure 3. GC chromatograms (first column) corresponding to, (a) Extraction from spent medium of strain TM23, expressing tHMG1, AtLUP1, AtATRI and ML257. (b) Extraction from spent medium of strain TM22, expressing tHMG1, AtLUP1, AtATRI and CYP716A12. (c) Extraction from spent medium of control strain TM28, expressing tHMG1, AtLUP1 and AtATRI. (d) A betulin standard. Columns two and three show the mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively.
Combinatorial biosynthesis in *Saccharomyces cerevisiae* 6

Oxidation of other triterpene sapogenin backbones by C-28 oxidases.

In an attempt to generate a dammarenediol producing yeast strain we expressed the plasmid pESC-URA[\textit{GAL10/tHMG1; GAL1/CaDDS}], harboring a dammarenediol synthase gene (\textit{CaDDS}) from *Centella asiatica* (GenBank accession number AY520818; (Kim et al., 2009)), in the sterol modified yeast strain \textit{TM1}, to generate strain \textit{TM33}. We cultured strain \textit{TM33} and the control strain \textit{TM5} in medium containing MβCD and analyzed the GC chromatograms obtained from the spent medium of both the strains. Unexpectedly, we found 3 new peaks in strain \textit{TM33} eluting at 27.2 min, 28.6 min and 33.5 min (Fig. 4a) that were absent in the control strain \textit{TM5} (Fig. 4b). We could confirm the identity of the peak eluting at 27.2 min as β-amyrin (Fig. 4c) based on its elution time and EI pattern. The peak at 28.6 min had a similar EI pattern as β-amyrin and was confirmed as α-amyrin by comparing to a standard (Fig. 4d). The EI pattern of the peak at 33.5 min could be interpreted to dammarenediol-II (Spencer, 1981), but was not confirmed due to the lack of an authentic standard. Although, the \textit{CaDDS} was initially reported as a putative \textit{bAS} (Kim et al., 2005) and later characterized as a dammarenediol synthase (Kim et al., 2009), in our yeast strain the gene was capable of cyclizing 2,3-oxidosqualene to both β-amyrin and dammarenediol in addition to a third product, α-amyrin. In our yeast strain the relative amounts of α-amyrin, β-amyrin and dammarenediol-II were in the ratio of 8.8:1.1:0.1, highlighting the very low dammarene synthase activity of \textit{CaDDS} as opposed to its current characterization. Therefore, we propose to characterize \textit{CaDDS} as a multi-functional oxidosqualene cyclase.

We continued with the strain \textit{TM33} expressing the multifunctional cyclase and used it for the generation of strains \textit{TM36}, \textit{TM34} and \textit{TM38}, expressing the plasmid pAG415[\textit{GAL1/AtATRI}] with either pAG423[\textit{GAL1/ML257}], pAG423[\textit{GAL1/CYP716A12}] or pAG423GAL-ccdB, respectively. GC-MS analysis of the spent medium of \textit{TM36}, \textit{TM34} and \textit{TM38} cultured in the presence of MβCD showed the presence of multiple new peaks corresponding to oxidation products in strains \textit{TM36} (Fig. 5a) and \textit{TM34} (Fig. 5b), but not the control strain \textit{TM38} (Fig. 5c).
Figure 4. GC chromatograms corresponding to, (a) Extraction from spent medium of strain TM33, expressing *tHMG1* and *CaDDS*. (b) Extraction from spent medium of control strain TM5, expressing *tHMG1*. (c) A β-amyrin standard. (d) An α-amyrin standard. Right panels show the mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively.

The peak eluting at 32.5 min (Fig. 5a and 5b) could be confirmed as erythrodiol (Fig. 5d), confirming the β-amyrin synthase functionality of *CaDDS*. Next to the erythrodiol peak, another high intensity peak with the same EI pattern as erythrodiol but a different elution
Combinatorial biosynthesis in *Saccharomyces cerevisiae* time of 34.2 min was observed, which could correspond to uvaol (28-hydroxy α-amyrin), thereby also reaffirming the α-amyrin synthase activity of *CaDDS*. We also observed higher concentrations of uvaol in both strains *TM36* and *TM34* compared to erythodiol.

**Figure 5.** GC chromatograms corresponding to, (a) Extraction from spent medium of strain *TM36*, expressing *tHMG1*, *CaDDS*, *AtATR1* and *ML257*. (b) Extraction from spent medium of strain *TM34*, expressing *tHMG1*, *CaDDS*, *AtATR1* and *CYP716A12*. (c) Extraction from spent medium of control strain *TM38*, expressing *tHMG1*, *CaDDS* and *AtATR1*. (d) An erythrodiol standard. Right panels show the mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively.
Oxidation of α-amyrin by a C-11 oxidase.

We supertransformed strain *TM33* with the plasmids pAG415[**GAL1/AtATR1**] and pAG423[**GAL1/CYP88D6**] to generate a strain *TM35*. GC chromatograms of extracts from the spent medium of *TM35* (Fig. 6a) cultured in the presence of MβCD showed the presence of four unique peaks, that were absent in the control strain *TM38* (Fig. 6b).

**Figure 6.** GC chromatograms corresponding to, (a) Extraction from spent medium of strain *TM35*, expressing *tHMG1*, *CaDDS*, *AtATR1* and *CYP88D6*. Right panels show the mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively. (b) Extraction from spent medium of control strain *TM38*, expressing *tHMG1*, *CaDDS* and *AtATR1*. (c) Mass spectra extracted from indicated (*) peaks of strain *TM35*.

Two of these peaks eluting at 25.4 min and 37.5 min were identified as 11-hydroxy β-amyrin and 11-oxo-β-amyrin based on their EI pattern (Seki et al., 2008), respectively. The remaining two peaks eluting at 29.4 min and 39.6 min had the same EI pattern (Fig. 6c) as 11-hydroxy and 11-oxo-β-amyrin, respectively, but a different elution time. These peaks could correspond to the C-11 oxidations of α-amyrin, but could not be confirmed due to the lack of authentic standards.
Oxidation of α-amyrin by a putative C-21 hydroxylase.

Strain TM37 was generated from strain TM33 by supertransforming with the plasmids pAG415[\textit{GAL1}/AtATR1] and pAG423[\textit{GAL1}/ML593]. The ML593 was characterized as a putative C-21 hydroxylase of β-amyrin (see chapter 5) and we determined the substrate specificity of this CytP450 by expressing it together with the multifunctional cyclase \textit{CaDDS}. We cultured strain TM37 and the control strain TM38 in the presence of MβCD and compared GC chromatograms for the presence of unique peaks. We identified two peaks eluting at 31.8 min and 33.2 min in the spent medium of TM37 (Fig. 7a), but not the control strain TM38 (Fig. 7b). The EI pattern of both these peaks were identical (Fig. 7a) and the peak at 31.8 min corresponded to (most likely) 21-hydroxy β-amyrin (see chapter 5). Therefore, the second peak at 33.2 min could correspond to 21-hydroxy α-amyrin, but was not confirmed due to the absence of an authentic standard.

\textbf{Figure 7.} GC chromatograms corresponding to, (a) Extraction from spent medium of strain TM37, expressing \textit{tHMG1}, \textit{CaDDS}, AtATR1 and ML593. Right panels show the mass spectra extracted from the indicated (*) peaks, where the X-axis corresponds to m/z values. (b) Extraction from spent medium of control strain TM38, expressing \textit{tHMG1}, \textit{CaDDS} and AtATR1. The X- and Y-axis of the GC chromatograms correspond to time in minutes and total ion current, respectively. Parts of structure in green indicate the putative hydroxylation position.
Combinatorial biosynthesis in *Saccharomyces cerevisiae*  

Table 1. List of yeast strains generated and used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>S288c BY4742</td>
<td>MATα; his3Δ1; leu2Δ0; ura3Δ0; lys2Δ0</td>
</tr>
<tr>
<td>TM1</td>
<td>S288c BY4742; P&lt;sub&gt;erg7&lt;/sub&gt;::P&lt;sub&gt;MET3&lt;/sub&gt;-ERG7</td>
</tr>
<tr>
<td>TM3</td>
<td>TM1; pESC-URA[GAL10/tHMG1; GALI/GgbAS] (36.2 mg/L β-aminry)</td>
</tr>
<tr>
<td>TM5</td>
<td>TM1; pESC-URA[GAL10/tHMG1]</td>
</tr>
<tr>
<td>TM6</td>
<td>TM1; pESC-URA[GAL10/tHMG1; GALI/AtLUP1] (46.3 mg/L lupeol)</td>
</tr>
<tr>
<td>TM18</td>
<td>TM3; pAG423[GALI/CYP88D6], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM21</td>
<td>TM3; pAG423[GALI/ML593], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM22</td>
<td>TM6; pAG423[GALI/CYP716A12], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM23</td>
<td>TM6; pAG423[GALI/ML257], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM27</td>
<td>TM3; pAG423, pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM28</td>
<td>TM6; pAG423, pAG425[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM32</td>
<td>TM3; pAG423[GALI/ML593-T2A-CYP88D6], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM33</td>
<td>TM1; pESC-URA[GAL10/tHMG1; GALI/CaDDS]</td>
</tr>
<tr>
<td>TM34</td>
<td>TM33; pAG423[GALI/CYP716A12], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM35</td>
<td>TM33; pAG423[GALI/CYP88D6], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM36</td>
<td>TM33; pAG423[GALI/ML257], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM37</td>
<td>TM33; pAG423[GALI/ML593], pAG415[GALI/AtATR1]</td>
</tr>
<tr>
<td>TM38</td>
<td>TM33; pAG423, pAG415[GALI/AtATR1]</td>
</tr>
</tbody>
</table>

Discussion

The triterpenoid saponins are a vast group of glycosides, whose structural diversity arises from the variable oxidation patterns on the sapogenin backbones that are composed of 30 carbon atoms arranged in ring structures. Although aglycones with oxidations on various positions of the 30 carbons have been described from different plants (Dinda et al., 2010), only 7 enzyme classes (or CytP450s) catalyzing specific modifications to C-6 (Han et al., 2012), C-11 (Seki et al., 2008), C-12 (Han et al., 2011), putative C-21 (see chapter 5), C-24 (Shibuya et al., 2006), C-28 (Carelli et al., 2011) and C-30 (Seki et al., 2011) have been identified to date. This great void in our understanding about the synthesis of
Combinatorial biosynthesis in *Saccharomyces cerevisiae* makes targeted biosynthesis highly complicated and sometimes impossible. To truly exploit the therapeutic potential of triterpenes it is essential to synthesize them in non-native hosts that can be engineered for high-level production. Hence, we set up a combinatorial approach to generate triterpene sapogenins, the building blocks of saponins in *S. cerevisiae* by random coexpression of known CytP450s.

We first attempted a targeted combinatorial biosynthesis approach by expressing enzymes *GgbAS*, *ML593* and *CYP88D6*, from *G. glabra*, *M. lanceolata* and *G. uralensis*, respectively. The resulting compounds, 11,21(?)-dihydroxy-β-amyrin and 11-oxo-21(?)-hydroxy-β-amyrin, are undescribed in any of the plant species from which the enzymes originated, but may be an intermediate in the biosynthesis of saikosapogenins produced by plants of the genus *Bupleurum*. This approach of coexpressing genes from different plants to generate a target molecule underscores the potential of combinatorial biosynthesis.

As a next step we generated triterpene alcohol backbones lupeol and dammarenediol-II in yeast. Our engineered strain *TM6* producing lupeol also cyclized oxidosqualene to high amounts of 3β,20-dihydroxylupane/lupanediol. Although unforeseen, this cyclization reaction catalyzed by *AtLUP1* generated a triterpene diol scaffold in yeast for further modification by CytP450s. We performed a targeted C-28 oxidation on both the lupane and lupanediol backbones by overexpressing C-28 oxidases, *ML257* and *CYP716A12* in the yeast strain *TM6*. The resulting 3β,20-dihydroxylupan-28-oic acid generated by the three-step C-28 carboxylation activity of *CYP716A12* on lupanediol is a known aglycone of the saponins Snatzkein C and D, from the plant *Arenaria filicaulis* (Elgamal et al., 1998). However, the 3β,20,28-trihydroxylupane resulting from the hydroxylation of C-28 of lupanediol is to the best of our knowledge a hitherto undescribed triterpene backbone and could serve as an interesting backbone for further enzymatic or semi-synthetic modifications. Lupeol derived sapogenins like betulin (3β,28-dihydroxylupane) and betulinic acid (3β-hydroxylupan-28-oic acid) have been valued for their potent antitumor activities (Fulda, 2008; Mullauer et al., 2009). Additional favourable modifications like hydroxylations on the backbone increase the hydrophilicity of the molecule and generate promising targets for clinical testing.
Similarly, in the course of generating a dammarenedi-diol-II producing yeast strain TM33, we discovered the multifunctionality of CaDDS which was reported as being only a dammarenediol synthase from *C. asiatica* (Kim et al., 2009). In our yeast strain, CaDDS cyclized oxidosqualene to three products, α-amyrin, β-amyrin and dammarenedi-diol-II, with α-amyrin being produced 8 times more than β-amyrin and dammarenedi-diol-II produced only as a minor product. The further combinatorial expression of a C-28 oxidase from *M. lanceolata* or *M. truncatula* resulted in the generation of erythrodiol and uvaol in the same strain. The higher concentration of uvaol over erythrodiol in the spent medium of strains TM36 and TM34 reaffirmed the preferred α-amyrin cyclization activity of CaDDS. This is the first report on the multifunctionality of CaDDS from *C. asiatica* and could explain its initial annotation as a putative β-amyrin synthase based on the presence of the conserved **W**CYCR β-amyrin production motif in the enzyme (Kim et al., 2005).

