Discovery and reconstitution of the secoiridoid pathway of *Catharanthus roseus*

Lemeng Dong
Thesis committee

Promotor
Prof. Dr G.H. Immink
Professor of Physiology of Flower Bulbs
Wageningen University

Co-promotor
Dr A.R. van der Krol
Associate professor at the Laboratory of Plant Physiology
Wageningen University

Other members
Prof. Dr W.J.H. van Berkel, Wageningen University
Dr M.C.R. Franssen, Wageningen University
Prof. Dr R. Verpoorte, Leiden University
Prof. P. Christou, University of Lleida, Spain

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.
Discovery and reconstitution of the secoiridoid pathway of 
*Catharanthus roseus*

Lemeng Dong

Thesis
submitted in fulfilment of the requirements for the degree of doctor 
at Wageningen University 
by the authority of the Rector Magnificus 
Prof. Dr M.J. Kropff, 
in the presence of the 
Thesis Committee appointed by the Academic Board 
to be defended in public 
on Friday 7 March 2014 
at 4 p.m. in the Aula.
Lemeng Dong
Discovery and reconstitution of the secoiridoid pathway of Catharanthus roseus, 202 pages.

With references, with summaries in English and Dutch

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter 1</strong></td>
<td>7</td>
</tr>
<tr>
<td>General introduction</td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 2</strong></td>
<td>29</td>
</tr>
<tr>
<td>Characterization of two geraniol synthases from <em>Valeriana officinalis</em> and <em>Lippia dulcis</em>: similar activity but difference in subcellular localization</td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 3</strong></td>
<td>65</td>
</tr>
<tr>
<td>Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the secoiridoid pathway</td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 4</strong></td>
<td>107</td>
</tr>
<tr>
<td>Monoterpene biosynthesis potential of plant subcellular compartments</td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 5</strong></td>
<td>127</td>
</tr>
<tr>
<td>Discovery and reconstitution of the secoiridoid pathway of <em>Catharanthus roseus</em></td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 6</strong></td>
<td>167</td>
</tr>
<tr>
<td>General discussion</td>
<td></td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>187</td>
</tr>
<tr>
<td><strong>Samenvatting</strong></td>
<td>191</td>
</tr>
<tr>
<td><strong>Acknowledgements</strong></td>
<td>195</td>
</tr>
<tr>
<td><em>Curriculum vitae</em></td>
<td>197</td>
</tr>
<tr>
<td><strong>Publication list</strong></td>
<td>198</td>
</tr>
<tr>
<td><strong>Education statement</strong></td>
<td>199</td>
</tr>
</tbody>
</table>
Chapter 1

General introduction
1. Plant secondary metabolites

Plants produce a vast number of metabolites which are not directly involved in primary processes of basic growth and development and which therefore are called secondary metabolites. Many plant species have been used for all kinds of medicinal purposes throughout the ages but the isolation of morphine by Friedrich Wilhelm Sertürner about 200 years ago is generally accepted as the beginning of the actual research on plant secondary products (Hartmann, 2007). Since then, secondary metabolite research has focused on many different aspects such as compound extraction and purification, elucidation of the molecular structure, isolation and characterisation of the related biosynthesis genes, transport of the secondary metabolites within and between cells, sequestration and storage, the biological function of these compounds for the plant and the evolution of biosynthesis pathways in different plant species. With regard to their medicinal effect, secondary metabolite research has focused on the uptake and stability of compounds and the effect on human physiology, for example their functioning as antioxidants, as specific protein functional inhibitors, etc. Already tens of thousands of plant secondary metabolites have been described, but these are still only a small fraction of the molecules actual present in nature, of which many likely will have potential medicinal applications (Kroymann, 2011).

For the plant, secondary metabolites play an important role in the interaction with their environment. They may serve as attractants for pollinators or seed dispersers or as deterrents for plant pathogens or they may be involved in alleviating abiotic stresses (Aharoni et al., 2005; Arimura et al., 2005; Gershenzon and Dudareva, 2007; McCall et al., 1994). Many plants, for example, produce UV absorbing flavonoids to protect themselves against potentially harmful UV-B irradiation (Tegelberg and Julkunen-Tiitto, 2001). In addition, some secondary metabolites act as antioxidants and radical scavengers, helping plants to cope with oxidative stress (Jacob et al., 2011; Oh et al., 2009). Flavonoids such as quercetin, for example, were reported to interfere with the generation of reactive oxygen species (ROS) via the Fenton reaction, thus contributing to a powerful antioxidant/antiradical performance (Leopoldini et al., 2006).

For humans, plant secondary metabolites are important because of their unique biochemical structure and properties. For instance, animals (and humans) need carotenoids as precursor for vitamin A biosynthesis but do not possess the biosynthetic pathway to make them. Both humans and animals are therefore totally dependent on plants to obtain these essential molecules. Plant secondary metabolites determine flavour and taste of food and they can have beneficial (but also detrimental!) effects on human health. Because of their highly diverse structures and often very specific action, plant secondary metabolites provide an important source of molecules which are screened in pharmaceutical research for different types of biological activity.
Based on their biosynthetic origins, the plant secondary metabolites can be divided into three major chemical classes: phenolics, terpenoids and alkaloids (Croteau et al., 2000). The phenolics are derivatives of the pentose phosphate, shikimate, and phenylpropanoid pathways in plants (Balasundram et al., 2006). They are molecules possessing at least one aromatic ring bearing one or more hydroxyl groups. Phenolic compounds show a large diversity of structures, varying from relatively simple molecules (e.g. vanillin, gallic acid, caffeic acid) to polyphenols such as stilbenes, flavonoids and even larger polymers derived from condensation of multiple phenolic groups.

Terpenoids are hydrocarbons resulting from the combination of one or more 5-carbon isoprene units. The classification of terpenoids is based on the numbers of isoprene units used to build the basic skeleton structure. For example, hemiterpenoids (1 isoprene unit, C5), monoterpenoids (2 isoprene units, C10), sesquiterpenoids (3 isoprene units, C15), diterpenoids (4 isoprene units, C20), triterpenoids and sterols (6 isoprene units, C30), etc up to polyterpenoid (with a larger number of isoprene units). In plant, these C5 isoprene units have two sources. One is from a cytosolic localized pathway which produces the C5 precursor farnesyl diphosphate (FPP) from mevalonic acid (therefore called the mevalonic acid or MVA pathway). The MVA pathway mainly serves to synthesize sesquiterpenoids, triterpenoids and sterols. Alternatively, C5 units are derived from geranyl diphosphate (GPP) which is produced in the plastids from 2-C-methyl-D-erythritol 4-phosphate (therefore named MEP pathway). The MEP pathway produces the precursors for hemiterpenoids, monoterpenoids and diterpenoids. Terpenoids are one of the major classes of secondary metabolites that are produced in response to both abiotic (Vickers et al., 2009) and biotic stresses, such as herbivore-attacks (Dicke, 1994; McCall et al., 1994). In order to protect themselves from photosynthetic hyperactivity under high light conditions certain tree species produce and emit large amounts of simple isoprene molecules (Gil et al., 2012; Holopainen, 2013). Indeed it has been shown that the production of isoprene under light stress helps to protect the thylakoid membranes of the chloroplasts (Possell and Loreto, 2013). But terpenoids also have important functions in the regulation of plant growth and development, as they form the direct precursor for plant hormones like brassinosteroids, abscisic acid, gibberellins and cytokinins (Bari and Jones, 2009; Vranová et al., 2013).

Alkaloids are derived from amino acids containing basic nitrogen atoms. Alkaloids are characterized by their great structural diversity and there is no uniform classification of alkaloids (Hesse, 2002). The structures of the different classes of secondary metabolites may already be quite complex but they get even more complex in combinations of molecules from these different classes. For instance, terpenenoids may combine with alkaloids to form terpene indole alkaloids. Different types of alkaloids, including these terpene indole alkaloids, are used in medicine, such as the analgesics morphine and codeine, the anticancer agents vinblastine and taxol, the gout suppressant colchicine,
the muscle relaxant (C)-tubocurarine, the antiarrhythmic ajmaline, the antibiotic sanguinarine, and the sedative scopolamine (Facchini, 2001). Because of their strong biological activity, the elucidation of the biosynthesis of these compounds would be of great value to mankind to improve the production of these highly specific medicinal compounds.

2. Terpene indole alkaloids

2.1 Importance of terpene indole alkaloids

The terpene indole alkaloids (TIAs) are a diverse class of secondary metabolites which are produced through the convergence of a secoiridoid biosynthesis pathway and an indole alkaloid biosynthesis pathway. After condensation of a molecule from each of these pathway the structures are further modified and rearranged, and up to date the collection of identified terpenoid indole alkaloids contains over 2000 members (O’Connor and Maresh, 2006) (Figure 1). TIAs are found in plant species belonging to the Apocynaceae, Loganiaceae, Rubiaceae and Nyssaceae. From these plant species, *Catharanthus roseus* is one of the best studied and therefore often serves as a model for TIA research. More than 130 TIAs have already been found in *C. roseus* (Jacobs et al., 2004; Wang et al., 2011). The natural role of TIAs in plants is probably related to defence against pests and diseases (Roepke et al., 2010). *C. roseus* or related species have been cultivated for many ages, both as ornamental plant but especially because of their medicinal properties. Root or shoot extracts were and are still used in Indian and Chinese traditional medicine against several diseases, including diabetes and malaria. The biologically most active compounds in *Catharanthus* extracts are vinblastine and vincristine, which are used in the treatment of leukemia and Hodgkin’s lymphoma (Brugères et al., 2009; Spagnuolo et al., 2010). Other active TIAs in *Catharanthus* are the anti-hypertensive drugs serpentine and ajmalicine and the anti-arrhythmic drug ajmaline (Sen, 2013).

2.2 Biosynthesis of terpenoids indole alkaloids

The biosynthesis of TIAs in *C. roseus* proceeds via approximately thirty enzymatic steps involving at least 35 known intermediates (Facchini and De Luca, 2008; Jacobs et al., 2004) and all of these steps are coordinated both in time and space as they may occur in different cell types (Figure 1). As mentioned above, TIA biosynthesis covers two convergent pathways: a secoiridoid biosynthesis pathway in which secologanin is synthesized from the precursor geranyl diphosphate (GPP) and the indole alkaloid
Figure 1: The terpene indole alkaloid (TIA) biosynthesis pathway and representative TIAs of Catharanthus roseus. Genes indicated in boxes were either published before (black background) or during (white background) the SmartCell program, or were discovered in the SmartCell program (this thesis: yellow background), where known at the onset of this thesis the subcellular localization of the enzyme is indicated (plastid, ER, cytosol, vacuole). The color of the box frame of the enzyme name indicates the cell specific expression of the gene encoding the enzyme (mRNA localization in the leaf internal phloem-associated parenchyma (IPAP): pink, in epidermis: blue). GPPS= geranyl diphosphate synthase; GES= geraniol synthase; IS= iridoid synthase; LAMT= loganic acid O-methyltransferase; SLS= secologanin synthase; STR= strictosidine synthase; TDC= tryptophan decarboxylase.
pathway, in which tryptamine is synthesized in one step from the amino acid tryptophan. Tryptamine and secologanin are condensed into strictosidine, which is the common precursor for all different types of TIAs (Figure 1). At the onset of this thesis work the full biosynthetic route had not been elucidated but the complex biosynthesis of strictosidine was thought to consist of four basic activities which divide the entire pathway in smaller sub-groups consisting of (1) biosynthesis of the monoterpene precursor GPP, (2) biosynthesis of secologanin from geraniol, (3) biosynthesis of tryptamine from tryptophan and finally (4) the condensation of tryptamine and secologanin to strictosidine.

2.2.1 Biosynthesis of monoterpene precursor GPP
Monoterpene precursors are generally assumed to be produced in the plastids through the MEP pathway. Indeed, Contin et al. proved that in C. roseus the initial steps for the biosynthesis of the monoterpene occurred via the MEP instead of the MVA pathway by feeding experiments in C. roseus cell cultures using isotopically labeled glucose (Contin et al., 1998). After this important breakthrough finding, a number of genes involved in the earlier steps of the MEP pathway in C. roseus were cloned and characterized, such as 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2C-methyl-d-erythritol-2,4-cyclodiphosphate synthase (MECS) and hydroxymethylbutenyl-4-diphosphate synthase (HDS) (Chahed et al., 2000; Veau et al., 2000). Elucidation of the entire MEP pathway shows that it consists of nine enzymatic steps (Vranová et al., 2012), but only recently some of the additional missing genes from the MEP pathway in C. roseus were characterized, such as isopentenyl diphosphate isomerase (IDI) and geranyl diphosphate synthase (GPPS) (Guirimand et al., 2012; Rai et al., 2013).

2.2.2 The secoiridoid biosynthesis pathway: from geraniol to secologanin
The class of terpenes that are combined with an alkaloid to form the TIAs are iridoid monoterpenes. Chemically, the iridoids usually consist of a cyclopentane ring fused to a six-membered oxygen heterocycle, while cleavage of a bond in the cyclopentane ring gives rise to a subclass known as the secoiridoid. For the secoiridoid pathway of C. roseus only the first two and last two enzymatic steps were characterized and for three of these the gene encoding the enzymes had been cloned at the onset of this thesis: geraniol-8-hydrolases (Collu et al., 2001), loganic acid methyltransferase (Murata et al., 2008) and secologanin synthase (Irmler et al., 2000) (Figure 1). The first step in the monoterpene biosynthesis pathway leading to secologanin must be catalysed by geraniol synthase (GES) which converts GPP to geraniol. At the onset of this thesis work the C. roseus gene for GES was not cloned yet, while at that time also the relevant sequence information from C. roseus was lacking. Therefore, in this project we used GES genes from other plant sources. The subsequent step in the secoiridoid biosynthesis
pathway is a cytochrome P450 enzyme, CYP76B6. This CYP76B6 was isolated from *C. roseus* and reported to catalyse hydroxylation of geraniol at the 8-position resulting in the formation of 8-hydroxygeraniol (Collu et al., 2001) and is therefore named G8O in this thesis. Subsequent steps in the pathway were not characterized at the onset of this thesis work (Figure 1). This means that when we started the work, the majority of the enzymatic steps from 8-hydroxygeraniol to loganic acid remained to be elucidated. Characterizing the last missing (supposedly) seven enzymatic steps in the secologanin biosynthesis pathway of *C. roseus* involved the main effort described in this thesis.

### 2.2.3 Condensation of tryptamine and secologanin to form strictosidine

The third important step in the indole branch of the pathway is the formation of tryptamine from tryptophan by the enzyme tryptophan decarboxylase (TDC). TDC is a pyridoxal-dependent enzyme and *TDC* was one of the first pathway genes cloned from *C. roseus* (de Luca et al., 1989). Finally, as fourth step in the pathway in strictosidine biosynthesis one molecule of tryptamine from the indole pathway and one molecule of secologanin from the secoiridoid pathway are combined by the strictosidine synthase (STR). Also the gene encoding STR was already cloned from *C. roseus* (McKnight et al., 1990).

### 2.3 Different subcellular and tissue localization of TIA biosynthesis in *C. roseus*

The secoiridoid biosynthesis branch of TIA production in plants is derived from the plastidial MEP pathway (Lichtenthaler, 1999; Oudin et al., 2007). During the course of this thesis the *C. roseus* GES was isolated and using fluorescent protein fusions it was shown that the GES protein localized to plastids and the stromules extending from the plastids which may be in close association with the membrane of the endoplasmic reticulum (ER) in *C. roseus* cells (Guirimand et al., 2009; Simkin et al., 2012). The G8O enzyme which uses geraniol as substrate is an ER-anchored protein due to the presence of a trans-membrane helix at the G8O N-terminal end (Guirimand et al., 2009). This also indicates that the geraniol which is produced in the plastids needs to be exported from the plastid to the ER anchored G8O for hydroxylation. This exchange from plastid to ER may occur through stromules that intercalate with ER membrane structures.

As discussed above, at the onset of this project the steps between hydroxygeraniol and loganic acid were not well characterized. However, LAMT and SLS catalyzing two steps after loganic acid were studied at the subcellular level. Transient expression of the YFP-LAMT fusion protein in *C. roseus* cells resulted in the detection of nucleocytosolic fluorescence signal of the fusion protein (Guirimand et al., 2011). SLS, just like G8O a cytochrome P450, is also anchored to the ER via an N-terminal helix (Guirimand et al., 2011). The product of SLS, secologanin is most likely produced on the cytosolic side of the ER membrane. Subsequently, secologanin together with tryptamine which
is produced in the cytosol (Guirimand et al., 2011) needs to be transported to the vacuole where strictosidine biosynthesis is completed, because STR is located in the vacuole (Guirimand et al., 2010) The different subcellular sites of the enzymes of the strictosidine pathway are illustrated in Figure 1.

Besides enzymes being active in different subcellular compartment, the different genes of the strictosidine pathway are also expressed in different cell types. The genes involved in the early steps of the strictosidine biosynthesis pathway (MEP pathway genes, GES and GBO) are expressed in internal phloem associated parenchyma (IPAP) cells (Burlat et al., 2004; Simkin et al., 2012) while the genes involved in the later steps of the pathway (SLS and STR) and indole branch (TDC) are all localized in the epidermis in the C. roseus aerial organs (Irmler et al., 2000; St-Pierre et al., 1999). After the common precursor strictosidine has been assembled in the epidermal cells, the different biosynthesis routes for different TIAs diverge structurally and spatially which results in final products of TIAs that are sequestered in different cell types (e.g. catharanthine in cell surface, vindoline in special cell types named idioblasts and laticifers cells) (St-Pierre et al., 1999; Verma et al., 2012).

Thus, TIA production in C. roseus therefore not only has a complexity at the cellular level with enzymes of the pathway being active in different subcellular compartments, but also at the tissue level with different intermediates being produced in different cell types of the leaf. Strictosidine production in C. roseus also requires transport of the intermediates between IPAP cells and epidermal cells. Since the steps between 8-hydroxygeraniol and loganic acid were unknown, in this project we hoped to find out which intermediate is transported between these two cell types.

3. Terpene indole alkaloid production in C. roseus

3.1 Classical breeding of TIA production in C. roseus

The anti-cancer compounds vinblastine and vincristine accumulate in extremely low amounts in C. roseus leaves (0.0003 to 0.01 % of the dry weight) (Rai et al., 2013), but industrial production still mainly relies on extraction from leaves. Due to the complexity of these molecules, it is very difficult to produce them by de novo chemical synthesis and therefore even for chemical synthesis natural precursors are mostly used, which are also extracted from C. roseus. But also the yield of these precursors is very poor (Ishikawa et al., 2008). Several options have already been explored to achieve higher production of precursors or final products in C. roseus. For example, C. roseus cultivars were screened for higher alkaloid content and breeding programs try to improve yield
(Dutta et al., 2005). Some of the features of *C. roseus* (it is diploid, has a short seed cycle and easily-controlled pollination) make it suitable for breeding and selection of plants with higher TIA content. However, such conventional breeding procedures are also time consuming and labour intensive and despite many such attempts only few novel cultivars have been obtained with marginally improved TIA yield (Chaudhary et al., 2012; Kulkarni et al., 1999; Mishra et al., 2001). Production level is therefore still far from the worldwide demand.

### 3.2 Engineering of TIA production in *C. roseus* cell suspensions or hairy roots

*C. roseus* tissue cultures (including cell and hairy root cultures) are the most common source for studying scalable TIA production. Manipulation of the production level of TIAS in cell and hairy roots cultures of *C. roseus* mainly focused on exogenous elicitation (including chemical and environmental induction, with eg. jasmonate and light (El-Sayed and Verpoorte, 2007; Vázquez-Flota et al., 2000) and genetic modification. Despite the significant efforts though the past years (Zhao and Verpoorte, 2007; Zhao et al., 2013), cell cultures or hairy roots cultures are still not an ideal platform for production. The problem is mostly that only part of the pathway is active in cells or hairy roots so that the main products that accumulate are monomeric TIAS such as ajmalicine, serpentine and tabersonine. Vinblastine and vincristine are not produced, apparently because this part of the pathway is not active in these cultures (Almagro et al., 2011; Vázquez-Flota et al., 2002). Metabolic reconstruction of this blocked pathway might solve this problem in the future. However, complex pathway compartmentalization and feedback regulatory control mechanisms may significantly increase the degree of unpredictability of metabolic engineering in *C. roseus*. For example, ORCA3 is a jasmonate responsive transcription factor that promotes transcription of TIA biosynthesis genes (Vom Endt et al., 2007). However, when ORCA3 is overexpressed in hairy roots, also repressor activity is induced, which in the long term actually causes a decrease in several TIA metabolites (Peebles et al., 2009). Because *C. roseus* is not easy to transform and grow the expression of the TIA pathway biosynthesis genes in a heterologous host that is easier to manipulate (e.g. microbes, tobacco) may provide a faster way towards engineering of the pathway once the relevant genes become available. A heterologous host may actually lack certain feedback regulation mechanisms which are present in *C. roseus*, but a heterologous host may have other types of problems, for example in dealing with the production of these potentially phytotoxic compounds.
3.3 Engineering of TIA production in a heterologous microbial host

Micro-organisms like yeast have been used for producing many different types of secondary metabolites (Komatsu et al., 2010; Verpoorte et al., 2002; Vogt and Jones, 2000). The degree of complexity in microorganisms is much lower than that of plant systems. This on the one hand may reduce the chance of (un)expected modifications of the engineered products by endogenous enzymes. On the other hand the unmodified product may be extremely harmful to the microbial cell and thus limit production. Therefore, extra steps are required to reduce toxicity to allow for continuous production or production to sufficiently high levels. Because the precursors for strictosidine biosynthesis are not available in micro-organisms, these have to be added. Production in yeast, for example, was tested by feeding tryptamine and secologanin to transgenic yeast expressing STR from C. roseus and indeed this did result in release of strictosidine into the medium (Geerlings et al., 2001).

3.4 Engineering of TIA production in a heterologous plant host

Plants are still one of the cheapest options to produce biomass and if the pathway of important secondary metabolites can be engineered into an easy to cultivate plant like tobacco production options would vastly increase. Moreover, a heterologous plant host would already possess the MEP precursor pathway. Already there are many examples of successful engineering of complex sesquiterpene biosynthesis pathways like that for artemisinin precursors in stable transformed tobacco plants (Farhi et al., 2011; Zhang et al., 2011) or by transient expression in N. benthamiana (Ting et al., 2013; van Herpen et al., 2010). Successful overexpression of monoterpene synthases has already been demonstrated for linalool synthase, and limonene synthase in tobacco, petunia and Arabidopsis (Aharoni et al., 2003; Aharoni et al., 2005; Degenhardt et al., 2003; Lücker et al., 2001; Ohara et al., 2003), but expression of more extensive monoterpene biosynthesis pathways was still lacking due to lack of pathway biosynthetic gene information. At the onset of this project only individual TIA pathway genes had been transformed to heterologous plant species (Hallard et al., 1997; Verpoorte and Memelink, 2002; Verpoorte et al., 2000). For example, tobacco and Morinda citrifolia cell suspension cultures expressing C. roseus TDC and STR genes can produce strictosidine when secologanin was fed to the cells (Hallard et al., 1997; Verpoorte et al., 2000). In this collaborative Smart cell project, C. roseus, Nicotiana benthamiana, Nicotiana tobaccum, Arabidopsis, and yeast were all used to evaluate enzyme function and to determine which host can produce the highest TIA end product or intermediate levels. In this thesis, we mainly focused on reconstitution of the pathway in the heterologous plant host N. benthamiana using transient expression mediated by agro-infiltration of leaves.
4. Tools for metabolic engineering in heterologous plant host

4.1 Strategies in TIA pathways gene discovery

At the onset of this thesis project, most of the terpene indole alkaloid biosynthetic genes were not identified, making heterologous expression of the full pathway impossible at that time. In the past, isolation of genes encoding TIA biosynthetic enzymes relied either on protein purification based on activity assays and protein amino acid sequencing. Alternatively, isolation of pathway genes was based on sequence homology based screening of cDNAs. While sequence information has become more readily available in the past decade, still identification of the gene encoding a certain enzymatic function of the TIA pathway may be a difficult task. However, thanks to the development of multiple high-throughput omics tools (transcriptomics, proteomics and metabolomics), it has become possible to correlate multiple data sets from which candidate genes can be predicted or at least prioritized from a long list of putative candidates genes.

The most common strategy is the integrated analysis of the transcriptome with metabolic profiling. This strategy is based on the assumption that a set of genes involved in the same pathway are co-regulated or co-expressed and that the expression of these correlates with the accumulation pattern of the pathway intermediates and end products. This strategy has already been used many times and was, for instance, used to select triterpene glycosyltransferases (GTs) from among 300 GTs expressed in *M. truncatula* (Achnine et al., 2005). In the case of *Papaver somniferum*, a cDNA encoding a short chain dehydrogenase/reductase involved in morphine biosynthesis was identified by comparing gene expression between morphine-containing and morphine-free *Papaver* species (Ziegler et al., 2006). With regard to *C. roseus*, Rischer et al. have started to establish a gene-to-metabolite network, which yielded a collection of known and previously undescribed transcript tags and metabolites associated with terpenoid indole alkaloids (Rischer et al., 2006). During this thesis work the *C. roseus* iridoid synthase, a crucial enzyme for TIA biosynthesis, was identified from a transcriptome assembly of *C. roseus* cDNA through co-expression analysis using the G8O expression profile as bait (Geu-Flores et al., 2012). Also in the SmartCell project discovery of the missing genes from the strictosidine biosynthesis pathway is based on metabolic profiling, selection of a candidate gene list based on sequence homology to known gene functions and comparison of expression profile of known genes of the pathway with candidate genes. This procedure will help to prioritize within the list of candidate genes (which can still be quite long) such that top candidate genes can be isolated and characterized in the different available expression platforms (*E. coli*, yeast, transient expression in *N. benthamiana* and stable transformation of tobacco or Arabidopsis).
4.2 Reconstruction of the biosynthesis pathway in heterologous plant hosts

Testing single candidate genes for function can be difficult when the putative substrates are not available. Also, once several or all genes of a pathway have been isolated the reconstruction of a (partial) pathway in a heterologous plant host can be quite time consuming. In the SmartCell project the group of Paul Christou of the University of Lleida in Spain attempted to introduce the pathway in stable transformed tobacco by co-bombardment of a set of genes. For these experiments it still takes considerable time before the results of the transformation can be tested. Recently, a transient expression system was established in *N. benthamiana* based on agro-infiltration in leaves. Instead of having to build expression vectors containing multiple genes, different agro-bacterium cultures, each carrying a different expression construct, can be mixed and co-injected into a leaf and results indicate that single cells are transformed by multiple agro strains. This expression system has been successfully used for metabolic engineering of the core parthway of artemisinic acid (van Herpen et al., 2010), indole glucosinolates (Pfalz et al., 2011), costunolide (Liu et al., 2011). The transient expression of multiple genes by agro-infiltration was used in this project to test whole groups of candidate genes together with a precursor pathway gene. If within the group of candidate genes the right gene(s) is/are present the analysis of the products will show and additional conversion step in the pathway. This is an efficient way to quickly test the activity of a whole group of unknown genes. Leaves can be extracted about one week after infiltration of the agrobacteria with the expression constructs with the appropriate buffer and metabolites in the extract are separated either by high-performance-liquid-chromatography (HPLC) or by gas chromatography (GC) depending on chemical properties of the expected products from the TIA pathway. After separation by HPLC or GC, mass spectrometry (MS) is used to (tentatively) identify and to quantify metabolites.

5. Outline of this thesis

This work was part of the Smartcell project, which is an EU Seventh Framework Program large collaborative project with 18 academic partner universities or institutes from nine European countries and two end-user companies. The main goal of SmartCell was to develop fundamental knowledge and enabling technology to rationally engineer and process plants and plant cells towards improved economical production of secondary metabolites (TIAs as a model) for industrial use. In total, 10 work packages were distributed over the different EU partners or beneficiaries and our activities were mainly in candidate gene discovery and validation and developing an analytical platform for metabolites analysis. At the beginning of this study, between seven steps in the hypothetical pathway from geraniol to secologanin were unidentified (Figure 1).
Candidates genes involved in these unknown steps were identified by other partners using transcriptomics and proteomics analysis of *C. roseus* tissues, using co-expression with known enzymatic steps of the pathway and matching product accumulation profiles as important criteria. In addition, candidate genes were identified from RNA sequencing information based on sequence homology with known enzyme functions. In the SmartCell project, gene discovery in *C. roseus* by transcriptomics and proteomics were conducted by VIB (Belgium) and Leiden university. The different candidate genes that were selected during this project were cloned into different expression construct for expression in *E. coli*, yeast or plants for initial characterization. Validated candidate genes were used in (partial) pathway reconstitution experiments in plants, either by transient expression or by stable transformation and methods to enhancing precursor input into the pathway by overexpression of a GPP synthase were explored. Eventually, the whole pathway from IPP and DMAPP to strictosidine (the common precursor for all TIAs) was reconstituted in *N. benthamiana*.

Selection of candidate genes was done by VIB (Belgium), University Catholique de Louvain (Belgium) and Leiden University. Functional analysis of candidate genes was done by Leiden University, CNRS (France) and Plant Physiology (PPH) Wageningen. Subcellular localization experiments with newly identified enzymes of the pathway were done by PPH Wageningen and VIB (Belgium). Reconstitution of the (partial) pathway into stable transformed tobacco was performed by University of Lleida, Spain, and testing partial pathway activity in cell suspensions or hairy roots was performed by VTT (Finland). Other partners in the SmartCell project were involved in the characterization of transporters, the testing of biological activities and the upscaling/optimisation of production platforms.

At the onset of the thesis the RNA sequence information from *Catharanthus* was not yet available and therefore work started by characterizing two GES genes from other plant species (GES from *Valeriana officinalis* and GES from *Lippia dulcis*) which were available. Chapter 2 describes cloning and functional characterization of these two geraniol synthase genes. *VoGES* and *LdGES* were expressed in *E. coli* to characterize the enzyme activities and in planta subcellular localization of the two GES proteins fused to the fluorescent marker was studied. Subsequently, *VoGES* and *LdGES* were transiently expressed in *N. benthamiana* to assess their in planta geraniol synthase activity. To further investigate which enzyme will result in higher geraniol production in planta, these two GES genes were stably transformed to *N. tabacum* which was analysed by metabolomics techniques using GC-MS and LC-MS to evaluate product formation.

Chapter 3 describes the characterisation of the gene(s) encoding the subsequent step in the pathway, geraniol hydroxylation. To be able to identify the most effective enzyme to be used for pathway reconstitution, CYP76 family members were studied using yeast
expression and an in vivo leaf disc assay. Transient expression of VoGES with G8O in N. benthamiana was used to establish whether it is possible to express a mini-pathway.

In Chapter 4 we explore the options to boost production of monoterpenes in plants by targeting enzymes to different subcellular compartments. Three different GPPS genes (previously characterized in vitro) were screened for in planta activity to identify a GPPS which can boost geraniol production. Subsequently, this GPPS was provided with an artificial plastid, cytosol or mitochondrial targeting signal. Similarly, the VoGES was provided with the same targeting signals and targeted VoGES and GPPS were co-expressed in all possible combinations by agro-infiltration of N. benthamiana leaves to assess which combination was most effective and to infer knowledge about the exchange of precursors and products between these compartments.

Chapter 5 describes the main goal of the Smartcell program: the reconstitution of the full secologanin biosynthesis pathway in planta, based on a collaborative effort from many of the Smartcell participants. It integrated transcriptomics and proteomics as a tool to select the candidate gene for the missing steps in secologanin biosynthesis. Four genes corresponding to these missing steps in the pathway were identified and the enzymes encoded by these genes were functionally characterized in in vitro enzyme assays. Tissue specificity of the expression of the genes was determined and for the proteins the subcellular localization was studied. Finally, we use transient expression system in N. benthamiana to reconstitute the whole (or partial) pathway to prove all genes’ function in vivo and at the same time evaluate whether N. benthamiana could be a platform for production of bioactive molecules.

Chapter 6 discusses the results from this thesis in a wider perspective. Bottlenecks in the metabolic engineering of complicated pathways in plants and options to improve methods are discussed.
References


Hesse, M., 2002. Alkaloids: nature’s curse or blessing?


Characterization of two geraniol synthases from *Valeriana officinalis* and *Lippia dulcis*: similar activity but difference in subcellular localization

Lemeng Dong\(^a\), Karel Miettinen\(^b\), Miriam Goedbloed\(^a\), Francel W.A. Verstappen\(^a\), Alessandra Voster\(^c\), Maarten A. Jongsma\(^c\), Johan Memelink\(^b\), Sander van der Klo\(^a\), Harro J. Bouwmeester\(^a\)

\(^a\) Laboratory of Plant Physiology, Wageningen UR, P.O. Box 658, 6700 AR Wageningen, The Netherlands

\(^b\) Institute of Biology, Sylvius Laboratory, Leiden university, P.O. Box 9505, 2300 RA Leiden, The Netherlands

\(^c\) Plant Research International, Wageningen UR, P.O. Box 619, 6700 AP Wageningen, The Netherlands

Published in *Metabolic Engineering* (2013) 20: 198–211
http://dx.doi.org/10.1016/j.ymben.2013.09.002
Abstract

Two geraniol synthases (GES), from *Valeriana officinalis* (VoGES) and *Lippia dulcis* (LdGES), were isolated and were shown to have geraniol biosynthetic activity with $K_m$ value of 32 µM and 51 µM for GPP, respectively, upon expression in *E. coli*. The *in planta* enzymatic activity and sub-cellular localization of VoGES and LdGES were characterized in stable transformed tobacco and using transient expression in *Nicotiana benthamiana*. Transgenic tobacco expressing VoGES or LdGES accumulate geraniol, oxidized geraniol compounds like geranial, geranic acid and hexose conjugates of these compounds to similar levels. Geraniol emission of leaves was lower than that of flowers, which could be related to higher levels of competing geraniol-conjugating activities in leaves. GFP-fusions of the two GES proteins show that VoGES resides (as expected) predominantly in the plastids, while LdGES import into the plastid is clearly impaired compared to that of VoGES, resulting in both cytosolic and plastidic localization. Geraniol production by VoGES and LdGES in *N. benthamiana* was nonetheless very similar. Expression of a truncated version of VoGES or LdGES (cytosolic targeting) resulted in the accumulation of 30% less geraniol glycosides than with the plastid targeted VoGES and LdGES, suggesting that the substrate geranyl diphosphate is readily available, both in the plastids as well as in the cytosol. The potential role of GES in the engineering of the TIA pathway in heterologous hosts is discussed.

**Keywords:** Engineering; Subcellular localization; geraniol synthase; metabolic profiling; enzyme assay
1. Introduction

Plants are estimated to produce more than 500,000 secondary metabolites of various classes (isoprenoids, phenylpropanoids, alkaloids) (Hadacek, 2002). Of these, the isoprenoids represent the largest family based on their diverse structural features which relate to numerous biological activities. Isoprenoids have been shown to affect many physiological processes such as respiration, signal transduction, cell division, membrane architecture, photosynthesis, and growth. In addition, isoprenoids have ecological significance as they play an important roles in the exchange of signals between plants and between plants and microorganisms or in defense against pathogens and herbivores. Also the applications of isoprenoids in foods, cosmetics and pharmaceutical drugs make specific terpenoids interesting commercial targets.

Although isoprenoids are extraordinarily diverse, they all originate from the condensation of the universal five-carbon precursors, isopentenyl diphosphate (IPP) and dimethyl allyl diphosphate (DMAPP). In higher plants, two independent pathways, located in separate intracellular compartments, are involved in the biosynthesis of IPP and DMAPP. In the cytosol, IPP is derived from the classic mevalonic acid (MVA) pathway that starts from acetyl-CoA (Porter et al., 1981), whereas in plastids, IPP is formed from pyruvate and glyceraldehyde 3-phosphate via the methylerythritol phosphate (MEP or non-mevalonate) pathway (Eisenreich et al., 2001; Lichtenthaler, 1999). Cytosolic IPP and DMAPP are converted to farnesyl diphosphate (FPP, C15), which serves as a precursor of sesquiterpene and triterpene biosynthesis in the cytosol. In contrast, the plastidial pool of IPP/DMAPP is converted to geranyl diphosphate (GPP, C10) and geranylgeranyl diphosphate (GGPP, C20) which serve as precursors for monoterpenes, and diterpenes and tetraterpenes, respectively, in the plastid (Lange et al., 2001; McConkey et al., 2000; Tholl and Lee, 2011; Turner et al., 1999).

Geraniol is an acyclic monoterpene alcohol that is synthesized in one step from GPP. Geraniol is a component of essential oils present in many fragrant plant species (Antonelli et al., 1997; Bakkali et al., 2008; Bayrak and Akgül, 1994; Sangwan et al., 2001; Yang et al., 2005). It has a rose-like odor and is commonly used in perfumes (Chen, 2006; Rastogi et al., 2003) and aromatic fragrance in wine (Herrero et al., 2008; Pedersen et al., 2003). Geraniol also has pharmaceutical properties, as it can inhibit the growth of human colon cancer cells (Carnesecchi et al., 2001) and interfere with membrane functions in *Candida albicans* and *Saccharomyces cerevisiae* (Bard et al., 1988). In some plant species geraniol is the precursor for terpenoid indole alkaloid (TIA) biosynthesis. For instance, in *Catharanthus roseus* the anticancer agents vinblastine and vincristine are synthesized from geraniol (monoterpenoid iridoid branch) and tryptophan (indole branch) in the TIA pathway.
Multiple approaches have been tested to increase TIA production. For example, overexpress gene 1-deoxy-D-xylulose synthase and geraniol-10-hydroxylase gene were shown to increase the flux towards vinblastine and vincristine in *C. roseus* hairy root (Peebles et al., 2011). Attempts to boost transcription of TIA biosynthetic genes in the hairy roots or suspension cells were only partially successful (Liu et al., 2011; Memelink and Gantet, 2007; Montiel et al., 2007). For example, ORCA3 is a jasmonate responsive transcription factor that promote transcription of TIA biosynthesis genes (Vom Endt et al., 2007). However, when ORCA3 is overexpressed, also repressor activity is activated, which in the long term actually caused a decrease in several TIA metabolites in *C. roseus* (Peebles et al., 2009). Expression of the TIA pathway biosynthesis genes in a heterologous host may provide a way to overcome such feedback regulation problems.

The objective of the present study was the efficient production of the monoterpene geraniol as the first step in a larger program to rebuild the complete monoterpene iridoid branch of the TIA biosynthesis pathway in a heterologous host. To achieve this, a geraniol synthase (GES) was cloned from *Valeriana officinalis* L. (Valerianaceae) (*VoGES*) and compared to the previously isolated *LdGES* from *Lippia dulcis* (Yang et al., 2011). Both proteins showed similar geraniol synthase activity *in vitro* and *in planta*. *VoGES* was subsequently used in a number of transient and stable metabolic engineering approaches to explore the possibility to reconstitute the monoterpene branch of TIA biosynthesis in tobacco.

2. Materials and methods

2.1 Cloning and sequence analysis of geraniol synthase gene

For the cloning of the geraniol synthase gene, *Valeriana officinalis* L. (*VoGES*) total RNA was isolated from *V. officinalis* leaves using SV Total RNA Isolation System (Promega). Based on conserved domains of known geraniol synthases, the degenerate primers (forward primer 5’-GAYGARAAYGGIAARTTYAARGA-3’ and reverse primer 5’-CCRTAIGCRTCRAAIGTRTCTC -3’) were designed to amplify partial cDNA fragment by reverse transcription PCR (RT-PCR). Full length sequences of the cDNAs were obtained by rapid amplification of cDNA ends (RACE).

Putative *VoGES* sequence was blasted against the GenBank ENTREZ database (NCBI Blast 2.2.23) (Altschul et al., 1997) and GES sequences were aligned using CLUSTALX 1.83 (Thompson et al., 1997) using standard settings. Prediction of the subcellular localization was from the targeting prediction programs PREDOTAR version 1.03 (http://urgi.versailles.inra.fr/predotar/) (Small et al., 2004) and TARGETP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al., 2000).
2.2 Heterologous expression of VoGES and LdGES protein in Escherichia coli

For the in vitro functional analysis of the putative geraniol synthase from *Valeriana officinalis* and comparison with LdGES, the truncated cDNAs ΔNVoGES (bp 178-1785) and ΔNLdGES (bp 139-1755) were subcloned into the multiple cloning site of the expression vector pRSET A (Invitrogen) to yield constructs pRSET-ΔNVoGES and pRSET-ΔNLdGES. Primer sequences for PCR amplification and restriction sites for each primer are listed in Table S1. After full re-sequencing to check integrity, constructs were transformed into *E. coli* BL21 (DE3) (Invitrogen) and expression was induced by isopropyl β-D thiogalactopyranoside (IPTG) in transformed *E. coli* BL21 (DE3) cell cultures. The His-tagged proteins were isolated by passing through Ni-NTA Spin columns according to the manufacturers recommendations (Qiagen). For quality analysis, the recombinant protein was confirmed with 12.5 % (w/v) SDS-PAGE gel electrophoresis followed by Western blotting using mouse monoclonal anti-His horseradish peroxidase (HPR) conjugate antibodies (5Prime, http://www.5prime.com). Antibody binding was detected by incubation in 250 μM sodium luminol, 0.1 M Tris-HCl (pH 8.6), 3 mM H₂O₂, 67 μM *p*-coumaric acid and exposure to X-ray film.

An enzyme assay was carried out for functional characterization, using geranyl diphosphate (GPP) and farnesyl diphosphate (FPP) as substrates. Enzymatic assays were done in 0.5 ml reaction buffer containing 50 mM Tris-HCl, 1 mM MgCl₂, 0.1 mM MnCl₂, and 10, 20, 50 or 100 μM GPP (or 62.5 μM FPP) and 0.5 μg (VoGES) and 2 μg (LdGES) of purified enzyme. The reaction mixture was incubated at 32 °C for 5 min. For quantitative analysis citronellol was added to a concentration of 50 μM as an internal standard into the reaction tube after incubation. The reaction was stopped by adding 1 volume of hexane, mixing thoroughly by vortexing and keeping on ice for 10 min. The tubes were centrifuged at 4000 g for 10 min, and the supernatant hexane phase was collected. The extraction was repeated with hexane (0.5 ml). Then the hexane phase was collected and dehydrated and then subjected to capillary gas chromatography-flame ionizing detector (GC-FID) and gas chromatography-mass spectrometry (GC-MS, supplementary method). For the latter, the hexane extract was separated on a Agilent GC 6890 series equipped with a DB-5 capillary column (30 m × 0.25 mm, film thickness of 0.25 μm) (J&W Scientific) using nitrogen as carrier gas at a flow rate of 1.2 ml min⁻¹. The separation conditions were: split mode 1: 5, injection volume 5 μl, injector temperature 230 °C, initial oven temperature 100 °C, then linear gradient to 140 °C at a rate of 10 °C min⁻¹ followed by a linear gradient to 240 °C at a rate of 35 °C min⁻¹. Amounts of geraniol formed in enzyme assays were calculated from the resultant GC/FID integral using the relative response factor with respect to the citronellol internal standard. Lineweaver-Burk plots of VoGES and LdGES activity were used to obtain the $K_m$ values for GPP.
For analysis of the subcellular targeting, the coding sequences of EGFP was fused to the N-terminus or C-terminus of full length VoGES and LdGES. In addition a truncated version of LdGES lacking the first 46 AA (ΔNLdGES: bp 139-1755) and a truncated version of VoGES lacking the first 56 AA (ΔNVoGES: bp 178 -1785) was made using standard cloning techniques and the C-terminal coding sequence of these genes was fused in frame to that of GFP. The VoGES-GFP, LdGES-GFP, GFP-VoGES, GFP-LdGES, ΔNVoGES-GFP and ΔNLdGES-GFP were cloned into impact vector plV2A 2.1 (www.pri.wur.nl/UK/products/ImpactVector/) under control of the CaMV 35S-promoter. In addition, the truncated versions of VoGES and LdGES were provided with a heterologous plastid import signal by cloning into impact vector pIV2A 2.4 which carries an artificial plastid targeting signal (www.pri.wur.nl/UK/products/ImpactVector/) (Wong et al., 1992). The fusion constructs were sequenced to check integrity, before transferring the fusion cassettes to the binary vector pBIN+ (van Engelen et al., 1995) using LR recombination (Gateway technology) (Karimi et al., 2002). Primer sequences for PCR amplification and restriction sites for each primer are listed in Table S1. Finally, the binary expression constructs were transformed into Agrobacterium tumefaciens strain AGL0 (Lazo et al., 1991) which was used for transient expression in Nicotiana benthamiana as described below. Expression and localization were analyzed at different days post-agroinfiltration in small leaf samples (~0.5 cm² leaf material from at least three independent agroinfiltrated plants) by confocal laser scanning.

Microscopy using an Axiovert 200 M with a Zeiss LSM 5 PASCAL laser scanning microscope (Carl Zeiss) and a 20 × (N.A. 0.5) Plan NeoFluar (Zeiss) or a 63 × (N.A. 1.4 oil) Plan Apochromat (Zeiss) objective. Samples were excited with 1 % of a 488 nm laser (emission from a 30 mW argon tube) for EGFP and chlorophyll excitation and green EGFP fluorescence and red chlorophyll fluorescence were collected using two emission filters. Band pass was set to 505 to 530 nm for EGFP fluorescence detection and long pass was set to > 560 nm for chlorophyll autofluorescence detection. Images were exported from .lsm files to .tif files using Zeiss LSM image browser version 3.5 without further processing.

2.4 Tobacco stable plant transformation

The VoGES and LdGES full length cDNA sequences were subcloned into the binary expression vector pBIN+ using standard cloning techniques. In this vector VoGES and LdGES are under the control of the constitutive CaMV-d35S promoter and a nopaline synthase terminator. Both the 35S::LdGES and 35S::VoGES pBIN+ constructs were introduced into A. tumefaciens AGL0 and in vitro grown wild-type N. tabacum 'Samsun NN' was transformed using the leaf disc method as described (Horsch et al., 1985).
Transformed shoots were selected on medium with 100 mg/l kanamycin. Primary transformed shoots were rooted on non-selective medium, checked by PCR for presence of the expression construct and positive T0 shoots were transferred to soil and grown until seed set. T1 plants were grown until seed set and from the T2 population five homozygous lines 35S::LdGES and three homozygous lines 35S::VoGES with single insert were selected.

2.5 Transient expression in Nicotiana benthamiana

A. tumefaciens infiltration (agro-infiltration) was performed according to the description of van Herpen et al. (van Herpen et al., 2010). Briefly, A. tumefaciens was grown at 28 °C for 24 hours in LB media with kanamycin (50 mg/L) and rifampicillin (34 mg/L). Cells were harvested by centrifugation for 20 min at 4000 g and 20 °C and then resuspended in infiltration buffer containing 10 mM MES (2-(N-morpholino) ethanesulfonic acid, Duchefa Biochemie), 10 mM MgCl$_2$ and 100 µM acetosyringone (4′-hydroxy-3′, 5′-dimethoxycetophenone, Sigma) to a final OD$_{600}$ of ~0.5, followed by incubation at room temperature under gentle shaking at 50 rpm for 150 min. In all experiments, A. tumefaciens harboring TBSV p19 was included to maximize protein production by suppression of gene silencing (Voinnet et al., 2003). A. tumefaciens harbouring constructs with GFP-GES or GES were infiltrated into leaves of five-week-old N. benthamiana plants by pressing a 1 mL syringe without metal needle against the abaxial side of the leaf and slowly injecting the bacterium suspension into the leaf. N. benthamiana plants were grown from seeds on soil in the greenhouse with a minimum of 16 hour light. Day temperatures were approximately 28 °C, night temperatures 25 °C. After agro-infiltration the plants remained under the same greenhouse conditions until further analysis.

2.6 Volatile GC–MS analysis

For analysis of the production of volatiles by 35S::LdGES or 35S::VoGES tobacco seedling, T2 seeds were sterilized by chlorine bleach and placed on ½ MS (Duchefa, Netherlands) medium. Petri dishes with ~50 seeds were incubated at 4 °C in the dark to ensure synchronous germination. Subsequently, petri dishes were transferred to growth cabinets with 16 L/8 D, 28 °C day, 25 °C night. After 2 weeks of growing, petri dishes were placed in 0.5 L glass jars, and volatiles were trapped for 5 hours in the light using Tenax TA (20/35 mesh, Alltech, Breda, the Netherlands).

Headspace analysis of stable transformed VoGES tobacco was done using detached leaves and flowers. Before they were enclosed in 1 L glass jars with a Teflon-lined lid equipped with in- and outlet, detached leaves and flowers from tobacco plants at different development stages were placed into small glass bottles with water in them.
while different parts of flowers were placed on wet filter paper. A vacuum pump was used to draw air through the glass jar at approximately 100 ml min\(^{-1}\), with the incoming air being purified through a glass cartridge (140 × 4 mm) containing 150 mg Tenax TA. At the outlet the volatiles emitted by the samples were trapped on a similar Tenax cartridge. After 5 h of trapping, the trapped volatiles were analyzed by Thermodesorption GC-MS using a thermal desorber (Unity, Markes International Limited) and a Trace GC Ultra (Thermo Electron Corporation) coupled with DSQ mass spectrometer (Thermo Electron Corporation). The tubes were first purged to remove water vapor and oxygen for 2 min at room temperature with helium flow of 50 ml min\(^{-1}\). Then trapped volatiles were desorbed from the Tenax in the thermal desorber at 250 °C for 5 minutes. Volatiles were collected in an electrically cooled sorbent trap (Unity; Markes, Llantrisant) at 10 °C and injected into the analytical column (ZB-5MSI, 30 m × 0.25 mm ID, 1.0 μm-film thickness, Zebron, Phenomenex). The temperature program of the gas chromatograph started at 40 °C (3 min hold) and rose to 280 °C at 12 °C min\(^{-1}\), with a hold at final temperature for 2 min. The mass spectrometer was set to scan from 45 to 300 m/z with a scan time of 5.4 scans s\(^{-1}\). The helium flow was constant at 1.0 ml min\(^{-1}\). Ionization potential was set at 70 eV. For quantification, a geraniol calibration curve was made with a series of standard solutions. The GC-MS results were analyzed using Xcalibur software (Thermo, Waltham). After headspace trapping for 5 h in the light, samples were frozen in liquid nitrogen and stored at -80 °C for further analysis.

### 2.7 GC-MS analysis of extracts

Volatile compounds accumulated in the plant material were extracted with dichloromethane (DCM) and measured by GC-MS as described above. Aliquots of 500 mg of frozen, powdered material in pre-cooled glass tubes were extracted twice with 3 ml DCM. Extracts were shortly vortexed, sonicated for 10 min, centrifuged for 10 min at 1200 g and filtered through a small glass column containing anhydrous Na\(_2\)SO\(_4\). The eluent was concentrated extracts under a flow of nitrogen and 1 μl concentrated extracts were injected into the GC-MS column. The initial oven temperature was 45 °C for 1 min, and was increased to 300 °C at a rate of 10 °C min\(^{-1}\) and held for 5 min at 300 °C. For quantification of geraniol and oxidised geraniol products, standards of geraniol, geranial and geranic acid were injected at different concentrations to establish calibration curves. Each sample was spiked with 2 μg/μl cis-nerolidol as internal standard.

### 2.8 Analysis of geraniol-derived conjugates by LC- QTOF-MS and LTQ-Orbitrap-MSn

Analysis of non-volatile compounds in transgenic tobacco extracts was done by liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC-QTOF-
Two geraniol synthases: similar activity but difference in subcellular localization

Aliquots of 200 mg of frozen, powdered material were extracted with 0.6 ml 99.9% MeOH / 0.133% formic acid in 1.5 ml Eppendorf vial. After short vortex and 15 min sonication, the extracts were centrifuged and filtered through 0.45 µm filters (SRP4, Sartorius, Germany) and 5 µl of the filtered extract was analysed using a Waters Alliance 2795 HPLC connected to a QTOF Ultima V4.00.00 mass spectrometer (Waters, MS technologies, UK). Measurements were in negative ionization mode and leucine encephalin ([M – H]⁻ = 554.2620) was used as a lock mass for online mass calibration.

Acquisition of LC-MS data was performed under MassLynx 4.0 (Waters). MassLynx was used for visualization and manual processing of LC-PDA-MS data. Mass data were processed using metAlign version 1.0 (www.metalign.nl). Baseline and noise calculations were performed from scan number 70 to 2,480. The maximum amplitude was set to 25,000 and peaks below three times the local noise were discarded. Multiple mass signals derived from the same compound were grouped using MSClust software (biotools.wurnet.nl) by Multivariate Mass Spectra Reconstruction (MMSR) (Tikunov et al., 2005). The selected mass intensities were normalized using log₂ transformation and standardized using range scaling, in which each value in a certain row, corresponding to the internal standard leucine encephalin, was divided by the intensity range observed for this row throughout all samples analysed. Each row was then mean centred. Finally, the normalized and log-transformed data matrix was used for Principal Components Analysis implemented in GeneMath XT version 2.1.

Significance of differences in intensity of each aligned mass signal between samples was assessed using student t-test (level of significance set at 0.05). Masses showing significant difference were manually checked in MassLynx. Putative identification of metabolites was by determining the best fit elemental composition using C, H and O with MassLynx software. For multiple possible molecular formulas (tolerance <5 ppm) the best matches were searched in the Dictionary of Natural Products and SciFinder databases for possible structures.

For further identification, selected compounds were targeted for fragmentation by an Accela HPLC tower connected to a LTQ Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific). The instrument settings of HPLC and LTQ Orbitrap were used as described earlier (van der Hooft et al., 2011). The LTQ was programmed to use a window of 10 D to isolate the mass of interest in MS1. The data-dependent fragmentation was set as follows: MS2 fragmentation of most intense ion in MS1; MS3 fragmentation of the 5 most intense fragment ions in MS2; MS4 fragmentation of the 5 most intense fragment ions in each MS3.


2.9 Glycosidase treatment

To quantify how much of the geraniol related products were modified by glycosylation, extracts were treated with glycosidase to release geraniol or geraniol related compounds. For this purpose, 200 mg infiltrated leaf material was ground in liquid nitrogen and extracted with 1 ml citrate phosphate buffer, pH 5.4. The extracts were prepared by brief vortexing and sonication for 15 min. 0.2 ml of Viscozyme L (Sigma) was added and overlaid with 1 ml of pentane to trap released volatiles and the samples were again vortexed. To trap released volatiles, 1 ml pentane was added on top of the extract. The mixture was incubated overnight at 37 °C, and subsequently extracted twice with 2 ml of pentane. Extracts were dehydrated using anhydrous Na$_2$SO$_4$ and concentrated to approximately 100 µl. An internal standard, cis-nerolidol was used to quantify the products. Samples were analyzed using GC-MS as described under 2.7.

2.10 Western blotting

Western blotting was used to check the integrity of the LdGES-GFP fusion proteins after transient expression in *N. benthamiana*. 10 mg liquid nitrogen grounded agro-infiltrated *N. Benthamiana* leaf was suspended in sample buffer (50 mM Tris pH 6.8, 2% SDS, 0.1% BFB, 10% glycerol, 4% β-me). The suspension together with sample was heated 10 min at 94 °C and centrifuged 1 min at 13000 rpm. 15 µl supernatant was first run in a 12% SDS-PAGE gel, subsequently was blotted on to a PDVF membrane (Biorad) for 1 hr at 100V. The membrane was blocked overnight at 4 °C in TBS-T (Tris buffered saline with 0.05% Tween 20) containing 3% non-fat dried milk. Next the blot was incubated with 1: 20000 rabbit anti-GFP (ab290, Abcam) for 1 hr at room temperature, and washed three times in TBS-T. Then the blot was incubated with donkey anti rabbit-HRP 1: 50000 dilution (Thermo Scientific) in TBS-T, followed by washing four times in TBS-T. 3 ml Supersignal® West Dura Extended Duration Substrate (Thermo scientific) was used.

2.11 GC-MS analysis of the enzymatic product from VoGES and LdGES expressed in E. coli

Products were analyzed by GC-MS using a gas chromatograph (7809A, Agilent Technologies, USA) equipped with a 30 m × 0.25 mm, 0.25 mm film thickness column (ZB-5, Phenomenex) and a Triple-Axis detector (model 5975C, Hewlett Packard, Agilent Technologies). The injection port (splitless mode), interface and MS source temperatures were 250 °C, 290 °C and 180 °C respectively. The injection volume was 2 µl. The oven was programmed at an initial temperature of 45 °C for 1 min, with a ramp of 10 °C min$^{-1}$ to 280 °C, and final time of 5 min. Scanning was performed from 45-300 atomic mass unit (amu). The helium inlet pressure was checked by electronic pressure control to achieve
Two geraniol synthases: similar activity but difference in subcellular localization

3. Results

3.1. Functional characterization of geraniol synthase from Valeriana officinalis and Lippia dulcis in vitro

The GES gene from Valeriana officinalis (VoGES) encodes a protein which shares 37-67% identity with GES proteins from other plant species (Table 1, Figure 1). Previous characterized GES gene from Lippia dulcis (LdGES) shares 63% identity with VoGES. The alignment shows that the Aspartate-rich DDxxD-motif for metal-dependent ionization of the prenyl diphosphate substrate (Bohlmann et al., 1998; Tarshis et al., 1996; Wendt and Schulz, 1998) is present in all GES sequences (Figure 1). Both the VoGES and LdGES protein with N-terminal truncation were produced in E. coli (see methods) to compare the in vitro activity. Analysis of both recombinant protein by SDS-PAGE and Coomassie Brilliant Blue staining and immunoprobing with anti-His antibodies showed the presence of one major band (Figure 2A). In the presence of GPP, both enzymes catalyzed the formation of the acyclic monoterpenic alcohol geraniol, which was identified based on comparison with the retention time of authentic geraniol (Figure 2) and the mass spectrum comparison with the NIST library (Figure S1). Truncated VoGES and LdGES catalyzed the conversion of GPP to geraniol with a \( K_m \) value of 32 \( \mu M \) and 51 \( \mu M \), respectively (Figure S2). No product was detected when VoGES or LdGES were supplied with FPP (data not show).

Table 1: Comparison of Geraniol synthase proteins from different plant species and their predicted location by Target P (http://www.cbs.dtu.dk/services/TargetP/).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Identities (%)</th>
<th>Full Length</th>
<th>Predicted Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoGES</td>
<td>100</td>
<td>594aa</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>LdGES</td>
<td>63</td>
<td>584aa</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>CrGES</td>
<td>67</td>
<td>589aa</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>ObGES</td>
<td>59</td>
<td>567aa</td>
<td>Mitochondrion</td>
</tr>
<tr>
<td>CtGES</td>
<td>42</td>
<td>603aa</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>PfGES</td>
<td>37</td>
<td>603aa</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>VvGES</td>
<td>57</td>
<td>595aa</td>
<td>Chloroplast</td>
</tr>
</tbody>
</table>

TP: targeting peptide
Figure 1: Alignment of deduced amino acid sequences of fObGES (*Ocimum basilicum* geraniol synthase, AY362553), PfGES (*Perilla frutescens* geraniol synthase, DQ234300), CtGES (*Cinnamomum tenuipilum* geraniol synthase, AJ457070), CrGES (*Catharanthus roseus* geraniol synthase, AFD64744), LdGES (*Lippia dulcis* geraniol synthase, GU136162), VvGES (*Vitis vinifera* geraniol synthase, ADR74218), and VoGES (*Valeriana officinalis* geraniol synthase) using CLUSTALX 1.83 program. The metal ion-binding motif DDXXD is double underlined. The black star indicates the putative cleavage site of the targeting signal.
Two geraniol synthases: similar activity but difference in subcellular localization

3.2. Different subcellular localization of VoGES and LdGES

Monoterpene synthases are generally believed to be localized in the plastids, and also their precursor GPP is supposed to be produced in the plastids. Target P (Emanuelsson et al., 2000) predicts plastid targeting for all GES proteins except ObGES, which is predicted to be targeted to the mitochondria (Table 1). Transient expression of the GES-GFP fusion proteins in N. benthamiana leaves showed that for VoGES-GFP the fluorescence signal was located in the plastids (Figure 3A). The fluorescence of VoGES-GFP was also detected in the stromules that emanate from the plastids, which was most clearly observed in epidermal cells (Figure 3B). However, for LdGES-GFP, the GFP fluorescence was observed in the cytosol and around the plastids (Figure 3C). No stromule labelling was observed with LdGES-GFP in epidermal cells, confirming impaired import of LdGES into plastids and suggesting preferred localization in the cytosol. The cytosolic
signal of GFP fluorescence can also be explained by instability of the LdGES-GFP fusion protein: if the GFP moiety is cleaved from the fusion protein in the cells, this would result in cytosolic localization. We therefore checked the integrity of the fusion proteins after transient expression in *N. benthamiana* on Western blots (Figure S3). The western blot shows that the signal detected by anti-GFP is of the predicted size of the LdGES-GFP fusion protein (91 kD), indicating that the fluorescence signal is derived from the intact fusion protein, which for LdGES-GFP is located in mostly in the cytosol. For the N-terminal fusion proteins, both GFP-VoGES and GFP-LdGES (full length and truncated one) located (as expected) to the cytosol (Figure 3D and E, result only show the GFP with full length). The fluorescence signal for LdGES-GFP and GFP-LdGES fusion proteins was consistently lower than that of VoGES-GFP, suggesting a lower stability for the LdGES protein.

### 3.3. The different subcellular localization of VoGES and LdGES does not affect transient in planta activity

GES is a monoterpene synthase which uses the substrate GPP, which is mostly synthesized in the plastidic MEP pathway. Because of the different subcellular location of VoGES and LdGES (Figure 3), it was of interest to compare their *in planta* activity. For this purpose, the native VoGES, LdGES and the VoGES-GFP and LdGES-GFP fusion proteins were expressed transiently in *N. benthamiana* leaves. Ten days after agro-infiltration the plastidic (VoGES-GFP) and cytosolic (LdGES-GFP) targeting of the fusion protein was again confirmed by confocal fluorescence microscopy. No free geraniol was detected in the headspace of leaves expressing VoGES or LdGES and their GFP fusions when assayed at ten days post-agroinfiltration (not shown). However, previous experiments with expression of LdGES in maize had shown that the LdGES product geraniol may be converted by endogenous enzymes to geranial, geranic acid and other geraniol related products, which subsequently are sequestered as glycosylated compounds (Yang et al., 2011). Therefore, *N. benthamiana* leaves expressing the different constructs were extracted with citrate phosphate buffer and either mock treated or treated with Viscozyme, to release any glycosylated geraniol compounds. The released geraniol, geranial and geranic acid were trapped in the pentane overlay. Indeed, without glycosidase treatment, the levels of free geraniol and geraniol derivatives were below the level of detection by GC-MS in the citrate phosphate leaf extracts. In contrast, with Viscozyme treatment of the leaf samples, geraniol and the geraniol related compounds geranial, geranic acid, nerol, and neral were detected by GC-MS (Table 2).
**Table 2:** GC-MS quantification of geraniol and its derivatives in *N. benthamina* leaves, *N. tobacum* leaves and flowers extracts

<table>
<thead>
<tr>
<th></th>
<th>Geraniol</th>
<th>Geranial</th>
<th>Geranic acid</th>
<th>Nerol</th>
<th>Neral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vis-</td>
<td>Vis+</td>
<td>Vis-</td>
<td>Vis+</td>
<td>Vis-</td>
</tr>
<tr>
<td><em>N. benthamiana</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VoGES</td>
<td>nd</td>
<td>47.82 ± 5.14</td>
<td>nd</td>
<td>1.05 ± 0.11</td>
<td>nd</td>
</tr>
<tr>
<td>LdGES</td>
<td>nd</td>
<td>45.10 ± 10.24</td>
<td>nd</td>
<td>0.58 ± 0.29</td>
<td>nd</td>
</tr>
<tr>
<td>VoGES-GFP</td>
<td>nd</td>
<td>46.23 ± 3.85</td>
<td>nd</td>
<td>0.95 ± 0.18</td>
<td>nd</td>
</tr>
<tr>
<td>LdGES-GFP</td>
<td>nd</td>
<td>72.07 ± 17.55</td>
<td>nd</td>
<td>0.53 ± 0.38</td>
<td>nd</td>
</tr>
<tr>
<td>empty vector</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td><em>N. tobacum</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VoGES leaf</td>
<td>4.25 ± 1.16</td>
<td>32.04 ± 9.53</td>
<td>nd</td>
<td>0.35 ± 0.05</td>
<td>nd</td>
</tr>
<tr>
<td>VoGES flower</td>
<td>1.27 ± 0.21</td>
<td>6.39 ± 0.79</td>
<td>nd</td>
<td>0.27 ± 0.04</td>
<td>nd</td>
</tr>
<tr>
<td>wild type</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
</tbody>
</table>

Results are the mean of 3 replicates ± standard deviation (µg/g fw). nd: not detected. Vis+ : with Viscozyme treatment; Vis-: without Viscozyme treatment.
No geraniol or geraniol related products were detected after Viscozyme treatment of *N. benthamiana* leaves infiltrated with empty vector (Table 2). There was no significant difference in geraniol and the different geraniol derived products from free GES or the GES-GFP fusion proteins, indicating that the C-terminal fusion of GFP does not impair the GES activity. Moreover, there was not much difference in GES related product accumulation for the full length VoGES and LdGES (Table 2). The presence of the products geranial, geranic acid, nerol, and neral indicates that geraniol was further converted to these products, likely by endogenous *N. benthamiana* enzymes.

**Figure 3:** Visualization of subcellular compartments of *N. benthamiana* cells using transient transformation with VoGES and LdGES fusion with epi-fluorescence GFP (e-GFP). A. VoGES-GFP (mesophyll cell), B. VoGES-GFP (epidermal cell), C. LdGES-GFP, D. GFP-VoGES, E. GFP-LdGES, F. ΔNLdGES-GFP, G. ΔNVoGES-GFP. CHL, chlorophyll autofluorescence (red), GFP, GFP fluorescence (green), Merge, merged green and red images. Arrows indicate stromules. Bar represents 5 µm.
Two geraniol synthases: similar activity but difference in subcellular localization

To determine if geraniol production is possible from a cytosolic localized GES we tested an N-terminal truncated version of both VoGES and LdGES (ΔNVoGES-GFP and ΔNLdGES-GFP, respectively), which lack the plastid import signal. As expected, both these constructs showed cytosolic localization when expressed in *N. benthamiana* (Figure 3F and 3G). In addition, we made expression constructs of the ΔNLdGES and ΔNVoGES with an N-terminal fusion to the same *A. thaliana* ribulose-1,5-bisphosphate carboxylase small subunit transit peptide (pΔNLdGES and pΔNVoGES, respectively) (Wong et al., 1992). The transient expression of ΔNLdGES and ΔNVoGES in *N. benthamiana* leaves resulted in similar levels of geraniol glycosides, but levels were about 30% lower than that produced by truncated genes fused to the heterologous plastid targeted signal (pΔNLdGES and pΔNVoGES) of the native VoGES and LdGES (Figure 4).

![Figure 4: Relative quantitative analysis of geraniol glycosides in agro-infiltrated *N. benthamiana* leaves with truncate (ΔN), artificial plastidic target (p) LdGES and VoGES and empty vector control (EV) by LC-MS. Bars represent the total mass intensity of all geraniol-derived glycosides (malonyl-hexosyl-geraniol, acetyl-hexosyl-hydroxy-dihydro-geranic acid, acetyl-dihexosyl-geraniol, hexosyl-carboxygeranic acid, hexosyl-hydroxy geraniol, pentosyl-hexosyl-geraniol and pentosyl-hexosyl-hydroxy-tetrahydro-geranic acid).](image)

3.4. Similar activity of VoGES and LdGES in stably transformed tobacco plants

Headspace analysis of stable transomants of *Nicotiana tabacum* L. ‘Samsun NN’ showed no geraniol in the headspace of wild type plantlets, but plantlets transformed with either VoGES or LdGES expression constructs did emit geraniol into the headspace (Figure 5A). Geraniol production differed between lines, but the production of geraniol was similar
in the highest producer of VoGES and highest producer of LdGES transformants (Figure 5B).

A

B
Two geraniol synthases: similar activity but difference in subcellular localization

Chapter 2

Figure 5: GC-MS headspace analysis of volatiles and principal component analysis of non-volatiles on transgenic and wild type tobacco plantlets.

A. GC chromatogram showing selected iron 69 of a geraniol standard, transgenic line VoGES18, LdGES10 and WT. Geraniol eluted at retention time 17.10 min.

B. Geraniol concentration in the headspace of tobacco plantlets (n= 3 groups ~50 plants)

C. Principal component analysis of the LC-MS data on tobacco plantlets. PC1, PC2 and PC3 describe 49.7%, 21.5% and 15.3% of the total metabolic variation, respectively. V5, V9 and V18 indicate independent transgenic lines of VoGES; L1, L7, L8, L10 and L11 indicate independent transgenic lines of LdGES (three biological replicates of each line).

To analyse the production of non-volatile geraniol-derived products produced in planta in VoGES or LdGES transformed plants, the semipolar, non-volatile metabolites were analysed by LC-QTOF-MS. PCA of reconstituted mass peaks (see materials and methods) shows clustering of the replicate groups and separation between WT and transgenic lines in the first component (49.7% of the variation; Figure 5C). There was no separation between VoGES and LdGES, confirming that both GES proteins have very similar in planta activity, despite their different subcellular localization. Because of strong similarity in chemical profiles between the VoGES and LdGES lines, subsequently only VoGES18 was analyzed in more detail.
3.5. Different geraniol related compounds in flowers and leaves of transformed plants expressing GES

To determine free geraniol emission capacity of different parts of the plant, the headspace of leaves and flowers of different developmental stages was analyzed. Leaves were harvested from the plant at the onset of flowering. The headspace volatiles were collected for 5 hours during the day. Analysis by GC-MS showed that the youngest leaves emit most geraniol and emission decreased in older leaves, while there was no significant difference between middle and old leaves (Figure 6A). Analysis of flowers in different stages of development showed that emission was highest in freshly opened flowers (Figure 6B). To understand which part of the flower contributes most to the geraniol emission, different parts of the tobacco flowers were isolated and separately analysed for headspace production. Figure 6C shows that petal, sepal and ovary contribute almost equally to the geraniol in the total flower headspace. However, we noted that after dissection of the flower the production of geraniol increased almost 10-fold compared with the intact flower, possibly as a result of a wound induced increase in the flux through the MEP pathway (Bonaventure and Baldwin, 2010; Cordoba et al., 2009). Alternatively, it is possible that wounding results in induction of glycosidases that could liberate glycosidically bound geraniol (Edwards and Wratten, 1983; Hoagland, 1990).

To determine the level of geraniol-derived glycosides in different tissues, leaves and flowers were treated with Viscozyme. Analysis of the released volatiles by GC-MS showed that without glycosidase treatment only geraniol is detected in the tissue extracts. Viscozyme treatment strongly increased geraniol production and now also geranial, and geranic acid were detected (Table 2). Combined, the level of glycosylated geraniol products in leaves (32.04 µg/g fw) were about 5-fold higher than in flowers (6.39 µg/g fw).