We further utilized strain TM33 harboring the multifunctional CaDDS as an α-amyrin producing yeast strain and explored possibilities for modification of the α-amyrin backbone by the expression of CYP88D6, a C-11 oxidase from *G. uralensis* or ML593, a putative C-21 hydroxylase from *M. lanceolata*. Through this combinatorial biosynthesis we generated 11-hydroxy-α-amyrin, 11-oxo-α-amyrin and 21(?)-hydroxy-α-amyrin producing yeast strains, that have not be reported earlier.

In conclusion, we established a combinatorial biosynthesis platform in *S. cerevisiae* for the generation of triterpene sapogenins and applied it for the production of both known compounds with documented biological activities, as well as novel scaffolds that might possess interesting therapeutic potential.
Materials and Methods

Chemicals. α-amyrin, β-amyrin, Lupeol, Erythrodiol and Betulin were purchased from Extrasynthese. Hexane, Pyridine and N-Methyl-N-(trimethylsilyl)trifluoroacetamide were purchased from Sigma-Aldrich. Methyl-β-cyclodextrin was purchased from CAVASOL.

Generation of plasmid vectors. The AtLUP1 was PCR amplified from a pENTR223-AtLUP1 vector (ABRC stock number U09402) using BamHI and XhoI containing primers P44 + P45 (Table 2) and cloned into Gateway™ vector pDONR221 for sequence verification prior to insertion into the MCS2 of the plasmid pESC-URA[GAL10/tHMG1] (described in chapter 4) to create the expression vector pESC-URA[GAL10/tHMG1; GAL1/AtLUP1]. Similarly, the CaDDS was amplified with XhoI and NheI containing primers P46 + P47 to generate pESC-URA[GAL10/tHMG1; GAL1/CaDDS].

The A. thaliana CPR (AtATRI, At4g24520) was cloned into the low-copy number yeast expression vector pAG415GAL-ccdB (Addgene plasmid 14145). The full-length open reading frame (FL-ORF) of CYP88D6 was PCR amplified from a pTOPO-CYP88D6 vector (Seki et al., 2008) using the primers P48 + P49 and cloned into pDONR221. The CytP450s ML257, CYP716A12, ML593 and CYP88D6 were Gateway™ recombined into the high-copy number expression vector pAG423GAL-ccdB (Addgene plasmid 14149).

The self-processing polyprotein of ML593 and CYP88D6, was generated by amplifying the FL-ORF of ML593 without a stop codon and having a 3’- overhang of the partial T2A sequence using the primer pair P41 + P50. The FL-ORF of CYP88D6 was amplified with a 5’- overhang of the partial T2A sequence using the primers P51 + P49 such that there was an overlap of 7 bp between the two amplified sequences. Since the primers P51 and P50 contain an Uracil each, the ML593 and CYP88D6 were PCR amplified using the Pfu Turbo Cx polymerase (Stratagene). The purified gel fragments were used for USER™ Cloning (New England Biolabs) to generate two fragments with complementary sticky ends which were ligated in vitro using the T4 DNA ligase (Invitrogen). The ligated DNA product was used as template for amplification with the primers P41 + P49. This
Combinatorial biosynthesis in *Saccharomyces cerevisiae* 6

The amplicon was Gateway™ recombined into pDONR221, sequence verified and further recombined into pAG423GAL-ccdB to generate pAG423[\(GAL1/ML593-T2A-CYP88D6\)].

**Culturing of yeast cells for metabolite analysis.** Precultures were grown at 30°C, 250 rpm for 18-20 h in 5 ml synthetic defined (SD) medium containing glucose (Duchefa, Clontech) and appropriate drop out (DO) supplements (Duchefa, Clontech). Cultures were induced by inoculating to a starting optical density of 0.25 in 15 ml SD Gal/Raf medium containing galactose and raffinose as carbon source (Duchefa, Clontech) and appropriate DO supplements, post washing of the preculture in water. The cultures were incubated at 30°C, 250 rpm for 24 h before addition of 2.25 ml of 10 mM methionine to a final concentration of 1.5 mM and subsequently incubated further for 48 h. MβCD treatment was carried out by adding 300 µl of a 250 mM MβCD stock to a final concentration of 5 mM, twice during culturing, at the time of methionine repression and 24 h after methionine addition.

**Metabolite extraction and GC-MS analysis.** An organic extract of the yeast cells or spent medium was prepared (as described in chapter 4) for identification and quantification of metabolites using GC-MS. In brief, first a hexane extraction was performed on alkali lyzed cell pellets or spent medium, which was concentrated by evaporation of the hexane and trimethylsilyl derivatized before GC-MS analysis. For identification of metabolites a full mass spectra was generated by scanning the m/z range of 60-800 with a solvent delay of 7.8 min. For quantification of lupeol a select ion mode was operated detecting the ions 498, 483, 393, 279, 218, 203 and 189. The areas of the peaks were calculated using the default settings of the AMDIS software (version 2.6, NIST, USA).
Table 2. Primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P41</td>
<td>GGGGACAAAGTTTGTAACAAAAAGCAGGCTTAATGTGGGTAGTGGGATTA</td>
</tr>
<tr>
<td>P44</td>
<td>GGGGACAAAGTTTGTAACAAAAAGCAGGCTTAatggatccATGTGGAAAGTTGAAG</td>
</tr>
<tr>
<td>P45</td>
<td>GGGGACACTTTGTACAAAGAAACCTGGGATActcgagTTAATTACAGATAAAC</td>
</tr>
<tr>
<td>P46</td>
<td>GGGGACAAAGTTTGTAACAAAAAGCAGGCTTAatcgagATGTGGAAAGCTGAAG</td>
</tr>
<tr>
<td>P47</td>
<td>GGGGACACTTTGTACAAAGAAAGCTGGGGTTgtcgacTCAATTGGGAGAGCCACAAGCG</td>
</tr>
<tr>
<td>P48</td>
<td>GGGGACAAAGTTTGTAACAAAAAGCAGGCTTAATGGAATGACATTGGGTTT</td>
</tr>
<tr>
<td>P49</td>
<td>GGGGACACTTTGTACAAAGAAAGCTGGGTACATCAAAGCAGGCTTAATGGAATGACATTGGGTTT</td>
</tr>
<tr>
<td>P50</td>
<td>ACCGGCAUGTTAGCAGACTTCTCCTCGCCCTCCTTGTCTTTTCTTTGGTGACCT</td>
</tr>
<tr>
<td>P51</td>
<td>ATGCACGGUGACCGTCCAGGAGAAATCCCTGGCCCAATTGGGAGATACATTGGGTTT</td>
</tr>
</tbody>
</table>

The sequences in lower case represent the restriction recognition site used for restriction enzyme mediated cloning. The underlined sequence corresponds to T2A partial sequences.

Acknowledgements

We thank Prof. Jan Van Bocxlaer (Universiteit Gent) for the fruitful discussions on GC-MS profiling. This work was supported by the Agency for Innovation by Science and Technology in Flanders (“Strategisch Basisonderzoek” project SBO040093) and the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. Tessa Moses is indebted to the VIB International PhD Fellowship Program for a predoctoral fellowship.
Combinatorial biosynthesis in *Saccharomyces cerevisiae*

References


Han, J.Y., Hwang, H.S., Choi, S.W., Kim, H.J., and Choi, Y.E. (2012). Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. Plant and Cell Physiology.


Combinatorial biosynthesis in Saccharomyces cerevisiae


The protein quality control system manages plant defense compound synthesis
Protein quality control in *Medicago truncatula*

Publication status:

**The protein quality control system manages plant defense compound synthesis.**
Manuscript submitted for publication

Author contribution:
Protein extractions and immunoblot analysis from hairy roots and yeast, yeast complementation assays, generation of yeast strains for localization, feeding assays with hairy roots, generation of constructs for protoplast assays and writing of the manuscript.
Abstract

In the defense against attackers, the jasmonate hormone regulates production of bioactive metabolites in plants. Knowledge on the conserved jasmonate perception machineries by plants is increasing, but the downstream mechanisms that regulate defense metabolism remain largely unknown. Here we show that in the model legume *Medicago truncatula*, jasmonate recruits the endoplasmic reticulum associated degradation (ERAD) quality control system to manage the production of triterpene saponins, widespread compounds that share a biogenic origin with sterols. An ERAD-type RING membrane-anchor E3 ubiquitin ligase is co-expressed with saponin synthesis enzymes to control the levels of the rate-limiting 3-hydroxy-3-methylglutaryl-CoA reductase to prevent unrestrained bioactive saponin accumulation and secure plant integrity. The down regulation of this key saponin regulatory gene results in accumulation of monoglycosylated saponin building blocks in transgenic *M. truncatula* hairy roots as a consequence of deficient, ectopic, or *de novo* biosynthesis of triterpene saponins. This control apparatus is equivalent to the ERAD system that regulates sterol synthesis in yeasts and mammals, but that employs distinct E3 ubiquitin ligases. Hence, the general principles for management of triterpene biosynthesis are conserved across eukaryotes but can be controlled by distinct regulatory cues.
Introduction

Jasmonates (JAs) are ubiquitous oxylipin-derived phytohormones essential for the regulation of multiple plant processes, encompassing development, growth and defense. Across the plant kingdom, JAs act as elicitors for the production of secondary metabolites, many of which serve as first-line defense against pathogen or herbivore attack (Wasternack, 2007; Pauwels et al., 2009; De Geyter et al., 2012). The early signaling machinery involved in JA-mediated secondary metabolite elicitation consists of a conserved module for JA perception and subsequent 'primary' signal transduction (Browse, 2009; Fonseca et al., 2009; Memelink, 2009; Pauwels and Goossens, 2011). Each plant, however, possesses a specific compendium of bioactive metabolites of a wide structural variety and different biochemical origin. During evolution, these species-specific secondary metabolic pathways evolved under JA control, downstream of the conserved hormone perception machinery (Pauwels et al., 2009; De Geyter et al., 2012).

Saponins comprise a large and very diverse group of plant natural products that serve in planta as anti-fungal or anti-insect agents, but also display a wide range of biological activities with different pharmaceutical properties (Augustin et al., 2011; Osbourn et al., 2011; Pollier et al., 2011a). Saponins are amphipathic glycosides with triterpene or steroid backbones that can be synthesized constitutively or inducibly, depending on species, developmental stage, or environmental conditions. Saponins share a common biogenic origin with the sterols, molecules that are ubiquitous in eukaryotic cells. They are both derived from 2,3-oxidosqualene and depend on the cytosolic pathway for isopentenyl diphosphate (IPP) supply that involves mevalonate (MVA) as a key intermediate (Chappell, 2002; Augustin et al., 2011; Osbourn et al., 2011; Pollier et al., 2011a). The two pathways diverge after 2,3-oxidosqualene, which can be cyclized either by cycloartenol synthase to yield cycloartenol, the sterol precursor, or by β-amyrin synthase (bAS) to yield β-amyrin, the precursor of many classes of triterpene saponins (Fig. 1). Both cyclases share extensive sequence homology, suggesting an evolutionary recruitment of (at least some) triterpene saponin biosynthesis genes from sterol metabolic 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. The endoplasmic reticulum (ER)-associated protein degradation (ERAD) quality control system directs the feedback-regulated destruction of the HMGR enzyme, both in yeast (Saccharomyces cerevisiae) and mammals, and involves the RING E3-ligase HMGR DEGRADATION 1 (HRD1/DER3) in yeast or its homologs Hrd1/SYVN1 and gp78 in humans (Hampton and Garza, 2009; Hirsch et al., 2009; Jo and Debose-Boyd, 2010; Burg and Espenshade, 2011). Both yeast and humans have other, distinct, E3-ligases, such as the human RING-FINGER PROTEIN WITH MEMBRANE ANCHOR 1 (RMA1) that participate in ERAD but not in the regulation of HMGR stability (Hirsch et al., 2009).