For analysis of the conjugated GES related products in the transformants the LC-MS chromatograms of WT and VoGES18 plants were compared and 19 compounds were identified with significantly higher signal in the transgenic line (Table 3). For identification of these putative geraniol-related products, the masses were targeted for fragmentation by LTQ Orbitrap MSn. According to the MSn spectra from 18 of the differential masses, the compounds are likely hexose, pentose and malonyl or acetyl conjugates of geraniol, hydroxy-geraniol, geranic acid, hydroxy-geranic acid and hydroxy-dihydro-geranic acid. One mass (709.3562), which is specific for sepal tissue could not be identified as a geraniol related product (Table 3). A malonylated hexosyl geranic acid conjugate was present in flower tissues but not in leaves, whereas a hexosyl hydroxyl-geranic acid conjugate was specific for leaves.
**Figure 6:** Headspace analysis of different tissues of transgenic VoGES tobacco plants.
A. Geraniol concentration of different developmental stage leaves. The numbers 3, 8 and 16 represent young (8 cm), middle (20 cm) and old (20 cm) age leaves, respectively.
B. Geraniol emission of different developmental stage flowers.
C. Geraniol emission of different parts of flowers.
Combined, the results suggest that the glycosylating activities for geraniol and geraniol derivatives are tissue-dependent. According to the mass intensity signal of the different glycoside conjugates in leaves, 89% were directly derived from geraniol, while 11% were apparently derived from geraniol after further modification to hydroxy-geraniol, geranic acid, hydroxy-geranic acid or hydroxy-dihydro-geranic acid (Figure 7). In flowers, the proportion of geraniol and further modified products was slightly different from that in leaves in sepal, petal and ovary (Figure 7), likely as a result of different endogenous enzyme activities in different tissues.
<table>
<thead>
<tr>
<th>Retention time</th>
<th>Observed mass</th>
<th>Putative identity</th>
<th>Calculated mass</th>
<th>ΔMass (ppm)</th>
<th>Elemental formula</th>
<th>MS(n) fragment</th>
<th>Tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.04</td>
<td>345.1563</td>
<td>Hexosyl-hydroxy-geranic acid</td>
<td>345.1555</td>
<td>2.35</td>
<td>C_{26}H_{35}O_{10}</td>
<td>131, 143, 149, 161, 167, 179, 183, 206, 219, 261, 282, 299, 307, 327, 331</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>30.95</td>
<td>435.1862</td>
<td>Acetyl-hexosyl-hydroxy-dihydro-geranic acid (formic acid adduct)</td>
<td>435.1866</td>
<td>-0.92</td>
<td>C_{27}H_{36}O_{11}</td>
<td>157, 185, 205, 231, 249, 267, 291, 309, 315, 355, 375, 389, 417</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>32.98</td>
<td>435.1869</td>
<td>Acetyl-hexosyl-hydroxy-dihydro-geranic acid (formic acid adduct)</td>
<td>435.1866</td>
<td>0.69</td>
<td>C_{27}H_{36}O_{11}</td>
<td>131, 157, 185, 205, 231, 249, 267, 291, 309, 315, 373, 391, 417</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>40.20</td>
<td>447.2242</td>
<td>Pentosyl-hexosyl-geraniol</td>
<td>447.2235</td>
<td>1.52</td>
<td>C_{21}H_{35}O_{10}</td>
<td>131, 143, 149, 161, 179, 191, 233, 251, 293, 311, 357, 401, 430</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>39.88</td>
<td>447.2243</td>
<td>Pentosyl-hexosyl-geraniol</td>
<td>447.2235</td>
<td>1.87</td>
<td>C_{21}H_{35}O_{10}</td>
<td>131, 143, 149, 161, 179, 191, 221, 239, 251, 275, 281, 293, 311, 315</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>39.17</td>
<td>447.2245</td>
<td>Pentosyl-hexosyl-geraniol</td>
<td>447.2235</td>
<td>2.28</td>
<td>C_{21}H_{35}O_{10}</td>
<td>131, 143, 149, 161, 179, 191, 217, 221, 233, 251, 281, 293, 311, 347, 420</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>38.45</td>
<td>461.2036</td>
<td>Pentosyl-hexosyl-geranic acid</td>
<td>461.2023</td>
<td>2.82</td>
<td>C_{21}H_{35}O_{11}</td>
<td>149, 191, 211, 239, 293, 311, 339, 373, 389</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>42.41</td>
<td>461.2193</td>
<td>Deoxyhexosyl-hexosyl-geraniol</td>
<td>461.2192</td>
<td>0.21</td>
<td>C_{21}H_{35}O_{10}</td>
<td>161, 179, 213, 267, 295, 307, 323, 373, 383, 395, 415, 447</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>23.54</td>
<td>463.2178</td>
<td>Pentosyl-hexosyl-hydroxy-geraniol</td>
<td>463.2185</td>
<td>-1.40</td>
<td>C_{26}H_{35}O_{11}</td>
<td>143, 149, 161, 191, 221, 251, 275, 293, 311, 349, 387, 425</td>
<td>L, S</td>
</tr>
<tr>
<td>36.50</td>
<td>477.2335</td>
<td>Dihexosyl-geraniol</td>
<td>477.2331</td>
<td>0.75</td>
<td>C_{22}H_{37}O_{11}</td>
<td>161, 179, 221, 251, 281, 307, 323, 341, 361, 421, 449</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>25.37</td>
<td>611.2547</td>
<td>Dipentosyl-hexosyl-hydroxy-dihydro-geranic acid</td>
<td>611.2551</td>
<td>-0.65</td>
<td>C_{26}H_{43}O_{16}</td>
<td>149, 161, 191, 211, 251, 281, 293, 317, 347, 395, 429, 441, 461, 479</td>
<td>L, S, P</td>
</tr>
<tr>
<td>24.95</td>
<td>611.2560</td>
<td>Dipentosyl-hexosyl-hydroxy-dihydro-geranic acid</td>
<td>611.2551</td>
<td>1.47</td>
<td>C_{26}H_{43}O_{16}</td>
<td>149, 185, 221, 251, 275, 293, 311, 333, 360, 377, 435, 449, 461, 479</td>
<td>L, S, P</td>
</tr>
<tr>
<td>44.84</td>
<td>489.2329</td>
<td>Acetyl-pentosyl-hexosyl-geraniol</td>
<td>489.2336</td>
<td>-1.43</td>
<td>C_{23}H_{37}O_{11}</td>
<td>131, 149, 161, 191, 203, 233, 315, 327, 335, 367, 447, 463</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>47.39</td>
<td>503.2486</td>
<td>Acetyl-deoxyhexose-hexose-geraniol</td>
<td>503.2492</td>
<td>-1.19</td>
<td>C_{24}H_{39}O_{11}</td>
<td>143, 161, 243, 247, 307, 329, 443, 461, 477</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>37.06</td>
<td>579.2661</td>
<td>Dipentosyl-hexosyl-geraniol</td>
<td>579.2653</td>
<td>1.40</td>
<td>C_{24}H_{43}O_{14}</td>
<td>233, 293, 352, 372, 401, 425, 463, 493, 511, 567</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>34.30</td>
<td>609.2729</td>
<td>Pentosyl-dihexosyl-geraniol</td>
<td>609.2753</td>
<td>-3.94</td>
<td>C_{27}H_{45}O_{15}</td>
<td>191, 242, 267, 299, 315, 355, 387, 413, 447, 455, 475, 499, 537, 571</td>
<td>L, S, P, O</td>
</tr>
<tr>
<td>34.08</td>
<td>709.3562</td>
<td>Unknown 1</td>
<td>709.3588</td>
<td>-3.67</td>
<td>C_{46}H_{53}O_{11}</td>
<td>161, 173, 399, 445, 449, 467, 495, 541, 583, 609, 663, 699</td>
<td>S</td>
</tr>
<tr>
<td>47.95</td>
<td>803.3712</td>
<td>Malonyl-hexosyl-geraniol (dimer)</td>
<td>803.3701</td>
<td>1.31</td>
<td>C_{38}H_{59}O_{18}</td>
<td>131, 149, 161, 221, 329, 357, 401, 450, 501, 597, 643, 717</td>
<td>L, S, P</td>
</tr>
<tr>
<td>48.72</td>
<td>831.3299</td>
<td>Malonyl-hexosyl-geranic acid (dimer)</td>
<td>831.3287</td>
<td>1.44</td>
<td>C_{38}H_{55}O_{20}</td>
<td>161, 167, 191, 265, 307, 349, 371, 415</td>
<td>S, P, O</td>
</tr>
</tbody>
</table>

Δmass (ppm), deviation between the observed mass and calculated mass in ppm
L, S, P, O, leaf, sepal, petal, ovary.
4. Discussion

4.1. Localisation of VoGES in the plastids and LdGES mostly in the cytosol has no effect on in planta activity

In this study, we show that both VoGES and LdGES only produce the product geraniol in vitro as specific product from GPP (Figure 2). Both VoGES and LdGES contain a putative plastid signal peptide (Table 1) and monoterpene synthases are generally believed to be active in the plastids (Bouvier et al., 2000; Dudareva et al., 2005; Turner et al., 1999). For instance, CrGES from C. roseus has been shown to be localized in the plastids (Simkin et al., 2012). Our localization experiments showed that VoGES directs the GFP protein to the plastids, as expected, and no cytosolic signal was observed from transient expression of VoGES-GFP in agroinfiltrated leaves of N. benthamiana, indicating that import into the plastids was not saturated. The plastidic localization of VoGES-GFP was also evident from GFP signal in the stromules extending from the plastids (Figure 3A-B). In contrast, the signal of LdGES-GFP was mostly in the cytosol. Although some fluorescence in or around the plastids was observed, no LdGES-GFP signal was localized to the stromules of the plastids (Figure 2C). Control experiments show that the cytosolic signal from LdGES-GFP does not result from cleavage of GFP from the fusion protein (Figure S3). The GFP localization studies thus suggest a predominantly cytosolic localization of LdGES. Because product analysis (Table 2 and Figure 5) shows that native VoGES and LdGES have similar activity in planta, this suggests that the substrate pool of GPP is not limiting for native LdGES in the cytosol. However, we also note that activity of the truncated VoGES and LdGES proteins (both localized to the cytosol) was only 30% of plastid localized VoGES and LdGES, suggesting that (although for these cytosolic GES proteins the GPP substrate is available) GPP is not as abundant available as in plastids. Indeed, it recently was shown that in tomato fruits cytosolic monoterpene biosynthesis can be supported by plastid-generated geranyl diphosphate substrate (Gutensohn et al., 2013). The higher activity of native (mostly) cytosolic LdGES and truncated ΔNLdGES could be related to the partial plastidic localization of native LdGES.

4.2. Multiple oxidized and glycosylated forms of geraniol produced by endogenous enzymes

In our experiments approximately 11% of the produced geraniol was converted to the oxidized products geranial, hydroxy-geraniol, geranic acid, hydroxy-geranic acid and hydroxy-dihydro-geranic acid, which were mostly present as conjugates (Figure 7, Table 3). These conversions of geraniol are likely catalyzed by endogenous tobacco enzymes. For examples, geranic acid is likely to be synthesized from geranial by an NAD+ dependent dehydrogenase (Boyer and Petersen, 1991; Campos-García, 2010; Davidovich-Rikanati.
et al., 2007; Lüddeke et al., 2012). Geranic acid was only present in glycosylated forms in the transgenic plants and not as the free form (Table 2). Because low levels of geranial were detected only after and not before de-glycosylation, presumably this product is formed from oxidation of geraniol during the de-glycosylation reaction.

4.3. Biotransformation of geraniol in different hosts

In this study, VoGES and LdGES were expressed in both a microbial, \textit{E. coli}, and plant host (\textit{N. benthamiana} and \textit{N. tobacum}). With \textit{E. coli} protein, we could only detect geraniol as product, while geraniol, geranial and geranic acid were observed in transiently expressing \textit{N. benthamiana} as well as stable transformed \textit{N. tobacum}. In addition, in \textit{N. benthamiana} small amounts of nerol and neral were detected (Table 2). Previously it was shown that LdGES expressed in maize also results in geranyl acetate production in leaves (Yang et al., 2011), while this product was absent in \textit{N. benthamiana} or tobacco plants expressing the same LdGES (this paper). This shows that different heterologous GES hosts metabolize the produced geraniol differently, and that especially plants tend to produce a wider spectrum of geraniol derivatives. Similar observation were made with expression of \textit{Ocimum basilicum} geraniol synthase (ObGES) in yeast, \textit{E. coli}, grape and tobacco (Fischer et al., 2012) and tomato (Davidovich-Rikanati et al., 2007). As expected, in all hosts the main product of ObGES was geraniol, but minor product (linalool, citronellol, nerol, α-terpineol...) formation differed between the heterologous hosts.

4.4. Different glycosylation patterns of geraniol products in different tissues and stages of development

Geraniol and geraniol derivatives accumulated in glycosylated form, some of which were further modified by endogenous malonyltransferases or acetyltransferases. The glycosides were mono-glycoside, di-glycosides and tri-glycosides. The hexoside, ubiquitous in nature, which can potentially be attached to geraniol is glucopyranoside, whereas the pentosides could be arabinofuranoside, arabinopyranoside, apiofuranoside, xylopyranoside or rhamnopyranoside (Genovés et al., 2005; Maicas and Mateo, 2005; Mateo and Jiménez, 2000). Presumably, the glycosylated geraniol products are sequestered in the vacule (Wink, 2010; Winterhalter and Skouroumounis, 1997). The profile of glycosylated geraniol products differed between the agro-infiltrated \textit{N. benthamiana} leaves and leaves of the stable transformed tobacco plants (Table 2). The glycosylation profile also differed between tissues and developmental stages within the same plant (Table 3). For instance, in the mature leaves of stable transformed tobacco plants expressing GES, the acetylated and malonylated glycosides of geraniol are ~5-fold higher than in seedlings (data not shown), suggesting that acetylation and malonylation
are induced later during development. In total, the bioconversion by endogenous enzyme activities distributed the product of VoGES, geraniol, over at least 19 different compounds. Clearly, for effective reconstruction of a multiple step terpene biosynthesis pathway, such diversion of products needs to be avoided in order to raise the desired product yield. The use of different promoters, driving expression in tissues more specialized for the production and/or storage of terpenoids, could be one way to achieve this.

4.5. No pleiotropic effects of GES overexpression

Introduction of GES into tobacco could potentially result in competition for the substrate GPP with other monoterpene synthases or result in depletion of the substrates isopentenyl diphosphate and dimethylallyl diphosphate which are important for the formation of other terpenes, carotenoids and gibberellins, and this could have a dramatic effect on the plant phenotype (Aharoni et al., 2003; Davidovich-Rikanati et al., 2007). However, no leaf or flower phenotype was observed in our stable transformed tobacco plants expressing VoGES or LdGES, despite high transgenic product levels suggesting that sufficient GPP is available in both leaves and flowers. Also, in the LC-MS and GC-MS analysis of transgenic plants or after transient expression of GES genes in N. benthamiana, only new products were detected but no obvious reduction in endogenous products, which suggests there is indeed no competition for substrate.

In the selected transgenic lines no pleiotropic effect of GES overexpression was observed. We cannot exclude that in the regeneration of stable transformed tobacco plants with 35S-VoGES or 35S-LdGES a selection occurred against very high GES expression, resulting in either toxic levels of geraniol or creating a drain of GPP required for other processes. Boosting of the GES production in the stable transformed tobacco plants (e.g. by combining with GPPS overexpression) could result in toxic effects of geraniol, but these effects may be eliminated by simultaneous co-expression with P450 enzymes which convert geraniol to less reactive intermediates/end products and/or by using tissue-specific promoters to drive geraniol production in tissues more specialized for terpene production and storage (for example trichomes).

4.6. Higher fluxes through geraniol sequestration than through geraniol emission

We note that emission of geraniol in stable transformed tobacco plants changed during development. For instance, in young seedlings emission was ~8 ng/g h, while in mature plants emission by leaves declined from young to old leaves from 55 to 5 ng/g h (Figure 5 and Figure 6). In flowers of stable transformants emission of geraniol increased until flowers were fully open from 12 to 28 ng/g h (Figure 6). Based on the average headspace emission, the loss of geraniol from leaves would be ~17µg/g month and from flowers ~13µg/g month, while 37 µg/g fwof geraniol (derived) products...
Two geraniol synthases: similar activity but difference in subcellular localization

Chapter 2

(determined by de-glycosylation) are stored in 1 month old tobacco leaves (Table 2). *N. benthamiana* leaves at four days post-agroinfiltration show an average emission rate of 0.13 ng/g h (not shown), while no geraniol was emitted into the headspace ten days post-agro-infiltration (data not shown). This might suggest that after ten days not only has the T-DNA with the expression construct degraded, but that also the GES protein is no longer active in the cells, indicating a relatively high turn-over of terpene synthase protein.

Assuming that (1) effective expression in transient assays is from day 4-7 and that (2) geraniol production is at least 0.13 ng/g h in *N. benthamiana* leaves and that (3) flux through the pathway is 24 hr/day (which may not be the case if precursors are preferably produced in the light (Aharoni et al., 2003), the cumulative production of emitted geraniol over 4 days of transient expression would be ~0.012 µg/g fw. However, at ten days post agroinfiltration, the combined level of accumulated geraniol (derived) products (determined by de-glycosylation) was around 56 µg/g fw (Table 2).

Assuming that geraniol derived products in *N. benthamiana* accumulate during four days of continuous activity, this suggests a geraniol production of ~ 0.58 mg/ g h fw. Total level of geraniol derived products in the stable transformants is ~37µg/g fw and that is produced in 1 month, the geraniol production rate of stable transformants in tobacco is only 0.05 µg/ g h fw. The high production in the transient assay is most likely due to an extremely high gene dosage number of the infiltrated GES expression construct compared to the low copy number in stable transformants. At present it is not clear why the high GES activity in the transient expression assay is not matched by a higher emission of geraniol, but one explanation may be that the conjugation capacity of infiltrated *N. benthamiana* leaves is highly efficient.

4.7. Towards a monoterpene iridoid pathway in heterologous plants

Expression of GES in tobacco is the first step in a larger effort to introduce the terpene indole alkaloid (TIA) biosynthesis into tobacco plants, consisting of a monoterpene iridoid branch and an indole branch (Mahroug et al., 2007; Simkin et al., 2012). The second step of this pathway is the conversion of geraniol to 10-hydroxy-geraniol by geraniol 10-hydroxylase (G10H), but (as described above) this conversion may already take place by endogenous tobacco enzymes (Table 3). We are currently testing whether co-expression of VoGES with a G10H can effectively compete for the endogenous unwanted oxidation and glycosylation of the primary product geraniol. In the context of multiple-step monoterpene pathway reconstruction in heterologous plant hosts, elimination of the conversion by endogenous enzyme activities may become an important target for efficient channeling towards the desired product(s).
Acknowledgements

This research was funded by the European Community’s Framework VII Program FP7/2007-2013 to the SMARTCELL project KBBE–2007–3–1–01. Lemeng Dong was supported by a PhD fellowship from China Scholarship Council. We thank Bert Schipper and Ric de Vos for their help with LC–MS analysis.

References


Two geraniol synthases: similar activity but difference in subcellular localization


Figure S1: GC-MS profile of volatiles produced in vitro from GPP by N-terminal truncated VoGES and LdGES. A. GC chromatogram, B. Mass spectrum of the peak at 11.16 min, C. Mass spectrum of geraniol from the NIST library.
Figure S1: GC-MS profile of volatiles produced in vitro from GPP by N-terminal truncated VoGES and LdGES. A. GC chromatogram, B. Mass spectrum of the peak at 11.16 min, C. Mass spectrum of geraniol from the NIST library.

Figure S2: Michaelis-Menten (left) and Lineweaver-Burk (right) plots of the geraniol production from GPP as catalyzed by truncated VoGES and LdGES. The data are means ± standard errors of three replicates.

Figure S3: Western blot analysis of LdGES-GFP fusion protein in agro-infiltrated leaves. N. benthamiana leaves infiltrated with LdGES-GFP (lane 1) with water (lane 3). Lane 2: purified GFP protein.
**Table S1:** Primer sequences for PCR amplification and restriction site for *E. coli* expression and GFP fusion constructs.

<table>
<thead>
<tr>
<th>Name</th>
<th>Primer (5’-3’)</th>
<th>Restriction site</th>
</tr>
</thead>
<tbody>
<tr>
<td>LdGES-pRSETAF</td>
<td>TATAGGATCCTCACTCATCAGTATCAGCTGCTACTGCT</td>
<td>BamHI</td>
</tr>
</tbody>
</table>
| LdGES-pRSETAR | TATAAGATCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA

*N, N terminal; C, C terminal*
Chapter 3

Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the secoiridoid pathway

René Höfer¹, Lemeng Dong², François André³, Jean-François Ginglinger¹, Raphael Lügan¹, Carole Gavira¹, Sébastien Grec⁴, Gerhard Lang⁵, Johan Memelink⁶, Sander Van Der Krol², Harro Bouwmeester², Danièle Werck-Reichhart¹

¹ Institut de Biologie Moléculaire des Plantes, CNRS UPR 2357-Université de Strasbourg, France
² Laboratory of Plant Physiology, Wageningen University, 6700 AR Wageningen, The Netherlands
³ CEA, iBiTecS, Service de Bioénergétique Biologie Structurale et Mécanismes (SB2SM), Gif-sur-Yvette, France
⁴ Université Lille Nord de France, Lille 1, France
⁵ Philip Morris Products S.A., Research & Development, Quai Jeanrenaud 5, CH-2000 Neuchâtel, Switzerland
⁶ Institute of Biology, Leiden University, Sylvius Laboratory, PO Box 9505, 2300 RA Leiden, The Netherlands.

Published in Metabolic Engineering (2013) 20: 221-232
http://dx.doi.org/10.1016/j.ymben.2013.08.001
Abstract

The geraniol-derived secoiridoid skeleton is a precursor for a large group of bioactive compounds with diverse therapeutic applications, including the widely used anticancer molecule vinblastine. Despite of this economic prospect, the pathway leading to iridoid biosynthesis from geraniol is still unclear. The first geraniol hydroxylation step has been reported to be catalyzed by cytochrome P450 enzymes such as CYP76B6 from Catharanthus roseus and CYP76C1 from Arabidopsis thaliana. In the present study, an extended functional analysis of CYP76 family members was carried-out to identify the most effective enzyme to be used for pathway reconstruction. This disproved CYP76C1 activity and led to the characterization of CYP76C4 from A. thaliana as a geraniol 9- or 8-hydroxylase. CYP76B6 emerged as a highly specialized multifunctional enzyme catalyzing two sequential oxidation steps leading to the formation of 8-oxogeraniol from geraniol. This dual function was confirmed in planta using a leaf-disc assay. The first step, geraniol hydroxylation, was very efficient and fast enough to outcompete geraniol conjugation in plant tissues. When the enzyme was expressed in leaf tissues, 8-oxogeraniol was converted into further oxidized and/or reduced compounds in the absence of the next enzyme of the iridoid pathway.

Keywords: Catharanthus roseus; cytochrome P450; geraniol oxidation; multifunctional enzyme; in planta enzyme assay; pathway reconstitution
Geraniol hydroxylase activities of the CYP76 family of cytochrome P450 enzymes

1. Introduction

*Catharanthus roseus* or Madagascar periwinkle (formerly known as *Vinca rosea*) contains a complex mixture of about 130 terpenoid indole alkaloids (TIAs) or Vinca alkaloids with a broad range of biological activities (van der Heijden et al., 2004; O’Connor and Maresh, 2006). Among them, vinblastine and the related compound vincristine, or the synthetic derivatives vindesine and vinorelbine are major drugs used for the treatment of various cancers and as leads for further drug development (van Der Heijne et al., 2004; Roepke et al., 2010). Despite their low *in planta* concentration and the presence of a complex mixture of other alkaloids, these compounds and their monomeric precursors are still extracted from the periwinkle plant, with low yield and at high cost. For these reasons *C. roseus* became one of the most studied medicinal plants and a model for biotechnological developments in plant secondary metabolism (see e.g. Montiel et al., 2006; Peebles et al., 2009; 2011; O’Connor, 2012; Zhao et al., 2013).

The formation of the complex vinblastine molecule is thought to be catalyzed by around 30 enzymes with intricate subcellular localization and tissue-specific expression patterns (Guirimand et al., 2011; Verma et al., 2012). All TIAs found in plants are derived from a common precursor, strictosidine, which is formed by condensation of tryptamine and the iridoid secologanin. While tryptamine biosynthesis and many of the steps in the formation of the vinblastine precursor vindoline have been described, the pathway that leads to secologanin is still partly unclear. This is surprising since iridoids are quite widespread in dicotyledenous plants and exhibit a range of interesting bioactivities that make them candidates for therapeutic treatments as anti-inflammatory, anticancer, antimicrobial, neuroprotective, immunomodulating, hypoglycemic, hypolipidemic, antispasmodic and anti-oxidative compounds (Dinda et al., 2011; Tundis et al., 2008; Viljoen et al., 2012). Moreover, iridoids could be of great interest as agrochemicals to repel insects (Birkett et al., 2011; Kahn et al., 2012) and are one of the main classes of compounds responsible for the bitter taste of plants or plant extracts, as exemplified by olive oil (Mateos et al., 2004). With the objective to improve secologanin and vinblastine production in *C. roseus* or to introduce this pathway in heterologous engineered host plants, we focused on the iridoid segment of the pathway. The first dedicated step in this pathway is the formation of geraniol from geranyl diphosphate (Simkin et al., 2013), followed by 8-hydroxylation (commonly misnamed geraniol 10-hydroxylation). The cytochrome P450 enzyme CYP76B6 isolated from *C. roseus* was previously reported to catalyze geraniol 8-hydroxylation (Collu et al., 2001). The same activity was reported for the closely related CYP76B10 from *Swertia mussotii* (Gentianaceae) that accumulates the iridoid monoterpenoid swertiamarin (Wang et al., 2010). Geraniol 8-hydroxylase activity was also claimed for CYP76C1 of *Arabidopsis thaliana* in a patent by Otah and Mizutani (1998), which suggests that geraniol hydroxylation could be a common
property of CYP76s from different plant species. These reports led us to investigate and compare geraniol hydroxylase activities of a set of CYP76s from different plant species in order to identify the most suitable candidate for metabolic engineering and optimal production of 8-hydroxygeraniol in planta. In contrast to what was reported in the patent by Otah and Mizutani (1998), we show that CYP76C1 does not have geraniol hydroxylase activity. However, we identify CYP76C4, also from A. thaliana, as an active geraniol 8-, 9-hydroxylase. In addition, our data show that C. roseus CYP76B6 is a highly efficient and multifunctional geraniol-8-oxidase, that catalyzes the first two committed oxidation steps of the iridoid pathway, the regioselective geraniol 8-hydroxylation and the further oxidation to 8-oxogeraniol, both in vitro and in planta. In the absence of the next enzyme in the iridoid pathway, engineering of the first two steps of the pathway in Nicotiana benthamiana leads to the accumulation of 8-oxogeraniol, and particularly reduced, further oxidized, and conjugated metabolites of 8-oxogeraniol, which suggests that intermediate channeling will be critical to reconstruct the entire TIA pathway in heterologous plant hosts.

2. Material and Methods

2.1 Generation of expression vectors

The CYP76C1, CYP76C2, CYP76C5 and CYP76C7 constructs, were generated by PCR amplification of the desired fragment from cDNA prepared from different tissues of A. thaliana with primers harboring appropriate enzyme restriction sites. The digested PCR fragments were then integrated into the yeast expression vector pYeDP60 (Table S3). For CYP76C3, CYP76C4, CYP76C6 and CYP76B6 the USER™ cloning technique (New England Biolabs) was used according to Nour-Eldin et al. (2006). Their coding sequences were amplified from cDNA prepared from A. thaliana flowers (CYP76Cs) or C. roseus (CYP76B6) using PCR primers designed to introduce a USER cloning site (Table S3). The resulting fragments were integrated into pYeDP60u2. The USER compatible vector pYeDP60u2 was prepared as described in Hamann and Lindberg Møller (2007) by integration of complementary oligonucleotides (5'-phos-AATTTCGCTGAGTCTAATTAGGATCCTTAATTAACCTCAGCG-3' and 5'-phos-GATCCGCTGAGGCTTAATTAAGGATCCTTAATTAAGCCTCAGCG-3') into a BamHI/EcoRI digested pYeDP60. The CYP76B1::pYeDP60 construction is described in Didierjean et al. (2002). Plant expression vectors for CYP76C4 and CYP76B6 were produced by integration of the PCR fragments into pCAMBIA330035Su (kindly provided by B. A. Halkier, University of Copenhagen). Correct amplification of all constructs was confirmed by sequencing.
2.2 Heterologous expression in yeast and enzyme activity

The yeast expression constructs were transformed into the *Saccharomyces cerevisiae* strain WAT11 (Pompon et al., 1996) as described by Gietz and Schiestl (2007). One transformed yeast colony was grown and protein expression induced as described by Pompon et al. (1996). Microsomal membranes were prepared as described in Heitz et al. (2012) and their cytochrome P450 content was measured by differential spectrophotometry according to Omura and Sato (1964). Standard enzyme assays were carried out in 100 µL of 20 mM phosphate citrate buffer (pH 7.4) containing varying concentrations of substrate, 600 µM NADPH and adjusted amounts of P450 enzyme. After the addition of NADPH, samples were incubated at 27 °C for 6 to 20 min. The reaction was stopped by addition of 10 µL of 1 M HCl and 500 µL ethyl acetate. After centrifugation, the ethylacetate phase was transferred to a new vial and extraction was repeated with another 500 µl ethyl acetate. The combined organic phase was dried over anhydrous Na₂SO₄, concentrated under argon and analyzed by GC-FID and GC-MS. To generate amounts of products large enough for NMR analysis, the standard enzyme assay was scaled-up to a volume of 10 mL containing 400 µM substrate. After a first incubation for 15 min at 27 °C, a second aliquot of P450 enzyme was added and incubated for another 15 min. The reaction was stopped by adding 1 mL of 1 M HCl, vortexing and cooling on ice. Several scaled-up assays were pooled to ensure proper NMR detection of the products. For the extraction of the products, SPE columns (Oasis® HLB extraction cartridges, Waters) were equilibrated with CHCl₃, methanol and water prior to gradual extraction of up to 15 mL of combined samples. After drying, the columns were eluted with CDCl₃ and the combined organic phase was dried over anhydrous Na₂SO₄ and concentrated under argon before NMR analysis.

2.3 GC-FID and GC-MS analysis

Capillary GC was performed on a Varian 3900 gas chromatograph (Agilent technologies) equipped with a flame ionization detector and a DB-5 column (30 m, 0.25 mm, 0.25 µm; Agilent technologies) with splitless-injection, at 250°C injector temperature, and with a temperature program of 0.5 min at 50°C, 10°C/min to 320°C, and 5 min at 320°C. Terpenoids were identified on the basis of their retention time and using electron-impact MS spectra (70 eV, m/z 50-600) with a PerkinElmer Clarus 680 gas chromatograph coupled to a PerkinElmer Clarus 600T mass spectrometer (PerkinElmer). Capillary GC-MS was performed as described above, prior to a second GC-MS run optimized for better separation of the products and enhanced spectra quality with splitless-injection at 250°C injector temperature, and 0.5 min at 50°C, 10 °C/min to 100°C, 4°C/min to 160°C, 20°C/min to 320°C and 4 min at 320°C in the GC-oven. Reference standards of 8-hydroxygeraniol, 8-hydroxygeranial, 8-oxogeraniol and 8-oxogeranial
were synthesized by Chiralix B.V. (Nijmegen, The Netherlands), 8-carboxygeraniol and 8-carboxygeranic acid by Synthelor (Vandoeuvre-les-Nancy, France).

2.4 NMR analysis

NMR was conducted on a 500 MHz Bruker Avance spectrometer equipped with a 5 mm DCH dual cryoprobe or a 5 mm BBI probe with z-gradient operating at 500.13 MHz for $^1$H and 125.758 MHz for $^{13}$C. A number of different spectra including 1D $^1$H, $^1$H-$^1$H COSY, edited $^1$H-$^{13}$C HSQC and $^1$H-$^{13}$C HMBC were recorded for each sample, adding $^1$H-$^1$H NOESY and 1D $^{13}$C when required. Pulse sequences were taken from the Bruker library. All experiments were acquired at 300 K with a minimal relaxation delay of 2 s and a mixing time of 600 or 800 ms for NOESY experiments. Coupling constants were assumed to be around 145 Hz and 8 Hz for $^1J(^{13}$C-$^1$H) and $^nJ(^{13}$C-$^1$H) respectively. Acquisition parameters were adjusted when necessary but typically spectral windows were set to 7 KHz for $^1$H and 27 or 31 KHz for $^{13}$C. For 2D spectra, data size was at least 2048 points in the direct dimension and varied between 128 and 1024 points in the indirect dimension according to the required resolution.

2.5 Leaf disc assay

For transient expression of the genes of interest in *N. benthamiana* leaves, the plant expression constructs (pCAMBIA330035Su derived) were introduced into the *Agrobacterium tumefaciens* strain LBA4404 (vir GN54D) (van der Fits et al., 2000) by electroporation as described by Weigel and Glazebrook (2006). Transformed *A. tumefaciens* strains were grown in 5 mL Luria broth containing the appropriate antibiotics. Following overnight growth at 28 °C, bacteria were pelleted by centrifugation for 20 min at 4000 g, washed twice with water and resuspended in tap-water to OD$_{600}$ = 0.2. An *A. tumefaciens* strain carrying a binary vector for expression of the viral p19 silencing suppressor protein was prepared in the same way (Voinnet et al., 2003). The bacterial suspensions were mixed in a ratio 1/5 (v/v; p19/gene of interest). Young, fully expanded leaves of 6-week-old *N. benthamiana* plants, cultivated in a growth chamber at 20 °C with a photoperiod of 16/8 h light/dark cycle, were infiltrated with the bacterial mix on the abaxial side using a needleless syringe. After 5 days, 15 mm diameter discs were excised from the infiltrated leaves with a cork borer. The leaf discs (20 per sample in three replicates) were placed on 10 mL 20 mM phosphate citrate buffer pH 7.4 containing 400 µM geraniol in a Petri dish, exposed to vacuum twice, and placed in a growth chamber for two hours. After removal of the leaf discs, terpenes were extracted from the incubation medium with two times 5 mL ethyl acetate. The combined organic phase was dried over anhydrous Na$_2$SO$_4$, concentrated under argon and analyzed by GC-FID and GC-MS.
2.6 UPLC-MS analysis

The leaf discs were directly ground with mortar and pestle in 5 mL methanol and incubated at 20 °C for one hour. The extracts were centrifuged 1 min at 2000 rpm and the supernatant was transferred into a new vial. The cleared samples were concentrated under argon and analyzed by UPLC-MS. Analyses were performed using a Waters Quattro Premier XE (Waters) equipped with an electrospray ionization source (ESI) and coupled to an Acquity UPLC system (Waters). Chromatographic separation was achieved using an Acquity UPLC BEH C_{18} column (100 x 2.1 mm, 1.7 µm; Waters) and pre-column. The mobile phase consisted of (A) water and (B) methanol, both containing 0.1 % formic acid. The run started by 2 min of 95 % A. Then a linear gradient was applied to reach 100 % B at 12 min, followed by isocratic run using B during 2 min. Equilibration to initial conditions was achieved in 3 min, giving a total run time of 17 min. The column was operated at 35 °C with a flow-rate of 0.35 mL/min, injecting 3 µL samples. Nitrogen was used as the drying and nebulizing gas. The nebulizer gas flow was set to approximately 50 L/h, and the desolvation gas flow to 900 L/h. The interface temperature was set at 400 °C and the source temperature at 135 °C. The capillary voltage was set to 3.4 kV and the cone voltage to 25 V, the ionization was in positive or negative mode. Low mass and high mass resolution were 15 for both mass analyzers, ion energies 1 and 2 were 0.5 V, entrance and exit potential were 50 V and detector (multiplier) gain was 650 V. Data acquisition and analysis were performed with the MassLynx software (version 4.0). Full-scan and Selected Ion Recording (dwell time 0.1 s) modes were used for qualitative analyses.

2.7 Homology modeling of CYP76C4 and CYP76B6 enzymes and docking

3D models of CYP76B6 and CYP76C4 were obtained using Modeller 9v8. For structural template identification, PSI-BLAST of CYP76B6 and CYP76C4 was performed against the PDB database, and several multi-template rebuilt models were tested. Final models were based on an alignment of the crystal structures of CYP1B1 (αNF-bound, PDB code 3pm0), CYP2C9 (flurbiprofen-bound, PDB 1r9o), CYP2C8 (free, PDB 1pq2), CYP1A2 (αNF-bound, PDB 2hi4), and CYP2A13 (indole-bound, PDB 2p85) as templates for CYP76C4, and an alignment of the crystal structures of CYP1A2 (αNF bound, PDB 2hi4), CYP2C9 (warfarin-bound, PDB 1og2), CYP2C5 (free, PDB 1dt6), CYP2A6 (coumarin bound, PDB 1z10), and CYP2E1 (4-Methylpyrazole-bound, PDB 3e4e) as templates for CYP76B6. Structures with bound substrates in the active site were included in the template selection to allow for flexible active site topology. Multiple alignments were performed using MUSCLE (Multiple Sequence Comparison by Log-Expectation) within Jalview 2.7 or using MUSTANG (Multiple Structural Alignment Algorithm) for structure-based alignments. For each enzyme, 500 models were generated and evaluated by their
DOPE score (Discrete Optimized Protein Energy generated by Modeller). The best model according to the DOPE score was then submitted to optimization steps consisting in several cycles of minimization and molecular dynamic (MD) simulations (20 ns runs, truncated octahedron periodic boundary conditions, 17800 water molecules). Counter ions were added to the solvent bulk of the protein water complex in order to maintain neutrality of the complex. The AMBER db99 force field parameters were used for the amino acids and TIP3P model employed for explicit water molecules. The parameters applied for the heme were obtained from A. Oda (Oda et al., 2005). A time step of 2 fs was used in all MD runs, which were carried out with the AMBER 9.0 program suite. Prior to starting MD simulations, three consecutive minimizations of the entire ensembles were performed, setting the maximum number of cycles to 1000, the cut off to 8 Å, and switching from steepest descent to conjugate gradient method after 100 minimization cycles. After heating to 300K (50 ps), 50 ps equilibration of the water shell at 300K, 2 ns of equilibration of the entire ensemble at 300K, two steps of 10 ns of MD simulation were performed in the NTP ensemble for both models, in periodic boundary conditions.

Molecular docking experiments of geraniol in the active site were performed using AutoDock 4.2 in the semi-flexible mode. The docking box, in which grid maps were computed, included the active site with the protoporphyrin group on one edge, and part of the access channels. Ligand molecules were prepared under Maestro 9.2 (Schrödinger, LLC, New York) molecular modeling suite and saved under mol2 format as input files for Autodock. Partial charges were assigned using OPLS 2005 force field, for ligand and receptor structures. Ten different conformers of geraniol were generated using ConfGen module. The docking procedure showed that the ligand and some selected residues of the protein active site are flexible, whose torsion angles were identified to generate their PDBQT files using AutoDockTools (ADT). For each ligand conformer, 100 conformations were generated using the recent Lamarckian genetic algorithm, generating a total of 1000 conformations.

The poses were assigned a score calculated by Autodock that can be considered as an estimated free energy of ligand binding (indicative of binding affinity), then clustered as a function of the closeness of their positions and conformations (with RMSD set at 2.0 Å), and finally ranked by their binding score (for the best pose in the cluster). The grid built by AutoGrid included 126, 126, and 126 points in x, y, and z directions, with a grid spacing of 0.20 Å to allow a good compromise between resolution of the explored surface and the size of the binding area. The default settings were used for all other parameters. The best clusters, corresponding to the lowest (i.e. the most negative) energetic binding scores, were generally well-resolved in the histogram of scores.
2.8 Transient expression of multiple genes in *N. benthamiana*

Agro-infiltration of whole *N. benthamiana* leaves was performed as described in (van Herpen et al., 2010). A suspension of different combinations of *A. tumefaciens* cultures harboring expression constructs (35S:GES, 35S:CYP76C4, 35S:CYP76B6, or empty vector) were infiltrated into leaves of six-week-old *N. benthamiana* plants by pressing a 1 mL syringe without metal needle against the abaxial side of the leaf and slowly injecting the bacterium suspension into the leaf. After agro-infiltration the plants were grown under greenhouse conditions until further analyses (16/8h light/dark cycle, 28 °C day, 25 °C night). Three samples were harvested ten days post-agroinfiltration consisting each of several leaves.

2.9 Analysis by LC-PDA-QTOF-MS and LTQ-Orbitrap-MSn

Products in agro-infiltrated *N. benthamiana* leaves were analyzed by liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC-QTOF-MS) (De Vos et al., 2007). Frozen leaf samples were ground to powder in liquid nitrogen. An aliquots of 200 mg powdered material was extracted in a 1.5 ml Eppendorf vial with 0.6 ml 99.867% MeOH / 0.133% formic acid. After short vortexing and 15 min sonication, the extracts were centrifuged and filtered through 0.45 µm filters (SRP4, Sartorius) and 5 µl of the filtered extract was injected into the LC-QTOF-MS. Measurements were in negative ionization mode and leucine encephalin ([M – H]⁻ = 554.2620) was used as a lock mass for online mass calibration.

For further identification, selected compounds were targeted for fragmentation by an Accela HPLC tower connected to a LTQ Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific). The instrument settings of HPLC and LTQ Orbitrap were as described (van der Hooft et al. 2011). The LTQ was programmed to use a window of 10 D to isolate the mass of interest in MS1. The data-dependent fragmentation was set as follows: MS2 fragmentation of most intense ion in MS1; MS3 fragmentation of the 5 most intense fragment ions in MS2; MS4 fragmentation of the 5 most intense fragment ions in each MS3.

2.10 LC-QTOF-MS data processing and multivariate analysis

MassLynx 4.0 (Waters) was used for visualization and manual processing of LC-QTOF-MS data. For comparison between samples, mass data were automatically aligned using MetAlign version 1.0 (www.metalign.nl). Baseline and noise calculations were performed from scan number 70 to 2,480. The maximum amplitude was set to 25,000 and peaks lower than three times the local noise were discarded. Multiple mass signals derived from the same compound were grouped using MSClust (biotools.wurnet.nl)
with the Multivariate Mass Spectra Reconstruction (MMSR) model (Tikunov et al., 2005). The selected mass intensities were log2 transformed and normalized using range scaling, in which each value in a row is divided by the intensity range observed for this row throughout all samples analyzed. Each row was then mean centered. Finally, the normalized and log-transformed data matrix was used for Principle Component Analysis (PCA) implemented in GeneMath XT version 2.1 (Applied Mathematics).

### 2.11 Annotation of metabolites

Significant differences between transgenic and wild type plant tissues in intensity of each aligned mass signal were determined using the Student t-test (level of significance set at 0.05). In addition, the masses which showed significant difference were checked manually in MassLynx. Metabolites were identified by calculating the elemental composition, selected for the presence of C, H and O using the MassLynx software. A list of possible molecular formulas was obtained using a Mass tolerance of 5 ppm. Best matches were searched in the Dictionary of Natural Products and SciFinder databases for possible structures.

### 3. Results

#### 3.1 Candidate screening identifies a novel geraniol hydroxylase

As geraniol 8-hydroxylases reported so far belong to the CYP76 family of cytochrome P450 enzymes (Collu et al., 2001; Wang et al., 2010; Otah and Mizutani, 1998), we assumed that other and possibly more active geraniol 8-hydroxylases could be found among related proteins. Based on available data, we selected candidates from the CYP76B and CYP76C subfamilies. Besides the reference enzyme CYP76B6 from *C. roseus*, we included the closely related CYP76B1 from *Helianthus tuberosus* - so far described as a herbicide detoxifying enzyme (Robineau et al., 1998; Didierjean et al., 2002) - as well as the set of seven members of the CYP76C subfamily in *A. thaliana* in our study. The ten genes were expressed in yeast under the control of a galactose-inducible promoter in the strain WAT11 also expressing the P450-reductase ATR1 from *A. thaliana*. Enzyme expression and activity were tested with the microsomal fractions. Evaluation of protein expression by differential spectrophotometry (Omura and Sato, 1964), confirmed that CYP76B1, CYP76B6, CYP76C1, CYP76C2, and CYP76C4 were expressed at significant levels in yeast (> 400 nmol/L). Geraniol hydroxylase activity in the presence of NADPH was detected as expected for CYP76B6, but not for CYP76C1 (Figure 1), despite repeated attempts with different microsomal preparations and despite the fact that enzyme activity was detected with other monoterpenols. No activity was detected for the other
candidates, except for CYP76C4. GC-FID and GC-MS analysis of the reaction products showed that CYP76B6 catalysed the formation of one major compound from geraniol with a retention time and fragmentation pattern identical to the reference standard 8-hydroxygeraniol and minor product (x), while CYP76C4 generated two products in a ratio 1.3/1, the minor likely 8-hydroxygeraniol, the major with a different retention time but identical mass spectrum (Figures 1, S1). For unambiguous identification, we scaled-up the microsome assay to allow NMR analysis of the products. Their structures were determined as 9- (major) and 8- (minor) hydroxygeraniol (Figures S3, S4, S6-S9).

**Figure 1:** Geraniol metabolism by CYP76C1, CYP76C4 and CYP76B6

GC-FID chromatograms of the reaction mixtures of geraniol incubated with yeast-expressed CYP76C1, CYP76C4 or CYP76B6. Microsomal preparations from recombinant yeast transformed with the CYP76C1, CYP76C4, CYP76B6 expression vector or with the empty vector (empty-control) were incubated with 200 µM geraniol for 20 minutes in the presence of NADPH. No NADPH was added to the negative control (neg-control). Compounds were identified based on their retention time and GC-MS spectra. With g = geraniol; a = 8-hydroxygeraniol; b = 9-hydroxygeraniol.

### 3.2 CYP76B6 is a bifunctional enzyme that catalyzes two oxidation steps in the iridoid pathway

To compare the relative efficiencies of CYP76B6 and CYP76C4 for geraniol hydroxylation, their catalytic parameters were compared. Typical Michaelis-Menten kinetics were observed for CYP76C4 (Figure 2) allowing calculation of a $K_{m,app}$ of $21 \pm 6$ µM and a $k_{cat}$ of $20 \pm 1$ min$^{-1}$. CYP76B6 however provided aberrant kinetics whatever the concentration and incubation time range used in the experiments (Figure 2). This suggested that 8-hydroxygeraniol was further converted and therefore competed with geraniol hydroxylation. To test this hypothesis, we first reinvestigated the minor products obtained after longer incubations and higher substrate concentrations. With a refined analytical method, we detected two additional products with CYP76B6 (Figure
3A), while no additional product was detected with CYP76C4. To test if the additional products resulted from further oxidation of 8-hydroxygeraniol, the latter was assayed as a substrate of yeast-expressed CYP76B6. We observed an efficient conversion of 8-hydroxygeraniol into a compound with a GC retention time and fragmentation pattern similar to reference 8-oxogeraniol (Figure 3B, S1) and small amounts of the unknown product x that was also detected upon incubation of the enzyme in the presence of geraniol. Scaling-up the enzyme assay allowed NMR analysis of the products and confirmed 8-oxogeraniol as the main product (Figures S3, S5-S9), but the amounts of x were not sufficient for structure determination. CYP76B6 did not catalyze any further conversion when 8-oxogeraniol was used as substrate. Hence, x most likely derives from 8-hydroxygeraniol. In the presence of 8-hydroxygeraniol, the CYP76B6 reaction showed biphasic kinetics (Figure S2) with estimated $K_{\text{m,app}}$ ranging between 3.5 and 151 µM and $k_{\text{cat}}$ between 15 and 37 min$^{-1}$. CYP76B6 did not metabolize other potential downstream products such as 8-hydroxygeranial and 8-oxogeranial or geranic acid. No significant inhibition of geraniol hydroxylation was observed in the presence of 8-oxogeraniol (data not shown).

To enable a rough comparisons of CYP76B6 and CYP76C4 turnovers and of the conversion rates of geraniol and 8-hydroxygeraniol by CYP76B6, we performed short enzyme assays at saturating substrate concentration (200 µM), assuming linear rates for the reactions. In these assays, estimated CYP76C4 turnover rate was 16 ± 1 min$^{-1}$ for production of 8- and 9-hydroxygeraniol, and CYP76B6 catalyzed the formation of 8-hydroxygeraniol with a turnover rate of 170 ± 6 min$^{-1}$ and the formation of
8-oxogeraniol from 8-hydroxygeraniol with a turnover of $64 \pm 1 \text{ min}^{-1}$.

**Figure 3:** Metabolism of geraniol and 8-hydroxygeraniol by CYP76B6 and CYP76C4. GC-FID chromatograms of A) geraniol and B) 8-hydroxygeraniol conversion by microsomal preparations from recombinant yeast expressing CYP76C4, CYP76B6 or transformed with the empty expression vector (empty-control). Incubations contained 200 µM geraniol or 8-hydroxygeraniol and NADPH. No NADPH was added to the negative control (neg-control). Based on retention time and fragmentation pattern, compound (c) and (d) was identified as 8-oxogeraniol, compound (x) could not be identified. With g = geraniol; a = 8-hydroxygeraniol; b = 9-hydroxygeraniol. c = 8-oxogeraniol; x = unknown compound.

### 3.3 Binding of geraniol in CYP76C4 and CYP76B6 active sites

While CYP76C4 and CYP76B6 are functionally related P450 isoforms, they share a rather moderate similarity in protein sequence (46%, alignment shown in Figure S10). It was therefore interesting to locate the structural differences in their respective active sites using homology models of the two enzymes. Homology modeling was based on the crystal structures of the five most similar P450s available in PDB for each of them. Their divergence in sequence led to the selection by PDB-BLAST of slightly different sets of structural templates (see alignments if Figures S11 and S12). The resulting structures, after refinement by MD simulations, revealed a strikingly conserved active site environment consisting mainly of hydrophobic residues (white-colored in Figure S13 B, C, E, F), including a conserved tryptophan residue in close proximity of the heme (TRP115 in CYP76B6 and TRP123 in CYP76C4). Two aspartate residues (ASP293/300 in CYP76B6 and 307/314 in CYP76C4) are also conserved and positioned similarly above the heme plane.

Despite this similarity of the active site environment, the shape and size of the two active sites, as characterized by VOIDOO, were different with CYP76C4 exhibiting a more
narrow active site than CYP76B6 (Figure S13 A, D). As a result, the docking experiments of geraniol into the active sites by Autodock revealed different positioning in the vicinity of heme (Figure 4). In CYP76C4, geraniol is positioned above the heme plane, almost perpendicular, presenting either the C8 or the C9 carbon to the heme iron, while in CYP76B6, the geraniol molecule is positioned closer to protoporphyrin with the double bond C6-C7 parallel to the heme plane.

![Figure 4: Docking of geraniol in the active site of CYP76C4 and CYP76B6. The heme catalytic center is represented with iron in orange. The orientation of geraniol relative to the heme of CYP76C4 and CYP76B6 is shown. A and B show the two possible positions of geraniol in the active site of CYP76C4. C illustrates the positioning of geraniol in the active site of CYP76B6. Distances of the substrate C6 and C8 carbon atoms to the heme iron and of the C1-OH to the side chain of Asp300 are indicated in Angströms.](image)

In CYP76C4, as shown in Figure 4A, the most favored cluster (clusters were ranked as a function of their energy) of poses of geraniol in the active site (two poses were considered the same when the calculated RMSD was less than 1.5 Å) contained positionings with equivalent distances from C8 and C9 to the heme iron, alternating with distances C(8/9)-Fe typically found at around 3.5 Å. The hydroxyl terminal group of geraniol was not found stabilized by H-bonding, although Asp307 could act as a potential H-bonding partner. Trp123 lines the cavity, forming a hydrophobic contact along the rest of the geraniol molecule.

Conversely, in the best ranked cluster of poses in CYP76B6, geraniol preferentially exposed the double bond C6-C7 to the heme iron at an average distance of 4.5 Å, which suggests a favored attack on C6, while C8 and C9 were at about 5.2 Å (Figure 4B). As in CYP76C4, Trp115 is in hydrophobic contact with geraniol, but, in CYP76B6, geraniol positioning is stabilized by H-bonding with Asp300 (the equivalent of Asp314 in CYP76C4). The regiospecific 8-hydroxylation observed with CYP76B6 thus suggests that either a more appropriate conformation is achieved upon oxygen binding, or that
regioselectivity results from easier oxidation of the allylic C8. The docking with the C8 position not directly pointing to the heme allows further oxidation of 8-hydroxygeraniol by CYP76B6, which is not observed in the case of CYP76C4.

3.4 Characterization of products formed in planta

3.4.1 Leaf disc substrate feeding assay

As our ultimate goal is reconstruction of the TIA pathway in a plant for production of anticancer drugs, the activity of CYP76B6 and CYP76C4 was also tested in planta. To this end, we developed a fast leaf-disc assay for evaluation of individual substrate conversion and fate in plant tissues. Each P450 was transiently expressed by agroinfiltration in leaves of *N. benthamiana*. Leaf discs were excised 5 days post infiltration, vacuum-infiltrated with and incubated in buffer containing substrate. Products were analyzed in the ethyl acetate extract of the incubation medium, and in the leaf-disc methanol extract.

GC analysis of the incubation buffer extract confirmed the formation of 9- and 8-hydroxygeraniol by CYP76C4 in the plant in a 2/1 ratio (Figures 5, S14). In the case of CYP76B6-transformed leaves, 8-hydroxygeraniol and 8-oxogeraniol were detected in the incubation medium together with the unknown compound x (Figures 5, S14). The 9/8-hydroxygeraniol and the 8-oxogeraniol/x ratios observed in vivo were different from those recorded in yeast microsome incubations. This can be the result of a differential membrane permeability or transport of the products, or might reflect differential conversion/conjugation and storage in the plant tissues.

To clarify this point, leaf disc methanol extracts were analyzed by UPLC-MS/MS, providing a more complex picture. In a first approach, targeted analysis by single-ion recording (SIR) was performed. Target ions were selected after analysis of commercial or synthetic standards of geraniol, 8-hydroxygeraniol and 8-oxogeraniol. The most prominent peak in all samples was geraniol (main fragment with *m/z* 137), with the highest abundance in leaf discs infiltrated with the empty vector, and lower levels in CYP76C4- and CYP76B6-infiltrated discs (Figure 6A). Also in the control, infiltrated with empty vector, geraniol was metabolized to some extent, likely by *N. benthamiana* enzymes, to form several compounds with a fragment of *m/z* 137 and RT between 9.5 and 10.3 min (y in Figure 6A). In the CYP76C4-transformed discs, free 9- and 8-hydroxygeraniol (main fragment of *m/z* 135) could be clearly identified, while in the CYP76B6 expressing leaf discs only 8-hydroxygeraniol was detected (a and b in Figure 6A). In addition, we detected 8-hydroxygeranyl hexose in similar amounts in CYP76C4 and CYP76B6. This may indicate that the formation of this compound was limited by the activity of the glycosyl transferase catalyzing this conversion and not by the amount hydroxygeraniol produced. In the CYP76B6-expressing discs, only small amounts of
8-oxogeraniol were detected (c in Figure 6A), and two peaks d and e not present in the control (Figure 6A).

**Figure 5:** Products of geraniol conversion in extracts of agro-infiltrated *N. benthamiana* leaf-disc incubation medium.

*N. benthamiana* leaf discs transiently expressing CYP76C4, CYP76B6 or the empty pCAMBIA330035Su plant expression vector (neg-control) were incubated in phosphate buffer containing 400 µM geraniol. The ethyl acetate extract of the incubation medium was analyzed by GC-FID and the products were identified by GC-MS: a = 8-hydroxygeraniol; b = 9-hydroxygeraniol; c = 8-oxogeraniol; x = unknown product.

To identify these peaks and other related products, we performed untargeted UPLC-MS analysis (Figure S15). Based on available standards and the untargeted UPLC-MS analysis, we were able to show the presence of free foliamenthoic acid (or 8-carboxygeraniol) (h in Figure 6B) and 8-carboxygeranic acid (j in Figure 6B) in the extract of the CYP76B6-transformed leaves. Taking these compounds as starting points to look for related products, we were able to propose the tentative structures for most of the major compounds detected in disc extracts (Fig S16). They include the reduced forms 8-hydroxy-2,6-dimethyloct-2-enoic acid (d in Figure 6B), 8-hydroxy-2,6-dimethyloctanoic acid (f in Figure 6B) and their glycosylated derivatives. The 3,7-dimethyloct-2-ene-1,8-diol (e in Figure 6) and 8-hydroxy-2,6-dimethyloct-2-enoic acid (d in Figure 6) correspond to the peaks visible in the SIR chromatogram shown in Figure 6A. Complex metabolism of 8-oxogeraniol including oxidation of the alcohol and aldehyde at the 1 and/or 8 positions, double bond reduction and glycosylation thus occurs in *N. benthamiana* leaves (Figure S18). Taken together, these results confirm the *in planta* geraniol 8-oxidase activities of CYP76B6 and geraniol 9- and 8-hydroxylase
activity of CYP76C4. The enzyme products do not accumulate in the plant but are further converted by endogenous enzymes of *N. benthamiana*.

**3.4.2 Pathway reconstitution in *N. benthamiana***

In addition to leaf disc assays with substrate feeding, the first two steps of the iridoid pathway were also reconstituted *in planta* for determination of product formation during prolonged *in planta* gene expression. CYP76C4 and CYP76B6 were transiently co-expressed with a geraniol synthase (*VoGES*) from *Valeriana officinalis* (Dong et al., submitted) via agro-infiltration in 6-week-old *N. benthamiana* leaves. To be able to discriminate products generated from endogenous substrates, infiltrated plants with single genes (*VoGES, CYP76C4, CYP76B6* and empty vector (*EV*)) and non-infiltrated plants (wild type) were used as control. Previous experiments had shown that 10 days after agro-infiltration of *N. benthamiana* leaves with *VoGES*, the levels of free geraniol and geraniol derivatives were below the level of detection by GC-MS. Therefore, the semi-polar, non-volatile metabolites were extracted with methanol and analyzed by LC-QTOF-MS. Figure S17A shows a representative example of the chromatograms obtained for leaves infiltrated with *VoGES+CYP76B6, VoGES+CYP76C4, VoGES* and the *EV* control. For comparison of the different samples, the LC-MS chromatograms were processed and the resulting mass lists used for PCA. This shows clustering of the replicate groups and separation between the extracts from *GES, VoGES+CYP76C4* and *VoGES+CYP76B6* agro-infiltrated plants in the first component (59.3% of the variation; Figure S17B). The samples of *N. benthamiana* leaves agro-infiltrated with only *CYP76C4, CYP76B6* or *EV* also showed separation in the PCA plot, indicating that *CYP76C4* and *CYP76B6* catalyzed the oxidation of endogenous *N. benthamiana* compounds (Figure S17B). Comparison of the chromatograms identified in total 19 geraniol related masses (compounds) with significant differential peak intensities in the extracts of *VoGES* infiltrated leaves when compared with the controls of *CYP76C4, CYP76B6* or *EV* infiltrated leaves (Table S2). For identification, these masses were targeted for fragmentation by LTQ Orbitrap MSn. The resulting MSn spectra show that these compounds most likely are conjugates of geraniol, hydroxygeraniol, foliamenthoic acid or carboxygeranic acid with hexose, pentose, malonyl or acetyl groups (Table S2).

In *N. benthamiana* agro-infiltration with *VoGES*, eight geraniol glycosides and three hydroxygeraniol glycosides were found. When *CYP76C4* was co-infiltrated with *GES*, the level of geraniol glycosides decreased by half compared with those in leaves only infiltrated with *VoGES* (Figure 7). In addition, the levels of the three hydroxygeraniol glycosides were increased, while also new conjugates of hydroxygeraniol glycosides with an acetyl group were formed. The geraniol glycosides had completely disappeared in leaves infiltrated with *VoGES+CYP76B6*, suggesting that the geraniol is converted more efficiently by *CYP76B6* than by *CYP76C4*.
Figure 6: Products of geraniol conversion detected in the methanol extract of agro-infiltrated *N. benthamina* leaf-discs.

*N. benthamiana* leaf discs transiently expressing CYP76C4, CYP76B6 or the empty pCAMBIA330035Su plant expression vector (neg-control) incubated with geraniol, were extracted with methanol and the extracts were analyzed by UPLC-MS. (A) SIR chromatograms showing variation in geraniol (g), geraniol glycosides (y), 8-hydroxygeraniol (a) and 9-hydroxygeraniol (b) contents in the leaf tissues. (B) Representative MS fragmentation spectra of the differentially accumulated compounds identified. The spectra of g = geraniol, a = 8-hydroxygeraniol, b = 9-hydroxygeraniol, c = 8-oxogeraniol, h = foliamenthoic acid and j = 8-carboxygeranic acid are similar to those of authentic standards. The molecular formulas determined for d = C_{10}H_{18}O_{3}, e = C_{10}H_{20}O_{2}, f = C_{10}H_{20}O_{3} and i = C_{10}H_{16}O_{4} were tentatively allocated the following structures according to their UPLC-MS fragmentation patterns: d = 8-hydroxy-2,6-dimethyloct-2-enoic acid, e = 3,7-dimethyloct-2-ene-1,8-diol, f = 8-hydroxy-2,6-dimethyloctanoic acid and i = 2,6-dimethyl-3-oxooct-6-enoic acid. The main fragments for each molecule are detailed in Table S1.
For both VoGES+CYP76C4 and VoGES+CYP76B6 transformed tissues, the glycosides of further oxidized products foliamenthoic acid and carboxygeranic acid were formed, while these were not present in leaves infiltrated with VoGES alone. Again, the levels of these products were higher in the plants co-infiltrated with VoGES+CYP76B6 than in the plants expressing VoGES+CYP76C4 (Figure 7), suggesting that geraniol is oxidized more efficiently by CYP76B6 is more efficiently than by CYP76C4.

Figure 7: LC-QTOF-MS analysis of oxidized derivatives of geraniol and their conjugates in GES+CYP76B6, GES+CYP76C4 and GES agro-infiltrated N. benthamiana leaves.
4. Discussion

4.1 CYP76B6 is a regiospecific geraniol oxidase

Previous reports of geraniol hydroxylase activity of cytochrome P450 enzymes belonging to different CYP76 subfamilies (CYP76B6/B10 and CYP76C1) suggested that this function might be shared by a fairly large set of CYP76 enzymes. Our data show that this is not the case. We detected geraniol hydroxylase activity for the CYP76B6 and CYP76C4 enzymes, but not for the other members of the CYP76B and CYP76C subfamilies that we investigated. This rather suggests that the geraniol hydroxylase function evolved in different phyla as the result of a convergent evolution, favored by a common functional background inherited from a common ancestor. This is supported by the different regiospecificities and catalytic properties of the two enzymes (Figure 8), and by the different volumes and properties of their active sites (Figure S13).

CYP76C4 from *A. thaliana* is a geraniol hydroxylase oxidizing the 9 as well as the 8 position. Such relaxed regiospecificity, together with moderate enzyme turnover rates, suggest that geraniol is likely not the physiological substrate of CYP76C4 in arbadopsis. So far, no geraniol synthase has been reported in arbadopsis. In addition, geraniol was not detected in the head-space collected from arbadopsis plants and isolated organs (Rohloff and Bones, 2005), so it is possible that geraniol does not occur in this plant species. Hydroxygeraniol is not further converted by CYP76C4. Homology modeling, based on the resolved structures of mammalian P450 enzymes, indicates that geraniol can bind the CYP76C4 active site in varios similar positions that are found in the same dominant cluster of poses, placing either C8 or C9 at distances suitable for oxidative attack, but with no obvious anchoring of the alcohol function (Figure 4A). In all cases, the C8 or C9 of geraniol points straight to the heme. In this orientation, likely due to steric hindrance constraints, hydroxygeraniol binding for further oxidation would not be favored. Alternatively, coordination of the alcohol oxygen with the heme iron would be expected, resulting in enzyme activity inhibition (Ortiz de Montellano and Correia, 1995). Modeling thus provides a reasonable explanation for the observed CYP76C4 activity and the duality of its regiospecificity. Conversely, CYP76B6 is a highly regiospecific and effective geraniol 8-hydroxylase with a catalytic turnover rate of 150 min⁻¹ in the range of CYPs with selection-optimized physiological functions. In addition to this high regioselective geraniol 8-hydroxylase activity, CYP76B6 efficiently converts 8-hydroxygeraniol into 8-oxogeraniol. Homology modeling of CYP76B6 and docking of geraniol in its active site suggests that, in this active site, the substrate is stabilized in the active site by the alcohol H-bonding with Asp300 (Figure 4B). The oxidation at C8 appears favored by both the allylic position and its proximity to the heme. As geraniol lies almost parallel to the heme, no steric hindrance is expected for 8-hydroxygeraniol binding and further oxidation. The 8-hydroxygeraniol oxidase activity, which was not
previously detected, corresponds to the next step in the proposed pathway to secologanin. This activity was so far attributed to a soluble NADPH-dependent oxidoreductase, unspecifically oxidizing allylic primary alcohols at the 1 and 8 positions (Ikeda et al., 1991; Hallahan et al., 1995). This would suggest that either the two enzymes cooperate to the formation of iridodial or that the described oxidoreductase, which so far has not been characterized at the molecular level, has another function in the plant.

![Diagram](image)

**Figure 8:** Summary of reactions catalyzed by CYP76B6 and CYP76C4. In a first step CYP76B6 acts as G8H to convert geraniol into 8-hydroxygeraniol, in a second step CYP76B6 is a 8HGO to produce 8-oxogeraniol from 8-hydroxygeraniol. The G8H and G9H activities of CYP76C4 result in the production of 8- and 9-hydroxygeraniol from geraniol.

We were not able to accurately determine the catalytic parameters of the geraniol hydroxylation step by CYP76B6 due to the further conversion of the product and resulting competition. It was not possible either to reliably evaluate the affinity of the CYP76B6 enzyme for geraniol through spectral changes induced by substrate binding, due to the rather low expression of CYP76B6 in yeast. A $K_m$ of 16 µM and a $k_{cat}$ of 8 min$^{-1}$ were previously reported for the enzyme expressed in insect cells also expressing the *A. thaliana* P450 reductase ATR1 (Sung et al., 2011). The $k_{cat}$ measured with insect cell membranes is about 20 times lower than the approximate turnover rate that we measured when the enzyme was expressed in yeast, which points to an importance of the membrane context (yeast or insect) for achieving optimal activity. Taking as a reference the $K_m$ for geraniol measured in insect cells, a strong competition between geraniol and 8-hydroxygeraniol would be expected. The very low geraniol hydroxylation obtained with insect cell membranes, most likely explains why no further conversion of the product was observed. The high turnover rate measured for geraniol hydroxylation with yeast microsomes would however favor the hydroxylation reaction over further oxidation of 8-hydroxygeraniol, explaining the accumulation of 8-hydroxygeraniol observed in the *in vitro* enzyme assays.
4.2 CYP76B6 metabolites are further processed in N. benthamiana

Analysis of the free forms of geraniol and oxidized geraniol products formed in the CYP76B6-transformed leaf-disc feeding assay and the transient expression of VoGES with CYP76B6 in intact leaves of N. benthamiana suggest a preferential formation of products as shown in Figure S18. A number of different products were detected in the leaf-disc assay (2 hours geraniol conversion) and the intact-leaf infiltration (ten days expression of VoGES and CYP76B6). The leaf-disc results show that early 8-oxogeraniol conversion mainly consists of double bond reduction (formation of 8-hydroxy-2,6-dimethyloct-2-enoic acid, 8-hydroxy-2,6-dimethyloctanoic acid and 2,6-dimethyl-3-oxooct-6-enoic acid and the corresponding glycosylated forms; Figure S18). These compounds are not detected after ten days of VoGES and CYP76B6 expression in planta (Figure 7). Instead, the main long-term accumulation products are 8-carboxygeranic acid and several different conjugates. It is thus likely that the 8-hydroxy-2,6-dimethyloct-2-enoic acid, 8-hydroxy-2,6-dimethyloctanoic acid and 2,6-dimethyl-3-oxooct-6-enoic acid derivatives are further processed in the plant. The leaf disc assay thus provides a picture of the early modifications by endogenous enzymes of geraniol, whereas pathway reconstitution in the plant rather identifies final steady-state products.

4.3 Towards TIA production in heterologous hosts

In a first step of the TIA pathway reconstruction, the N. benthamiana leaves were agro-infiltrated with VoGES alone, resulting in production of geraniol, which was mostly accumulated as geraniol glycosides after conversion by endogenous glycosyltransferases (Dong et al., submitted). The second step of the TIA pathway is the conversion of geraniol to 8-hydroxygeraniol by a geraniol 8-hydroxylase (G8H). CYP76B6 is here confirmed to have a high G8H activity and the different experiments in planta show that it can effectively compete with the glycosylation - by endogenous N. benthamiana glycosyl transferases - of the first product of the pathway, geraniol. Even when the geraniol concentration was very high, e.g. upon feeding of geraniol to leaf discs expressing CYP76B6, only hydroxygeraniol glycosides and no geraniol glycosides were found (Figure S16). In contrast, geraniol glycosides were still present upon expression of CYP76C4, which indicates, in good agreement with the in vitro data, that the enzyme turnover is not high enough to outcompete the N. benthamiana glycosyl transferases that use geraniol as substrate.

From the in planta expression results we can in principle not conclude whether CYP76B6 acts only on free geraniol or also on the geraniol glycosides. However, it's very high turnover rates measured with free geraniol in vitro strongly supports the former. In addition, the difference in glycosylation profile of geraniol glycosides and oxidated geraniol glycosides seems to rule out a sequential product relationship, but
rather suggests independent pathways (Figure 7). Moreover, in the short term leaf-disc assay, oxidized geraniol is found with only one hexose (Figures S15, S16), suggesting that CYP76B6 acts only on free geraniol. While this is consistent with the evolution of CYP76B6 of *C. roseus* as a very efficient geraniol hydroxylase, the subcellular localization and proximity to GES in the plant cell may also contribute to preferential geraniol oxidation over conjugation. For instance, VoGES was shown to be located on plastidial stromules (Dong et al., submitted), and stromules have been suggested to interact with the ER (Schattat et al., 2011).

The next step towards TIA production in *N. benthamiana* would be the conversion of 8-oxogeraniol to 8-oxogeranial. In *N. benthamiana* this enzymatic step needs to compete with step 3 and step 6 in the pathway (Figure S18). At present we cannot exclude that some 8-oxogeraniol is already converted to 8-oxogeranial in our *in vivo* experiments as this double aldehyde may be very reactive and therefore not detectable as intermediate. This compound (lacking free OH groups) can also not be sequestered and stabilized as a glycoside. The intermediate 8-carboxygeranial may be formed via foliamenthoic acid or via 8-oxogeranial, but was not detected, presumably because of rapid conversion to 8-carboxygeranic acid. In the same way, 8-oxogeraniol may be readily converted into foliamenthoic acid in the plant in the absence of a competing enzyme. Further introduction of the iridoid synthase which converts 8-oxogeranial to iridodial (Geu-Flores et al., 2012) will be the next step and will tell if an intermediate enzyme is still required.
Acknowledgments

The research described here was funded by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. J-FG PhD was supported by a BDI from CNRS and the Région Alsace. CG PhD was supported by CIFRE a grant from the Agence Nationale de la Recherche Technique and MANE fils. We are grateful to L. Miesch and M. Miesch from the Laboratoire de Chimie Organique Synthétique, UMR 7177, CNR-University of Strasbourgh for very helpful discussion.

References


Supplemental data

Figure S1: GC-MS mass spectra of the metabolic products from geraniol and 8-hydroxygeraniol formed by CYP76C4 and CYP76B6.

A) Mass spectra of an authentic 8-hydroxygeraniol standard, and of the products formed by (a) CYP76C4 (Figure 3A, a) and (b) CYP76B6 (Figure 3A, a). The second product of (c) CYP76C4 (Figure 3A, b) was identified as 9-hydroxygeraniol. B) Mass spectra of an authentic 8-oxogeraniol standard, and the products formed by (d, e) CYP76B6 (Figure 3A, c; Figure 3B, c). The second product of (f) CYP76B6 (Figure 3B, x) was not identified.
Figure S2: Enzyme kinetics of CYP76B6 catalyzed conversion of 8-hydroxygeraniol into 8-oxogeraniol. Eadie-Hofstee diagram obtained with microsomal preparations from recombinant yeast expressing CYP76B6 incubated with different concentrations of 8-hydroxygeraniol in the presence of NADPH. The amount of products was determined by GC-FID. Incubations were performed in triplicate, error bars represent the standard deviation.

Figure S3: Overview of the molecules identified by NMR analysis. The molecule structure including the carbon numbering is shown for A) geraniol, B) 8-hydroxygeraniol, C) 9-hydroxygeraniol and D) 8-oxogeraniol. The table shows the $^1$H- and $^{13}$C-NMR assignments of each molecule analyzed.
Figure S4: $^1$H NMR spectra in CDCl$_3$ (500 Mhz) from CYP76B6 and CYP76C4 incubations with geraniol. Some characteristic peaks of the identified molecules are shown, with 1 = geraniol; 2 = 8-hydroxygeraniol; 3 = 9-hydroxygeraniol; 4 = 8-oxogeraniol.

Figure S5: $^1$H NMR spectra in CDCl$_3$ (500 Mhz) from CYP76B6 incubation with 8-hydroxygeraniol. Some characteristic peaks of the identified molecules are shown, with 1 = 8-hydroxygeraniol; 2 = oxogeraniol.
Geraniol hydroxylase activities of the CYP76 family of cytochrome P450 enzymes

Chapter 3

Figure S7: $^1$H-13C HMBC spectra in CDCl$_3$ (500 MHz) for product identification. The spectra from CYP76B6 and CYP76C4 incubations with geraniol (A) and 8-hydroxygeraniol (B) are shown.

Figure S6: $^1$H COSY spectra in CDCl$_3$ (500 MHz) for product identification. The spectra from CYP76B6 and CYP76C4 incubations with geraniol (A) and 8-hydroxygeraniol (B) are shown.

Figure S8: $^1$H-13C HSQC spectra in CDCl$_3$ (500 MHz) for product identification. The spectra from CYP76B6 and CYP76C4 incubations with geraniol (A) and 8-hydroxygeraniol (B) are shown.

Figure S9: $^1$H NOESY spectra in CDCl$_3$ (500 MHz) for product identification. The spectra from CYP76B6 and CYP76C4 incubations with geraniol (A) and 8-hydroxygeraniol (B) are shown.
Figure S10: Alignment of the protein sequences of CYP76C4 and CYP76B6 showing the residues located in the active site. Black stars: residues located close to the heme plane (first crown). Cyan star: residues located above the heme plane (second crown).
Geraniol hydroxylase activities of the CYP76 family of cytochrome P450 enzymes

Chapter 3

Figure S11: Templates and alignment used for building the model of CYP76C4.

Templates (PDB codes): CYP1B1 (αNF bound, 3pm0), CYP2C9 (flurbi bound, 1r9o), CYP2C8 (free, 1pq2), CYP1A2 (αNF bound, 2hi4), CYP2A13 (indole bound, 2p85)
Figure S12: Templates and alignment used for building the model of CYP76B6.

Templates (PDB codes): CYP1A2 (αNF bound, 2hi4), CYP2C9 (warfarin bound, 1og2), CYP2C5 (free, 1dt6), CYP2A6 (coumarin bound, 1z10), CYP2E1 (4-Mepy bound, 3e4e)
Geraniol hydroxylase activities of the CYP76 family of cytochrome P450 enzymes

Chapter 3

Figure S13: Models of the active sites of CYP76C4 and CYP76B6. (A, B) First crown of residues above the heme plane, (C, D) first and second crown of residues in the active site of CYP76C4 and CYP76B6, respectively. With A) residues 122 (B'C loop), 375-376-379-380 (K-β1 loop), 307-311-315 (I helix); B) Ile492, Thr493 (β-hairpin of β4 sheet), Asp314 on I helix, Trp123 on B'C loop; C) 115 (B'C loop), 361-362-365-366 (K-β1 loop), 293-297-301(I helix); D) Lys475, Phe476 (β-hairpin of β4 sheet), Asp300 (I helix), Pro363 (K-β1 loop). White : hydrophobic, green : polar; red: positively charged, blue: negatively charged. (E, F) Volume of the active site cavities calculated by VOIDOO and visualized with PyMOL.

Figure S14: Products of geraniol conversion in the extract of Agro-infiltrated N. benthamiana leaf-disc incubation medium: same data as in Figure 5 but at the same scale. N. benthamiana leaf discs transiently expressing CYP76C4, CYP76B6 or the empty pCAMBIA330035Su plant expression vector (neg-control) were incubated in phosphate buffer containing 400 µM geraniol. The ethyl acetate extract of the incubation medium of was analyzed by GC-FID and the products were identified by GC-MS: a = 8-hydroxygeraniol; b = 9-hydroxygeraniol; c = 8-oxogeraniol; x = unknown product.
Figure S15: Products of geraniol conversion detected in the methanol extract of Agro-infiltrated *N. benthamiana* leaf-disc tissues.