Triterpene saponin accumulation in plant cells is stimulated in the defense responses that are triggered by JAs, but as for most other plant natural products, the underlying molecular mechanisms remain largely unknown. Here, we identified a RMA1-like E3 ubiquitin ligase, called MAKIBISHI1 (MKB1), that is co-induced with the triterpene saponin biosynthetic enzymes following JA elicitation in the model legume Medicago truncatula. MKB1 operates in an ERAD-like manner to regulate the triterpene saponin biosynthetic pathway and is essential for the self-protection of M. truncatula tissues against the bioactive saponins.

Results

Co-regulation of biosynthetic and regulatory genes.

We used genome-wide transcript profiling to identify new regulators of plant triterpene saponin biosynthesis. Using cDNA-AFLP, we monitored over time the methyl jasmonate (MeJA)-modulated transcriptome of suspension-cultured M. truncatula cells, in which MeJA-inducible triterpene saponin biosynthesis has been reported before (Suzuki et al., 2002).
Figure 1. The triterpene saponin biosynthesis pathway in *Medicago truncatula*. HMG, 3-hydroxy-3-methylglutaryl and P450s, cytochrome P450 monooxygenases.
The expression of a total of 8,462 transcripts was visualized and 282 MeJA-responsive tags were identified. Average linkage hierarchical clustering showed that gene tags corresponding to HMGs, geranyl pyrophosphate synthase, squalene synthase, squalene epoxidase, bAS and CYP93E2, which encode enzymes catalyzing steps in triterpene saponin biosynthesis (Fig. 1) displayed an almost identical, MeJA-induced, expression pattern, suggesting a tight co-regulation (Fig. 2a).

**Figure 2.** Expression of the RMA-type E3 ubiquitin ligase, MKB1 is coregulated with that of triterpene saponin enzymes. (a) Subcluster of the Mt transcriptome comprising tags corresponding to genes reported to be involved in triterpene biosynthesis or jasmonate signal transduction. 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. (b) Screenshot of the coexpression analysis of *M. truncatula* MKB1 (green), HMGRI (red), HMGRI4 (orange), and HRD1 (black) with the MtGEA software.

Several tags corresponding to transcription factors, protein kinases and E3 ubiquitin ligases had maximal transcriptional upregulation within 30 min to 2 h after elicitation, prior to or concurrent with the onset of genes involved in triterpene saponin synthesis, and hence represent potential regulators of this metabolic pathway. Within this gene set a MYC-like bHLH protein and several homologs of the JAZ repressor proteins, known elements of the core JA signaling module were present (Browse, 2009; Fonseca et al., 2009; Memelink, 2009; Pauwels et al., 2010; Pauwels and Goossens, 2011), as well as a tag (MT067) corresponding to an RMA-like E3 ubiquitin ligase (Hirsch et al., 2009), and denominated
Protein quality control in *Medicago truncatula*

MKB1 (Fig. 2a, Fig. 3a-c). The early MeJA response of MKB1 and HMGR was verified and confirmed in the *M. truncatula* Gene Expression Atlas (MtGEA; (Kushiro et al., 2006) http://bioinfo.noble.org/gene-atlas/ (He et al., 2009)) (Fig. 2b).

**Figure 3.** Sequence and structural analysis of eukaryotic RMA and HMGR proteins. (a) Phylogenetic analysis of MKB1 and other RMA E3 Ub-ligases. The percentage of replicate trees that clustered together in the bootstrap test is shown next to the branches. The scale bar indicates the number of amino acid substitutions per site. *Arabidopsis thaliana* (At), *Capsicum annuum* (Ca), *Caenorhabditis elegans* (Ce) and *Homo sapiens* (Hs) amino acid sequences were retrieved from GenBank. Amino acid sequences of *M. truncatula* MKB1 and homologous proteins (prefix TC) were retrieved from the *Medicago truncatula* Gene Index following BLAST searches. (b) Schematic representation of the MKB1 protein and its domain structure. (c) Comparison of the amino acid sequence of MKB1 with that of RMA proteins from *A. thaliana*, *C. annuum*, *C. elegans* and *H. sapiens*. Conserved amino acids that are identical in the seven proteins are indicated with an asterisk. (d) Kyte & Doolittle hydropathy plot of MKB1 and Arabidopsis Rma2 (At4g27470) with window size 7. Red bars indicate the membrane anchor.
Toxic saponins cause the ‘makibishi’ phenotype.

To assess the potential role of MKB1 in JA signaling or the regulation of triterpene synthesis, we generated transgenic *M. truncatula* hairy roots in which *MKB1* was overexpressed (MKB1\(^\text{OE}\) lines) or silenced by means of a hairpin RNAi construct (MKB1\(^\text{KD}\) lines) (Fig. 4a). The MKB1\(^\text{KD}\) lines showed a striking phenotype, in particular when they were transferred to liquid medium for upscaling. In its most pronounced form, silencing of the *MKB1* gene caused ‘dissociation’ of the hairy roots in ‘caltrop-like’ structures (Fig. 4b-c), hence the name ‘makibishi’, which is the Japanese term for caltrop. No such caltrop phenotypes were observed in MKB1\(^\text{OE}\) lines or hairy root lines transformed with control constructs. Microscopic analysis of the MAKIBISHI phenotype in the MKB1\(^\text{KD}\) lines revealed severe root growth and epidermal deficiencies (Fig. 4d-f). Notably, the MKB1\(^\text{KD}\) root phenotype strongly resembled the stunted root morphology and epidermal defects of the oat (*Avena sativa*) mutants *sad3* and *sad4*, that accumulate the incompletely glycosylated form of the saponin avenacin A-1 (Mylona et al., 2008).

To verify whether the MKB1\(^\text{KD}\) root phenotype was correlated with altered metabolism, we performed metabolite profiling by liquid chromatography electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (LC-ESI-FT-ICR-MS) (Pollier et al., 2011b). Processing of the mass spectrometry data yielded 16,278 m/z peaks. Principal component analysis indicated that the MKB1\(^\text{KD}\) samples were clearly different from those of the control and MKB1\(^\text{OE}\) lines, whereas no significant differences were detected between control and MKB1\(^\text{OE}\) roots (Fig. 5a-b). These observations were also confirmed by pairwise comparative analysis (data not shown).
Protein quality control in *Medicago truncatula*  

**Figure 4.** The MAKIBISHI phenotype. (a) *MKB1* transcript steady-state levels in transgenic *M. truncatula* control (CTR), *MKB1*<sub>OE</sub> (OE) and *MKB1*<sub>KD</sub> (KD) lines. (b) Roots of control (CTR), OE and KD lines grown on solid medium. (c) Fluorescent microscopy analysis of roots grown in liquid medium. (d) Longitudinal optical sections obtained by confocal microscopy of roots grown in liquid medium. (e) Scanning electron microscopy analysis of roots grown in liquid medium. (f) Transmission electron microscopy analysis of cross sections of roots from CTR and KD lines grown in liquid medium. Arrows indicate wavy/thickened appearance of cell margins in *MKB1*<sub>KD</sub> roots.

A supervised partial least squares discriminant analysis model to separate the *MKB1*<sub>KD</sub> samples was created (Fig. 5c) and used to generate an S-plot to identify the peaks responsible for the observed differences (Fig. 5d). 83 down-regulated and 37 up-regulated m/z peaks, corresponding to 41 and 21 compounds, respectively, were identified as significantly different, among which were triterpene saponins and phenolic compounds, the two main classes of metabolites known to accumulate in *M. truncatula* hairy roots (Dixon and Sumner, 2003). The most striking trend was observed with the saponins. First, they were most represented within the down- and upregulated compounds (15 and 7, respectively). Second, the majority of the upregulated saponins were monoglycosylated compounds, such as Hex-medicagenic acid, Hex-bayogenin and Hex-hederagenin, whereas higher glycosylated forms were more represented among the downregulated peaks (Fig. 5a,d-e and Table 1).
Protein quality control in *Medicago truncatula* 7

**Figure 5.** *MKB1* silencing causes ectopic accumulation of monoglycosylated triterpene saponins. (a) Detail of the full MS scan of root extracts from a control line (black) and a *MKB1*<sup>KD</sup> line (red). The peak at *t*<sub>R</sub> 26.71 represents Hex-medicagenic acid. (b) Principal component analysis and (c) partial least squares discriminant analysis (PLS-DA) of samples from *MKB1*<sup>KD</sup> (red), *MKB1*<sup>OE</sup> (blue) and control (black) lines. (d) S-plot derived from PLS-DA. Peaks in the lower left and upper right quadrants (marked by dotted red lines) correspond to metabolites that are higher and lower abundant in the *MKB1*<sup>KD</sup> samples, respectively. Peaks with an absolute covariance value above 0.03 and an absolute correlation value above 0.6 were considered as significantly different. (e) Average total ion current of the peaks colored in red in the S-plot and corresponding to soyasaponin I (upper) and Hex-medicagenic acid (lower). Error bars indicate the standard error (n=4). (f) Detail of the full MS scan of the medium from a control line (black) and a *MKB1*<sup>KD</sup> line (red). The peak at *t*<sub>R</sub> 27.95 represents Hex-medicagenic acid.

Remarkably, LC-ESI-FT-ICR-MS analysis of the growth medium of *MKB1*<sup>KD</sup> roots revealed the presence of tens of compounds, whereas in the chromatograms of the growth medium from control or *MKB1*<sup>OE</sup> roots, metabolites were totally absent (Fig. 5f), indicating release of compounds from the *MKB1*<sup>KD</sup> roots, including of the three monoglycosylated saponins. These findings suggest that silencing of *MKB1* primarily causes deficient, ectopic, or *de novo* biosynthesis of triterpene saponins, which leads to the overaccumulation and release of incompletely glycosylated saponins, which correlates with the occurrence of the *sad3/sad4*-like makibishi phenotype.
Table 1. LC-ESI-FT-ICR-MS analysis of MKB1\textsuperscript{KD} roots.

<table>
<thead>
<tr>
<th>Downregulated compounds</th>
<th>Upregulated compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Rha-Gal-GlcA-soyasapogenol B (Soyasaponin I)</td>
<td>Afrormosin + malonyl hexose</td>
</tr>
<tr>
<td>Hex-Hex-medicagenic acid</td>
<td>Hex-medicagenic acid</td>
</tr>
<tr>
<td>Formononetin + malonyl hexose</td>
<td>Hex-bayogenin</td>
</tr>
<tr>
<td>Naringenin + hexose</td>
<td>dHex-Hex-Hex-hederagenin</td>
</tr>
<tr>
<td>Hex-Hex-Hex-medicagenic acid</td>
<td>Hex-isoflavone (biochanin A?)</td>
</tr>
<tr>
<td>Malonic acid + pentose + hexose derivate</td>
<td>?-Hex-isoflavone (biochanin A?)</td>
</tr>
<tr>
<td>Malonic acid + pentose derivate?</td>
<td>Hederagenin</td>
</tr>
<tr>
<td>Malonyl-Hex-HexA-bayogenin</td>
<td>Tryptophan??</td>
</tr>
<tr>
<td>Rha-Gal-GlcA-soyasapogenol E</td>
<td>Afrormosin</td>
</tr>
<tr>
<td>Malonyl-dHex-Hex-HexA-soyasapogenol B</td>
<td>Malonyl-Hex-bayogenin</td>
</tr>
<tr>
<td>Naringenin + hexose + malonic acid</td>
<td>Hex-methyl hydroxy benzoic acid</td>
</tr>
<tr>
<td>Hex-HexA-hederagenin</td>
<td>Medicarpin</td>
</tr>
<tr>
<td>Malonyl-HexA-aglycone A</td>
<td>Vanillic acid conjugate</td>
</tr>
<tr>
<td>Hex-HexA-bayogenin</td>
<td>Malonyl-Hex-malonyl-Hex-medicagenic acid</td>
</tr>
<tr>
<td>Hex-HexA-aglycone A</td>
<td>Vanillic acid conjugate</td>
</tr>
<tr>
<td>Hex-HexA-soyasapogenol E</td>
<td>Vanillic acid conjugate</td>
</tr>
<tr>
<td>Hydroxymethoxybenzoic acid +</td>
<td>Malonyl-Hex-malonyl-Hex-hederagenin</td>
</tr>
<tr>
<td>hydroxybenzoic acid + 276 Da</td>
<td></td>
</tr>
<tr>
<td>Hex-Hex-bayogenin (+ 46 Da adduct)</td>
<td></td>
</tr>
<tr>
<td>Hex-HexA-hederagenin</td>
<td></td>
</tr>
<tr>
<td>Malonyl-Hex-dHex-medicagenic acid</td>
<td></td>
</tr>
<tr>
<td>Lariciresinol + hexose + malonic acid?</td>
<td></td>
</tr>
<tr>
<td>Malonyl-? + carboxybenzoic acid (phthalic acid or isomer)</td>
<td></td>
</tr>
<tr>
<td>Hex-HexA-aglycone A</td>
<td></td>
</tr>
<tr>
<td>Dihydroxyflavone or hydroxyflavonol derivate?</td>
<td></td>
</tr>
</tbody>
</table>

To establish that the accumulation of the incompletely glycosylated saponins was the cause and not the consequence of the root defects, as was also the case with the \textit{sad3} and \textit{sad4} oat mutants (Mylona et al., 2008), we tried to ‘phenocopy’ the MKB1\textsuperscript{KD} phenotype by adding different metabolites to the culture medium of control roots. First, aliquots of the culture medium of MKB1\textsuperscript{KD} roots containing the incompletely glycosylated saponins were applied resulting in epidermal loosening of control roots, whereas similar or higher doses of medium from control roots devoid of saponins did not have an effect (Fig. 6a). Second, we purified 3-O-Glc-medicagenic acid and medicagenic acid from \textit{M. sativa}.  

---

---
roots (Oleszek et al., 1990) and applied them to control roots. A transient phenocopy was obtained after addition of 3-O-Glc-medicagenic acid (Fig. 6b) but not of purified medicagenic acid (data not shown). No phenocopy was observed after treatment with phytohormones, such as auxins, brassinosteroids, gibberellins, cytokinins, and JAs, or commercially available compounds such as the flavonoid quercetin or the fully glycosylated soyasaponin I, the most abundant compound in control hairy roots and also abundantly present in the medium of MKB1\(^{KD}\) roots, supporting 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 6.** Phenocopy of the MKB1\(^{KD}\) phenotype. (a) Light and fluorescent microscopy analysis of control hairy roots incubated for 1 week in medium supplemented with medium from control (CTR) or MKB1\(^{KD}\) root lines. (b) Light microscopy analysis of control hairy roots incubated for 1 day in medium supplemented with purified 3-O-Glc-medicagenic acid (GMA) or an equivalent amount of the solvent (EtOH).
Protein quality control in *Medicago truncatula*

These observations are also in agreement with structure-activity relationship studies that indicate that the monodesmoside derivatives of medicagenic acid possess a higher hemolytic activity, as well as a higher anti-fungal and allelopathic activity than the related bi- and tridesmosides (Oleszek, 1990; Augustin et al., 2011). Hemolytic activity is generally attributed to membrane bursts caused by the interaction between the saponins and the sterols of the erythrocyte membrane (Augustin et al., 2011; Osbourn et al., 2011).

**Transcriptional feedback for plant survival.**

We assessed whether the effects of loss of MKB1 function on the metabolite profiles, were reflected in the profile of gene transcripts corresponding to enzymes involved in the triterpene saponin pathway. Remarkably, in the MKB1\(^{KD}\) roots the expression of all of the hitherto known specific triterpene saponin biosynthetic genes (*bAS*, *CYP93E2*, *CYP716A12*, *UGT73F3* and *UGT73K1*) was strongly downregulated (Fig. 7). A similar downregulation was also observed with *HMGR1* and *HMGR4*, the two HMGRs of *M. truncatula* that are most coregulated with MKB1. In contrast, expression of none of the remaining *M. truncatula* HMGR genes, nor of the *SQS*, *SQE* genes, corresponding to enzymes catalyzing triterpene synthesis up to the oxidosqualene precursor, was markedly altered in the MKB1\(^{KD}\) lines (Fig. 7).

Similarly, none of the putative *M. truncatula* orthologs of genes corresponding to sterol biosynthesis enzymes from *Arabidopsis*, was differentially expressed in the MKB1\(^{KD}\) lines (Fig. 8), indicating that alteration in the expression of MKB1 did not affect the transcriptional regulation of sterol biosynthesis or triterpene synthesis in general. In agreement with this, we did not detect any differences in sterol levels (data not shown). The expression of none of the tested triterpene pathway genes differed in MKB1\(^{OE}\) lines either (Fig. 7 and Fig. 8), which is in accordance with the lack of any observable (metabolic) phenotype. These findings point towards the occurrence of a saponin specific feedback mechanism, caused by the absence of MKB1 activity and its requirement to cope with the consequent ectopic accumulation of the toxic monoglycosylated saponins.
Figure 7. Expression of saponin synthesis genes in transgenic *Medicago truncatula* hairy roots. qRT-PCR analysis of saponin biosynthetic genes in control (CTR), MKB1\textsuperscript{OE} (OE) and MKB1\textsuperscript{KD} (KD) lines.
Figure 8. Expression of sterol synthesis genes in transgenic *M. truncatula* hairy roots. qRT-PCR analysis of sterol biosynthetic genes in control (CTR), MKB1^OE^ (OE) and MKB1^KD^ (KD) lines. A schematic overview of the sterol biosynthesis pathway and the qRT-PCR patterns are shown (left and right, respectively).
The RMA-type MKB1 possesses HRD-type activity.

A. thaliana and mammalian RMA proteins localize to the ER and possess E3 ubiquitin ligase activity \textit{in vitro} (Matsuda et al., 2001; Lee et al., 2009). Similar to the \textit{Arabidopsis} RMA proteins, the MKB1 protein contains a single N-terminal C3HC4-type RING domain and a putative membrane anchor at the C-terminus of the protein (Fig. 3d). To verify the activity of MKB1, a recombinant, truncated version of the MKB1 protein, lacking the membrane anchor domain (MKB1ΔC), was shown to possess self-ubiquitination activity in an \textit{in vitro} assay (Fig. 9a), whereas a mutated version of the recombinant MKB1ΔC protein, in which the essential amino acid residues Cys\textsuperscript{37} and Cys\textsuperscript{40} were substituted by Ser residues (Fig. 3b) did not.

To verify the localization of MKB1, we produced GFP-tagged versions of both the full-length (FL) MKB1 and the truncated MKB1ΔC in bombarded onion (\textit{Allium cepa}) cells. The GFP-MKB1 protein was visible in a network pattern, as \textit{Arabidopsis} RMA1 (Matsuda et al., 2001; Lee et al., 2009), whereas this pattern was lost and replaced by a cytosolic distribution pattern when the GFP-MKB1ΔC protein was produced (Fig. 9b). The fluorescence of the FL GFP-MKB1 fusion protein was always very weak, presumably because of the self-ubiquitination activity, which hampered confocal microscopy analysis. Therefore, the localization of GFP-MKB1 was also determined in yeast cells, in which co-localization with a known ER protein was confirmed (Fig. 9c). Thus, MKB1 corresponds to an active, ER-localized, E3 ubiquitin ligase.

\textit{Arabidopsis} RMA1 protein has recently been shown to regulate aquaporin levels by ubiquitination and subsequent proteasomal degradation (Lee et al., 2009). 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). Humans also have two other, non-related, HRD-type E3 ubiquitin ligases that besides being involved in the ERAD quality control system also play a crucial role in the control of sterol synthesis through the regulation of HMGR levels (Hampton and Garza, 2009; Hirsch et al., 2009; Jo and Debose-Boyd, 2010; Burg and Espenshade, 2011). To our knowledge, a role for human RMA proteins in the latter process has not been reported.
yet. Similarly, yeast does not possess RMA-type proteins and only employs HRD-type E3 ubiquitin ligases for sterol control. Despite this and the apparent lack of sequence similarity between the two types of E3 ubiquitin ligases and the different membrane topology of the HMGR enzymes in plants, yeasts and mammals (Fig. 10), we reasoned that the RMA-type MKB1 might survey triterpene saponin synthesis in *M. truncatula* by targeting HMGR.

**Figure 9.** MKB1 is an ER-localized protein with autoubiquitination 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 with 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). (c) Subcellular localization of MKB1 in yeast cells. The pictures show the signal of GFP-MKB1, SEC13-RFP, and the merged image, respectively.
Two lines of indirect evidence supported this hypothesis. Besides the correlation between the MKB1\textsuperscript{KD} root phenotype and the ectopic accumulation of products from the HMGR-dependent triterpene saponin pathway, we observed an extraordinarily high expression correlation between MKB1 and two of the \textit{M. truncatula} HMGR genes. Such a correlation in expression was not observed between these HMGR isoforms and the putative \textit{M. truncatula} homolog of yeast HRD1 (Fig. 2b). To assess whether MKB1 functionality affects HMGR stability, we checked the accumulation of HMGR proteins in \textit{M. truncatula} roots using immunoblot analysis with polyclonal antibodies raised against the conserved catalytic domain of \textit{Arabidopsis} or melon (\textit{Cucumis melo}) HMGR proteins (Kobayashi et al., 2002; Leivar et al., 2011). With both antibodies, a small but consistent
increase in HMGR proteins was detected in the \(\text{MKB}1^{\text{KD}}\) roots as compared to control roots (Fig. 11a-b), pointing towards a potential inverse correlation between \(\text{MKB}1\) and HMGR proteins. Unexpectedly however, HMGR activity was markedly decreased in the \(\text{MKB}1^{\text{KD}}\) roots as compared to control roots (Fig. 11c), indicating that higher HMGR protein levels did not lead to higher enzymatic activity. We speculate that this effect is caused by a negative feedback control system triggered by the loss of \(\text{MKB}1\) activity, the resulting decreased HMGR turnover and the consequent ectopic accumulation of harmful triterpene saponins. A similar inverse correlation between HMGR activity and triterpene levels was recently observed in transgenic \(\text{Taraxacum brevicorniculatum}\) plants in which the synthesis of rubber (mainly a polymer of \(\text{cis}-1,4\)-isoprene) was targeted by an RNAi approach. Besides a significant reduction in rubber biosynthesis, a 50% increase in the levels of pentacyclic triterpenes was observed, accompanied by a 50% decrease in HMGR activity, which was postulated to reflect feedback inhibition from oidosqualene-derived products or precursors thereof (Post et al., 2012). Multilevel control of HMGR in yeast and mammals has been well documented (Hampton and Garza, 2009; Jo and Debose-Boyd, 2010; Burg and Espenshade, 2011), recently also reported in plants (Leivar et al., 2011), and may, for instance, involve phosphorylation.

To further support that \(\text{MKB}1\) can target HMGR proteins for degradation, we performed three sets of additional experiments. In the first, we assessed HMGR transcript and HMGR protein levels before and after MeJA application. MeJA application dramatically enhanced HMGR transcript steady state levels, in particular of \(\text{HMGR}1\) (Fig. 11d), whereas HMGR protein levels remained stable (Fig. 11e), supporting the existence of a negative posttranscriptional control of HMGR levels.