Shown is the full scan UPLC-MS chromatogram with ES+ ionization of *N. benthamiana* leaf discs transiently expressing CYP76C4, CYP76B6 or the empty pCAMBIA330035Su plant expression vector (neg-control) incubated with geraniol and extracted with methanol. With g = geraniol, a = 8-hydroxygeraniol, b = 9-hydroxygeraniol, c = 8-oxogeraniol, d = 8-hydroxy-2,6-dimethyloct-2-enoic acid, e = 3,7-dimethyloct-2-ene-1,8-diol, f = 8-hydroxy-2,6-dimethyloctanoic acid, h = 8-foliamenthoic acid, i = 2,6-dimethyl-3-oxooct-6-enoic acid, j = 8-carboxygeranic acid, k = 8-carboxygeranic acid conjugated to hexose, l = 2,6-dimethyl-3-oxooct-6-enoic acid conjugated to hexose, m = 8-hydroxy-2,6-dimethyloct-2-enoic acid conjugated to hexose and n = 8-hydroxygeraniol conjugated to hexose.
Geraniol hydroxylase activities of the CYP76 family of cytochrome P450 enzymes

<table>
<thead>
<tr>
<th>RT (min)</th>
<th>Compound</th>
</tr>
</thead>
<tbody>
<tr>
<td>g 10.78</td>
<td>geraniol</td>
</tr>
<tr>
<td>a 8.33</td>
<td>8-hydroxygeraniol</td>
</tr>
<tr>
<td>b 8.44</td>
<td>9-hydroxygeraniol</td>
</tr>
<tr>
<td>c 8.18</td>
<td>8-oxogeraniol</td>
</tr>
<tr>
<td>d 8.61</td>
<td>8-hydroxy-2,6-dimethyloctanoic acid</td>
</tr>
<tr>
<td>e 8.81</td>
<td>3,7-dimethyloct-2-ene-1,8-diol</td>
</tr>
<tr>
<td>f 9.14</td>
<td>8-hydroxy-2,6-dimethyloctanoic acid</td>
</tr>
<tr>
<td>h 8.37</td>
<td>8-foliamenthoic acid</td>
</tr>
<tr>
<td>i 8.52</td>
<td>2,6-dimethyl-3-oxooct-6-enoic acid</td>
</tr>
<tr>
<td>j 8.21</td>
<td>8-carboxygeranic acid</td>
</tr>
<tr>
<td>k 7.21</td>
<td>8-carboxygeranic acid conjugated to hexose</td>
</tr>
<tr>
<td>l 7.50</td>
<td>8-hydroxy-3-[5-hydroxy-3-methylpent-3-en-1-yl]-2-methyloxirane-2-carboxylic acid or 2,6-dimethyl-3-oxooct-6-enoic acid conjugated to hexose</td>
</tr>
<tr>
<td>m 7.56</td>
<td>8-hydroxy-2,6-dimethyloct-2-enoic acid conjugated to hexose</td>
</tr>
<tr>
<td>n 7.65</td>
<td>8-hydroxygeraniol conjugated to hexose</td>
</tr>
</tbody>
</table>

Figure S16: A list and comparison of the relative amounts of all compounds identified in the methanol extracts of Agro-infiltrated N. benthamiana leaf-discs. A list of all compounds identified by UPLC-MS of a methanol extract of N. benthamiana leaf discs transiently expressing CYP76C4, CYP76B6 or the empty pCAMBIA33003SSu plant expression vector (neg-control) incubated with geraniol is shown (A). The peak areas in the UPLC-MS chromatogram of a given compound were compared between the samples to compare the relative amounts (B).
Figure S17: Analysis of non-volatile compounds in agro-infiltrated *N. benthamiana* leaves. A) LC-MS chromatograms of agro-infiltrated leaves with GES+CYP76B6, GES+CYP76C4, GES and empty vector (EV). B) Discriminant analysis of the LC-MS data from *N. benthamiana* leaves agro-infiltrated with GES, GES+CYP76B6, GES+CYP76C4, only CYP76B6, only CYP76C4, EV and wild type. PC1, PC2 and PC3 describe 59.3%, 14.7% and 10.2% of the total metabolite variation, respectively.
Figure S18: Proposed metabolic fate of 8-oxogeraniol produced by CYP76B6 in *N. benthamiana* leaves. All the compounds shown but not the two highlighted in gray were detected in the leaf disc tissues expressing CYP76B6 and incubated with geraniol. The three underlined compounds were also detected in *N. benthamiana* leaves, as glycoside, expressing *GES* and *CYP76B6*. Solid arrows indicate the flux from geraniol. Dashed arrows indicate putative reactions for which the product has not been detected.
Table S1: Fragmentation patterns of products detected in *N. benthamiana* leaf-discs Agro-infiltrated with CYP76B6 upon incubation with geraniol. Structures were deduced from molecular formula and main m/z fragments obtained in UPLC-MS (Figure 6).

<table>
<thead>
<tr>
<th>Structure</th>
<th>Formula</th>
<th>m/z in ES+</th>
<th>Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Structure" /></td>
<td>C_{10}H_{18}O</td>
<td>81</td>
<td>[M+H-2H_{2}O-(CH_{2}-C(CH_{3})-CH_{3})]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95</td>
<td>[M+H-H_{2}O-(CH_{2}-C(CH_{3})-CH_{3})]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>137</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>273</td>
<td>[2(M+H-H_{2}O)-H]</td>
</tr>
<tr>
<td><img src="image2.png" alt="Structure" /></td>
<td>C_{10}H_{18}O_{2}</td>
<td>93</td>
<td>[M+H-2H_{2}O-(CH_{2}-C(CH_{3})-CH_{3})]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>107</td>
<td>[M+H-H_{2}O-CH_{3}-(CH_{2}-OH)]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>135</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>193</td>
<td>[M+Na]</td>
</tr>
<tr>
<td><img src="image3.png" alt="Structure" /></td>
<td>C_{10}H_{16}O_{2}</td>
<td>93</td>
<td>[M+H-2H_{2}O-(CH_{2}-C(CH_{3})-CH)]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>123</td>
<td>[M+H-H_{2}O-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>151</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>191</td>
<td>[M+Na]</td>
</tr>
<tr>
<td><img src="image4.png" alt="Structure" /></td>
<td>C_{10}H_{16}O_{3}</td>
<td>121</td>
<td>[M+H-2H_{2}O-CO]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>167</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>207</td>
<td>[M+Na]</td>
</tr>
<tr>
<td><img src="image5.png" alt="Structure" /></td>
<td>C_{10}H_{18}O_{3}</td>
<td>m/z in ES-183</td>
<td>[M-H]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>123</td>
<td>[M+H-2H_{2}O-CO]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>151</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>169</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>187</td>
<td>[M+H]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>209</td>
<td>[M+Na]</td>
</tr>
<tr>
<td><img src="image6.png" alt="Structure" /></td>
<td>C_{10}H_{16}O_{4}</td>
<td>m/z in ES-199</td>
<td>[M-H]</td>
</tr>
<tr>
<td><img src="image7.png" alt="Structure" /></td>
<td>C_{10}H_{20}O_{2}</td>
<td>137</td>
<td>[M+H-2H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>155</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>195</td>
<td>[M+Na]</td>
</tr>
<tr>
<td><img src="image8.png" alt="Structure" /></td>
<td>C_{10}H_{20}O_{3}</td>
<td>125</td>
<td>[M+H-2H_{2}O-CO]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>153</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>171</td>
<td>[M+H-H_{2}O]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>211</td>
<td>[M+Na]</td>
</tr>
<tr>
<td>Retention time</td>
<td>Observed mass</td>
<td>Putative identity</td>
<td>Calculated mass</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>23.53</td>
<td>359.1340</td>
<td>Hexosyl carboxygeranic acid (2)</td>
<td>359.1342</td>
</tr>
<tr>
<td>20.51</td>
<td>521.1884</td>
<td>Dihexosyl carboxygeranic acid (2)</td>
<td>521.187</td>
</tr>
<tr>
<td>21.81</td>
<td>359.1341</td>
<td>Hexosyl carboxygeranic acid (1)</td>
<td>359.1342</td>
</tr>
<tr>
<td>18.82</td>
<td>521.1871</td>
<td>Dihexosyl carboxygeranic acid (1)</td>
<td>521.187</td>
</tr>
<tr>
<td>22.73</td>
<td>183.1028</td>
<td>Foliamenthoic acid</td>
<td>183.1021</td>
</tr>
<tr>
<td>19.09</td>
<td>345.1552</td>
<td>Hexosyl f oliamenthoic acid (1)</td>
<td>345.1549</td>
</tr>
<tr>
<td>19.72</td>
<td>345.1556</td>
<td>Hexosyl f oliamenthoic acid (2)</td>
<td>345.1549</td>
</tr>
<tr>
<td>22.07</td>
<td>493.2289</td>
<td>Dihexosyl Hydroxygeraniol</td>
<td>493.2285</td>
</tr>
<tr>
<td>29.41</td>
<td>417.1772</td>
<td>Malonyl hexosyl hydroxygeraniol</td>
<td>417.1761</td>
</tr>
<tr>
<td>28.71</td>
<td>373.1871</td>
<td>Acetyl hexosyl hydroxygeraniol</td>
<td>373.1862</td>
</tr>
<tr>
<td>23.56</td>
<td>463.2180</td>
<td>Pentosyl hexosyl hydroxygeraniol</td>
<td>463.2185</td>
</tr>
<tr>
<td>47.73</td>
<td>803.3708</td>
<td>Malonyl hexosyl geraniol (dimer)</td>
<td>803.3701</td>
</tr>
<tr>
<td>36.43</td>
<td>477.2341</td>
<td>Dihexosyl geraniol (1)</td>
<td>477.2331</td>
</tr>
<tr>
<td>40.91</td>
<td>519.2449</td>
<td>Acetyl dihexosyl geraniol (1)</td>
<td>519.2442</td>
</tr>
<tr>
<td>34.30</td>
<td>609.2762</td>
<td>Pentosyl dihexosyl geraniol</td>
<td>609.275</td>
</tr>
<tr>
<td>31.50</td>
<td>639.2867</td>
<td>Trihexosyl geraniol</td>
<td>639.2864</td>
</tr>
<tr>
<td>39.12</td>
<td>447.2241</td>
<td>Pentosyl hexosyl geraniol</td>
<td>447.2235</td>
</tr>
<tr>
<td>37.17</td>
<td>477.2337</td>
<td>Dihexosyl geraniol (2)</td>
<td>477.2331</td>
</tr>
<tr>
<td>41.52</td>
<td>519.2444</td>
<td>Acetyl dihexosyl geraniol (2)</td>
<td>519.2442</td>
</tr>
</tbody>
</table>
### Table S3: Oligonucleotides used for the construction of yeast and plant expression vectors.

<table>
<thead>
<tr>
<th>Oligo-name</th>
<th>Sequence</th>
<th>Enzyme restriction side</th>
<th>Target gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP76C1-f</td>
<td>5'-ggatccATGGACATAATCTCAGGGCAAG-3'</td>
<td>BamHI</td>
<td>CYP76C1</td>
</tr>
<tr>
<td>CYP76C1-r</td>
<td>5'-gaattcCTAATTAATTTGCGCGTTTCTT-3'</td>
<td>EcoRI</td>
<td>CYP76C1</td>
</tr>
<tr>
<td>CYP76C2-f</td>
<td>5'-ggatccATGGATACATTGAACAGCTT-3'</td>
<td>BamHI</td>
<td>CYP76C2</td>
</tr>
<tr>
<td>CYP76C2-r</td>
<td>5'-ggtaccCTAATTACCGGCAGTTTCTTT-3'</td>
<td>KpnI</td>
<td>CYP76C2</td>
</tr>
<tr>
<td>CYP76C3-f</td>
<td>5'-ggcttaauATGGACCTCTCTACTTAATGAGGCA-3'</td>
<td>USER</td>
<td>CYP76C3</td>
</tr>
<tr>
<td>CYP76C3-r</td>
<td>5'-ggttaaaaaaaaaTTAATAAGAGAGATATTGTAGGGTTT-3'</td>
<td>USER</td>
<td>CYP76C3</td>
</tr>
<tr>
<td>CYP76C4-f</td>
<td>5'-ggcttaauATGACGCTCTCTGCAGGCA-3'</td>
<td>USER</td>
<td>CYP76C4</td>
</tr>
<tr>
<td>CYP76C4-r</td>
<td>5'-ggttaaaaaaaaaTTAATAAGAGAGATATTGTAGGGTTT-3'</td>
<td>USER</td>
<td>CYP76C4</td>
</tr>
<tr>
<td>CYP76C5-f</td>
<td>5'-ggatccATGAGTTTACTCAGCTGTTC-3'</td>
<td>BamHI</td>
<td>CYP76C5</td>
</tr>
<tr>
<td>CYP76C5-r</td>
<td>5'-gaattcTCAGAGGCAAGTAAAACGAG-3'</td>
<td>EcoRI</td>
<td>CYP76C5</td>
</tr>
<tr>
<td>CYP76C6-f</td>
<td>5'-ggcttaauATGGACATCTCATCTCAAGGCAAC-3'</td>
<td>USER</td>
<td>CYP76C6</td>
</tr>
<tr>
<td>CYP76C6-r</td>
<td>5'-ggttaaaTTAATTTGCTGTTTCTTTCA-3'</td>
<td>USER</td>
<td>CYP76C6</td>
</tr>
<tr>
<td>CYP76C7-f</td>
<td>5'-ggatccATGAGTTTACTCAGCAGTGTCT-3'</td>
<td>BamHI</td>
<td>CYP76C7</td>
</tr>
<tr>
<td>CYP76C7-r</td>
<td>5'-gaattcTCAGAGGCAAGTAAAACGAG-3'</td>
<td>EcoRI</td>
<td>CYP76C7</td>
</tr>
<tr>
<td>CYP76B1</td>
<td>5'-ggcgtccATGGAGTTTAAATAGGTAGTA-3'</td>
<td>BamHI</td>
<td>CYP76B1</td>
</tr>
<tr>
<td>CYP76B1</td>
<td>5'-cgctgagcGTCTCAATGGGATAAGGCAACAC-3'</td>
<td>Sacl</td>
<td>CYP76B1</td>
</tr>
<tr>
<td>CYP76B6-f</td>
<td>5'-ggttaaaATGGAGTTTAAATAGGTAGTA-3'</td>
<td>USER</td>
<td>CYP76B6</td>
</tr>
<tr>
<td>CYP76B6-r</td>
<td>5'-ggttaaaTTAAGGGGCTTGGTACAGC-3'</td>
<td>USER</td>
<td>CYP76B6</td>
</tr>
</tbody>
</table>
Monoterpenes biosynthesis potential of plant subcellular compartments

Lemeng Dong¹, Sander Van Der Krol¹, Harro Bouwmeester¹

¹Laboratory of Plant Physiology, Wageningen University, 6700 AR Wageningen, The Netherlands

Submission in preparation
Abstract

Monoterpenes are C10 isoprenoids and are mostly produced in plastids from the common precursor geranyl diphosphate, GPP. In the present study, we compared the monoterpene biosynthesis potential of the plastids with that of cytosol and mitochondria. To achieve this, a truncated geraniol synthase cDNA was provided with an artificial targeting signal to target the encoded enzyme to the plastids, cytosol or mitochondria. Transient expression of these three different constructs in *Nicotiana benthamiana* resulted in the production of geraniol upon targeting to all three compartments, indicating there is local GPP biosynthesis or GPP exchange from other compartments. A comparison of three different *in vitro* characterized GPP synthase genes, showed that one, from *Picea abies* (*PaGPPS1*), could enhance geraniol production. Subsequently, the local potential for GPP and geraniol biosynthesis was probed by ectopic transient expression of *PaGPPS1* with artificial plastidic, cytosolic or mitochondrial targeting in all possible combinations with the plastidic, cytosolic or mitochondrial targeted GES. The headspace and leaf extracts were analyzed for geraniol and geraniol-derived products by GC-MS and LC-MS analysis, respectively. Results show that plastids have the highest potential for monoterpene production, followed by mitochondria and the cytosol the lowest. In addition, our study suggests that there is GPP exchange between all subcellular compartments, except from the cytosol to the plastids. Because the IPP consuming activity of a cytosolic *PaGPPS1* reduces plastidial monoterpene production, this suggests a preferred drain of IPP from the plastids when cytosolic IPP levels are low and that cytosolic produced GPP is not readily available for GES activity in the plastids. GPP produced in the mitochondria boosts GES activity in the cytosol and plastids. Because GPP in the cytosol does not boost geraniol production in the plastids, this suggests a direct exchange of GPP between mitochondria and plastids.

**Keywords:** Monoterpene; geranyl diphosphate synthase; plastids; cytosol; mitochondria
1. Introduction

Plants produce an array of different isoprenoids which are synthesised either in plastids, mitochondria or in cytosol. The types of isoprenoids synthesized in each cellular compartment largely depend on the substrates, geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) that are available. For example, in plastids the C10 GPP is used for monoterpene biosynthesis and the C20 GGPP for carotenoid (two GGPP units), chlorophylls, diterpene and prenyl lipid biosynthesis. In the cytosol, the C15 FPP is used for triterpene and steroid (two FPP units), and sesquiterpene biosynthesis and GGPP for polyprenol and diterpene biosynthesis. In the mitochondria, GGPP is used for diterpene and ubiquinone and the FPP is used for triterpene and polyprenol biosynthesis. The availability of these substrates is a function of the prenyltransferases localized within each organelle, geranyl diphosphate synthase (GPPS) in the plastids and geranylgeranyl diphosphate synthase (GGPPS) in the plastids, mitochondria and cytosol and farnesyl diphosphate synthase (FPPS) in the mitochondria and cytosol. Despite their differential localisation, all three enzymes exclusively utilize isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) as substrates. Isoprenoid biosynthesis capacity in the different subcellular compartments can be postulated to be a function of local GGPPS, GPPS and FPPS activity, local IPP and DMAPP pools and the capacity of substrate exchange between the compartments.

Monoterpenoids play an important role in a broad range of ecological and biological processes, such as direct defence against insects and pathogens (Keeling and Bohlmann, 2006; Unsicker et al., 2009), attraction of the enemies of herbivores (Arimura et al., 2004; Poecke and Dicke, 2004) or attraction of pollinators (Borg-Karlson et al., 1993). Monoterpenoids are generally volatile and can be stored in glandular trichomes or are emitted into the plant headspace. Monoterpenes may get modified by oxidation and/or glycosylation and monoterpene glycosides can be stored in the plant vacuole (Janson, 1993; Lücker et al., 2001; Mateo and Jiménez, 2000; Nogués and Loreto, 2013). Monoterpenes are widely used as fragrance additives in perfume (Curtis and Williams, 1994), wine (Câmara et al., 2004) and foods (Lewinsohn et al., 2001), but are also used as pesticide (Yang et al., 2012) and some have useful pharmaceutical properties (Bezerra et al., 2013; Lappas and Lappas, 2012). Exploring the monoterpenes biosynthesis potential of the different subcellular compartments in plants can give insight in the regulation of monoterpenes production in plants and could provide useful information for the optimisation of metabolic engineering of terpenoids.

The biosynthesis of monoterpenes is controlled by three main steps: (1) building the basic units IPP and DMAPP; they have been reported to be present in plastid, cytosol and mitochondria (Aharoni et al., 2005). (2) condensation of IPP and DMAPP by GPPS to form GPP; GPPSs are mainly active in the plastids although a low GPP pool has been
reported for the cytosol as well (Vranová et al., 2012; Wu et al., 2006). (3) conversion of GPP to the monoterpene parent skeleton catalyzed by a monoterpene synthase; monoterpene synthases are mostly targeted to the plastids, but some were also shown to be active in the cytosol (Dong et al., 2013). Here we tested the monoterpene biosynthesis capacity of the different subcellular compartments by ectopic expression of a geraniol synthase from *Valeriana officinalis* (*VoGES*) targeted to plastid, cytosol or mitochondria. To probe the potential IPP and DMAPP precursor pool we used over expression of different targeted *GPPS*. Supposedly, this makes local conversion of IPP and DMAPP to GPP only a function of local IPP/DMAPP levels. Moreover, to explore the exchange of GPP between subcellular compartments, we tested VoGES activity in one compartment with and without boosting GPPS activity in other compartments. We show that the highest potential for terpene biosynthesis resides in the plastids and there is considerable (bi)directional GPP exchange between the different subcellular compartments. Implications for the metabolic engineering of terpene biosynthesis in plants are discussed.

2. Material and methods

2.1 Construction of different targeted GES and GPPS

A geraniol synthase from *Valeriana officinalis* was isolated and characterized as previous described (Dong et al., 2013). To determine the potential of monoterpene biosynthesis in different subcellular compartments, the *VoGES* was provided with a plastid, cytosol or mitochondrial targeting signal. Firstly, a truncated version of *VoGES* lacking the first 56 AA (ΔN*VoGES*: bp 178 -1785) was made using standard cloning techniques and then was cloned into impact vector pIV2B 2.1 (www.pri.wur.nl/UK/products/ImpactVector/) under control of the CaMV 35S-promoter. In addition, the truncated versions of *VoGES* was provided with a heterologous plastid import signal by cloning into impact vector pIV2B 2.4 which carries an artificial plastid targeting signal (www.pri.wur.nl/UK/products/ImpactVector/) (Loyola-Vargas et al., 2007). The truncated *VoGES* was also cloned into impact vector pIV2B 2.5 which contains a yeast CoxIV secretion signal which was shown to deliver proteins to the mitochondrial matrix (Köhler et al., 1997).

Truncated *GPPS1* from *Picea abies* lacking of first 67 amino acids was individually cloned into impact vector pIV2A 2.1 (cytosol), pIV2A 2.4 (plastid), and pIV2A 2.5 (mitochondrial) under the control of the CaMV 35S-promoter. *AtGPPS* (*At2G34630.1*) cDNA (Bouvier et al., 2000) was cloned into impact vector pIV2A 2.1 pIV2A 2.4 (plastid).
All the constructs above were sequenced to check integrity, before transferring the impact vector to the binary vector pBIN+ (van Engelen et al., 1995) using LR recombination (Gateway technology) (Karimi et al., 2007). Full length of \textit{PaGPPS1} and \textit{PaGPPS2} were isolated from \textit{Picea abies} (Schmidt and Gershenzon, 2008; Schmidt et al., 2010) and cloned into the pCAMGW (a pCAMBIA derived vector) under the control of the CaMV 35S-promoter and a maize ubi1 promoter, respectively, using Gateway technology. Primer sequences for PCR amplification and restriction sites for each primer are listed in supplementary Table S2. Finally, the binary expression constructs were transformed into \textit{Agrobacterium tumefaciens} strain AGL0 (Lazo et al., 1991) which was used for transient expression in \textit{Nicotiana benthamiana} as described below.

\textbf{2.2 Transient expression in \textit{Nicotiana benthamiana}}

Transient expression of pathway genes using \textit{A. tumefaciens} infiltration (agro-infiltration) of leaves of five week old \textit{N. benthamiana} plants was performed as described (Dong et al., 2013). The amounts of cell suspension for each expression construct were kept constant by adding the corresponding amount of \textit{A. tumefaciens} carrying an empty vector. Leaves were harvested for metabolite analysis 5 days post agro-infiltration.

\textbf{2.3 Volatile GC-MS analysis and non-volatile LC-QTOF-MS analysis}

Headspace analysis of detached \textit{N. benthamiana} leaves was done by volatile trapping for 2.5 h in the light. Samples were analysed by GC-MS as described before (Dong et al., 2013). For quantification, a geraniol calibration curve was made with a series of standard solutions. After headspace trapping, samples were frozen in liquid nitrogen and used for LC-MS analysis as described (Dong et al., 2013).

Acquisition of Headspace GC-MS data and LC-MS data was performed using Xcalibur software (Thermo, Waltham) and MassLynx 4.0 (Waters), respectively. Mass data were processed using MetAlign version 1.0 (www.metalign.nl). Baseline and noise calculations were performed from scan number 500 to 9759 for GC-MS data and 70 to 2,480 for LC-MS. The maximum amplitude was set to 10,000,000 for GC-MS data and 25,000 for LC-MS data and peaks below three times the local noise were discarded. Multiple mass signals derived from the same compound were grouped using MSClust (biotools.wurnet.nl). The selected mass intensities were normalized and implemented in GeneMath XT version 2.1 as described before (Dong et al., 2013).
3. Result

3.1 Intrinsic monoterpenoid biosynthesis capacity of plastid, cytosol and mitochondria

To determine the potential of monoterpenoid biosynthesis in different subcellular compartments, the geraniol synthase from *Valeriana officinalis* (VoGES) (Dong et al., 2013) was provided with a plastid, cytosol or mitochondrial targeting signal and each was placed under control of the CaMV 35S promoter in a binary expression vector (see Methods). Each construct was transiently expressed using *Agrobacterium* infiltration of *N. benthamiana* leaves. Headspace analysis showed similar geraniol emission by *N. benthamiana* leaves expressing cytosol (C-VoGES) and plastid targeted VoGES (P-VoGES) and lower emission from the leaves expressing the mitochondrial targeted VoGES (M-VoGES) (Figure 1A). However, LC-MS analysis revealed significant difference in geraniol glycosides between leaves expressing C-VoGES and P-VoGES (Figure 1B), with much higher levels being produced upon plastid targeting. These glycosides are produced by the action of *N. benthamiana* glycosyl transferases on the heterologous geraniol. Besides geraniol glycosides also low levels of oxidized geraniol glycosides were detected, indicating that also oxidising *N. benthamiana* enzymes can use the geraniol as substrate. The putative identification for geraniol glycosides and and oxidized geraniol glycosides are listed in Table 1. The free geraniol and geraniol-derived glycosides upon VoGES targeting to the different compartments suggests that all compartments have some level of GPP available for GES activity, although this level is clearly highest in the plastids.

![Figure 1](image)

*Figure 1:* Geraniol production in agro-infiltrated *N. benthamiana* leaves with mitochondrial (M), cytosolic (C) and plastidic (P) targeted *Valeriana officinalis* geraniol synthase, VoGES.
A. GC-MS headspace analysis of geraniol emission
B. Relative quantitative analysis of geraniol (or oxidized geraniol) glycosides by LC-MS
Bars represent the total mass intensity of all geraniol and (or oxidized geraniol) glycosides, see Table 1. Bars with different letters differ significantly (*P*<0.05). Oxidized geraniol glycosides are indicated with capital letters. Geraniol glycosides and free geraniol are indicated with small letters.
Table 1: LC-QTOF-MS analysis of geraniol (or oxidized geraniol)glycosides in *N. benthamiana* leaves agroinfiltrated with *PaGPPS* and *VoGES*.

<table>
<thead>
<tr>
<th>Retention time (min)</th>
<th>Putative identity</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.89</td>
<td>Dihexosyl carboxygeranic acid</td>
<td>521.1986</td>
</tr>
<tr>
<td>23.22</td>
<td>Malonyl dihexosyl geranic acid (dimer)</td>
<td>1155.4382</td>
</tr>
<tr>
<td>27.85</td>
<td>Malonyl-hexosyl geranic acid (dimer)</td>
<td>831.3294</td>
</tr>
<tr>
<td>14.77</td>
<td>Hexosyl carboxygeranic acid</td>
<td>359.1049</td>
</tr>
<tr>
<td>19.10</td>
<td>Hesoxyl hydroxy geranic acid formic acid</td>
<td>391.1982</td>
</tr>
<tr>
<td>24.83</td>
<td>Dihexosyl carboxygeranic acid</td>
<td>521.2817</td>
</tr>
<tr>
<td>17.09</td>
<td>Dihexosyl dihydro-8-carboxygeraniol</td>
<td>499.2000</td>
</tr>
<tr>
<td>19.94</td>
<td>Malonyl dihexosyl pentosyl hydroxy geranic acid</td>
<td>725.2865</td>
</tr>
<tr>
<td>15.53</td>
<td>Hexosyl carboxygeranic acid</td>
<td>359.1352</td>
</tr>
<tr>
<td>23.01</td>
<td>Dihexosyl carboxygeranic acid</td>
<td>521.2249</td>
</tr>
<tr>
<td>15.21</td>
<td>Hesoxyl hydroxy geranic acid formic acid</td>
<td>391.1602</td>
</tr>
<tr>
<td>10.46</td>
<td>Hexosyl tetrahydro-carboxygeranic acid formic acid</td>
<td>409.1720</td>
</tr>
<tr>
<td>20.97</td>
<td>Dihexosyl geranic acid</td>
<td>491.2139</td>
</tr>
<tr>
<td>17.91</td>
<td>Malonyl hexosyl hydroxy geraniol (dimer)</td>
<td>835.3620</td>
</tr>
<tr>
<td>17.98</td>
<td>Trihexosyl geranic acid formic acid</td>
<td>699.2722</td>
</tr>
<tr>
<td>18.18</td>
<td>Malonyl hexosyl hydroxy geraniol</td>
<td>417.1781</td>
</tr>
<tr>
<td>22.37</td>
<td>Pentosyl-hexosyl geranic acid formic acid</td>
<td>507.2076</td>
</tr>
<tr>
<td>21.48</td>
<td>Dihexosyl geranic acid formic acid</td>
<td>537.2191</td>
</tr>
<tr>
<td>27.30</td>
<td>Malonyl-hexosyl geranic acid (dimer)</td>
<td>831.3295</td>
</tr>
<tr>
<td>15.77</td>
<td>Hexosyl geraniol formic acid</td>
<td>377.1834</td>
</tr>
<tr>
<td>18.78</td>
<td>Trihexosyl geraniol formic acid</td>
<td>685.2924</td>
</tr>
<tr>
<td>21.36</td>
<td>Dihexosyl geraniol</td>
<td>477.2345</td>
</tr>
<tr>
<td>20.06</td>
<td>Pentosyl dihexosyl geraniol formic acid</td>
<td>655.2819</td>
</tr>
<tr>
<td>21.65</td>
<td>Dihexosyl geraniol</td>
<td>477.2344</td>
</tr>
<tr>
<td>23.54</td>
<td>Acetyl Dihexosyl geraniol</td>
<td>519.2450</td>
</tr>
<tr>
<td>22.61</td>
<td>Pentosyl-hexosyl geraniol formic acid</td>
<td>493.2294</td>
</tr>
<tr>
<td>26.89</td>
<td>Malonyl-hexosyl geraniol (dimer)</td>
<td>803.3709</td>
</tr>
</tbody>
</table>

3.2 Only *PaGPPS1* increases geraniol formation in *N. benthamina* leaves

*VoGES* activity in the different subcellular compartments could be limited by the local conversion of IPP and DMAPP to GPP, as local subcellular GPPS activity may differ. To maximize local conversion of IPP and DMAPP to GPP we wanted to use a *GPPS* with targeting to either plastid, cytosol or mitochondria. We therefore set out to identify a functional *GPPS*. For this two *GPPS* from *Picea abies* (*PaGPPS1* and *PaGPPS2*) (Schmidt and Gershenzon, 2008; Schmidt et al., 2010) and one *GPPS* from *Arabidopsis* (*AtGPPS*) (Bouvier et al., 2000) were isolated and cloned into a binary expression vector. The *AtGPPS* was fitted with the same plastid targeting signal as the plastid targeted *VoGES*, for the *PaGPPSs* the native plastid targeting signal was used during the screen. Each GPPS expression construct was co-expressed with the *P-VoGES* in *N. benthamiana* leaves. Leaves were harvested 5 days post-agroinfiltration for headspace analysis by GC-MS and leaves were extracted for analysis of non-volatile products by LC-MS. Data analysis
using Metalign resulted in 6818 masses detected by GC-MS and 5057 masses detected by LC-MS. MSClust was used to combine masses belonging to the same compounds. For both volatiles (GC-MS) and non-volatiles (LC-MS), expression of either PaGPPS1, AtGPPS or PaGPPS2 alone did not have a significant effect on the product profile when compared with expression of empty vector (Figure 2A and B). When co-expressed with the P-VoGES, only the PaGPPS1 was able to boost product formation from P-VoGES activity (Figure 2). After data processing 145 volatiles and 155 non-volatile compounds were determined to be different in the P-VoGES plus PaGPPS1 co-expressing leaves compared to leaves which only express the empty vector (Figure 2). Geraniol was the only volatile that was different between the leaves expressing PVoGES and the empty vector (Figure 2A). However, when GPPS production was boosted by PaGPPS1 (co-expressing P-VoGES and PaGPPS1), in addition to geraniol 47 other volatiles were significantly increased (P<0.05; Figure 2A). From these 48 compounds, 18 were not detected in the N. benthamiana leaves which were agro-infiltrated with the empty vector (control), while 29 were present in the control, but were up-regulated upon co-expression of PaGPPS1 (Figure 2A). Besides the primary volatile product geraniol in the plant also conversion to oxidized geraniol derivatives and glycosides takes place. These geraniol-derived products were analysed by LC-MS. In the N. benthamiana leaves which co-expressed PaGPPS1 and P-VoGES, glycosides of geraniol and its further oxidized derivatives were significantly increased compared with expression of P-VoGES alone (Figure 2B). Interestingly, when P-VoGES was co-expressed with AtGPPS, geraniol glycosides significantly decreased compared with P-VoGES expression alone (Figure 2B). PaGPPS2 did not cause any change in geraniol glycosides (Figure 2B).

In these experiments we also identified a large cluster of compounds that are upregulated in empty vector agro-infiltrated leaves compared with non-infiltrated leaves, indicating a metabolic response to the presence of Agrobacterium in leaves (Figure 2).

To test if conversion of IPP and DMAPP to GPP is limiting geraniol production in the plastids, the plastid targeted P-VoGES was co-expressed with plastid targeted P-PaGPPS1. This resulted in a 118-fold increase in free geraniol in the headspace of leaves (Figure 3A), a small increase in the geraniol related glycosides and an approximately 5-fold increase in glycosides of oxidized geraniol derivatives (Figure 3B, Table 1). The results show that indeed the endogenous GPPS activity is limiting the substrate availability for P-VoGES. When this limitation is relieved by co-expressing P-PaGPPS1, the production capacity of the plastid is too large relative to the endogenous processing capacity (oxidation and glycosylation). As a consequence, the bulk of the geraniol is being emitted to the headspace.
Figure 2: Hierarchical cluster analysis of the metabolites extracted from agro-infiltrated *N. benthamiana* leaves with *AtGPPS, PaGPPS1, PaGPPS2* with and without plastid targeted *Valeriana officinalis* geraniol synthase (P-VoGES). Metabolites extracted from leaves agro-infiltrated with empty vector and non-agro-infiltrated leaves are shown as controls.

A. Volatiles analysed by headspace GC-MS. All volatiles, except geraniol, were only present in the leaves agro-infiltrated with *PaGPPS1* and *P-VoGES*.

B. Non-volatile metabolites as analysed by LC-MS.

Colour code indicates the quantitative change in mass intensity of all compounds.
When expression of plastid-targeted *P-VoGES*, was combined with an increase in GPP production in the mitochondria by co-expression of *M-PaGPPS1*, free geraniol production was boosted 76-fold (Figure 3A), suggesting that GPP produced in the mitochondria can be readily transported to the plastids. The increase in geraniol production in the plastids by *M-PaGPPS1* co-expression did not result in a significant increase in geraniol related glycosides, but glycosides of oxidized geraniol derivatives increased 4-fold compared with expression of *P-VoGES* alone (Figure 3B). Finally, when *P-VoGES* expression was combined with boosting of GPP production in the cytosol by *C-PaGPPS1* expression, the free geraniol levels in the headspace did not increase, while geraniol and geraniol-derived glycosides decreased compared with leaves expressing *P-VoGES* alone (Figure 3B). The decrease in geraniol (or derived) glycoside levels suggest that *C-PaGPPS1* activity in the cytosol actually lowers the IPP levels in the cytosol to such an extent that this increases the transfer of IPP from the plastids to the cytosol, thus competing with geraniol production in the plastid.

**Figure 3**: Geraniol production by *N. benthamiana* leaves agro-infiltrated with cytosolic (C), mitochondrial (M) or plastidic (P) *PaGPPS1* in combination with plastidic *VoGES*.
A. GC-MS headspace analysis of geraniol emission
B. Relative quantitative analysis of geraniol (or oxidized geraniol) glycosides by LC-MS
Bars represent the total mass intensity of all geraniol and (or oxidized geraniol) glycosides, see Table 1.
Bars with different letters differ significantly (*P*<0.05). Oxidized geraniol glycosides are indicated with capital letters. Geraniol glycosides and free geraniol are indicated with small letters.
3.4 Monoterpene biosynthesis potential of the cytosol when GPP biosynthesis in different compartments is increased

In the cytosol the MVA pathway supposedly provides the bulk of the IPP and DMAPP. Native conversion of cytosolic IPP and DMAPP to GPP may be limiting geraniol production as it was shown that there is just a small pool of cytosolic GPP (Wu et al., 2006). Indeed, when \textit{C-VoGES} was co-expressed with \textit{C-PaGPPS1}, free geraniol levels increased 30-fold compared with \textit{C-VoGES} alone (Figure 4A). However, increased GPP production in the cytosol through expression of \textit{C-PaGPPS1}, had little effect on the level of glycosides of geraniol and geraniol-derivatives (Figure 4B, Table 1). When \textit{C-VoGES} was combined with increased GPP production in the mitochondria by co-expressing \textit{M-PaGPPS1}, geraniol levels in the headspace increased 12-fold. This is less than for GPP production in the cytosol, but the total of free geraniol and geraniol-derived glycosides was higher with mitochondrial GPP production than with cytosolic GPP production (Figure 4B, Table 1). Apparently GPP can be transferred from mitochondria to the plastid. In contrast, boosting GPP production in the plastid did not result in a significant increase in free geraniol and geraniol-derived glycosides produced by the cytosolic geraniol synthase (Figure 4A,-B), suggesting inefficient transfer of GPP from the plastids to the cytosol.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Geraniol production by \textit{N. benthamiana} leaves agro-infiltrated with cytosolic (C), mitochondrial (M) or plastidic (P) \textit{PaGPPS1} in combination with cytosolic \textit{VoGES}.}
\begin{enumerate}
\item A. GC-MS headspace analysis of geraniol emission
\item B. Relative quantitative analysis of geraniol (or oxidized geraniol) glycosides by LC-MS
\end{enumerate}
Bars represent the total mass intensity of all geraniol and (or oxidized geraniol) glycosides, see Table 1. Bars with different letters differ significantly ($P<0.05$). Oxidized geraniol glycosides are indicated with capital letters. Geraniol glycosides and free geraniol are indicated with small letters.
\end{figure}
3.5 Monoterpene biosynthesis potential of mitochondria when GPP biosynthesis in different compartments is increased

When \textit{M-VoGES} was co-expressed with \textit{M-PaGPPS1} geraniol production was increased 70-fold (Figure 5A), while geraniol-derived glycosides increased 3.3-fold and oxidized geraniol glycosides 82-fold (Figure 5B, Table 1). This suggests that conversion of IPP and DMAPP to GPP in the mitochondria normally (only \textit{M-VoGES} expression) is not saturated. When \textit{M-VoGES} expression was combined with \textit{C-PaGPPS1} expression, free geraniol increased more than by co-expression with \textit{M-PaGPPS1} (Figure 5A), while geraniol-derived glycosides also increased. This indicates that an efficient exchange of GPP from cytosol to mitochondria is possible. When GPP production was boosted in the plastids, geraniol production in the mitochondria also increased, although not as much as by boosting GPP production in cytosol or mitochondria (Figure 5A and B). This indicates that a limited transport of GPP from plastid to mitochondria is possible (Figure 5).