Secondly, we produced firefly luciferase (fLUC)-tagged versions of \(\text{M. truncatula}\) HMGR proteins in tobacco (\(\text{Nicotiana tabacum}\)) protoplasts combined or not with \(\text{MKB}1\). Cotransfection with \(\text{MKB}1\) decreased the HMGR1-fLUC and HMGR3-fLUC signal by almost 2-fold, whereas it did not reduce the levels of HMGR2-fLUC or free fLUC activity (Fig. 11f). Hence, tagging of fLUC with particular \(\text{M. truncatula}\) HMGR isoforms converts it into a target of \(\text{MKB}1\)-mediated protein degradation.
Figure 11. The RMA-like MKB1 E3 ubiquitin ligase targets the HMGR enzyme for degradation. Immunoblot analysis of 15µg and 10µg total protein was performed with polyclonal antibodies against the conserved catalytic domain of (a) melon and (b) Arabidopsis HMGR proteins in control (CTR) and MKB1\textsuperscript{KD} lines (KD), respectively. (c) Specific HMGR activity in \textit{M. truncatula} roots, relative to the activity in CTR1 line. (d) HMGR transcript steady state levels in mock or MeJA-treated \textit{M. truncatula} roots. (e) HMGR protein levels in 15µg loading, represented as average of two biological repeats in mock or MeJA-treated \textit{M. truncatula} roots. (f) Degradation of \textit{M. truncatula} HMGR-flLUC fusion proteins in tobacco protoplasts. Percentage stability relative to the flLUC value measured in the absence (+) or presence (+M) of MKB1. Error bars indicate the standard error (n=24). Statistical significance was determined by Student’s t-test (** P<0.01). (g) Complementation of \textit{HRD1} (H) or \textit{hrd1} (h) yeast cells transformed with MKB1 (+M) or MKB1mut (+m), spotted in a 10-fold dilution series on SD medium lacking URA supplemented (SD-U+LOV) or not (SD-U) with lovastatin. Empty vector control (-). (h) Immunoblot analysis for 6myc-tagged Hmg2p and Coomassie Blue staining of total protein extracts from RHY400 (H) and RHY401 (h) transformed with MKB1 (+M) or MKB1mut (+m). The destination vector was used as a control (-).
Finally, we checked whether MKB1 could target yeast HMGR and thereby complement a yeast strain devoid of functional HRD1, despite the lack of sequence and topology similarity between both the E3 ubiquitin ligases and the HMGR enzymes from yeasts and plants (Fig. 10). Complementation was scored by determining the ability of transformed strains to grow on medium containing lovastatin, an inhibitor of HMGR (Fig. 1). Wild-type yeasts with normal HRD1 activity have low HMGR levels and therefore are sensitive to this compound, whereas hrd1 yeasts with higher HMGR levels can overcome this growth inhibition (Hampton and Garza, 2009). The lovastatin tolerance of the hrd1 strains was returned to wild-type 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 (Fig. 11g). Similar results were obtained independent of the yeast strain used, either containing only the native yeast HMGR proteins, or in which both native HMGRs had been replaced by a single 6myc-tagged Hmg2p copy (Fig. 11g). The latter genetic background also allowed monitoring the levels of yeast HMGR proteins in the strains transformed with the different MKB1 constructs. In the wild-type HRD1 yeast background, the tagged Hmg2p was nearly absent, whereas in the hrd1 strain, it could still be detected. In agreement with the lovastatin growth complementation assay, expression of wild-type MKB1, but not the ligase-dead MKB1mut, restored the Hmg2p levels to those of the HRD1 strain (Fig. 11h). Together, these results indicate that MKB1 can functionally complement the yeast HRD1 and, thereby, target the yeast HMGR for ubiquitin-mediated degradation.

**Discussion**

Plants respond to insect or pathogen attack by activating defense programs that include the production of bioactive secondary metabolites to eliminate the attackers. These compounds are often also toxic to the plants themselves (Goossens et al., 2003; Sirikantaramas et al., 2009), hence their biosynthesis and accumulation should be strictly regulated, spatially and temporally, quantitatively and qualitatively, to optimize plant survival ‘in good and bad times’. Ectopic accumulation of bioactive or phytotoxic
intermediates of the triterpene saponin and steroidal alkaloid pathways in oat and tomato (*Solanum lycopersicum*) mutants, respectively, has dramatic effects on growth performance (Mylona et al., 2008; Itkin et al., 2011). Our findings reveal an efficient, hitherto undetected, metabolite surveillance system that controls the pace of bioactive triterpene saponin synthesis and thereby safeguards plant integrity. In the model legume *M. truncatula*, this surveillance system has been recruited from the ERAD protein quality control system, which is a conserved molecular machinery in eukaryotes, but in this species has evolved to control the activity of the HMGR enzymes.

Triterpene saponins share a common biogenetic origin with the sterols and are derived from oxidosqualene which is generated by the cytosolic MVA pathway. In yeast and mammals the activity of HMGR is regulated from the transcriptional to the post-translational level and many of the regulatory mechanisms involved have been characterized. However, in plants little is known about the regulatory mechanisms that control HMGR activity. Regulation at the transcriptional and posttranslational level in response to environmental and developmental factors, or to perturbations of the metabolic flux through the sterol pathway has been reported (Nieto et al., 2009). Accordingly, in MeJA-elicited *M. truncatula* cells, transcriptional upregulation of the MVA pathway was reflected by the co-activation of different isoforms of the *HMGR* gene with known saponin biosynthetic genes, such as *bAS* and several *UGTs*. Similarly, posttranslational regulation has been demonstrated by an increase in HMGR activity without changes in the HMGR transcript level (Nieto et al., 2009). The study of plant SNF1-related protein kinases (SnRKs) that phosphorylate and thereby inactivate Arabidopsis HMGR reports on a possible posttranslational regulatory mechanism of plant HMGR (Dale et al., 1995). Here, we show that MeJA-modulated (transcriptional) elicitation of saponin biosynthesis in *M. truncatula* co-regulates with the transcriptional induction of a RING E3 Ub ligase, MKB1, that controls the stability of HMGR through a mechanism that is analogous to the ERAD quality control system that directs the sterol-regulated destruction of the HMGR enzyme 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
Protein quality control in *Medicago truncatula*

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. Nonetheless, MKB1 can complement a yeast *hrd1* mutant, implicating 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 employ the molecular machinery employed by HRD1 to do so.

There are common mechanistic principles between the yeast/mammalian HRD-type and the *M. truncatula* RMA-type ERAD ligases in the control of HMGR stability. In mammalian cells, coexpression of Insig proteins is needed to restore the resistance of overexpressed HMGR to sterol-induced ubiquitination and degradation, which suggests that ectopic reductase saturates the endogenous ERAD machinery. Accordingly, RNAi against Insigs 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 have any effect on 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, is apparently capable of provoking 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 toxic to the cells as exemplified by the dramatic phenotype of MKB1<sup>KD</sup> roots. Hence, co-induction of *MKB1* by MeJA treatment might be considered as a means to prevent saturation of the ERAD system and to control or restrict the flux to the saponin pathway to ensure proper advance of the triterpene biosynthetic process.

Plant SnRKs phosphorylate and inactivate HMGR e.g. Arabidopsis HMG1 at a site (Ser-577) equivalent to that at which mammalian AMPK, a major metabolic regulator in mammalians, phosphorylates mammalian HMGR (Dale et al., 1995). Hence, conservation
between the activity of the modifier and the primary sequence of the physiological substrate between plants and mammalians might occur, although the signaling molecule that activates plant SnRK is not AMP as it is for the mammalian AMPK. In case of the ERAD control of HMGR stability, it appears unlikely that such conservation of the primary sequence of the substrate exists. 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 mammalians, 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 HMGRs is highly divergent: little primary sequence conservation can be observed, the amino acid residues important for the sterol-sensing in animal HMGRs are not conserved, and the N-terminal domain is truncated and lacks the complex membrane-spanning architecture, i.e. plant HMGRs only possess 2 membrane spanning domains whereas yeast and human HMGRs contain 7 and 8 transmembrane domains, respectively (Learned and Fink, 1989). Hence, recognition of the plant HMGRs by RMA-type ligases such as MKB1 might have evolved alternative recognition and ubiquitinylation sites. Alternatively, 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 was found for yeast Hmg2p (Gardner and Hampton, 1999; Hampton, 2002). Though plant HMGRs have a truncated membrane domain, such structural features may have been conserved. The latter hypothesis is further supported by the observation that despite the structural resemblance, only limited primary sequence resemblance between the membrane domains of yeast and mammalian HMGR proteins was observed (<20% identity over 340 amino acids; DeBose-Boyd, 2008).

Remarkably, *M. truncatula* uses an ERAD system different from the one that directs the sterol-regulated destruction of the HMGR enzymes in yeast and mammals. Nonetheless, despite the marked differences in sequence and topology between the plant RMA and yeast HRD1 proteins, as well as between the respective target HMGR enzymes, the
MKB1 RMA-type E3 ligase can take over at least some of the functions of the yeast HRD-type E3 ligase protein in the absence of the latter, and thus seems compatible with the connected molecular machinery. In addition to sterols in mammals, both yeast and mammalian cells appear to use IPP-derived non-sterols to regulate HMGR stability or degradation (Sever et al., 2003; Garza et al., 2009). The divergent sequences of the plant proteins might have allowed the evolution of a plant-specific gateway to the ERAD-mediated control of HMGR stability, for instance regulated by specific saponin intermediates. Possibly MKB1 may manage the activity of more ER-localized enzymes that catalyze steps in triterpene saponin synthesis, such as the Cytochrome P450 monooxygenases that hydroxylate the β-amyrin backbone, and in this way guarantee self-protection of the plant from its own arms.

Materials and Methods

*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.

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 BstYI+1/MseI+2 primer combinations. 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 qRT-PCR total RNA was extracted with the RNasy 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).

**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, HMGR2 and HMGR3 sequences (Genbank accession EU302813, EU302814 and EU302815, respectively) (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 in onion cells, Gateway recombination was carried out with the p2GW7 and pK7WGF2 vectors, respectively.

For recombinant protein production, the MKB1 sequences were recombined in pDEST15 expression vector and the resulting clones transformed to E. coli BL21 (DE3) cells. For the yeast complementation and localization 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 HMGR ORFs were fused at their C-terminus with the firefly luciferase ORF by a fusion PCR and Gateway recombined in the p2GW7 vector.

**Phylogenetic analysis.** The protein sequences were aligned with ClustalW and the resulting alignments were manually adjusted. 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).

**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 (Pollier et al., 2011b). The generated hairy root lines 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.

**Phenotypic analysis of transgenic M. truncatula hairy roots.** Samples for scanning electron microscopy (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 (OsO₄) solution (Fluka) for 2 hours, and subsequently subjected to a dehydration 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 transmission electron microscopy (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) OsO₄ 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.

**Metabolite 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. Processing and metabolite extraction from hairy root tissue was performed as described (Pollier et al., 2011b).
Protein quality control in *Medicago truncatula*

To remove salts from the samples of the culture medium, 1 ml of medium was brought on a 100 mg Extract-Clean™ 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 eluate was evaporated to dryness under vacuum and the residue dissolved in 200 µl water for analysis.

LC ESI FT-ICR MS was carried out as described (Pollier et al., 2011b). Elucidation of the MS^n spectra was according to (Pollier et al., 2011b) for the saponins and (Morreel et al., 2006; Morreel et al., 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, mzdif=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.

**Ubiquitination assay.** Recombinant GST-MKB1 fusion proteins (truncated with or without mutation) were purified from transformed *E. coli* cells pretreated for 2 hours with isopropyl-β-D-1-thiogalactopyranoside (IPTG), with Glutathione Sepharose™ 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 MgCl₂, 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 RGS/penta/tetra His antibody (Qiagen) and α-GST (GE Healthcare) antibodies.

**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 before analysis by confocal microscopy.

**Confocal microscopy.** Transformed BY4742 yeast cells (with an integrated RFP tagged Sec13 protein for ER and Golgi marking) (Huh et al., 2003) and bombarded onion slices were analyzed by confocal microscopy with the FV10 ASW Olympus Confocal with a water immersion 63× objective.

**Analysis of *M. truncatula* HMGR protein levels and activity.** Protein extraction from *M. truncatula* hairy roots was carried out as described (Leivar et al., 2011). Determination of HMGR protein levels by immunoblot analysis with polyclonal antibodies raised against the conserved catalytic domain of *Arabidopsis* or melon (*Cucumis melo*) HMGR proteins was performed essentially as described (Kobayashi et al., 2002; Leivar et al., 2011). Determination of HMGR-specific activity was carried out as described (Leivar et al., 2011).

**HMGR degradation assays in tobacco 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).
Protein quality control in *Medicago truncatula*

**Yeast complementation and protein degradation assays.** Two sets of *S. cerevisiae* strains were used for the complementation and protein degradation assays, namely strains YWO1167 (W303 *Mat a, ura3-1, his3-11,15, leu2-3,112, trp1-1, ade2-1ocre, can1-100, prc1-1, doa10::KanMX*) and its *hrd1* (*Δ*der3/*hrd1::HIS3*) knock-out (YWO1528), on the one hand, and RHY400 (*Mat a, ade2-101, his3Δ200, lys2-801, met2, hmg1::LYS2, hmg2::HIS3, ura3-52::6MYC-HMG2*) expressing 6myc-Hmg2p and its *hrd1*-*1* mutant RHY401 (Hampton et al., 1996) on the other hand. 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* mutant phenotype complementation assays, minimal medium supplemented with 100 µg/ml or 175 µg/ml lovastatin was used for the RHY and YWO strains, respectively. 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.