\textbf{Figure 5:} Geraniol production by \textit{N. benthamiana} leaves agro-infiltrated with cytosolic (C), mitochondrial (M) or plastidic (P) \textit{PaGPPS1} in combination with mitochondrial \textit{VoGES}.

A. GC-MS headspace analysis of geraniol emission
B. Relative quantitative analysis of geraniol (or oxidized geraniol) glycosides by LC-MS

Bars represent the total mass intensity of all geraniol and (or oxidized geraniol) glycosides, see Table 1. Bars with different letters differ significantly (\(P<0.05\)). Oxidized geraniol glycosides are indicated with capital letters. Geraniol glycosides and free geraniol are indicated with small letters.

There seems to be no direct correlation between the free geraniol emission level and the level of the accumulated geraniol related glycosides. For instance, co-expression
of C-PaGPPS1 and M-VoGES resulted in the highest level of emitted geraniol, while the highest geraniol / oxidized geraniol glycoside levels occurred in leaves expressing the combination of M-PaGPPS1 and M-VoGES (Figure 5).

4. Discussion

4.1 Highest monoterpene biosynthesis potential in the plastids

Here we show that in each subcellular compartment the pool of available IPP/DMAPP cannot be fully used by the introduced monoterpene synthase but that in each compartment the level of monoterpene production is limited by GPPS activity. When GPPS is targeted to the same compartment where GES is active, in each case geraniol and geraniol-derived glycoside production are boosted (summarized in Figure 6). In the plastid IPP is synthesized in the MEP pathway (Bick and Lange, 2003; Rodríguez-Concepción and Boronat, 2002), while in the cytosol IPP is synthesized by the MVA pathway (Bick and Lange, 2003; Rodríguez-Concepción and Boronat, 2002). Some research also suggests that IPP in the cytosol may be (partially) provided by the MEP pathway and that trafficking of IPP occurs unidirectionally from plastids to cytosol (Dudareva et al., 2005). For the mitochondria no IPP biosynthesis pathway has been described, however, the mitochondrial localization of IPP isomerases (Guirimand et al., 2012; Phillips et al., 2008), FPS (Martín et al., 2007; Vranová et al., 2013) and a GGPPS (Okada et al., 2000) suggest that IPP is available in mitochondria to support terpene biosynthesis. It has been suggested that IPP is imported from the cytosol (Disch et al., 1998; Lütke-Brinkhaus et al., 1984). Overall both un-boosted and boosted monoterpene production was highest in the plastids, followed by the mitochondria while lowest production occurred in the cytosol. As discussed above, it is assumed that isoprenoid production in the mitochondria depends on substrate (IPP) import from other compartments. However, the lower boosted production capacity of the cytosol than that of boosted production in the mitochondria suggests that the capacity to boost production in the mitochondria is mostly derived from substrate imported from the plastids rather than from the cytosol. Since the emission of the geraniol only take very small part (1% on a monthly basis) of the total geraniol production (Gershenzon et al., 2000), therefore the combined relative peak intensities of geraniol glycosides and geraniol-derived glycosides are taken as a measure for the available GPP in each compartment. The overall capacity (local boosting by GPPS) of the plastid is ~180 units (103 relative units for glycosides of geraniol and 77 for glycosides of geraniol derivatives; see Figure 3, 4, 5, and summarized in Figure 6). Measured in a similar way the overall capacity of mitochondria ~56 units and of the cytosol ~36 units (Figure 3, 4, 5, and summarized in Figure 6). The increase in the geraniol glycosides in the plastid by
plastid targeted GPPS1 is ~90 units, while this only results in a small increase in cytosolic or mitochondrial geraniol glycosides (~12 units), suggesting that only a small fraction of the GPP produced in the plastids is available in the cytosol or the mitochondria. In contrast, when GPPS is targeted to the mitochondria, the production of geraniol-related glycosides by mitochondrial targeted GES is boosted by 41 units and almost all of the GPP produced in the mitochondria seems to be available for geraniol production in the plastids, suggesting a very efficient transfer of GPP from mitochondria to plastids. When GPP production is boosted in the mitochondria, only 18 units of the potential 56 units in the mitochondria are available for geraniol production in the cytosol, suggesting that transfer of GPP from mitochondria to the cytosol is far less efficient than transfer to the plastids. When GPP production is boosted in the cytosol, almost all of the 36 units of GPP are available for boosting geraniol production in the mitochondria, suggesting a very efficient transfer of GPP from cytosol to mitochondria. Finally, when GPP production is boosted in the cytosol, the geraniol production in the plastid was drop 45% when compared with agroinfiltration with plastid targetd *P-VoGES* alone (Figure 3B). These could be due to the potential GPP production is drained from the plastids. This drain is most likely at the level of IPP and indicates that the cytosolic pathway when it utilizes higher amounts of IPP, directly competes with the plastidial pathway.

**Figure 6:** Summary of GPP exchange between different subcellular compartments
The units of GPP in each compartment are based on quantification of the geraniol glycosides plus / oxidized geraniol glycosides (relative units). The red boxes show endogenous local GPP units for each compartment without local boosting of GPP biosynthesis, the blue boxes show the additional GPP units produced by targeted *PaGPPS1*. The red plus blue boxes show the endogenous local GPP plus the additional GPP produced by PaGPPS1. Green boxes next to arrows show the relative units of GPP that are exchanged between compartments. Note that for plastids to cytosol, plastid to mitochondria and mitochondria to cytosol these are only a fraction of the potentially available units in these compartments. For the exchange from mitochondria to plastids and cytosol to mitochondria almost all available units can be exchanged, while there is no exchange of GPP units from cytosol to plastids.
4.2 **GPP exchange between mitochondria and plastids**

The results suggest that when GPPS is expressed in the mitochondria, almost all produced GPP is available for geraniol production in the plastids and cytosol. Because (according to our results (Figure 3)) there is no exchange of GPP from cytosol to plastids, this suggests that the boosting of plastid geraniol production by GPP produced in the mitochondria is mediated through a direct exchange of GPP between mitochondria and plastids. It has been shown that plastids and mitochondria may be spatially closely associated with each other as some metabolites from the photorespiration pathway in the plastids require recycling in the mitochondria (e.g. glycolate is recycled to glycine in the mitochondria). Also, ammonium, which is produced during photorespiration in the mitochondria must be re-assimilated by the glutamine synthase/glutamate synthase cycle in the plastids (Hodges et al., 2013). This raises the question how metabolites are exchanged between the mitochondria and plastids.

It has been shown that plastids may display structures called stromules, dynamic tubular membranes that have been shown to interact with the membrane of the endoplasmic reticulum (ER) (Schattat et al., 2011), nuclear envelope (Deslandes and Rivas, 2011) and with themselves (Gray, 2013). However, stromules may also interact with mitochondria (Gray, 2013; Hanson and Sattarzadeh, 2013). Gunning (Gunning, 2005) reported that also mitochondria can sometimes form transient tubular extensions and that these mitochondrial equivalents of stromules interacted with the streaming cytoplasm to develop a stromule like protrusion. Also, these structures protruding from mitochondria are dynamic, they can retract and may be newly produced along a different stream (Gunning, 2005). Presumably, at the contact side of stromules and mitochondrial tubular extensions, small molecules may be exchanged between these organelles, possibly including GPP.

In literature membrane contact sites (MCS) between different organelles of the cell have been described. These contact sites are close appositions between two organelles that facilitate both signalling and the passage of molecules from one cellular compartment to another. Although discovered many decades ago, still very little is known about the molecular machinery required to create them or their structure, function and regulation. The best-characterized contact sites to date are the nucleus-vacuole junction, the mitochondria-ER, plasma membrane-ER and plastid-ER contact sites (Elbaz and Schuldiner, 2011). In yeast, Kornmann et al. identified a protein complex that is located at the interface of the ER and mitochondria, and serves to zip them together (Kornmann and Walter, 2010). The tethering complex is composed of proteins resident of both ER and mitochondria and is functionally connected to phospholipid biosynthesis and calcium-signalling genes. Such protein complexes have not been described for plants yet. Plastid-ER contact sites, however, have been described and are
thought to be involved in the transfer of lipids between the organelles, especially under phosphate limitation when phospholipids of the plasma membrane are exchanged for plastid derived galactolipids (Xu et al., 2010). Because it has been shown that there is contact between the ER and both plastids and mitochondria, it could be that exchange of molecules between plastid and mitochondria occurs via the ER.

### 4.3 Monoterpene production capacity in *N. benthamiana*

Many of the primary monoterpene products are emitted as volatiles (Janson, 1993; Nogués and Loreto, 2013). However, when they are converted into non-volatile compounds, the evaporative losses were shown to be small (1% on a monthly basis) of the total monoterpene production (Gershenzon et al., 2000). When ectopically produced in the heterologous host *N. benthamiana* the volatile geraniol is mostly converted to non-volatile geraniol glycosides. In addition, the geraniol is first further oxidized to geranic acid, hydroxyl geraniol, hydroxyl geranic acid and carboxyl geranic acid which may subsequently be glycosylated as well. Oxidation and glycosylation of geraniol in a heterologous host may be a response to the potentially phytotoxic geraniol and provides the means to keep free geraniol levels low. We assume that glycosylated geraniol products are sequestered into the vacuole and that therefore oxidation takes place on free geraniol and not on geraniol-glycosides. After oxidation, the geraniol-derivatives are also further inactivated by glycosylation. Judging from the high production of geraniol glycosides, glycosylation of geraniol is apparently the preferred detoxification reaction. However, the capacity of geraniol glycosylation seems to be limited and in case of high geraniol production also further oxidized geraniol glycosides are produced (Figure 3, 4, 5). There was no relationship between the emitted free geraniol and the level of conjugated geraniol for the different combinations of targeted *VoGES* and targeted *PaGPPS1*. This could be an indication that the flow of geraniol towards the apoplastic space or towards conjugation and/or oxidizing activity may be different for plastids, mitochondria and cytosol.

In conclusion, results show that plastids have the highest potential for monoterpene production, followed by mitochondria and the cytosol the lowest. Terpene precursor GPP and IPP are able to move between different subcellular compartments. This work provided novel insight into the flux capacity for terpene precursor biosynthesis, which has broad implications for the production of terpenoid products in plants.
Acknowledgments

The research described here was funded by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 - SMARTCELL. PaGPPS1 and PaGPPS2 of pCAMGW constructs were provided by Prof. Jonathan Gershenzon (Max Planck Institute for Chemical Ecology).

References


Chapter 5

Discovery and reconstitution of the secoiridoid pathway from Catharanthus roseus

Karel Miettinen†, Lemeng Dong‡, Nicolas Navrot§, Thomas Schneider¶, Vincent Burlat¶, Jacob Pollier¶, Lotte Woittiez¶, Sander van der Krol§, Raphaël Lugan§, Tina Ilc¶, Robert Verpoorte†, Kirsi-Marja Oksman-Caldentey†, Enrico Martinoia¶, Harro Bouwmeester¶, Alain Goossens§, Johan Memelink† and Danièle Werck-Reichhart§

†Institute of Biology Leiden, Sylvius Laboratory, Sylviusweg 72, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands.
‡Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.
§Institut de Biologie Moléculaire des Plantes, Unité Propre de Recherche 2357 du Centre National de la Recherche Scientifique, Université de Strasbourg, 28 rue Goethe, 67000 Strasbourg, France.
¶Institute of Plant Biology, University Zurich, Zollikerstrasse 10, CH-8008 Zurich, Switzerland.
§CNRS; UMR 5546, Université de Toulouse ; UPS ; UMR 5546, Laboratoire de Recherche en Sciences Végétales ; BP 42617 Auzeville, F-31326 Castanet-Tolosan, France.
¶Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Gent, Belgium.
§Bio and Process Technology, VTT Technical Research Centre of Finland, P.O.Box 1000, FI-02044 VTT (Espoo), Finland.

† These authors contributed equally to this work.
Submitted to Nature Communications (under review)
Abstract

The secoiridoids and their derivatives, the monoterpenoid indole alkaloids (MIAs), form two large families of plant-derived bioactive compounds with a wide spectrum of high-value pharmacological and insect-repellent activities. Vinblastine and vincristine, MIAs used as anti-cancer drugs, are produced by Catharanthus roseus in extremely low levels, leading to high market prices and poor availability. Their biotechnological production is hampered by the fragmentary knowledge of their biosynthesis. Here we report the discovery of the last four missing steps of the secoiridoid biosynthesis pathway. Expression of the eight genes encoding this pathway together with two genes boosting precursor formation and two downstream alkaloid biosynthesis genes in an alternative plant host allow the heterologous production of the complex MIA strictosidine. This confirms the functionality of all enzymes of the pathway and highlights their utility for synthetic biology programs towards a sustainable biotechnological production of valuable secoiridoids and alkaloids with pharmaceutical and agricultural applications.

Keywords: Catharanthus roseus; secoiridoid; strictosidine; transcriptomics; proteomics; subcellular localization; tissue localization; pathway reconstitution
1. Introduction

Monoterpenoid indole alkaloids (MIAs) are a large group of plant-derived natural products with a range of pharmacological properties. Examples of MIAs are camptothecin used to treat cancer and quinine, the antimalarial drug of choice till the mid of the last century. Madagascar periwinkle, Catharanthus roseus, the best-characterized MIA-producing plant species, is the source of the valuable MIAs vincristine and vinblastine, which are used directly or as derivatives for the treatment of several cancer types. Because of the extremely low concentrations (0.0002 % fresh weight), production of vincristine and vinblastine is expensive (3000 USD/g) and availability of the drug is sensitive to environmental and political instability in the production countries. Therefore, biotechnology-based production of MIAs in microorganisms or alternative plant hosts has been proposed as a sustainable substitute but progress has been hampered by the lack of knowledge of the enzymes responsible for MIA biosynthesis, particularly in the secologanin pathway (Figure 1). Secologanin is the monoterpenoid (also called iridoid or secoiridoid) branch end point and is coupled to tryptamine by strictosidine synthase (STR) to form strictosidine, the universal MIA precursor in plants. The secologanin pathway has broad importance as many plant species accumulate iridoids and secoiridoids (including secologanin) as end products without incorporating them into complex alkaloids. Many secoiridoids are bioactive themselves, with among others anticancer, anti-microbial and anti-inflammatory activities (Dinda et al., 2007a; Dinda et al., 2007b; Tundis et al., 2008; Viljoen et al., 2012). Iridoids are also pheromones in some insect species, and as such can be employed for pest management in agriculture and the control of insect-related disease vectors (Birkett et al., 2011; Birkett and Pickett, 2003).

The secoiridoid pathway is still largely unresolved. It starts with geraniol and is thought to comprise approximately 10 enzymes catalyzing successive oxidation, reduction, glycosylation and methylation reactions (Figure 1). Although the pathway has been investigated for decades (Loyola-Vargas et al., 2007; Oudin et al., 2007), only the first step (geraniol 10-hydroxylase/8-oxidase, G8O) (Burlat et al., 2004) and the two last steps (loganic acid O-methyltransferase, LAMT; and secologanin synthase, SLS) (Irmler et al., 2000; Murata et al., 2008) are well established. Only recently, two additional enzymes were identified, iridoid synthase (IS) responsible for the reductive cyclisation step (Geu-Flores et al., 2012) and geraniol synthase (GES) (Simkin et al., 2012a). A complicating factor for gene discovery as well as biotechnological production is that the MIA pathway in C. roseus is organized in a complex manner, with the enzymes localized in different cell types and subcellular compartments (Supplementary Figure S1) (Mahroug et al., 2007; Verma et al., 2012).

Here we report the characterization of the last missing steps of the C. roseus seco-
iridoid pathway. We use an integrated transcriptomics and proteomics approach for gene discovery, followed by biochemical characterization of the isolated candidates. Furthermore, we reconstitute the entire MIA pathway up to strictosidine in the plant host *Nicotiana benthamiana*, by heterologous expression of the newly identified genes in combination with the previously known biosynthesis genes. This work provides essential tools that will allow development of synthetic biology platforms for the production of bioactive iridoids, seco-iridoids and complex MIAs with a wide range of agricultural and pharmaceutical applications, including the treatment of cancer.

### 2. Material and Methods

#### 2.1 Chemicals

The substrates 8-carboxygeranial, 8-carboxygeranic acid, 8-oxogeranic acid and 8-hydroxygeranic acid were synthesized on order by Synthelor (Vandoeuvre-Lès-Nancy, France), whereas 8-OH-geraniol, 8-oxogeraniol, 8-OH-geranial and 8-oxogeranial were synthesized by Chiralix B.V. (Nijmegen, Netherlands). Iridodial-glucoside, iridotrial-glucoside and 7-deoxyloganic acid were synthesized from aucubin extracted from *Aucuba japonica* leaves by Chiralix B.V. as described (Jensen et al., 1987; Rosendal et al., 1989). The aglucone iridoid pathway intermediates were produced by incubation with almond β-glucosidase (Sigma Aldrich) in 50 mM acetate buffer (pH 5). Loganetic acid and loganetin were produced by the deglucosylation of loganic acid and loganin (Extrasynthese). Aglycones were extracted with diethyl ether, evaporated under N₂ and quantified by ¹H-NMR.

#### 2.2 Transcriptomic analysis

Transgenic derivatives of *C. roseus* cell line MP183L (overexpressing the ORCA transcription factors) were generated by particle bombardment (van der Fits and Memelink, 1997) with derivatives of the pER8 plasmid (Zuo et al., 2000) carrying either the ORCA2 or ORCA3 ORFs. Selected transgenic lines were treated for 24 h with 10 µM estradiol, and RNA was isolated as described (Van Moerkercke et al., 2013). Illumina HiSeq2000 RNA sequencing, assembly, annotation and mapping of the RNA-Seq reads onto the reference transcriptome was carried out as described (Jensen et al., 1987). Complete linkage hierarchical cluster analysis was achieved using the CLUSTER and TREEVIEW software (Eisen et al., 1998) and the log₁₀ transformed values of the normalized FPKM values were used as input for CLUSTER.
2.3 Proteomic analysis

2.3.1 Plant material
*Catharanthus roseus* var. Little Bright Eyes seeds were sown in sterilized soil and covered with transparent plastic until germination. The soil was kept humid. The plants were fertilized weekly with 0.2% liquid Wuxal (29 g/l nitrate-N; 46 g/l ammonium-N; 25 g/l carbamide-N; 100 g/l P2O5 total phosphate; 75 g/l K2O; 124 mg/l B; 50 mg/l Cu; 248 mg/l Fe; Cu, Fe, Mn and Zn as EDTA-chelate). The plants were re-potted twice during further growth.

2.3.2 Isolation of epidermal protoplasts
Young, light-green *C. roseus* leaves (length 4-7 cm) from side branches (without buds or flowers) of 8 to 11-week-old plants were harvested for protoplast isolation. The mid-vein was removed and the leaves were cut into 1-2 mm strips. Protoplasts were isolated as described (Tohge et al., 2011). To collect epidermal protoplasts, a layer of 1 ml betaine solution (0.5 M betaine, 1 mM CaCl$_2$, 10 mM MES, pH 5.6 with KOH) was added on top and the tubes containing the protoplast mix. After centrifugation for 7 min at 1500 rpm and 4°C, epidermal protoplasts were collected from the upper interphase. The suspension was mixed with 4 ml protoplast solution and 25% Percoll (pH 6), and overlaid with 1.5 ml of betaine solution for a second purification step. The tubes were centrifuged at 700 rpm for 30 min and the epidermal protoplasts were again collected from the upper interphase. Protoplasts were pelleted in the betaine solution.

2.3.3 Isolation of mesophyll protoplasts
Mesophyll protoplasts were isolated as described for the epidermal protoplasts with replacement of the MCP solution with TEX solution (3.2 g/l Gamborg’s B5 medium, 3.1 mM NH$_4$NO$_3$, 5.1 mM CaCl$_2$, 2.6 mM MES, 0.4 M sucrose, pH 5.6-5.8 with 0.5 M KOH), and the betaine solution was replaced with mannitol/W5 (1 mM D-glucose, 30 mM NaCl, 25 mM CaCl$_2$, 1 mM KCl, 0.3 mM MES, 320 mM α-mannitol, pH 5.6-5.8 with KOH). For the first gradient, 5–10% Percoll (pH 6) was added instead of 3–5%.

Collection of protoplast microsomes. The protoplast pellet was resuspended in 2–4 ml buffer A (20 mM HEPES pH 7.2, 1 tablet/10 ml Roche Protease Inhibitor, 1 mM PMSF). The mixture was pressed 10–20 times through a syringe with needle and centrifuged for 10 min at 3000 rpm and 4°C. The samples were ultra-centrifuged for 1 h at 30,000 rpm. The membrane pellets were dissolved in 1 ml washing solution (0.3 M NaCl, 20 mM HEPES-KOH pH 7.2, 1 tablet/10 ml Roche Protease Inhibitor, 1 mM PMSF) by stirring with a small brush, and then vortexed for 30 s. The microsomes were frozen at -80°C for at least 1 h, thawed and then centrifuged for 1 h at 30,000 rpm and 4°C. The pellets were dissolved in 50–100 µl buffer A for further analysis. The protein concentration was determined using the Bio-Rad DC protein assay according to the manufacturer’s instructions.
2.3.4 Mass spectrometry

Approximately 50 µg of protein was separated by one-dimensional gel electrophoresis (Laemmli, 1970) in 10% polyacrylamide gels to reduce sample complexity. The gels were stained with Coomassie Brilliant Blue, the lanes were cut into ten slices, the proteins were reduced with tris(carboxyethyl)phosphine hydrochloride (TCEP) and sulphydryl groups were blocked with iodoacetamide. In-gel digestion with sequencing-grade modified trypsin (Promega V5111) was carried out overnight at 37°C (Shevchenko et al., 1996). The resulting peptides were recovered by adding 40 mM Tris-HCl (pH 8.0) and 50% acetonitrile/1% formic acid. The peptide mixtures were desalted by solid-phase extraction on C_{18} reversed-phase columns and analyzed on an LTQ Orbitrap mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to an Eksigent Nano HPLC system (Eksigent Technologies, Dublin, CA, USA) as previously described (Schneider et al., 2013).

2.3.5 Database searching and protein identification

Protein databases were searched using Mascot v2.3. Raw data were searched against a composite database consisting of all entries in the NCBI Viridiplantae database (released in November 2010), all publicly available C. roseus expressed sequence tags (downloaded from NCBI in November 2010) and the reference transcriptome released on CathaCyc and ORCAE (Van Moerkercke et al., 2013) (database contained forward and reverse protein entries, total number of protein entries 1,166,013). The parameters for precursor and fragment ion mass tolerance were set to 5 ppm and 0.8 Da, respectively. One missed trypsin cleavage was allowed. Carboxyamidomethylation of cysteine was specified as fixed modification, and oxidation of methionine and pyroglutamate formation from glutamine were selected as variable modifications.

2.3.6 Data processing and protein quantitation

Scaffold v3.0 (Proteome Software, Portland, OR, USA) was used to validate and quantify MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they were established at greater than 95% probability as specified by the Peptide Prophet algorithm (Keller et al., 2002). Protein identifications were accepted if they were established at greater than 90% probability and at least one peptide was uniquely assigned to a corresponding protein in a minimum of two of our samples. Protein probability was assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that were identified with the same set of peptides and could not be differentiated by MS/MS analysis were grouped to protein clusters to satisfy the principles of parsimony. Peptide and protein false-discovery rates (FDR) were determined by the Scaffold software. A peptide FDR of 0.01% and a protein FDR of 0.2% were computed. The Scaffold software was also used to determine protein abundance in the mesophyll and epidermal factions based on the number of spectra assigned to each protein. An F-test was applied to assess significant differences in protein abundances.
2.4 Gene isolation

Open reading frames (ORFs) were amplified by PCR from a pACT2 cDNA library of a *C. roseus* cell culture elicited with yeast extract (Menke et al., 1999) using the primers listed in Supplementary Table S2. For expression in plants, the ORFs were transferred to vector pRT101 (Töpfer et al., 1987) to bring them under the control of the Cauliflower Mosaic Virus 35S promoter, and the expression cassettes were then transferred to the binary vector pCAMBIA1300 (Cambia). For expression in *E. coli*, the ORFs were transferred to vector pASK-IBA45plus (IBA) and/or pET16-H (Novagen pET-16b derivative). Probes for *in situ* hybridization were prepared from the same ORFs cloned in pBluescript II SK+. For localization analysis the ORFs were transferred to vector pTH2 (Chiu et al., 1996; Niwa et al., 1999) and/or pTH2BN (a derivative of pTH2). The marker for nucleocytosolic localization (pRT101-mCherry) was prepared by amplifying the mCherry ORF from plasmid ER-rk (Nelson et al., 2007) (The *Arabidopsis* Information Resources, TAIR, clone CD3-959).

2.5 Isolation of His-tagged recombinant proteins

Recombinant proteins carrying a His$_6$ tag were expressed using plasmid pASK-IBA45plus and/or pET16-H in *E. coli* strain BL21 (DE3) pLysS and purified by Ni-NTA agarose chromatography (Qiagen). For quality analysis, the recombinant proteins were separated by 12.5% (w/v) SDS-PAGE, transferred to Protran nitrocellulose membranes (Whatman) by semidy electroblotting, and western blots were probed with mouse monoclonal anti-His horseradish peroxidase (HRP) conjugate antibodies (5Prime). Antibody binding was detected by incubation in 250 μM sodium luminol, 0.1 M Tris-HCl (pH 8.6), 3 mM H$_2$O$_2$ and 67 μM pcoumaric acid, followed by exposure to X-ray film.

2.6 Enzymatic assays of UGT and oxidoreductases

UGT activity was detected in 0.1 ml reaction buffer containing 50 mM potassium phosphate (pH 7.5), 2 mM UDP-glucose, 5–1000 μM 7-deoxyloganetic acid or 2 mM of other tested compounds and 50–1000 ng of purified enzyme. Reactions were incubated at 32°C for 15 min and stopped by adding 1 volume of methanol, mixed by vortexing and kept on ice for 10 min. The tubes were centrifuged at 4000 x g for 10 min, and the supernatants were passed through 0.22-μm nylon filters.

Oxidoreductase activity was detected in 1 ml reaction buffer containing 50 mM bis-tris propane (pH 9 for oxidation and pH 7.5 for reduction), 2–1000 μM 8-OH-geraniol, 8oxogeraniol, 8-OH-geranial, 8-oxogeranial or other tested compounds, 2–2000 μM NAD$^+$ or NADH and 50–1000 ng of purified enzyme. Reactions were incubated for 15 min at 32°C, stopped by adding 0.2 volumes of 1 M sodium citrate (pH 3) and centrifuged and
filtered as above. Quantitative assays were carried out by measuring NADH production at 340 nm in a Nanodrop2000c (Thermoscientific).

2.7 Chromatographic analysis of 8-HGO and 7-DLGT products

Identification of 8-HGO enzyme products was done by capillary gas chromatography mass spectrometry. The ethyl acetate extract of the reaction mixture was separated on an Agilent GC 7890A series equipped with a 5975C MSD and DB-5 capillary column (30 m x 0.25 mm, film thickness of 0.25 μm) (JpW Scientific). Helium gas was used as carrier at a flow rate of 1.2 mL/min. The separation conditions were: split mode 1:5, injection volume 5 μL, injector temperature 230 °C, initial oven temperature 60 °C, then linear gradient to 100 °C at a rate of 20 °C/min followed by a linear gradient to 160 °C at a rate of 2 °C/min (run time 32 min). Analysis of 7-DLGT products was carried out using an Agilent series 1200 HPLC with a diode array detector (DAD) and a Polymer Laboratories PL-ELS 2100 ICE evaporative light scattering detector (ELSD) and a Phenomenex Luna 5 micron 150 x 4.6 mm C₁₈ column. The injection volumes were 10 or 100 μl. The binary solvent system consisted of acetonitrile and 0.1% trifluoroacetic acid in water. The elution program was: 5 min isocratic 10% acetonitrile and then 25 min gradient until 95% acetonitrile. Peak areas were calculated using Agilent ChemStation.

2.8 NMR

Structures of enzyme products were analyzed by nuclear magnetic resonance (NMR) spectroscopy in 750 μl of acetone-d₆ or methanol-d₄ in a 5 mm NMR glass tube (Kim et al., 2010).

2.9 Subcellular localization studies

C. roseus MP183L cell suspension cultures were maintained by weekly 10-fold dilution in 50 ml Linsmaier & Skoog (LS) medium containing 88 mM sucrose, 2.7 μM 1-NAA and 0.23 μM kinetin (LS-13). The cells were grown at 25°C with an 18/6h light dark cycle. The cells were bombarded (Hallahan et al., 1995) using the plasmids pTH2-IO, pTH2-7-DLH, pTH2-7-DLGT, pTH2BN-7-DLGT, pTH2-8-HGO and pTH2BN-8-HGO. The first two were combined with equal amounts of the ER marker ER-rk, and the others with the nucleocytosolic marker mCherry (Nelson et al., 2007). Bombarded cells were placed on Petri dishes with LS-13 medium and viewed after 24 h using a Zeiss Observer laser scanning microscope.
2.10 In situ hybridization

pBluescript plasmid derivatives containing the cDNAs for 8-HGO, IO, 7-DLGT and 7-DL were used for the synthesis of antisense and sense digoxigenin-labeled riboprobes as previously described (Burlat et al., 2004). G8O antisense probes (Burlat et al., 2004) and SGD antisense probes (Guirimand et al., 2010) were used as internal phloem-associated parenchyma and epidermis markers, respectively. Paraffin-embedded serial longitudinal sections of young developing leaves were hybridized with digoxigenin-labeled riboprobes and localized with antidigoxigenin-alkaline phosphatase conjugated antibodies (Mahroug et al., 2006).

2.11 P450 expression in yeast and enzyme assays

P450 coding sequences were amplified from pRT100 source vectors using specific primers to introduce USER™ sites at the ends of each sequence. The genes were subsequently transferred to the plasmid pYeDP60 using the USER™ cloning technique (New England Biolabs, Ipswich, UK)(Nour-Eldin et al., 2006). The resulting recombinant plasmids were introduced into *S. cerevisiae* strain WAT11, cultivated at 28°C and P450 expression was induced as described (Höfer et al., 2013). Cells were harvested by centrifugation and manually broken with 0.45 mm glass beads in 50 mM Tris-HCl buffer (pH 7.5) containing 1 mM EDTA and 600 mM sorbitol. The homogenate was centrifuged for 10 min at 10,000 x g and the supernatant was centrifuged for 1 h at 30,000 x g. The pellet, comprising microsomal membranes, was resuspended in 50 mM Tris-HCl (pH 7.4), 1 mM EDTA and 30% (v/v) glycerol with a Potter-Elvehjem homogenizer and stored at −20°C. All procedures for microsomal preparation were carried out at 0-4°C. P450 expression was evaluated as described (Omura and Sato, 1964) and enzymatic activities were determined in a standard 0.1 ml assay comprising for IO 2.3 nmol cytochrome P450, 0.6 mM NADPH and substrate in 20 mM sodium phosphate (pH 7.4). The reaction was initiated by the addition of NADPH and was stopped after 5 min by the addition of 10 µl 1M HCl. Iridoids were extracted in 1 ml ethyl acetate, and the organic phase was concentrated to 200 µl prior to GC-FID analysis on a Varian 3900 gas chromatograph (Agilent technologies) equipped with a flame ionization detector and a DB-5 column (30 m, 0.25 mm, 0.25 µm; Agilent technologies) with splitless-injection, at 250°C injector temperature, and with a temperature program of 0.5 min at 50°C, 10°C/min to 320°C, and 5 min at 320°C. For 7-DLH, 10 µl of yeast microsomes expressing 7-DLH (130 µg of microsomal protein) were incubated for 20 min at 27°C, in 0.1 ml of 20 mM Na-phosphate (pH 7.4) containing 0.6 mM NADPH and substrate. The reaction was initiated by the addition of NADPH and was stopped after 20 min on ice. After addition of 50 µl of 50 % acetic acid, tubes were vortexed and centrifuged. The supernatant was run on reverse-phase HPLC (Alliance 2695 Waters system, NOVA-PAK C18 4.6 x 250 mm
column) with photo-diode array detection at 236.5 nm. Peak areas of the product(s) were used to calculate the catalytic parameters of each enzyme.

2.12 Leaf disc assays

Five-week-old *N. benthamiana* leaves were infiltrated with *A. tumefaciens* transformed with vector pC1300 containing the relevant candidate genes, plus the helper plasmid p19. Five days post-infiltration, leaves were used in a leaf disc assays as previously described (Van Moerkercke et al., 2013).

2.13 Pathway reconstruction in *N. benthamiana*

The pathway genes were transiently expressed in the leaves of five-week-old *N. benthamiana* plants by agroinfiltration as previously described (24). Briefly, bacteria carrying the relevant expression constructs (PaGPPS, VoGES, G8O, 8-HGO, IS, IO, 7-DLGT, 7-DLH, LAMT, SLS, TDC, STR, empty vector or TBSV p19 (Voinnet et al., 2003)) were grown individually at 28°C for 24 h. Cells were harvested by centrifugation and then resuspended in infiltration buffer containing 10 mM MES (Duchefa Biochemie), 10 mM MgCl2 and 100 µM acetosyringone (4′-hydroxy-3′,5′-dimethoxyacetophenone, Sigma) to a final OD600 of ~0.5. For all gene combinations that compared subsequent steps in the pathway, the amounts of cell suspension for each expression construct were kept constant by adding the corresponding amount of *A. tumefaciens* carrying an empty vector. All infiltrations were done in three replicates. In several experiments, pathway intermediates were injected into the same leaves three days after agroinfiltration. Compounds used for infiltration were diluted to a final concentration of 400 µM in methanol/water (1:19), with the same ratio of methanol/water alone as a negative control and 400 µM 7-deoxyloganic acid, 8-carboxygeranic acid, secologanin or strictosidine as positive controls. Leaves were harvested for metabolite analysis 5 days after agroinfiltration. Frozen, powdered *N. benthamiana* leaves (200 mg aliquots) were extracted in 0.6 ml 99.867% methanol, 0.133% formic acid and 5 µl of the extract was analyzed using a Waters Alliance 2795 HPLC connected to a QTOF Ultima V4.00.00 mass spectrometer (Waters, MS technologies, UK). Measurements were taken in negative ionization mode. LC-MS data were acquired using MassLynx 4.0 (Waters) and processed using MetAlign version 1.0. The normalized and log-transformed data matrix was used for principal component analysis implemented in GeneMath XT v 2.1. ANOVA was used to evaluate the statistical significance of differences in metabolite levels between all treatments.
3. Results

3.1 Iridoid gene discovery based on transcriptome and proteome analysis

We previously reported the assembly of CathaCyc, a C. roseus metabolic pathway database based on Illumina HiSeq2000 RNA sequencing data (Guirimand et al., 2011). The dataset was derived from C. roseus suspension cells and shoots treated with the plant hormone methyl jasmonate (MeJA) (Van Moerkercke et al., 2013). Here, we complemented this dataset with RNA-Seq data from cell suspensions overexpressing either ORCA2 or ORCA3, transcription factors that regulate the expression of LAMT, SLS and several other genes in the MIA biosynthesis pathway (van der Fits and Memelink, 2000), but not GES and G8O (Figure 2a). These four and all other known MIA genes are induced by MeJA both in cell suspension cultures and whole C. roseus plants, although GES/G8O and LAMT/SLS show different induction characteristics (Figure 2b). Furthermore, GES, G8O and IS are expressed in the internal phloem-associated parenchyma (IPAP) cells, whereas LAMT and SLS are expressed in the leaf epidermis (Geu-Flores et al., 2012; Guirimand et al., 2011; Irmler et al., 2000; Murata et al., 2008; Simkin et al., 2012a; Tundis et al., 2008). The differential induction and in situ expression data suggested that the first part of the pathway (possibly up to 7-deoxyloganin acid) corresponds to one transcriptional regulon, whereas the subsequent steps up to the synthesis of secologanin would comprise a second regulon.

Based on the hypothetical pathway and predicted enzyme activities catalyzing the hitherto missing steps (Figure 1), we screened our dataset for genes encoding NAD(P)-binding Rossmann fold domain-type oxidoreductases, cytochrome P450 monooxygenases (P450s) and UDP-glycosyltransferases (UGTs) that display co-expression with GES/G8O, and for P450s that show co-expression with LAMT/SLS. Three genes encoding putative oxidoreductases showed a high degree of co-expression with GES/G8O (Figure 2a). The first (accession number Caros008267) was annotated as a progesterone reductase, the second (Caros003452) as an aldehyde dehydrogenase and the third (Caros009903) as a 12-oxophytodienoate reductase. We also identified four P450s (Caros020058, Caros001222, Caros018961, and Caros005234) and one UGT (Caros009839) that showed close co-expression with GES/G8O (Figure 2a). The first oxidoreductase (Caros008267) was found to encode the recently described iridoid synthase (Geu-Flores et al., 2012), thus confirming the validity of our screening strategy. The others were selected for further functional analysis. We also verified the co-expression of these candidate genes in the publicly available dataset from the Medicinal Plant Genomics Resource consortium (http://medicinalplantgenomics.msu.edu), which has been integrated into the ORCAE website (http://bioinformatics.psb.ugent.be/orcae (Van Moerkercke et al., 2013). This analysis strongly supported our selection of candidate genes (Figure 2b).
Figure 1: The secologanin-strictosidine pathway.
Genes indicated in boxes were published before (black background) or during (white background) the present study, or are reported here (yellow background). Frames indicate mRNA localization in the leaf internal phloem-associated parenchyma (IPAP) (pink) or epidermis (blue). Numbers indicate predicted enzyme classes in the initial gene discovery strategy. 1: oxidoreductase, 2: cytochrome P450, 3: UDP-glycosyltransferase (UGT). GPPS, geranyl diphosphate synthase, GES, geraniol synthase; GBO, geraniol 8-oxidase; 8-HGO, 8-hydroxygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7-DLGT, 7-deoxyloganetic acid glucosyl transferase; 7-DLH, 7-deoxyloganetic acid hydroxylase; LAMT, loganic acid O-methyltransferase; SLS, secologanin synthase; STR, strictosidine synthase; TDC, tryptophan decarboxylase. Iridotrial indicated in brackets was a previously proposed intermediate that we did not detect in vitro or in vivo.
Discovery and reconstitution of the secoiridoid pathway from *Catharanthus roseus*  

**Figure 2:** Gene discovery strategy. 
(a-b) Complete-linkage hierarchical clustering of early MIA pathway gene expression in *C. roseus* based on our data (a) or the Medicinal Plant Genomics Resource consortium (http://medicinalplantgenomics.msu.edu) (b). Colors indicate transcriptional activation (blue) or repression (yellow) relative to untreated samples. Tissues: Fl, flower; ml, mature leaves; il, immature leaves; St, stem; Ro, root; Sdlg, seedling. Suspension cells (CellSus): Wt, wild-type; O2, ORCA2; O3, ORCA3. Hairy roots (HairRt): Wt, wild-type; Td, TDG; RebH, RebH_F. Treatments: Not, no treatment; MeJA, methyl jasmonate (6, 12 or 24 h); Con, mock; YE, yeast extract. 
(c) Candidate P450 proteins in epidermis and mesophyll from proteomics analysis. GES, geraniol synthase; G8O, geraniol 8-oxidase; 8-HGO, 8-hydroxygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7-DLGT, 7-deoxyloganetic acid glucosyl transferase; 7-DLH, 7-deoxyloganic acid hydroxylase; LAMT, loganic acid O-methyltransferase; SGD, strictosidine β-D-glucosidase; SLS, secologanin synthase; STR, strictosidine synthase (1-3: three related genes); TDC, tryptophan decarboxylase.
The tissue localization of the enzymes in the leaf was used as a second criterion to pinpoint the most promising candidate genes. This was investigated using proteomics on epidermal and mesophyll protoplasts isolated from *C. roseus* leaves. The proteomics analysis resulted in the identification of 2200 proteins. Three P450s (Caros005234, Caros006766 and Caros020058) and one UGT (Caros020739) were enriched in the mesophyll fraction, whereas one P450 (Caros007986), one UGT (Caros004449) and nine oxidoreductases (Caros022489, Caros002459, Caros017236, Caros002170, Caros012730, Caros006689, Caros007544, Caros021570 and Caros003491) were enriched in the epidermal fraction. One P450 (Caros003164) was present in both tissues (Figure 2c and Supplementary Table S1).

### 3.2 *Caros003452* is the missing 8-hydroxygeraniol oxidoreductase

The three oxidoreductases (Caros008267, Caros003452 and Caros009903) were produced in *Escherichia coli* and purified for *in vitro* enzyme assays. Confirming a previous report\(^\text{12}\), the putative progesterone reductase (Caros008267) was shown to possess iridoid synthase (IS) activity in the presence of 8-oxogeranial (data not shown), yielding a mixture of *cis-trans*-iridodial and *cis-trans*-nepetalactol. Caros003452 was active with the substrates 8-OH-geraniol, 8-OH-geranial and 8-oxogeraniol in the presence of NAD\(^+\), yielding mixtures of the three compounds and 8-oxogeranial in varying relative amounts depending on the combination and the incubation time (Figure 3). The enzyme was therefore coined 8-hydroxygeraniol oxidoreductase (8-HGO). With the co-factor NAD\(^+\) it did not convert 8-oxogeranial (Figure 3), and it was not active with any of the substrates listed above in the presence of NADP\(^+\)/NADPH. Relatively high activity was observed with some other primary alcohols such as geraniol (Supplementary Table S3). Given the complex kinetics, with four interconvertible compounds and eight possible reactions, the reaction kinetics could not be determined.

### 3.3 *CYP76A26* is the missing iridoid oxidase

Six candidate P450 genes (*CYP76A26*, *CYP81Z1*, *CYP81Q32*, *CYP72A224* and *CYP71AY1*, *CYP71AY2*) were transferred to a yeast expression vector and co-expressed in *Saccharomyces cerevisiae* together with the P450 reductase *ATR1* from *Arabidopsis thaliana* (Pompon et al., 1996) (Supplementary Fig. S2). Functional screening was carried out as described (Höfer et al., 2013) with geraniol, 8-hydroxygeraniol, 8-hydroxygeranial, 8-oxogeraniol, 8-oxogeranial, iridodial, iridotrial, 7-deoxyloganic acid and 7-deoxyloganetic acid as potential substrates. CYP76A26 converted both iridodial and iridotrial into 7-deoxyloganetic acid. The *cis*-iridodial and *trans*-iridodial freely interconverted with *cis-trans*-nepetalactol (Geu-Flores et al., 2012), and although CYP76A26 seemed to use the bicyclic nepetalactol as the preferred substrate, the
monocyclic cis- and trans-iridodial were also utilized, possibly after spontaneous conversion into nepetalactol (Figure 4). The interconversion and sequential metabolism of nepetalactol and the iridodials in aqueous solution prevented reliable evaluation of the catalytic parameters with these substrates. We never detected iridotrial as an intermediate in the iridodial to 7-deoxyloganetic acid conversion even after short incubations at low temperature. Although the reaction appeared to occur without detectable release of the iridotrial intermediate, the latter was very efficiently converted into 7-deoxyloganetic acid with a $K_{m,\text{app}}$ of 25 µM and $k_{\text{cat}}$ of 5.2 s$^{-1}$ (Supplementary Fig. S3). The $K_{m,\text{app}}$ value for iridotrial observed in these experiments is not particularly low. Together with the impossibility to detect iridotrial, this strongly suggests that the reaction occurs without release of this intermediate.

Figure 3: Functional characterization of recombinant 8-hydroxygeraniol oxidoreductase (8-HGO). Affinity-purified enzyme expressed in E. coli was incubated with 8-OH-geraniol and NAD$^+$. (a) GC-MS profile of the reaction extract compared with authentic standards. (b) Product identity was confirmed by comparison of MS spectra with authentic standards. 8-HGO catalyzes the stepwise and reversible conversion of 8-OH-geraniol into 8-oxogeraniol or 8-OH-geranial and then into 8-oxogeranial.

Because the related G80 (CYP76B6) was previously shown to oxidize several (other) monoterpenoid alcohols in addition to geraniol (Collu et al., 2001), these compounds were also tested as potential substrates for CYP76A26. The enzyme converted 8-oxogeraniol
into an unidentified product (Supplementary Fig. S4), albeit with a low efficiency, and catalyzed the hydroxylation of linalool, nerol, citronellol and lavandulol, but not geraniol (Supplementary Fig. S4; Supplementary Table S4). CYP76A26 thus catalyzes the direct conversion of iridodial into 7-deoxyloganetic acid, and was consequently named iridoid oxidase (IO) (Figure 1).

![Figure 4: Functional characterization of iridoid oxidase (IO).](image)

The cytochrome P450 enzyme was expressed in yeast and assayed as a microsomal preparation by incubation with cis-trans-nepetalactol and NADPH. (a) GC-MS profile of the reaction extract compared to the negative control and authentic standards. (b) Product identity was confirmed by comparison of the MS spectrum with the authentic standard. Cis-trans-nepetalactol is the peak 3\(^{13}\), in the substrate standard, peaks 1 and 2 are its open dialdehyde forms (iridodials). These compounds are in equilibrium in water\(^{13}\). Peak 3 is the first converted by IO, but ultimately, possibly after interconversion, all tree compounds are metabolized.

### 3.4 UGT709C2 is the missing 7-deoxyloganetic acid glucosyl transferase

The UGT (Caros009839 or UGT709C2), produced in *E. coli*, catalyzed the glucosylation of 7-deoxyloganetic acid to form 7-deoxyloganetic acid using UDP-glucose as the sugar donor (Figure 5). The enzyme had a \(K_{\text{mapp}}\) of 9.8 \(\mu\)M and a \(k_{\text{cat}}\) of 1.25 s\(^{-1}\) (Supplementary Fig. S5). The enzyme was inactive with loganetic acid, loganetin, iridodial, iridotrial, 8-OH-geraniol, jasmonic acid, gibberellic acid, indole acetic acid, salicylic acid, abscisic acid, zeatin and luteolin. It thus behaved as a selective 7-deoxyloganetic acid glucosyl transferase (7-DLGT) (Figure 1).
Discovery and reconstitution of the secoiridoid pathway from *Catharanthus roseus*  

Chapter 5

Figure 5: Functional characterization of 7-deoxyloganetic acid glucosyl transferase (7-DLGT).
Affinity-purified enzyme expressed in *E. coli* was incubated with 7-deoxyloganetic acid and UDP-glucose. (a) HPLC-DAD profile of the reaction extract compared with authentic standards. (b) Product identity was confirmed by comparison of the mass spectrum with authentic standard.

3.5 CYP72A224 is the 7-deoxyloganic acid hydroxylase

In despite of its poor expression in yeast, CYP72A224 catalyzed the conversion of 7-deoxyloganic acid into loganic acid in yeast microsomes (Figure 6) with a $K_{\text{mapp}}$ of 400 µM and a $V_{\text{max}}$ of 0.5 pmol.min$^{-1}$.mg$^{-1}$ microsomal protein (Supplementary Fig. S6). $K_{\text{cat}}$ could not be determined due to low expression in yeast preventing precise quantification of the enzyme concentration. Because of the low expression of CYP72A224 in yeast, a *N. benthamiana* leaf-disc assay was also carried out to further confirm its role in loganic acid formation (Supplementary Fig. S7). The aglycone derivative of 7-deoxyloganic acid (7-deoxyloganetic acid) was not a substrate for CYP72A224, confirming that glycosylation precedes hydroxylation of the cyclopentane ring. CYP72A224 was thus named 7-deoxyloganic acid hydroxylase (7-DLH).

CYP72A224 belongs to the same P450 subfamily as secologanin synthase (SLS; CYP72A1), which catalyzes the conversion of loganin into secologanin (Irmler et al., 2000). Both of these P450s use glycosylated substrates and have similar regiospecificities, suggesting that SLS evolved from 7-DLH thus extending the iridoid pathway to the seco-iridoids. The overlap in their substrate specificities was therefore investigated (Supplementary Fig. S8). Since SLS shows no 7-DLH activity and vice versa, the hypothesized evolutionary process apparently resulted in an exclusive shift leading to highly specific enzyme activities.
Figure 6: Functional characterization of 7-deoxyloganic acid hydroxylase (7-DLH). The cytochrome P450 enzyme was expressed in yeast and assayed as a microsomal preparation by incubation with 7-deoxyloganic acid and NADPH. (a) HPLC-DAD profile of the reaction extract compared to the negative control and authentic standards. (b) Product identity was confirmed by comparison of the mass spectrum with authentic standard.

3.6 Tissue-specific and subcellular localization of the pathway

Efficient pathway engineering requires precise knowledge of the cellular and subcellular organization of different pathway components. IO, 7-DLH, 8-HGO and 7-DLGT were therefore expressed as green fluorescent protein (GFP) fusions in C. roseus cells together with mCherry markers for nucleus, cytosol, endoplasmic reticulum (ER), plastids, mitochondria or peroxisomes. This revealed that IO and 7-DLH are ER-associated as predicted, whereas 8-HGO and 7-DLGT are soluble proteins found in the cytosol and the nucleus (Supplementary Fig. S9). Previous in situ transcript analysis (Guirimand et al., 2011) suggested that a pathway intermediate is translocated from IPAP cells (the site of the early biosynthetic steps) to the epidermis, where the final steps of the pathway occur. In situ hybridization showed that 8-HGO, IO, 7-DLGT and 7-DLH are expressed in the same tissue as G8O (Figure 7). The separate clustering of the expression of these genes (Figure 2a) and the similar tissue-specific transcript localization suggest that these genes constitute a transcriptional regulon for the production of loganic acid in IPAP cells. The epidermis-specific positive control SGD confirms the localization of the next transcriptional regulon, consisting of LAMT, SLS, STR and TDC, in the epidermis.
The existence of this regulon is supported by the highly similar co-expression profiles of these genes as shown in Figure 2a.

**Figure 7:** Expression of the transcriptional regulon required for loganic acid biosynthesis in the internal phloem-associated leaf parenchyma (IPAP) cells. *In situ* hybridization on serial longitudinal sections of young developing leaves was carried out with antisense (AS) probes and sense (S) probes as controls. Geraniol 8-oxidase (*G8O*) and strictosidine β-D-glucosidase (*SGD*) AS probes were used as IPAP and epidermis markers, respectively. Sense probe controls gave no signals. 8HGO, 8-hydroxygeraniol oxidoreductase; IO, iridoid oxidase; 7-DLGT, 7-deoxyloganetic acid glucosyl transferase; 7-DLH, 7-deoxyloganic acid hydroxylase; ipap, internal phloem associated parenchyma; ep, epidermis. Scale bar = 100 μm.
3.7 Reconstitution of strictosidine synthesis in Nicotiana benthamiana

The identification of the four enzymes described above allowed us to propose a complete secologanin pathway (Figure 1), which we tested by stepwise combinatorial transient expression of the corresponding genes in N. benthamiana (Figure 8). To boost substrate availability for the pathway a geranyl diphosphate synthase from Picea abies (PaGPPS) (Schmidt et al., 2010) and a geraniol synthase from Valeriana officinalis (VoGES) (Dong et al., 2013) were used because the C. roseus orthologs were not available at the onset of these studies. The transient expression of PaGPPS+VoGES in N. benthamiana resulted in the formation of geraniol, but also several additional oxidized and glycosylated derivatives were detected, presumably produced by endogenous N. benthamiana enzymes as previously reported (Dong et al., 2013). The stepwise addition of G8O and IS resulted in the generation of new compounds, including new derivatives of the anticipated pathway intermediates (combinations 1, 2 and 4 in Figure 8a and 8b; Supplementary Fig. S10). In contrast, the addition of 8-HGO to the combination PaGPPS+VoGES+G8O only modified the existing product profile but did not generate any new compounds (combination 3 in Figure 8a and 8b; Supplementary Fig. S10). This is probably due to the fact that endogenous N. benthamiana enzymes have similar activity as 8-HGO (Supplementary Fig. S11). When IO was co-expressed with PaGPPS+VoGES+G8O+8-HGO+IS (combination 5), the metabolic profile did not change, and 7-deoxyloganetic acid and its derivatives were not detected (Figure 8a and 8b). However, reconstitution of the pathway up to 7-DLGT (combination 6) was successful and resulted in the production of 7-deoxyloganic acid and putative acetylated 7-deoxyloganic acid (Figure 8a and 8c). Without IO, these products were not detected (Supplementary Fig. S12), indicating that IO is functional in N. benthamiana and is an essential part of the biosynthesis pathway. These findings also illustrate the importance of full pathway coverage for functional analysis of individual enzymes.

Subsequently, the entire postulated secologanin pathway (PaGPPS to SLS) was introduced by agroinfiltration, but this only yielded products up to 7-deoxyloganic acid. Therefore, we increased the input halfway into the pathway by infiltrating the intermediates iridodial, iridotrial or 7-deoxyloganic acid in combination with the pathway genes. In all cases this resulted in the production of secologanin, indicating that the second half of the pathway is also functional (Figure 8d). Finally we tested whether the biosynthesis pathway up to secologanin can be functionally combined with the tryptamine branch of the MIA pathway. When the secologanin biosynthesis pathway genes were co-infiltrated with the tryptophan decarboxylase (TDC) and STR genes, and the flux through the pathway was boosted by co-infiltration of the intermediates iridodial, iridotrial or 7-deoxyloganic acid, strictosididine was indeed produced (combination 10 in Figure 8d).
Figure 8: Reconstitution of the strictosidine pathway in N. benthamiana.
(a). Gene combinations infiltrated in leaves in triplicate. (b). Principal component analysis. PC1 and PC2 describe 36.2% and 31.1% of the total mass variation, respectively. (c). LC-MS analysis showing selected masses 401 and 359 representing (acetylated) 7-deoxyloganic acid (7-DLA) from infiltrations with 8-carboxygeranic acid (CGA), 7-DLA or gene combinations 6 or 7 (negative control). The two peaks likely represent 7-DLA acetylated at two different positions in the glucose moiety. (d). LC-MS analysis showing selected masses 433 (formic acid adduct of secologanin) and 575 (formic acid adduct of strictosidine) from infiltrations with secologanin or strictosidine, or with gene combinations 9 or 10, with or without iridodial. *Identical profiles with iridotrial or 7-DLH. Hex = hexosyl.

4. Discussion

Whereas feeding experiments clearly indicated that secologanin is derived from geraniol (Uesato et al., 1986), the exact sequence of intermediates and enzymes leading to the formation of its penultimate precursor loganic acid was still obscure. We report here 4 novel enzymes that together with two previously reported enzymes (Geu-Flores et al., 2012; Simkin et al., 2012b) fill the existing gaps and thereby provide a full description of the core secoiridoid pathway. Geraniol is converted to secologanin by the sequential
action of four different cytochrome P450 enzymes, two different oxidoreductases, one glucosyltransferase and one methyltransferase. The missing enzymes and corresponding genes were identified by a combination of transcriptomic and proteomic approaches exploiting the current knowledge of the spatiotemporal regulation of the seco-iridoid pathway.

Our results also address longstanding questions in the field. First, as previously proposed for oxidoreductase proteins purified from *Rauwolfia serpentina* and *Nepeta racemosa* (Hallahan et al., 1995; Ikeda et al., 1991), 8-HGO catalyzes two successive and reversible oxidation steps for the formation of 8-oxogeranial. Therefore two enzymes can contribute to the formation of the intermediate 8-oxogeraniol. G8O was recently shown to also produce 8-oxogeraniol from geraniol (Höfer et al., 2013), thus G8O and 8-HGO appear to catalyze partially overlapping (and in the case of G8O, monodirectional) oxidation reactions that result in the production of 8-oxogeraniol from 8-hydroxygeraniol (Figure 1). Second, a single cytochrome P450 enzyme IO/CYP76A26 can convert *cis-trans-*iridodial and *cis-trans-*nepetalactol into 7-deoxyloganetic acid without release of an iridotrial intermediate. Third, our data demonstrate that 7-deoxyloganetic acid glycosylation precedes its further oxygenation by 7-DLH. This answers the longstanding question which intermediate is transferred from IPAP cells to the epidermis. The expression of 7-DLH in IPAP cells indicates that loganic acid is the mobile intermediate transferred to the epidermis, and hence that glycosylation by 7-DLG is not sufficient for mobility but that further hydroxylation by 7-DLH is also required. The tissue-specific expression of pathway sections may increase the flux through the pathway by alleviating feedback inhibition by intermediates and products, and/or it may segregate the iridoid and seco-iridoid/MIA pathways, allowing them to fulfill tissue-specific functions. The high $K_m$ of 7-DLH for 7-deoxyloganetic acid probably results in substantial accumulation of 7-deoxyloganetic acid in IPAP cells where it may have a specific function.

Pathway reconstitution experiments in a heterologous host validated the enzyme sequence leading to the formation of strictosidine (Figure 1). We however observed a low flux through the pathway in the leaves of *N. benthamiana*. Previous experiments carried out with the upstream segment of the pathway indicated substantial conversion of intermediate products to other oxidized, acetylated, malonylated and glycosylated derivatives by endogenous *N. benthamiana* enzymes (Höfer et al., 2013). Despite downstream channeling of intermediates upon reconstruction of the full pathway, competing conversions are still observed and are probably responsible for the low flux. This would, for example, explain why we cannot see the product of IO upon infiltration of pathway combination 5 (Figure 8; Supplementary Fig. S12). Nevertheless, we know that IO is functional because infiltration of combination 9 or 10 with iridodial did result in production of secoliganin or strictosidine, respectively (Figure 8d). Side reactions
could possibly be eliminated by further metabolic engineering. The side products obtained in our experiments may also provide valuable information on the enzyme networks naturally present in *N. benthamiana* and could be leads for combinatorial biochemistry programs to produce chemical diversity and novel bioactive compounds.

In conclusion, we have identified the four missing enzymes in the secoiridoid pathway in *C. roseus*. In combination with the genes previously identified, the genes encoding these four enzymes are sufficient to engineer secoiridoid production and, together with *TDC* and *STR*, biosynthesis of the complex alkaloid strictosidine in the heterologous plant *N. benthamiana*. Although different segments of the strictosidine pathway are localized in different cell types in *C. roseus*, our results show that the entire pathway can be successfully reconstituted in a single *N. benthamiana* organ. This paves the way for the biotechnological production of valuable secoiridoids and derived compounds, such as the MIAs vincristine and vinblastine, making these important anti-cancer drugs available to more people and at a lower price.
Acknowledgments

We are grateful to Richard Twyman for critical reading of the manuscript and to Søren Rosendal Jensen for advice on iridoid substrate synthesis. David Nelson is acknowledged for naming P450 enzymes. The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 222716 – SMARTCELL. TI was supported by the Marie Curie ITN action P4Fifty.

References


Discovery and reconstitution of the secoiridoid pathway from *Catharanthus roseus*  

**CHAPTER 5**


Table S1: Candidate enzymes in mesophyll and epidermal cells.
A proteomic approach with microsomes isolated from enriched epidermal or mesophyll protoplast fractions was carried out as described in Methods. Shown are candidate enzymes found among 2200 identified proteins and their distribution between mesophyll and epidermis. n.s.: not significant.

<table>
<thead>
<tr>
<th>Enzyme class</th>
<th>Enzyme name</th>
<th>Accession number/locus</th>
<th>Proteomics hits</th>
<th>Statistics fishe's exact test</th>
<th>Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Epidermis</td>
<td>Mesophyll</td>
<td>Statistics fishe's exact test</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean ± SD</td>
<td>Mean ± SD</td>
<td>Significance level (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mesophyll</td>
<td>n.s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyt P450</td>
<td>CYP72A224</td>
<td>Caros005234</td>
<td>0.0±0.0</td>
<td>2.3±2.5</td>
<td>97</td>
</tr>
<tr>
<td>Cyt P450</td>
<td>CYP76B6</td>
<td>Caros006766</td>
<td>1.0±1.0</td>
<td>14.7±9.9</td>
<td>97</td>
</tr>
<tr>
<td>Cyt P450</td>
<td>CYP81Q32</td>
<td>Caros003164</td>
<td>6.3±4.2</td>
<td>7.7±7.2</td>
<td>0</td>
</tr>
<tr>
<td>Cyt P450</td>
<td>CYP76A26</td>
<td>Caros003676</td>
<td>0.67±1.2</td>
<td>11.0±8.4</td>
<td>95</td>
</tr>
<tr>
<td>Cyt P450</td>
<td>CYP71A2</td>
<td>Caros007986</td>
<td>8.0±2.6</td>
<td>3.0±2.6</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (old yellow enzyme)</td>
<td>CrOYE1</td>
<td>Caros006173</td>
<td>1.7±0.6</td>
<td>1.0±1.0</td>
<td>0</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH2</td>
<td>Locus_547</td>
<td>5.3±4.2</td>
<td>0.7±1.2</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH3</td>
<td>Locus_12</td>
<td>15.0±10.8</td>
<td>3.3±3.2</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH5</td>
<td>Locus_1687</td>
<td>11.7±7.8</td>
<td>4.3±3.2</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH8</td>
<td>Locus_26759</td>
<td>27.3±17.1</td>
<td>20.0±16.1</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH9</td>
<td>Locus_15</td>
<td>16.0±4.6</td>
<td>3.7±6.4</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH11</td>
<td>Caros007544</td>
<td>36.0±6.2</td>
<td>18.7±15.1</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH12</td>
<td>Caros005057</td>
<td>3.7±2.1</td>
<td>3.0±1.7</td>
<td>0</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH14</td>
<td>Locus_605</td>
<td>48.0±13.7</td>
<td>11.0±9.5</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase (alcohol dehydrogenase)</td>
<td>CrADH15</td>
<td>Locus_197</td>
<td>14.0±5.0</td>
<td>3.0±5.2</td>
<td>95</td>
</tr>
<tr>
<td>oxidoreductase</td>
<td>CrBBE1</td>
<td>Caros003491</td>
<td>2.0±2.6</td>
<td>0.0±0.0</td>
<td>95</td>
</tr>
<tr>
<td>UGT</td>
<td>CrUGT1</td>
<td>Locus_7621</td>
<td>2.3±4.0</td>
<td>1.0±1.7</td>
<td>95</td>
</tr>
<tr>
<td>UGT</td>
<td>CrUGT2</td>
<td>Locus_438</td>
<td>16.3±9.8</td>
<td>5.3±4.5</td>
<td>95</td>
</tr>
<tr>
<td>UGT</td>
<td>CrUGT3</td>
<td>Caros020739</td>
<td>0.0±0.0</td>
<td>10.3±14.6</td>
<td>95</td>
</tr>
</tbody>
</table>
Table S2: Sequences of primers and cloning methods used for plasmid constructions. Primers are displayed from 5’ to 3’ end.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Restriction site</th>
<th>Primer F</th>
<th>Primer R</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRT101 IO</td>
<td>XhoI,KpnI</td>
<td>GGCCCTTCTCGAGATGGCGACCATCACTTTCG</td>
<td>AAGGCGGCTACCTTAGATATGAACTCTTTCTTATGGGATG</td>
</tr>
<tr>
<td>pRT101 7-DLH</td>
<td>XhoI,KpnI</td>
<td>GGCCCTTCTCGAGATGGAATTGAACTTCAAATCAATTA</td>
<td>GGCCCTTGATACCTTAGATTTGTGCAGAATCACAAATGAG</td>
</tr>
<tr>
<td>pRT101 7-DLGT</td>
<td>KpnI,SpeI</td>
<td>CGGGTACCAGTTAGGGTTTCTCAAGAAACACAAATTTTG</td>
<td>CCACTAGGTTCAATTACGTTAGTTTTATGTAATCAAAC</td>
</tr>
<tr>
<td>pRT101 8-HGO</td>
<td>KpnI,XbaI</td>
<td>CGGTTACCAGTACAGTTGGATTGCTAGTTAGTTATGGAAGAGG</td>
<td>ACCAGGACTTCTATGAGTTTTATATGGTAATCAAAC</td>
</tr>
<tr>
<td>pRT101 CYP81Z1</td>
<td>XhoI,SalI</td>
<td>GCCGCTGAGGAGATGCATGATAGTTTTATGGAAGAGG</td>
<td>ACCAGGACTTCTATGAGTTTTATATGGTAATCAAAC</td>
</tr>
<tr>
<td>pRT101 CYP71AY1</td>
<td>XhoI,EcoRI</td>
<td>CGGGTACCAGTTAGGGTTTCTCAAGAAACACAAATTTTG</td>
<td>CCACTAGGTTCAATTACGTTAGTTTTATGTAATCAAAC</td>
</tr>
<tr>
<td>pRT101 IS</td>
<td>XhoI,KpnI</td>
<td>GCCGCTTCTCGAGATGGCGACCATCACTTTCG</td>
<td>AAGGCGGCTACCTTAGATATGAACTCTTTCTTATGGGATG</td>
</tr>
<tr>
<td>pRT101 SLS</td>
<td>XhoI,KpnI</td>
<td>GCCGCTTCTCGAGATGGCGACCATCACTTTCG</td>
<td>AAGGCGGCTACCTTAGATATGAACTCTTTCTTATGGGATG</td>
</tr>
<tr>
<td>pASKIBA45plus 7-DLGT</td>
<td>XhoI,KpnI</td>
<td>CGGGTACCAGTTAGGGTTTCTCAAGAAACACAAATTTTG</td>
<td>CCACTAGGTTCAATTACGTTAGTTTTATGTAATCAAAC</td>
</tr>
<tr>
<td>pET16H 8-HGO</td>
<td>KpnI,SalI</td>
<td>GCCGCTGAGGAGATGCATGATAGTTTTATGGAAGAGG</td>
<td>ACCAGGACTTCTATGAGTTTTATATGGTAATCAAAC</td>
</tr>
<tr>
<td>pASKIBA45plus IS</td>
<td>XhoI,SalI</td>
<td>GCCGCTGAGGAGATGCATGATAGTTTTATGGAAGAGG</td>
<td>ACCAGGACTTCTATGAGTTTTATATGGTAATCAAAC</td>
</tr>
<tr>
<td>pET16H IS</td>
<td>XhoI,KpnI</td>
<td>GCCGCTGAGGAGATGCATGATAGTTTTATGGAAGAGG</td>
<td>ACCAGGACTTCTATGAGTTTTATATGGTAATCAAAC</td>
</tr>
<tr>
<td>pTH2 IO</td>
<td>SalI</td>
<td>CCGGCTGAGAATAATGGCGACCATCATCATTTCG</td>
<td>CCGGCTGAGATATGAACTCTTTCTTATGGGATG</td>
</tr>
<tr>
<td>pTH2 7-DLH</td>
<td>SalI</td>
<td>GCCGCTGAGAATAATGGCGACCATCATCATTTCG</td>
<td>CCGGCTGAGATATGAACTCTTTCTTATGGGATG</td>
</tr>
<tr>
<td>pTH2BN 7-DLGT</td>
<td>XhoI,SpeI</td>
<td>GCCGCTGAGGAGATGCATGATAGTTTTATGGAAGAGG</td>
<td>ACCAGGACTTCTATGAGTTTTATATGGTAATCAAAC</td>
</tr>
<tr>
<td>pTH2 BN 8-HGO</td>
<td>SalI</td>
<td>GCCGCTGAGAATAATGGCGACCATCATCATTTCG</td>
<td>CCGGCTGAGAATAATGGCGACCATCATCATTTCG</td>
</tr>
<tr>
<td>pRT101 mCherry</td>
<td>XhoI,KpnI</td>
<td>GCCGCTGAGGAGATGCATGATAGTTTTATGGAAGAGG</td>
<td>ACCAGGACTTCTATGAGTTTTATATGGTAATCAAAC</td>
</tr>
<tr>
<td>pYeDP60 CYP81Z1</td>
<td>USER</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
</tr>
<tr>
<td>pYeDP60 CYP71AY1</td>
<td>USER</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
</tr>
<tr>
<td>pYeDP60 10/CYP76A26</td>
<td>USER</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
</tr>
<tr>
<td>pYeDP60 7-DLH/CYP72A224</td>
<td>USER</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
</tr>
<tr>
<td>pYeDP60 G8O/CYP76B6</td>
<td>USER</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
<td>GGGCTTAUAATGGGAGTTTCTTTTCTTACAC</td>
</tr>
</tbody>
</table>
Table S3: Relative conversion rates for 8-HGO substrates.
Values are the average of 4 replicates ± standard error. Compounds were prepared as 100x stock solutions in acetone. Reactions proceeded for 5 min at room temperature with 100 µM of specific substrate, 200 µM NAD⁺ and 200 ng enzyme in 50 mM Na-phosphate buffer pH 9. Rates were calculated based on NADH production measured in a spectrophotometer at 340 nm.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Relative conversion rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-OH-geraniol</td>
<td>100 ± 11</td>
</tr>
<tr>
<td>geraniol</td>
<td>75 ± 8</td>
</tr>
<tr>
<td>trans-2-hexenol</td>
<td>72 ± 12</td>
</tr>
<tr>
<td>8-oxogeraniol</td>
<td>53 ± 3</td>
</tr>
<tr>
<td>farnesol (mix of isomers)</td>
<td>48 ± 7</td>
</tr>
<tr>
<td>nerol</td>
<td>43 ± 8</td>
</tr>
<tr>
<td>8-OH-geranial</td>
<td>33 ± 4</td>
</tr>
<tr>
<td>4-isopropylbenzyl alcohol</td>
<td>29 ± 4</td>
</tr>
<tr>
<td>octanol</td>
<td>10 ± 2</td>
</tr>
<tr>
<td>3,7-dimethyl octanol</td>
<td>9 ± 1</td>
</tr>
<tr>
<td>(±)-β-citronellol</td>
<td>8 ± 1</td>
</tr>
<tr>
<td>heptanol</td>
<td>7 ± 1</td>
</tr>
<tr>
<td>cis-4-heptenol</td>
<td>0</td>
</tr>
<tr>
<td>(±)-linalool</td>
<td>0</td>
</tr>
</tbody>
</table>

Table S4: Relative conversion rates for IO substrates.
Values are the average of 3 replicates ± standard error. The different substrates (100 µM) were incubated for 5 minutes at 27 °C with 0.25 µM iridoid oxidase. Ethyl acetate extracts were analyzed by GC-MS (see Methods). Activity was quantified based on peak area of products. Minor conversion of linalool, 8-oxogeraniol, and 8-oxogeranial was observed using higher enzyme concentrations.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Relative conversion rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>iridotrial</td>
<td>100 ± 5</td>
</tr>
<tr>
<td>(±)-β-citronellol</td>
<td>20 ± 4.5</td>
</tr>
<tr>
<td>nerol</td>
<td>2 ± 0.5</td>
</tr>
<tr>
<td>lavandulol</td>
<td>2.5 ± 0.5</td>
</tr>
<tr>
<td>linalool (mix of isomers)</td>
<td>0</td>
</tr>
<tr>
<td>8-oxogeraniol</td>
<td>0</td>
</tr>
<tr>
<td>8-oxogeranial</td>
<td>0</td>
</tr>
<tr>
<td>geraniol</td>
<td>0</td>
</tr>
<tr>
<td>7-deoxyloganic acid</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure S1: Overview of the MIA pathway and the cell-specific localization of the branches leading to vinblastine and vincristine. Solid arrows represent single enzymatic steps, dashed arrows multiple enzymatic steps. The arrow bearing a black circle represents the transport of loganic acid from one type of cells to the other.
Figure S2: Evaluation of the expression of P450 candidates in transformed yeast microsomes. Differential absorbance of the CO-saturated-reduced versus reduced microsomes was recorded between 400 and 500 nm and cytochrome P450 concentration was determined according to X. x-axis: wavelength in nm. Iridoid oxidase (IO, CYP76A26, Caros003676), 7-deoxyloganic acid hydroxylase (7-DLH, CYP72A224, Caros005234), Geraniol-8-oxidase (G8O, CYP76B6, Caros006766), CYP81Z1 (Caros001222), CYP71AY1 (Caros018961), CYP71AY2 (Caros007986), CYP81Q32 (Caros003164).
**Figure S3:** Kinetics of the iridoid oxidase (IO) reaction. Lineweaver–Burk plot of the reaction rates measured by product formation. The data are means ± standard errors of three replicates.

**Figure S4:** Activity of iridoid oxidase (IO; CYP76A26) with different substrates. Substrates (100 µM) were incubated for 20 min at 27 °C with 23 µM of CYP76A26 in absence (top) or in presence (bottom) of NADPH. Samples were then extracted with ethyl acetate and analyzed on GC-FID. s: substrate peak(s); p: peak(s) of product(s).
Figure S5: Kinetics of the 7-deoxyloganetic acid-O-glucosyltransferase (7-DLGT) reaction. Lineweaver-Burk plot of the reaction rates measured by product formation. The data are means ± standard errors of three replicates.

Figure S6: Kinetics of 7-deoxyloganic acid hydroxylase (7-DLH) reaction. Lineweaver-Burk plot of the reaction rates measured by product formation. The data are means ± standard errors of three replicates.
Figure S7: Evaluation of 7-deoxyloganic acid conversion into loganic acid by 7-DLH in *N. benthamiana* leaf-disc assay.

Discs from leaves agro-infiltrated with a binary vector containing the 7-deoxyloganic acid hydroxylase (7-DLH; CYP72A224) sequence were excised 5 days post-infiltration and incubated for 3 hours on buffer containing 7-deoxyloganic acid. A leaf methanol extract was analyzed on UPLC-MS. Multiple reaction monitoring in positive mode with the transition 215.1>108.9 is shown. EV: extracts of discs agro-transfected with the empty vector incubated with 7-deoxyloganic acid.
Discovery and reconstitution of the secoiridoid pathway from *Catharanthus roseus*

**Figure S8:** Activity of CYP72A224 (7-DLH) and CYP72A1 (SLS) with 7-deoxyloganic acid and loganin. Microsomes from yeast expressing CYP72A224 or CYP72A1 were tested for activity with 7-deoxyloganic acid and loganin. No 7-deoxyloganic acid conversion was observed with SLS (a). No loganin conversion was observed with 7-DLH (b). Identity of the SLS products was determined using MS and authentic secologanin (c). Product 1 corresponds to an oxygenated derivative of secologanin \((m = 404.13)\), product 2 corresponds to reference secologanin \((m = 388.14)\).