Whole cell protein extracts were prepared from yeast cells by washing them with minimal medium containing 0.1% NaN₃ 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). 10 µg of the clear liquid lysate was loaded on SDS-PAGE gels for protein separation, followed by immunoblotting with the 9E10 monoclonal anti-myc antibody.
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, Annick Bleys for help in preparing the manuscript, Richard Dixon (Samuel Roberts Noble Foundation, Ardmore, USA) for providing the *M. truncatula* cell line, and Randy Hampton (UC San Diego, USA), Alexandra Stolz and Dieter Wolf (Universitaet Stuttgart, Germany) for providing yeast strains. The research leading to these results has received funding from the Agency for Innovation by Science and Technology in Flanders (“Strategisch Basisonderzoek” Combiplan project SBO040093) and the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. Tessa Moses is indebted to the VIB International PhD Fellowship Program for a predotoral fellowship.

References


Protein quality control in *Medicago truncatula* 7


Protein quality control in *Medicago truncatula*


Protein quality control in *Medicago truncatula*


Outlook
A major deliverable of this PhD was the establishment of a *Saccharomyces cerevisiae* based yeast platform for the production of triterpenoid sapogenins. As a first step towards achieving this goal, yeast strains producing five different triterpene sapogenin backbones, namely β-amyrin, α-amyrin, lupeol, lupanediol and dammarenediol were generated. These starter strains served as tools for the characterization of novel CytP450s involved in triterpene saponin biosynthesis, resulting in the identification of three new enzymes. Additionally, a combinatorial biosynthesis platform was established for the synthesis of both known and potentially novel sapogenins in yeast. Although the main goals of this PhD were achieved, several problems were encountered that need to be addressed to improve the existing platform and allow the full exploitation of its potential. In the following discussion several of these issues are pinpointed with possible ways of overcoming them.

**Can the production of sapogenins be enhanced further in the present yeast strains?**

Our *S. cerevisiae* strain was rationally engineered to express a feedback-free cytosolic version of *HMGR* and down regulate the expression of *ERG7*, to accumulate 2,3-oxidosqualene. In addition, flux through the ergosterol pathway could be further enhanced (a) by overexpressing every single gene of the pathway upto the production of 2,3-oxidosqualene, or (b) by expressing a mutant version of the transcription factor (*upc2-1*) to globally deregulate sterol biosynthesis or alternatively (c) by employing both (a) and (b). These approaches have been engineered in *S. cerevisiae* to increase amorphadiene and taxadiene titers substantially.

In addition to conventional pathway specific engineering, the modulation of distant and/or non rate-limiting genes can also influence metabolic phenotypes. For instance, a genotype-to-phenotype linking study led to the identification of two sterol biosynthetic genes and one fatty acid metabolic gene with non-silent single-nucleotide polymorphisms (SNPs), to be responsible for increased ergosterol and fatty acid levels in *S. cerevisiae* CEN.PK 113-7D. The overexpression of these three genes in a *S. cerevisiae* S288c strain
tremendously improved its heterologous triterpene production (Madsen et al., 2011). Similarly, using a carotenoid-based screening, gene deletions that improved isoprenoid production were identified. By combining these gene deletions with other mevalonate pathway modifications the production of the terpene-based biofuel bisabolene was increased by over 20-fold (Özaydın et al., 2012). Besides analyzing metabolic phenotypes, a genome-wide microarray analysis can help identify stress response factors that are activated upon heterologous pathway expression. Model-based engineering can further assist recognize the gene manipulations to overcome the stress response and can positively influence the production of the exogenous compound of interest.

In our yeast strains expressing CytP450s, we also observed a great variance in the amount of oxygenated products generated from the substrate, β-amyrin for instance, thereby representing the different catalytic efficiencies of the CytP450s involved (Fig. 1). However, the very low amounts of product, particularly with the strain expressing CYP93E2, indicate a possible reaction bottleneck (compound 6, Fig. 1). Although the actual cause of this low efficiency is currently unknown, a plausible reason could be the incompatibility of the CytP450 with the reductase partner employed. This could be addressed by either coupling the CytP450 with its native CPR or by employing an ubiquitous electron transporter like cytochrome b5 as the redox partner. Additionally, protein misfolding, protein instability and enzyme half-life could be other factors attributing to the low product concentration. Amino acid substitutions resulting from SNPs can positively influence the folding, stability and catalytic activity of proteins and have been suggested to be key players in the process of natural evolution (Tokuriki and Tawfik, 2009). Hence, through modeling and thermodynamic predictions the inter-residue changes that influence protein-folding and therefore protein stability can be determined to enhance a protein’s (or CytP450’s) function (Bloom et al., 2005).
Figure 1. An overlay of GC chromatograms obtained from the supernatant of strains cultured with MβCD. The black chromatogram corresponds to the β-amyrin producing control strain not expressing any CytP450. The colored chromatograms correspond to β-amyrin producing strains expressing a CytP450 each as indicated in the legend. Compound 1, ergosterol; 2, unknown; 3, 11-hydroxy β-amyрин; 4, unknown; 5, β-amyрин; 6, 24-hydroxy β-amyрин; 7, putative 21-hydroxy β-amyрин; 8, 28-hydroxy β-amyрин (erythrodol); 9, 28-carboxy β-amyрин (oleanolic acid); 10, putative oleanolic aldehyde; 11, 11-oxo β-amyрин.

However, the physical properties of proteins are generally not known and in such cases it is more feasible to apply a random approach to obtain enzyme variants. Directed evolution of enzymes which involves mutagenesis and selection of enzyme mutants is an alternative way of obtaining variants of a lead enzyme with desirable changes. For instance, the reaction specificity of a γ-humulene synthase from Abies grandis that cyclizes FPP to 52 different sesquiterpenes was evolved by site-saturation mutagenesis of its promiscuous active site to generate seven independent synthases (Yoshikuni et al., 2006). Similarly, directed enzyme evolution can also serve as a tool to enhance the catalytic activity of a triterpene saponin biosynthetic enzyme, such as the CYP93E2.

The yield of sapogenins reported in this thesis has been estimated from batch fermentations that have been performed devoid of optimization of culturing conditions. Hence, future efforts must be focussed on developing fed-batch fermentation-based production processes through the optimization of medium composition and culturing conditions to enhance the productivity of sapogenins. Additionally, by exploiting the
hydrophobic nature of sapogenins it might be feasible to develop a biphasic continuous cultivation method in which the cyclodextrin-sequestered sapogenins released into the medium can be captured in the organic phase that can be continuously recycled, thereby reducing any existing metabolic regulation and negative phenotypic effects like toxicity on the producer strain.

Next to optimizing culturing conditions strain optimization could also be performed by crossing our S288c BY4742 laboratory strain with an industrially robust strain to generate segregants that can be analyzed for enhanced sapogenin productivity.

Can an industrial yeast strain be engineered for sapogenin production?

The currently employed S. cerevisiae S288c BY4742 is a laboratory yeast strain that does not comply with the robust requirements of the industry. As a follow-up to this work, the goal of the PhD project of Philipp Arendt is to engineer the industrial yeast Yarrowia lipolytica for the synthesis of pharmaceutically active terpenes. Y. lipolytica is a non-conventional GRAS yeast extensively employed for recombinant protein production. In addition, Y. lipolytica can utilize fatty acids as a source of carbon for its growth and convert them to acetyl-CoA, the precursor for terpene synthesis. Therefore, one aim of this follow-up project will be to establish a triterpenoid sapogenin biosynthetic platform in Y. lipolytica for the production of known as well as novel compounds using combinatorial biosynthesis. Owing to the conversion of cheap feed-stock to valuable terpenoid products, the Yarrowia platform is more suited for industrial applications.

Next to Y. lipolytica, a S. cerevisiae based industrial strain, like CEN.PK2, could also be engineered for sapogenin production. Transferring the engineering knowledge gained from S. cerevisiae S288c to CEN.PK2, the final titer of amorphadiene was enhanced to over 40 g/L.
Can the sapogenins be easily purified from yeast cultures?

We reported the establishment of a novel strategy to sequester triterpene sapogenins from within yeast cells to the spent culture medium using MβCD. The extraction of sapogenins involves the transfer of, presumably MβCD-bound compounds from the aqueous medium to the highly apolar organic solvents, from where they are concentrated and separated using a TLC based approach. However, what is still unclear is whether the sapogenins once in the organic solvent still exist in a MβCD-bound form. Our initial attempts at NMR on the hydroxylated β-amyrin product of the CytP450 BF567, indicated the presence of sugar contaminants, presumably arising from the MβCD. Therefore, it is essential to develop a method to expel the MβCD post purification of MβCD-bound compound to obtain high purity triterpene sapogenins. Cyclodextrin degrading enzymes have been described that could be utilized to break down cyclodextrins to water-soluble maltose and glucose units (Turner et al., 2005), which can thus be separated from the water insoluble sapogenins to obtain compounds with high level of purity.

What is the future of combinatorial biosynthesis for novel sapogenins?

Despite the wide spread ability of plants to synthesize saponins, their biosynthetic pathways are poorly characterized, with not a single complete synthetic pathway having been elucidated from any plant species. Although multiple OSCs have been characterized, there are only seven CytP450 families identified of which the exact function of six are known, and six UGTs partially characterized (Augustin et al., 2011). This limits the pool of biosynthetic genes that can be currently utilized for combinatorial biosynthesis and emphasizes on the great need for gene discovery to identify saponin biosynthetic genes. Nonetheless, combinatorial biosynthesis could be attempted by utilizing genes not necessarily involved in saponin biosynthesis, but involved in triterpene biosynthesis and rely on their promiscuous nature to perform desired functionalities. For instance, the S. cerevisiae ERG25 gene product is a CytP450 that has a C-4 methyl sterol oxidase activity and is involved in ergosterol biosynthesis (Bard et
The reaction catalyzed by ERG25p results in a three step oxidation of a methyl group attached to C-4, to a carboxyl group (Fig. 2a). This carbon corresponds to the C-23 of an oleanane triterpene sapogenin backbone, for which a C-23 oxidase has not been identified to date. In the β-amyrin producing engineered *S. cerevisiae* strain it should be possible to bring about a C-23 carboxylation by overexpressing the *S. cerevisiae* ERG25 gene either in its native form or by expressing a variant of *ERG25* obtained through directed enzyme evolution (Fig. 2b). Such an approach could be extended to other plant and mammalian systems to identify CytP450s capable of catalyzing oxygenations on triterpene backbones.

![Figure 2. Hypothetical schematic for employing a yeast enzyme for triterpene sapogenin biosynthesis.](image)

(a) ERG25 catalyzes the conversion of 4,4-dimethylzymosterol to 4α-carboxy-4β-methyl-5α-cholesta-8,24-dien-3β-ol.

(b) ERG25 catalyzes the conversion of β-amyrin to 23-carboxy β-amyrin.

The recent progress in mining genetic information from non-culturable organisms (or metagenomics) has led to the discovery of novel enzymes which allow the development of compounds that benefit from enzyme-catalyzed chiral synthesis (Simon and Daniel, 2011). By applying either a function-based or sequence-based screening for DNA of
interest, metagenomics provides a vast pool of genetic information to screen for novel genes with triterpene (or any other compound) modifiable activity.

With bardoxolone methyl in phase II clinical trials, the triterpene sapogenins and its derivatives continue to be interesting molecules with pharmaceutical applications. However, there is a great need to generate large-scale combinatorial collections of sapogenins and its derivatives with high structural variability to screen for compounds with biological activity. The discipline of chemoinformatics which depicts the physicochemical properties of compounds with potential pharmaceutical properties emphasizes on the need for chiral centres, rotatable bonds, structural rigidity and aromatic rings in compounds to enhance their specificity and efficacy (Bade et al., 2010). These physicochemical properties are observed in natural products which are synthesized by enzyme catalyzed reactions, thereby emphasizing on the need for gene discovery to generate compounds with pharmaceutical potential. Hence, high-throughput sequencing of triterpene saponin producing plants and other triterpene producing organisms together with metagenomics will play an elementary role in the construction of large-scale sapogenin libraries. Furthermore, the expression of multiple genes in an easily scalable host like yeast will be mandatory to scale-up and obtain high amounts of relatively pure compounds. Additionally, applying multidisciplinary approaches including chemical modification of the reactive moieties on the enzyme-derived compounds will further increase the diversity of sapogenins. However, the possibility for chemical modifications is vast and therefore should be conjugated with in silico modeling to assess the feasibility of its synthesis. In parallel, computational structural biology predictions of the interaction of compound libraries with target proteins, associated with a target disease, will allow preliminary selection of chemical modifications that might influence the biological activity of the compound. Therefore, the successful generation of sapogenin libraries will depend on the future application of multiple disciplines including metabolic engineering in a heterologous host preferably yeast, genomics, in silico modeling and computational biology.

In conjunction with cell-based synthesis it is also essential to consider the application of in vitro synthetic biology in sapogenin biosynthesis. A major drawback to isolating
sapogenins from their natural source is the complexity associated with their purification. Although yeast-based production allows extraction of relatively pure compounds, there are multiple hurdles associated with extraction from culture medium and/or yeast cells. In this respect, in vitro systems can be the sapogenin production platform of the future, when gene and enzyme functionality is more resolved. A desired compound can be synthesized from a substrate through a series of enzyme catalyzed reactions in an artificial environment, thereby facilitating purification.

References


Triterpenoid saponins are glycosides of sapogenins produced by a wide array of plants. They constitute a structurally diverse group of biologically active compounds with general health-promoting and pharmaceutical applications. In the native hosts these compounds are produced in very low amounts, rendering extraction from plants commercially unsustainable. The complications of engineering plants for enhanced production of desired products has led to the heterologous synthesis of many terpenoids in *Escherichia coli* and *Saccharomyces cerevisiae*. In addition, the application of current approaches, for the design and engineering of pathways using synthetic biology tools, and for the production of novel terpenoids by evolving enzymes and combinatorial biosynthesis, has also been more compatible with heterologous microbial hosts to date (chapter 2).

The engineering of triterpenoid saponins in both plants and yeast has been limited, inspite of their potential applications. Hence, the main goal of this PhD research was to develop a heterologous yeast platform for triterpenoid sapogenin production. Producing high amounts of the precursor is an essential hurdle to overcome for establishing an efficient production system. Therefore, the native competing ergosterol biosynthesis of a *Saccharomyces cerevisiae* strain was made conditionally down-regulatable to accumulate 2,3-oxidosqualene and divert flux towards sapogenin backbone production. Using this engineered starter strain, yeast strains accumulating high amounts of the oleanane backbone β-amyrin, the ursane backbone α-amyrin and lupeol were generated (chapter 4 and 6). The engineered yeast strain also served as a platform for characterizing functional oxidosqualene cyclases. A β-amyrin synthase from *Maesa lanceolata* and the multifunctionality of the dammarenediol synthase from *Centella asiatica* were identified (chapter 5 and 6). The co-cyclization of 2,3-oxidosqualene to lupanediol by the lupeol synthase from *Arabidopsis thaliana* in addition to lupeol, generated an additional novel triterpene backbone in yeast (chapter 6).
Unlike mono- and sesquiterpenes, triterpenes are large and plausibly more toxic compounds that are not readily excreted from yeast cells. This complicates compound purification from yeast cultures, due to the strict requirement for cell harvesting and lysis prior to compound extraction. By exploiting their hydrophobic nature, β-cyclodextrins were established as a sequestering agent for triterpene sapogenins. For the first time the extraction of triterpenes from yeast cells was demonstrated using methylated β-cyclodextrins, which also positively influenced the production of sapogenins itself, perhaps a consequence of the lack of intracellular feedback regulation (chapter 4).

The engineered β-amyrin producing yeast strain served as a tool for the characterization of novel cytochrome P450 monoxygenases. A 28-oxidase and two putative 21β-hydroxylases were identified by employing this strain (chapter 4 and 5). Using the β-amyrin, α-amyrin and lupeol producing strains a proof of concept for combinatorial biosynthesis of triterpene sapogenins in yeast was delivered (chapter 6). Also, a first proof on the functional homology of two cytochrome P450 monooxygenases belonging to two different families was established in triterpene saponin (secondary metabolite) biosynthesis (chapter 4 and 5).

Finally, a triterpene saponin regulatory gene, called Makibishi 1 (MKB1), which encodes an RMA-type E3 ubiquitin ligase, was identified to be involved in the regulation of triterpene saponin biosynthesis in *Medicago truncatula* by targeting specific isoforms of 3-hydroxy-3-methylglutaryl-CoA reductase, the rate-limiting enzyme of the mevalonate pathway. Down regulating *MKB1* expression in transgenic *M. truncatula* hairy roots resulted in the accumulation of monoglycosylated saponins, thereby serving as a plant-based tool for the accumulation of saponin building blocks (chapter 7).

As a consequence, two *in vivo* production systems were established for the synthesis of known and novel saponin building blocks. The conventional yeast *S. cerevisiae* and the model legume *M. truncatula* could be engineered to accumulate unglycosylated sapogenins and glycosylated saponin building blocks, respectively.
Addendum 1

List of yeast strains generated in this study and the sapogenin backbones or sapogenins they produce.

<table>
<thead>
<tr>
<th>Name</th>
<th>Construct</th>
<th>Compounds produced</th>
</tr>
</thead>
<tbody>
<tr>
<td>S288c BY4742</td>
<td>MATa; his3Δ1; leu2Δ0; ura3Δ0; lys2Δ0</td>
<td></td>
</tr>
<tr>
<td>TM1</td>
<td>S288c BY4742; P&lt;sub&gt;erg7&lt;/sub&gt;:P&lt;sub&gt;MET3&lt;/sub&gt;-ERG7</td>
<td></td>
</tr>
<tr>
<td>TM2</td>
<td>TM1; pESC-URA[GAL10/hMG1; GAL1/MibAS] (19.3 mg/L β-amyrin)</td>
<td><img src="image" alt="β-amyrin" /></td>
</tr>
<tr>
<td>TM3</td>
<td>TM1; pESC-URA[GAL10/hMG1; GAL1/GgbAS] (36.2 mg/L β-amyrin)</td>
<td><img src="image" alt="β-amyrin" /></td>
</tr>
<tr>
<td>TM4</td>
<td>TM1; pESC-URA[GAL10/hMG1; GAL1/MibAS]</td>
<td><img src="image" alt="β-amyrin" /></td>
</tr>
<tr>
<td>TM5</td>
<td>TM1; pESC-URA[GAL10/hMG1]</td>
<td><img src="image" alt="β-amyrin" /></td>
</tr>
<tr>
<td>TM6</td>
<td>TM1; pESC-URA[GAL10/hMG1; GAL1/AtLUP1] (46.3 mg/L lupeol)</td>
<td><img src="image" alt="Lupeol" /></td>
</tr>
<tr>
<td>TM7</td>
<td>TM3; pAG423[GAL1/BF567], pAG425[GAL1/AtATR1]</td>
<td><img src="image" alt="20-hydroxy-lupeol" /></td>
</tr>
<tr>
<td>TM8</td>
<td>TM3; pAG423[GAL1/BF567], pAG305[GAL1/AtATR1]</td>
<td><img src="image" alt="21(?)-hydroxy-β-amyrin" /></td>
</tr>
<tr>
<td>TM9</td>
<td>TM3; pAG423[GAL1/BF567], pAG415[GAL1/AtATR1]</td>
<td><img src="image" alt="Erythrodiol" /></td>
</tr>
<tr>
<td>TM10</td>
<td>TM3; pAG423[GAL1/CYP716A12], pAG425[GAL1/AtATR1]</td>
<td></td>
</tr>
<tr>
<td><strong>TM17</strong></td>
<td>TM3: pAG423[<strong>GAL1/CYP716A12</strong>], pAG415[<strong>GAL1/AtATR1</strong>]</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image1" alt="Oleanolic aldehyde" /> <img src="image2" alt="Oleanolic acid" /></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TM18</strong></th>
<th>TM3: pAG423[<strong>GAL1/CYP88D6</strong>], pAG415[<strong>GAL1/AtATR1</strong>]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image3" alt="11-hydroxy-β-amyrin" /> <img src="image4" alt="11-oxo-β-amyrin" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TM20</strong></th>
<th>TM3: pAG423[<strong>GAL1/ML257</strong>], pAG415[<strong>GAL1/AtATR1</strong>]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image5" alt="Erythrodiol" /> <img src="image6" alt="Oleanolic aldehyde" /> <img src="image7" alt="Oleanolic acid" /></td>
</tr>
<tr>
<td>TM21</td>
<td>TM3; pAG423[\textit{GAL1/ML593}], pAG415[\textit{GAL1/AtATR1}]</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="21(?)-hydroxy-\beta-amyrrin" /></td>
</tr>
<tr>
<td>TM22</td>
<td>TM6; pAG423[\textit{GAL1/CYP716A12}], pAG415[\textit{GAL1/AtATR1}]</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Betulin" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="20-hydroxy-betulin" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="20-hydroxy-betulinic acid" /></td>
</tr>
<tr>
<td>TM23</td>
<td>TM6; pAG423[\textit{GAL1/ML257}], pAG415[\textit{GAL1/AtATR1}]</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Betulin" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="20-hydroxy-betulin" /></td>
</tr>
<tr>
<td>TM26</td>
<td>TM3; pAG423, pAG425[\textit{GAL1/AtATR1}]</td>
</tr>
<tr>
<td>TM27</td>
<td>TM3; pAG423, pAG415[\textit{GAL1/AtATR1}]</td>
</tr>
<tr>
<td>TM28</td>
<td>TM6; pAG423, pAG425[\textit{GAL1/AtATR1}]</td>
</tr>
<tr>
<td>TM30</td>
<td>TM3: pAG423[GAL1/BF567-T2A-CYP716A12], pAG415[GAL1/AtATR1]</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="21(?)-hydroxy-oleanolic acid" /></td>
</tr>
<tr>
<td>TM31</td>
<td>TM3: pAG423[GAL1/ML593-T2A-CYP716A12], pAG415[GAL1/AtATR1]</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="21(?)-hydroxy-oleanolic acid" /></td>
</tr>
<tr>
<td>TM32</td>
<td>TM3: pAG423[GAL1/ML593-T2A-CYP88D6], pAG415[GAL1/AtATR1]</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="11,21(?)-dihydroxy-β-amyrin" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="11-oxo-21(?)-hydroxy-β-amyrin" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="α-amyrin" /></td>
</tr>
<tr>
<td>TM33</td>
<td>TM1: pESC-URA[GAL10/HMG1; GAL1/CaDDS]</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="β-amyrin" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Dammarenediol" /></td>
</tr>
<tr>
<td><strong>TM34</strong></td>
<td><em>TM33; pAG423</em>[GAL1/CYP716A12], pAG415*[GAL1/AtATR1]*</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td><img src="Uvaol_structure.png" alt="Uvaol" /></td>
</tr>
<tr>
<td></td>
<td><img src="Erythrodiol_structure.png" alt="Erythrodiol" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TM35</strong></th>
<th><em>TM33; pAG423</em>[GAL1/CYP88D6], pAG415*[GAL1/AtATR1]*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="11-hydroxy-%CE%B1-amyris_structure.png" alt="11-hydroxy-α-amyris" /></td>
</tr>
<tr>
<td></td>
<td><img src="11-hydroxy-%CE%B2-amyris_structure.png" alt="11-hydroxy-β-amyris" /></td>
</tr>
<tr>
<td></td>
<td><img src="11-oxo-%CE%B1-amyris_structure.png" alt="11-oxo-α-amyris" /></td>
</tr>
<tr>
<td></td>
<td><img src="11-oxo-%CE%B2-amyris_structure.png" alt="11-oxo-β-amyris" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TM36</strong></th>
<th><em>TM33; pAG423</em>[GAL1/ML257], pAG415*[GAL1/AtATR1]*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="Uvaol_structure.png" alt="Uvaol" /></td>
</tr>
<tr>
<td>TM37</td>
<td>TM33; pAG423[\textit{GAL1/ML593}], pAG415[\textit{GAL1/AaATRI}]</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>TM38</td>
<td>TM33; pAG423, pAG415[\textit{GAL1/AaATRI}]</td>
</tr>
</tbody>
</table>
Cloning strategy and PCR fragments amplified for generation of strain TM1.

Plasmid pIP007 was generated by inserting the MET3 promoter (P_{MET3}) between the SpeI and SacII restriction sites of plasmid pUG6 harboring an ampicillin resistance cassette for bacterial selection. In vector pIP007 the P_{MET3} is inserted next to the kanamycin cassette (KanMX) which is driven by the TEF promoter (P_{TEF}) and TEF terminator and flanked by loxP sites for Cre-mediated recombination. The PCR fragments a and b were amplified from the plasmid vector pIP007 and fragments c and d from the yeast genomic DNA. In a second PCR the fragments e and f were amplified using the template pairs a+c and b+d, respectively. The fragments e and f were then transformed into an S288c BY4742 strain of Saccharomyces cerevisiae to replace the native ERG7 promoter (P_{ERG7}) with the methionine repressible P_{MET3} promoter.
Additional publications and patents

Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in *Arabidopsis* and tobacco


The phytohormones jasmonates (JAs) constitute an important class of elicitors for many plant secondary metabolic pathways. However, JAs do not act independently but operate in complex networks with crosstalk to several other phytohormonal signaling pathways. Here, crosstalk was detected between the JA and abscisic acid (ABA) signaling pathways in the regulation of tobacco (*Nicotiana tabacum*) alkaloid biosynthesis. A tobacco gene from the PYR/PYL/RCAR family, *NtPYL4*, the expression of which is regulated by JAs, was found to encode a functional ABA receptor. *NtPYL4* inhibited the type-2C protein phosphatases known to be key negative regulators of ABA signaling in an ABA-dependent manner. Overexpression of *NtPYL4* in tobacco hairy roots caused a reprogramming of the cellular metabolism that resulted in a decreased alkaloid accumulation and conferred ABA sensitivity to the production of alkaloids. In contrast, the alkaloid biosynthetic pathway was not responsive to ABA in control tobacco roots. Functional analysis of the *Arabidopsis* (*Arabidopsis thaliana*) homologs of *NtPYL4*, *PYL4* and *PYL5*, indicated that also in *Arabidopsis* altered *PYL* expression affected the JA response, both in terms of biomass and anthocyanin production. These findings define a connection between a component of the core ABA signaling pathway and the JA responses and contribute to the understanding of the role of JAs in balancing tradeoffs between growth and defense.


Contribution: Generation of *NtPYL4* Gateway constructs for PP2C enzyme assay & Y2H.
Combinatorial biosynthesis in plants: A (p)review on its potential and future exploitation

Jacob Pollier, Tessa Moses, and Alain Goossens

Combinatorial biochemistry, also called combinatorial biosynthesis, comprises a series of methods that establish novel enzyme–substrate combinations \textit{in vivo} and, 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 to create opportunities for the 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.

Published in: Natural Product Reports (IF 9.79), 2011, volume 28, pages 1897-1916.

Contribution: Writing parts of the manuscript.
A MultiSite Gateway™ vector set for the functional analysis of genes in the model Saccharomyces cerevisiae

Astrid Nagels Durand, Tessa Moses, Rebecca De Clercq, Alain Goossens, and Laurens Pauwels

**Background:** Recombinational cloning using the Gateway™ technology has been the method of choice for high-throughput omics projects, resulting in the availability of entire ORFeomes in Gateway™ compatible vectors. The MultiSite Gateway™ system allows combining multiple genetic fragments such as promoter, ORF and epitope tag in one single reaction. To date, this technology has not been accessible in the yeast Saccharomyces cerevisiae, one of the most widely used experimental systems in molecular biology, due to the lack of appropriate destination vectors.

**Results:** Here, we present a set of three-fragment MultiSite Gateway™ destination vectors that have been developed for gene expression in S. cerevisiae and that allow the assembly of any promoter, open reading frame, epitope tag arrangement in combination with any of four auxotrophic markers and three distinct replication mechanisms. As an example of its applicability, we used yeast three-hybrid to provide evidence for the assembly of a ternary complex of plant proteins involved in jasmonate signalling and consisting of the JAZ, NINJA and TOPELESS proteins.

**Conclusion:** Our vectors make MultiSite Gateway™ cloning accessible in S. cerevisiae and implement a fast and versatile cloning method for the high-throughput functional analysis of (heterologous) proteins in one of the most widely used model organisms for molecular biology research.


Contribution: Generation of entry clones and writing parts of the manuscript.
Triterpenoid sapogenin production in plant and microbial cultures

Assignees: VIB and UGent.

Inventors: Alain Goossens, Tessa Moses, and Lorena Almagro Romero

Field of invention: The current invention is situated in the fields of plant secondary metabolites with pharmacological or other industrial properties and metabolic engineering of these phytochemicals. More specifically, the invention relates to a method for enhancing the biosynthesis and/or secretion of sapogenins intermediates in the culture medium of plant and microbial cell cultures.

Abstract: The invention relates to a method for enhancing the biosynthesis and/or secretion of sapogenins in the culture medium of plant and microbial cell cultures.

Acknowledgements

Oh my! It definitely starts to sink in now that the end is near. Doesn’t writing acknowledgements mean I am nearly done with my Ph. D.? Oh mighty well it does! I am a happy and contended girl. I will cherish my Ph. D. experience for a long time to come and these memories of 5 years wouldn’t have been the same without the help, support, patience and just the physical presence of many of you.

I want to begin by thanking VIB and the whole team at the headquarters for not just the International Ph. D. program, but also the constant support in making my life fairly simple and straightforward in Belgium. If not for the International Ph. D. program I still would have been unaware of the geographical location of Belgium (oopsy daisies my geography is really bad!). Thank you Mark (Mark Veugelers) for your advice and scientific guidance all through. Marijke (Marijke Lein), I have not met someone more kind and considerate than you. You have always been there, you are an inspiration. Veerle (Veerle Quivreux), thank you for all the administrative paper work. Thank you for sorting out the mess with my tax papers year after year.

When I came for the International Ph. D. selection interview way back in 2007, my sole motivation was to work with yeast. The project Alain (Alain Goossens) and Johan (Johan Thevelein) coined was fascinating. I have to thank both of you for the opportunity and for accepting me into your groups. Alain, working with you is a constant learning process. Your hard working attitude is inspiring (quiet contagious actually!). I appreciate the independence you give me to plan my work, the listening ear you lend in scientific discussions, and all the extra little ‘side projects’ you let me get involved in. When I look back, those side projects have taught me much more than yeast engineering, and I thank you for letting me be part of them. Thank you Alain for the constant support, the understanding, the inspiration and for making me a wee bit cleverer than I was when I moved to the colder part of the planet. I tried hard not to sound ‘Shakespearean’ you
know 😊! If I ever get to writing something worth publishing someday, I promise to dedicate the most complicated of texts to you.

Johan, you were my mystery promoter. I still remember the first time I got to meet you was during the VIB seminar of 2008. I never got to work closely with you, but you have always been there with your valuable suggestions and useful comments. I am always amazed by how you come up with the most logical and simplest solutions to the most complicated problems. You inspire me to become scientifically wise. Thank you for being inspirational.

My Ph. D. wouldn’t have been technically as smooth without the help of the PSB administrative, ICT and logistic staff. Special thanks to Diane (Diane Hermie), Christine (Christine Tire), Delphine (Delphine Verspeel), Kristof (Kristof Verleye), Jackie (Jackie Vanden Driessche), Hendrik (Hendrik Labeeuw), Stefaan (Stefaan Vanderkerken), Tim (Tim Van de Woestyne), Wilson (Wilson Ardiles-Diaz), Nino (Raimundo Villarroel), An (An Bontinck), Bernard (Bernard Vanassche), Dirk (Dirk Van Akoleyen) and Karel (Karel Spruyt) who have always been willing to assist. Wilson, thank you for sheltering me when I landed in Belgium. You are so full of energy, you put me and many more to shame at the parties. Hendrik, I have enjoyed socializing with you outside the lab and the first PSB quiz will be memorable to me.

My dear fellow metabollers you guys are just awesome, such fun to work with. Gino, your ‘fulfil your dream’ attitude is inspirational. Vestrock truly rocks! Laurens, your broad view of science and the dedication to unraveling the mysteries of jasmonates is, in the deepest sense of the word, amazing. Andres, the man (I know that makes you happy!), please behave at my defense. Jacob, I love you and love working with you! Alex, I have never seen someone crack such hilarious jokes with the most innocent and serious look on their face. You are a cheerful little guy (pun intended). Rebecca, you are one person who comes smiling into the lab every morning, its such a pleasant sight. Sofie, trust me the diatom boys would be helpless without you. Way to go girl! Robin, you know what is most striking about you? The way you get all excited and happy about positive results. I enjoy working with you. Nathan, Azra, Astrid and Jonas, you guys
keep the group live and kicking. You are a cheerful bunch. **Amparo**, you are full of energy and great company both inside and outside the lab. **Michele**, I know you love working with yeast, and trust me, I honestly didn’t mean to get you high with the hexane. **Michiel**, thank you for the constant debate about everything under the sun, through the past years. Most of the discussions are enlightening and, although at first instance it doesn’t seem like, I do enjoy these conversations. **Sandra**, probably the best time I spent with you was when we were in Sweden for Terpnet, 2011. I will always remember how exhausted we were at the conference talks after chatting non stop in the nights. **Jan** and **Philipp**, thank you for expanding the yeast family in the group. I can vouch, working with you guys will be great fun. The team buildings wouldn’t be the same without you all.

Also a big thank you to all the students who did their projects with me during the past 5 years and contributed to this thesis in tits and bits. **Timo**, you were the first and the most memorable of them. It has been a great experience to work with you and to know you as a person. You are a great kid. **Assia, Saartje, Corinne** and **Nina** it has been a learning process for me whilst teaching you all. Sometimes the simplest of questions from you has sent me back to my basics to look for answers. Thank you girls.

Oh my Indian bunchies, what would the weekends have been like without our tea parties! **Anju** and **Anagha** thank you so much for getting me into the loop as soon as I landed in Gent. **Anju**, thank you for being our mommy in disguise and feeding us all whenever we ended up at your door. **Niloufer**, how is jelly belly? Your enthusiasm and craving for knowledge is so contagious. I have learnt a lot by just listening to you over a cup of coffee. Thank you my dearie. You are missed! **Deepti**, dip-tea, deepu where is the chai? Do you know since you have left Belgium, we havn’t had tea parties! **Rahul** baba you are such a sweet kid. You evoke sisterly instincts in me. **Vidhya**, you are such fun. **Sibu**, listen to your wife. **Elizabeth** darling don’t listen to Sibu and **Johan**, he is so charming already. **Vikram** and **Jurgen**, you both are sweethearts. Thank you for always being there for us. **Claire**, thank you for all the fun stories, ‘the match stick’will be my all-time favourite. You are such a stress-buster.
A big thank you to Jacob’s parents, Dorine and Patrick, for making me so comfortable and welcome in the family. Also a big thank you to Maya and Clovis for making my Saturday’s stress free. Marleen (Marleen De Keukelaere), I cannot thank you enough for all the care and hospitality you showered on me during the first two years. Thank you for taking care of me. Nelly (Nelly Van Hove), thank you for everything.

I wouldn’t have made it Belgium without the support of my papa bear, my ma and my chechy. Thank you for believing in me and letting me follow my dreams. Thank you daddy for making me believe I am a go-getter. Thank you ma for letting me come to Belgium, although it was hard for you. Chechy thank you for always being there for me, and for helping me get through my worries. You inspire me to be a self-less person.

Last but not the least, Jacob thank you my dear for being in my life. I dedicate this Ph. D. thesis to you. Without you I probably wouldn’t have made it through the worst of 2010. You constantly motivated me to hold on and fight back. Thank you for your patience, support, love, care and belief in me. Thank you for being my pillar of strength both at work and at home. I love you! And since it shouldn’t go unsaid, a big thank you to my doggy, my silent partner in crime.
Curriculum vitae

PERSONAL DETAILS

Name: Tessa Moses
Date of Birth: 4th October, 1983
Place of Birth: Thiruvananthapuram, Kerala, India.
Nationality: Indian.
Marital status: Married to Jacob Pollier.
Current address: Voskenslaan 453, 9000 Gent, Belgium
Permanent address: Hertooiebos 12A, 9052 Zwijnaarde, Belgium
Tel: +32 (0)475.98.73.57
E-mail: temos@psb.vib-ugent.be
tessa.moses@gmail.com

EDUCATION

2007-2012 PhD in Sciences: Biotechnology
Institute: VIB, Ghent University, Belgium
Thesis: Metabolic engineering for production of triterpenoid saponin building blocks in plants and yeast.
Promoter: Prof. Alain Goossens.
Co-promoter: Prof. Johan Thevelein

2004-2006 Master of Science in Microbiology (with Distinction)
Institute: University of Mumbai, India
Thesis: Development of seed transfer criteria in rifamycin B fermentation.
Promotor: Prof. Pramod Wangikar, Department of Chemical Engineering, Indian Institute of Technology, Mumbai, India.

2001-2004 Bachelor of Science in Biotechnology
Institute: University of Mumbai, India
Promotor: Prof. K. V. Venkatesh, Department of Chemical Engineering, Indian Institute of Technology, Mumbai, India.

PUBLICATIONS


Bapat, P.M., Sohoni, S.V., Moses, T., and Wangikar, P.P. A cybernetic model to predict the effect of freely available nitrogen substrate on rifamycin B production in complex media. Applied Microbiology and Biotechnology, 2006, 72(4), 662-670

MANUSCRIPTS IN PREPARATION


PATENTS


OTHER SCIENTIFIC COMMUNICATIONS