**Figure S9:** Subcellular localization of secoiridoid pathway enzymes. *C. roseus* cells were transiently co-transformed with plasmids expressing pathway enzymes fused to GFP (green; left) and mCherry organelle markers (red; middle). Co-localization of the fluorescent signals appears in the merged pictures in yellow (right). Iridoid oxidase (IO) and 7-deoxyloganic acid hydroxylase (7-DLH) were co-transformed with ER marker and 8-hydroxygeraniol oxidoreductase (8-HGO) and 7-deoxyloganetic acid-\(O\)-glucosyltransferase (7-DLGT) with cytosol/nucleus marker. Scale bars = 10 µM.
**Figure S10:** Masses and quantitative changes of compounds detected upon step-wise reconstitution of the secologanin pathway in *N. benthamiana*. Gene combinations are as listed in Fig. 8A. Heatmap shows relative mass intensity changes. 1. F, formic acid adduct; A, acetyl group adduct; M, malonyl group adduct; number 2, dimer; H, hexose, P, pentose. NI: not identified.
Figure S11: Modified pathway intermediates observed in *N. benthamiana*. 
Figure S12: Iridoid oxidase (IO) is an essential component of the pathway. LC-MS analysis on selected mass 359 (7-deoxyloganic acid; 7-DLA) and 401 (acetylated 7-DLA) of *N. benthamiana* leaves infiltrated with 7-DLA or 8-oxogeranial co-agro-infiltrated with gene combination *IS+IO+7-DLGT* or *IS+7-DLGT*. 
Chapter 6

General discussion
1. Secondary metabolites as targets for metabolic engineering

Many of the plant secondary metabolites possess a particular biochemical structure that can give them important physiological properties and which make them of interest for a broad range of industrial and medical applications. Thus, secondary metabolites are or may be used as pharmaceuticals, as replacement for chemicals currently produced from fossil fuels, as insecticides and as aroma ingredients for food and cosmetics. However, due to the often low natural concentration of secondary metabolites and their complex structure which makes chemical synthesis very difficult, production of many of the useful secondary metabolites cannot meet the industrial demand. For some complex secondary metabolites metabolic engineering in heterologous hosts may be a promising alternative to improve production. In the last two decades we have entered the ‘post-genomic’ era in which a staggering information about genes, proteins and metabolites can be acquired of any organism. For plant secondary metabolism this provide us a possibility to shift our focus from manipulation of individual genes to manipulation of entire pathways by metabolic engineering. In this thesis I showed that we can now design and engineer the key biosynthesis pathway to (seco-)iridoid, which are key precursors of the anticancer agents vinblastine and vincristine, and express this entire precursor pathway, involving 11 engineered steps, in a heterologous plant host. In this process, we not only developed fundamental knowledge about individual enzymatic steps, but also about the enabling technologies required for the rational engineering and processing of plants and plant cells. Still, the path of our engineering achievement has not been without problems and is still met by limitations. In the following sections I discuss these problems, and their potential solutions, that so far limit the engineering possibilities of the (seco-)iridoid pathway and other potentially interesting metabolic pathways.

2. Redundancy in candidate genes

In this thesis we describe the engineering of the (seco-)iridoid pathway for which at the onset of this thesis still several enzymatic steps had not been elucidated yet. We used advanced ‘omics’ data mining tools by integrating analysis of transcriptomics, proteomics and metabolomics to predict gene function and select candidate genes for the missing steps in the pathway. This concept is already a well proven approach and has been successfully applied in many studies on the elucidation of isoprenoid (Wille et al., 2004), aliphatic glucosinolate (Hirai, 2009; Hirai et al., 2007) and brassinosteroid (Lisso et al., 2005) pathways. In the present study, we used RNA sequence data generated from *C. roseus* suspension cells and shoots treated with the plant hormone methyl jasmonate (MeJA) (Van Moerkercke et al., 2013) and, in addition, RNA sequence data generated
from cell suspensions overexpressing either ORCA2 or ORCA3, transcription factors that regulate the expression of loganic acid methyltransferase (LAMT), secologanin synthase (SLS) and several other genes in the TIA biosynthesis pathway (van der Fits and Memelink, 2000). The cell suspension cells show a specific profile of the target compound with which the expression profile of known and new candidate genes of the pathway could be matched. Based on the putative biosynthesis pathway for the (seco-) iridoids, combined with predicted enzyme activities catalysing the missing steps, the sequence dataset was screened for genes that may encode the missing enzymes of the pathway. The expression profile of these genes was matched with that of the product accumulation in the samples and the expression profile of known genes of the pathway like geraniol synthase (GES), geraniol 8-oxidase (G8O), LAMT and SLS. This proved to be an excellent way for finding new genes of the biosynthesis pathway. However, during this study we also encountered some pitfalls and limitations to this approach. First of all, for many of the missing steps in the pathway the list of candidate genes initially was quite long, especially for those enzymatic steps thought to be performed by cytochrome P450 enzymes. For example, at least six candidate P450 genes were selected as iridoid oxidase (IO) and tested in yeast by one of the partners in the EU-project. Interestingly, these P450 genes were of four different subfamilies of the extensive P450 gene family in C. roseus. Eventually only one of them was functionally verified as a real iridoid oxidase (Chapter 5). So, although the expression profile of the other P450 enzymes did match that of the target pathway, they could not be connected to this pathway and therefore may link to another biosynthesis pathway which apparently responds to the same regulation as the (seco-)iridoid pathway. Alternatively, these enzymes are related to the pathway and may be involved in as yet unidentified modifications of intermediates of the pathway, possibly further downstream.

We also noted that not all genes belonging to the same pathway do necessarily show the same co-expression profile. For example, in this study, GES and G8O were shown not to be responsive to ORCA2 and ORCA3 overexpression, while downstream genes LAMT and SLS were. It could be that the products of GES and G8O are also involved in other biological processes which are constitutively active and that steps of LAMT and SLS are more dedicated to TIA production. Alternatively, or in addition, it was shown that the C. roseus GES and G8O are expressed in the internal phloem-associated parenchyma cells, while the genes for LAMT and SLS are expressed in epidermal cells (Oudin et al., 2007; Simkin et al., 2012). This difference in tissue localization may also be part of the reason why there is a differential regulation of the GES/G8O and LAMT/SLS genes, which nevertheless all function in the same pathway.
3. Stable or transient reconstitution of the (seco-)iridoid pathway in plants

Most of the interesting secondary metabolites are synthesized through multiple enzymatic steps, which may pose a problem for reconstitution of the entire pathway in a heterologous host. Reconstitution of multistep metabolic pathways has been achieved for the artemisinin pathway in stable transformed tobacco (Farhi et al., 2011) and through transient expression in *N. benthamiana* (Ting et al., 2013; van Herpen et al., 2010). For the indole pathway of *C. roseus*, reconstitution of the indole alkaloid branch in stable transformed tobacco has been achieved (Songstad et al., 1990). Reconstitution of a multistep pathway in stable transformed plants requires co-transformation of multiple expression constructs (e.g. by co-bombardment and use of multiple selection markers) or multiple genes need to be combined in a single expression construct, either with multiple independent genes or by making use of a so-called 2A construct in which multiple genes are under a single promoter and separated by 2A sequences which allows for ribosome skipping and re-initiation of translation on a single mRNA (van Herpen et al., 2010). A 2A expression construct was used in the reconstitution of the artemisinin pathway by combining 3 genes in one construct (amorphadiene synthase, 3-hydroxy-3-methylglutaryl-CoA reductase and farnesyl diphosphate synthase) (van Herpen et al., 2010). Making a multi-gene construct under the control of the same promoter is only useful for a limited set of genes as the efficiency of translation of the downstream genes decreases with the number of re-initiation steps. In this project stable tobacco transformants were made by one of our partners, using co-bombardment of expression constructs. Analysis of the transgenic plants showed, however, that with high frequency the co-transformed genes were not expressed. A problem of a step-wise reconstitution of a full pathway in stable transformed plants may be that expression of a partial pathway may lead to the production of a potential toxic compound and thus during the regeneration of plants there would be a selection for no or low expressers of the pathway. Another problem of co-expressing multiple genes under the same promoter in plants is that of promoter silencing.

In this study the target pathway has in total twelve genes, and it has not been tested yet whether co-bombardment and regeneration of stable transformed plants, expressing the full pathway with twelve genes is feasible. In future efforts, stable transformed plants may be constructed with the pathway up to the final enzyme, but lacking the first one or two enzymatic steps. This would render the pathway inactive, preventing selection against high expression of these downstream genes. The first two genes of the pathway may be transformed independently and eventually the whole pathway can be reconstituted by making crosses between the plants with the two parts of the pathway. However, we obtained some evidence that expression of a partial pathway
may be detrimental for plants from the transient expression studies in *N. benthamiana*. Co-expression of GES and G8O resulted in early leaf necrosis, indication that this partial pathway results in the accumulation of a toxic compound. When 8-hydroxygeraniol oxidoreductase (8-HGO) and iridoid oxidase (IO) (the next two steps of the pathway) were co-expressed with GES and G8O these necrotic symptoms were alleviated, indicating conversion of the toxic intermediate and formation of a product that was apparently less toxic to the plants.

In the transient expression assays in *N. benthamiana* leaves we used the viral gene silencing inhibitor P19 to limit gene-silencing effects that may be caused by the co-expression of multiple genes (Lakatos et al., 2004). For reconstitution of the partial pathway, seven genes responsible for the first half pathway up to the biosynthesis of 7-deoxyloganic acid and P19 were successfully expressed in *N. benthamiana*, resulting in the (low level) accumulation of free 7-deoxyloganic acid and acetylated 7-deoxyloganic acid. The biggest problem for product accumulation in this system was therefore not the expression of the genes, but the low flux through the pathway likely also caused by the diversion from the pathway at multiple points into modified (e.g. acetylated, glycosylated, malonylated, etc.) products. For whole pathway reconstitution (10 genes in total), this meant that no end product could be detected. However, by infiltrating pathway intermediates (e.g. iridodial) we could demonstrate that the full pathway was active in the leaves (Chapter 5). A big challenge for the future will therefore be to control the side reactions, which drain the flux from the main target pathway, in order to increase desired product yield.

### 4. Selection of engineering host

At the beginning of this study, two different plant host were used to evaluate engineering of the first step, geraniol production, by stable transformation: VoGES was stably transformed into *Arabidopsis thaliana* and *Nicotiana. tobaccum*. In *N. tobaccum* leaves expressing VoGES, we detected free geraniol in the headspace and also geraniol derived conjugation products in leaf extracts (Chapter 2). However, the *Arabidopsis* leaves expressing VoGES did not emit any geraniol and only very low levels of geraniol derived products were detected (data not shown). There are several possible explanations that could explain this difference between plant hosts: It could be that tobacco is better able to deal with the product geraniol than *Arabidopsis* and that in *Arabidopsis* a selection against high expression of VoGES has occurred. Interesting is also that geraniol processing by endogenous enzymes differs between tobacco and Arabidopsis. For example, in *N. tobaccum* leaves there are ~20 different geraniol derived products, while in *Arabidopsis* only one type of geraniol derived glycoside is present. Indeed,
this difference in processing of geraniol may relate to the difference in which these two plants can tolerate the ectopic production of geraniol. Alternatively, the difference may be caused by a difference in precursor availability in tobacco and Arabidopsis. As we showed in Chapter 4, when precursor levels are boosted by overexpression of a geraniol diphosphate synthase (GPPS), geraniol production can indeed dramatically be increased. This indicates that precursor availability is one of the variables that determines overall production and this may differ between plant hosts. Conversion of geraniol to other products has been demonstrated in many different hosts. For instance, expression of Ocimum basilicum geraniol synthase (ObGES) in yeast, E. coli, grape and tobacco (Niinemets et al., 2002) and tomato (Davidovich-Rikanati et al., 2007) resulted, as expected, in all hosts in formation of the main product geraniol, but also other, minor, products such as linalool, citronellol, nerol, α-terpineol, etc., depending on the host. Similarly, the linalool synthase from Clarkia breweri was expressed in petunia, carnation and tomato. Transgenic petunia plants expressing the Clarkia gene produced linalool, but all the product was conjugated to form the non-volatile linalyl glucoside (Lücker et al., 2001). In a separate study, the linalool synthase was expressed in carnation, and the transformed plants produced and emitted linalool and oxidized linalool derivatives. Finally, when the same linalool synthase was expressed in transgenic tomato, this led to the production of linalool and 8-hydroxylinalool in ripening fruits (Lewinsohn et al., 2001). Combined, these different experiments show that secondary metabolites may be produced ectopically in different heterologous systems through metabolic engineering but that the metabolism of the ectopically produced product by endogenous enzymes of the host is difficult to predict. Depending on the actual purpose of the metabolic engineering (improving agronomic traits, altering fragrance or flavor profiles or production of a target compound) a host plant should therefore be carefully selected. At present, experimental systems are still limited to a few selected plant species, mostly based on their easy transformation, while for high production of terpenes it may be worthwhile to screen additional potential host plant species. Because each individual species may have its own downsides, also control of host metabolism may be required for full optimization of target terpene production in plants (see below).

5. Metabolomics analysis

Untargeted metabolic profiling was used in this study to detect metabolic changes for each step of the (seco-)iridoid pathway reconstitution in N. benthamiana agro-infiltration experiments. This approach allowed quick gene activity identification and helped to narrow down the gene candidate list in a relatively short time. However, this approach also has some pitfalls:
1. In this study, we used MetAlign for pre-processing and comparison of full scan nominal or accurate mass data from GC-MS and LC-MS. Subsequently, all the identified masses in the replicate samples were subjected to MSClust software to remove redundancy in the data set by combining multiple isotopic forms and adducts of the same compound (LC-MS) and multiple fragments of the mass spectrum of one compound (GC-MS) into one peak area. Comparison of samples of *N. benthamiana* infiltrated with resolved pathway genes plus gene candidates with samples infiltrated only with the resolved genes allowed us to quickly identify new masses (presumably originating from the newly introduced enzyme activity) but also whether the mass signal of the precursor compound was decreased. However, when there are many side reactions the signal of the mass of the expected compound may be so low that it gets lost in the noise and may be filtered out in the stringent procedure to select differential masses. In such case it is helpful to have an authentic standard of the expected product to specifically screen the samples for the mass signal at the retention time at which the compound is expected. However, when the compound is fully converted by endogenous enzymes this will result in a novel mass and retention time and the signal may be lost. To help find the *in planta* converted compounds, *N. benthamiana* leaves may be infiltrated with the expected products as the bio-converted infiltrated substrate will have a higher signal intensity and thus may be easier to detect in subsequent analysis. We used this approach in our studies when we tried to identify the products of the 7-deoxyloganic acid pathway. Initially the end product was not detected, but when we infiltrated the authentic standard of 7-deoxyloganic acid into *N. benthamiana* leaves, we could detect the bioconversion of this compound to acetylated 7-deoxyloganic acid. Once the mass and retention time of the bio-converted product were known, targeted analysis of the leaf samples with the infiltrated pathway for the bio-converted product by mass and retention time showed it was indeed formed.

2. The untargeted metabolomics analysis generated many masses that seem to be the consequence of our pathway reconstitution as they were dependent on the presence of geraniol synthase or matched with the bioconversion products of standards infiltrated into *N. benthamiana* leaves. However, the identification of these compounds in most cases was not possible and only putative compound structures could be given, as actual standards for these different products were not available. Some hints of their identity can be derived from MS-MS experiments which may render some specific fragment masses (e.g. of sugars) next to the expected target compound mass, which was mostly interpreted as glycosylation of the target compound. However, most conversions involved multiple glycosylation steps (up to three hexoses and malonylation). Identification of such products was also done by treatment of the samples with viscozyme (multi-enzyme complex containing a wide range of carbohydrases, including arabanase, cellulase, β-glucanase, hemicellulase, and xylanase), which for the bulk of glycosylated
products resulted in the release of the free aglycon (for example geraniol) except for those glycosylated products which were capped with a malonyl group, in which case the viscozyme treatment could not release the free compound. Actual identification of the unknown compounds could be by NMR but since NMR is far less sensitive than MS this must be limited to those compounds that are present in high levels.

3. For some steps in the pathway reconstitution experiments we did not find any new products in the differential metabolite analysis. For example, after addition of iridoid oxidase (IO) to the previous engineered steps, the metabolite profile did not change: no new product was detected while also the signal of the precursor molecule was not changed. Nevertheless, we could show that IO was indispensable for the pathway activity as the activity of the next step in the pathway, catalyzed by 7-deoxyloganetic acid glucosyl transferase (7-DLG), depended on the presence of IO. We attribute this to the high reactivity of the iridotrial produced by IO, which can therefore not be detected, even though it is produced. Thus, results of metabolic profiling has to be interpreted with caution and seemingly negative results of screening certain enzyme activities should be re-evaluated when subsequent steps of the pathway become available.

6. Boosting (mono)terpene production

In this study, we have reconstituted all the steps of the (seco-)iridoid pathway in a heterologous plant host, using transient expression with each gene under a strong constitutive promoter and with high gene dosage. Nevertheless, the production level that was obtained in this transient expression system does not reach industrial requirements. However, there are still several options to improve end product accumulation such as elimination of competing side reactions, improved availability of precursors, the exploitation of multiple cellular compartments and optimization of the host plant performance.

6.1 Elimination of side reactions

In my thesis I show that the bulk of the products formed upon pathway expression in *N. benthamiana* are in unwanted side products. If these side reactions can be eliminated and all substrate can be converted to the target end product, the current transient expression system has the potential to yield 50 ug/g FW of (seco-)iridoid based on the total production of geraniol plus geraniol derived products obtained in Chapter 2 (but this can be likely further enhanced, as discussed below). The large number of side products is mainly due to the fact that the intermediates in the pathway can potentially serve as substrates for endogenous enzymes. For example, when we overexpress *GES* in *N. benthamiana*, the product geraniol was further oxidized to hydroxygeraniol.
by endogenous enzymes. Another cause for unwanted side-products, is that some enzymes of the pathway are also not very specific in terms of substrate they can act on. For example, in in vitro assays it was shown that iridoid synthase (IS) can use at least 8 analogs of 8-oxogeranial as substrates (Geu-Flores et al., 2012). In my PhD thesis work, upon expression in N. benthamiana, IS indeed reacted with many of these products (for example, 8-oxogeranial, 6,7-dihydro-8-oxogeranial, 8-oxogeraniol, 8-oxocitronellal, etc.) which were produced upon expression of the preceding steps of the pathway (GES, G8O and 8-HGO) likely also as a result of endogenous enzyme activity. Only the conversion of 8-oxogeranial by IS results in iridodial which is the specific substrate for the next step in the pathway, catalyzed by IO. Thus, a better flux through the pathway may be obtained by using less promiscuous enzymes. Interestingly, the multitude of side reactions occurring in N. benthamiana do not seem to occur in C. roseus, which raises the question how the highly-controlled flux through the pathway is regulated in the latter plant species. One possibility could be that in C. roseus the enzymes assemble into a metabolon in which the pathway intermediates are protected from side reactions. In principle metabolon formation could also occur in a heterologous host, such as N. benthamiana, but apparently that is not happening since we get many side products. Could it be that one or more of the genes that we cloned is not the real pathway gene? Is IS not the real iridoid synthase and/or are there other oxidoreductases which have better substrate specificity than the one used in this study? And would these improve the flux through the pathway upon transient expression in N. benthamiana?

We found that the amount of many of the intermediate products of (side reactions of) the pathway did not or hardly decrease when the next step of the pathway was added. For instance hexosyl carboxy geranic acid was produced upon transient expression of GES, G8O and 8-HGO (Chapter 5). When the next genes (IS and IO) were co-expressed with GES, G8O and 8-HGO, we expected the hexosyl carboxy geranic acid would dispear; however the product level only slightly decreased. One possibility is that if more A. tumefaciens strains are added to the mix that is infiltrated, at some point there are cells that are not infected with all the gene constructs anymore. When the number of different A. tumefaciens strains becomes too high, the percentage of cells with partial pathway expression increases and this could explain the presence of pathway intermediates. If this is indeed the cause of intermediate build-up, this is expected to be a function of the number of different A. tumefaciens strains that are used in the transient expression experiments. Alternatively, or in addition, the presence of pathway intermediates could be the result of different turnover of enzymes of the pathway. The total flux through the pathway is dependent on the activity of all the enzymes of the pathway. The efficiency of this flux could be affected when not all enzymes of the pathway have the same lifespan. If some of the down-stream enzymes have a higher turnover rate, their activity would decrease faster than that of the up-stream enzymes, resulting in an accumulation of
intermediates that are further converted by detoxifying enzymes of the plant host to the conjugates we detected. Further studies on the turnover rate of different enzymes and investigation of the efficiency of transient expression of all expression constructs as function of the \textit{A. tumefaciens} dosage will be needed to get the answers to these questions.

The total flux through the pathway also depends on how efficient the product from one enzyme can move to the next enzyme of the pathway. In the (seco-)iridoid pathway, biosynthesis starts in the plastids with the formation of GPP and geraniol by GPPS and GES, respectively. Geraniol then has to move from the plastids to the cytosolic face of the ER, where G8O is localized and where it is converted to 8-oxo-geraniol. This needs to move to the cytosol where 8-HGO converts it to 8-oxo-geranial. After a number of additional steps in the cytosol (IS, IO, 7-DLGT, LAMT and SLS) and on the cytosolic face of the ER (7-DLH), the last step of the pathway takes place in the vacuole (strictosidine synthase (STR)). For this last step the substrates secologanin and tryptamine need to be imported into the vacuole, where the STR enzyme is targeted. Apparently, \textit{N. benthamiana} does have transporters for import of these compounds into the vacuole as expression of the pathway genes was sufficient to produce strictosidine (after injection of the pathway intermediate iridodial). However, it could be that the influx into the vacuole can be further increased by also expressing specific transporter proteins from \textit{C. roseus}. Recently, two different genes have been isolated from \textit{C. roseus} encoding membrane transporter proteins which were shown to be involved in transport of loganic acid, loganin, and secologanin (Jacob Pollier, personal communication). It will be interesting to see what happens if these transporters are co-expressed, with the pathway biosynthesis genes, in \textit{N. benthamiana}.

\textbf{6.2 Boosting precursor level}

\textit{6.2.1 Selecting a geranyl diphosphate synthase for boosting GPP production}

The direct common precursor for monoterpen biosynthesis is GPP and this is produced by the condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) by geranyl diphosphate synthase (GPPS). In this study we screened different \textit{GPPS} genes that may help improve precursor influx into the pathway and thus boost accumulation of the target monoterpen (seco-)iridoid. A \textit{GPPS} from Arabidopsis (\textit{AtGPPS}) (Bouvier et al., 2000) and two \textit{GPPS}s from \textit{Picea abies} (\textit{PaGPPS1} and \textit{PaGPPS2}) (Schmidt and Gershenzon, 2008; Schmidt et al., 2010) have been described in the literature for their \textit{in vitro} GPPS activity. We tested these three genes \textit{in planta} for their capacity to boost geraniol production upon co-expression with \textit{VoGES}. Analysis of the production of geraniol showed that only co-expression with \textit{PaGPPS1} increased monoterpen production (Chapter 4). This demonstrates that
apparently the characterization of an enzyme reaction in vitro is not always predictive for the in vivo activity. It could be that in the different terpene biosynthesis pathways in plants, there is a certain product channeling, e.g. by specific association of a GPPS with a (mono)terpene synthase. Such product channeling is supported by the finding of dwarfed tomato plants when a Solanum lycopersicum GPPS was silenced (Van Schie et al., 2007). It turned out that only gibberellin production was significantly reduced suggesting that this SIGPPS is related to the production of gibberellins, which are diterpenes and hence are derived from geranylgeranyl diphosphate, GGPP. This suggest the produced GPP by SIGPPS are channeled to form diterpenes instead of monoterpenes as GGPP can be formed by condensing one unit of GPP and two units of IPP. Similarly, gene silencing of GGPP synthase from Nicotiana attenua (NaGGPPS) reduced only 17-hydroxygeranyllinalool diterpenoid glycosides while carotenoids and chlorophyll contents were not affected, demonstrating that NaGGPPS supplies substrates for GGPP biosynthesis for 17-hydroxygeranyllinalool diterpenoid glycosides, but not for phytoene or phytol (Jassbi et al., 2008).

It was shown that the PaGPPS1 that I used in my thesis may have a dual function and can produce both GGPP (major product) and GPP (only miner product) in vitro (Schmidt et al., 2010). GPPS enzymes are often not functional as monomer. They may be active as homodimers (Hsiao et al., 2008; Rai et al., 2013; Schmidt and Gershenzon, 2008), but often also as heterodimer, consisting of a large and small subunit (Burke et al., 1999; Wang and Dixon, 2009). Recently, three GPPS genes from C. roseus were characterized: two genes form the proteins of a heterodimer enzyme (CrGPPS.LSU; CrGPPS.SSU) and one gene encodes the protein for a homodimer GPPS (CrGPPS) (Rai et al., 2013). The GPP synthase large subunit normally shows a high level of amino acid sequence identity with GGPP synthases of plant origin and displays conserved catalytic amino acid residues essential for catalysis. The small subunit lacks these aspartate-rich functional motifs and on its own is inactive, but it might be capable of influencing the chain length specificity (Burke and Croteau, 2002; Tholl et al., 2004). The CrGPPS. LSU as homodimer is a bifunctional enzyme producing both GPP and geranyl geranyl diphosphate (GGPP), while CrGPPS.SSU by itself is inactive. However, co-expression of both subunits in Escherichia coli resulted in heteromeric enzyme with enhanced activity producing only GPP. CrGPPS is a homodimeric enzyme which only produces GPP. Because the PaGPPS1 sequence does not resemble the sequence of a GGPPS (indicative of a heterodimer GPPS large subunit), presumably PaGPPS1 is active as homodimer in N. benthamiana. Interestingly, localization studies indicated that the homodimer CrGPPS is mitochondrial, suggesting that the mitochondria can produce GPP (Rai et al., 2013). This would match with our GES localization study that showed that geraniol production occurs with mitochondrial targeted GES, suggesting that GPP is available in the plastids (Chapter 4; also see below). Moreover, transient overexpression of the
plastidal GPPS.SSU in *C. roseus* leaves resulted in increased levels of vindoline, the immediate monomeric precursor of vinblastine and vincristine, indicating that the expression levels of the GPPS.SSU, which gives specificity to the heterodimer GPPS is limiting for GPP biosynthesis in *C. roseus*.

The primary monoterpenene products are volatile and may be emitted into the headspace of the plant (Nogués and Loreto, 2013). However, this loss of monoterpenene molecules is only minor when the primary monoterpenene product is further oxidised and/or glycosylated or stored in specialised structures such as glandular trichomes. For instance, in peppermint only 1% of the monoterpenene production (on a monthly basis) is lost to the headspace, the rest is sequestered in the plant (Gershenzon et al., 2000). When monoterpenes are ectopically produced in novel hosts the modified terpenes are often stored in plant cells as sugar-conjugates (Lücker et al., 2001; Mateo and Jiménez, 2000). In our experiments with VoGES expression in *N. benthamina* with different subcellular targeting the headspace analysis showed that the emission of free geraniol was about the same for transient expression of cytosolic, mitochondrial and plastid-targeted VoGES. However, the levels of geraniol-derived glycosides with plastid-targeted VoGES were 3-fold higher than with cytosolic VoGES and 6-fold higher than with mitochondrial VoGES (Chapter 5), showing that there are substantial difference in GPP pools in these different subcellular compartments.

6.2.2 IPP and DMAPP availability in different subcellular compartment
IPP and its allylic isomer DMAPP are the basic five carbon units for terpenoids. Condensation of one unit of IPP and one unit of DMAPP results in the formation of GPP, the precursor for monoterpenene biosynthesis. Many studies have already demonstrated that the plastids can directly synthesize both IPP and DMAPP through the MEP pathway (Bick and Lange, 2003; Rodríguez-Concepción and Boronat, 2002). However, there is no consent on the origin of IPP and DMAPP in the cytosol. Some studies point out that the MVA pathway produces only IPP (Bick and Lange, 2003; Rodríguez-Concepción and Boronat, 2002) while others suggest the IPP is provided by the MEP pathway by unidirectional trafficking of IPP from the plastids to cytosol (Dudareva et al., 2005). It is unclear whether in plants DMAPP is produced in the cytosol. Since the isopentenyl diphosphate isomerases (Guirimand et al., 2012; Phillips et al., 2008), FPS (Martín et al., 2007), GGPPS (Okada et al., 2000) and GPPS (Rai et al., 2013) from different plant species were reported to be also located in the mitochondria, the mitochondrial compartment has the potential to produce different types of terpenes although precursor input for the isopentenyl diphosphate isomerases and downstream enzymes may be from the plastidial MEP and/or cytosolic MVA pathway.
To explore monoterpene production capacity for plastid, cytosol and mitochondria in relation to local GPP production, \( PaGPPS1 \) and \( VoGES \) were provided with targeting signals to either plastids, cytosol or mitochondria. For all three compartments the targeted \( PaGPPS1 \) was able to boost geraniol production of \( GES \) targeted to the same compartment. This suggests that in each subcellular compartment the local IPP/DMAPP pool was not fully utilized by the endogenous local GPPS activity and flux was enhanced by the ectopically expressed \( PaGPPS1 \). We also boosted GPP production in one compartment while expressing \( VoGES \) in the neighbouring compartment. This resulted in some interesting insights about GPP exchange between cytosol, mitochondria and plastids: most striking is the lack of exchange from cytosol to the plastids, while there is efficient exchange between cytosol and mitochondria and between mitochondria and plastids. At present not much is known about this exchange of compounds between the different organelles but both plastids and mitochondria can make tubular structures which extend from the organelles. For plastids these membrane tube like structures are called stromules while for mitochondria these are named matrixules (Zottini, 2013). Stromules have been shown to intercalate with the ER membrane, but at present it is not known whether mitochondrial matrixules interact with the plastidal stromules and whether or how metabolites may be exchanged between these structures. Apparently from all three subcellular compartments there is an efficient exchange of geraniol to the ER where the P450 enzyme of the next step in the pathway is located since several hydroxylated and carboxylated geraniol glycosides were found. Maximum production of geraniol may be obtained by simultaneous expression of \( GPPS \) and \( GES \) in all three compartments but since the combined capacity of cytosol and mitochondria is less than half that of the capacity of the plastids, this may boost production only less than two-fold compared with production in the plastids alone. Although this may deplete the pool of IPP for other activities in the cell, in a transient expression system this may be of limited consequences if the leaves would be harvested after one week anyway.

Boosting the input of the monoterpenoid pathway may still be of limited effect for end product accumulation when there are many side reactions draining the flux from the pathway. For instance, six enzymatic steps are involved in the production 7-deoxyloganic acid (seco-)iridoid from geraniol. We reconstituted this pathway with the six genes with and without boosting with \( PaGPPS1 \). When \( GPPS \) is co-expressed with plastidial \( VoGES \), geraniol production could be boosted 118-fold (Chapter 4), but in the 6-step pathway the production level of 7-deoxyloganic acid was only increased 3-fold (data not shown). Elimination of side reactions catalyzed by endogenous enzymes by knocking-down these enzymes or improved pathway flux through (artificial) metabolone formation will therefore likely provide a further increase in target compound production.
6.3 Optimizing the environment for production

Monoterpenes production has a positive correlation with photosynthesis (Maffei and Codignola, 1990; Ormeño et al., 2009; Porcar-Castell et al., 2009), since this process delivers the carbon, ATP and NADPH input required for their biosynthesis (Niinemets et al., 2002). Most genes of the plastidal MEP pathway are light-regulated and it is not unlikely that continuous light could boost the plastidial isoprenoid production capacity. For instance, in root cultures of *Artemisia annua* continuous light resulted in a substantial increase in transcript levels of deoxy-D-xylulose-5-phosphate synthase compared with dark-grown roots (Souret et al., 2002). We tested the effects of the growth environment on geraniol production with our transgenic *Arabidopsis* lines expressing *P-VoGES*. Plants were grown either under a diurnal photoperiod of 12 h light or 16 h light and the geraniol glycoside accumulation in leaves was compared. LC-MS analysis showed that the one third increase in light period (16 vs 12 hr) resulted in a 3-fold higher level of geraniol glycosides compared with plants grown under 12 h light (Figure 1). It remains to be tested if production can be boosted even further when such transgenic plants are grown under continuous light. Also for the transient expression method in *N. benthamiana* it would be of interest to see whether longer light exposure could increase terpenoid end-product accumulation.

Figure 1: Effect of photoperiod on the production of geraniol-related compounds in stable transformed Arabidopsis.
7. Future perspectives

In the SmartCell program we have used a combination of metabolomics, transcriptomics and proteomics data, to identify novel genes of the (seco-)iridoid biosynthesis pathway. With all the biosynthesis genes of the pathway up to strictosidine now being known we were able to reconstitute a fully active pathway in a heterologous host. However, clearly a better channeling of the pathway intermediates is needed to reach more significant end-product levels. Two main steps may be needed to obtain this goal: boosting the input of the pathway and eliminating side reactions. In addition to the technological solutions for these problems such as knocking down competing enzymes, there may also be things we can learn from strictosidine biosynthesis in *C. roseus*: how are side reactions avoided in this plant? Is there some form of product channeling between subsequent enzymes of the pathway and can we improve target product formation in a heterologous host by engineering of some form of enzyme scaffold which allows for efficient product channeling? Is the complicated transport of intermediates in the pathway between cells related to production capacity or protection of cells against potential toxic intermediates? Although we can be satisfied with the results that were obtained, there still remain many questions to be answered before economical production levels can be reached in heterologous plant hosts. Nevertheless, the options for such alternative production systems are getting more realistic, also thanks to the knowledge gained in this project.
References


Terpene indole alkaloids (TIAs) are plant-produced secondary metabolites that are very important for humans, since many of them have pharmaceutical properties and are e.g. used as anticancer drugs (for example vinblastine and vincristine). However, for the production of TIAs industry still fully relies on extraction from medicinal plant species. One of the major plant resources for vinblastine and vincristine is *Catharanthus roseus*, which only contains extreme low amounts of these compounds. Total chemical synthesis is difficult and expensive because of the molecular complexity and if chemical synthesis is used it is still based on naturally produced precursors. Strictosidine is the common precursor for all TIAs and is the target molecule that I tried to produce in my PhD project by metabolic engineering of a heterologous plant host. To express the pathway for strictosidine in a heterologous host, the genes encoding the enzymes involved in the biosynthetic pathway must be identified and cloned first. At the onset of my project 6 out of the presumed 12 genes of the pathway in *C. roseus* had not been discovered yet. Therefore, my thesis tells the story of gene discovery and characterization of the metabolic engineering of the strictosidine pathway in heterologous hosts.

At the onset of this project, the geraniol synthase (*GES*) gene, which catalyses the first step of the strictosidine biosynthetic pathway, had not yet been discovered in *C. roseus*. Because tobacco (the heterologous hosts that we used to reconstruct the pathway) does not produce geraniol, I used two *GES* genes from two other plant species (*Valeriana officinalis* and *Lippia dulcis*). I expressed both *GES* genes in *E. coli* and both of them enabled the production of geraniol as single product, using the general precursor geranyl diphosphate (GPP) as substrate (Chapter 2). Monoterpene synthases are generally believed to be located in the plastids because this is thought to be the place where GPP is produced. Subcellular localization studies with the two GES proteins fused to the fluorescent marker GFP showed that VoGES indeed mainly resides in the plastid, while for LdGES, however, the protein was mostly in the cytosol. Surprisingly, the transient expression of *VoGES* and *LdGES* in *N. benthamiana* leaves using agro-infiltration showed that both enzymes had similar *in planta* geraniol synthase activity, suggesting that GPP is equally available for monoterpenoid production in cytosol and plastids. Another interesting fact was that expression of GES in *N. benthamiana* not only resulted in geraniol formation but also in the production of a series of further oxidised and glycosylated geraniol-derived products. This further conversion of geraniol also occurred in stable GES transformants in *Nicotiana tabacum*. Because these stable transformants with either of the two GES genes contained similar amounts of free and conjugated geraniol and geraniol derivatives, only *VoGES* was used for further metabolic engineering.
The next step in the pathway, after GES, is catalysed by geraniol 8-hydroxylase (G8O), and candidate genes for this enzyme were isolated from *C. roseus* (*CrCYP76s*) and *Arabidopsis thaliana* (*AtCYP76s*). To determine which enzyme was most effective for pathway reconstitution, these CYP76 family members were extensively characterised in yeast expression assays and in an *in planta* leaf disc assay (Chapter 3). In the yeast expression assay, AtCYP76C1 did not have geraniol hydroxylase activity, but AtCYP76C4 was identified as a geraniol 8- and 9-hydroxylase. CrCYP76B6 was identified as a highly regio-specific enzyme catalyzing two sequential oxidations of C-8 of geraniol, leading to the formation of 8-oxogeraniol. This enzymatic activity was confirmed in the *in planta* leaf-disc assay. When ViGES and CYP76B6 were co-expressed *in planta* all geraniol was completely converted to the oxidized geraniol product and its derivatives. In contrast, AtCYP76C4 was not as efficient as CYP76B6, and only half of the *in planta* produced geraniol was converted to hydroxygeraniol and its derivatives (Chapter 3). Therefore, in the subsequent reconstitution of the strictosidine pathway, we used *CrCYP76B6*.

In Chapter 4, we tried to identify a geranyl diphosphate synthase (*GPPS*) which may boost the input of substrate into the monoterpene pathway and hence increase final product yield. Three *GPPS* candidate genes (two from *Picea abies*, one from Arabidopsis) were tested for boosting geraniol production by transient overexpression in *N. benthamiana* in combination with ViGES. Only *GPPS1* from *Picea abies* (*PaGPPS1*) showed boosting activity. To assess the boosting capacity of GPP biosynthesis in different subcellular compartments we made *PaGPPS1* expression constructs with plastidic, cytosolic and mitochondrial targeting of the *PaGPPS1* protein. Similarly, we made expression constructs of ViGES targeting the enzyme to the plastids, cytosol or mitochondria. All possible combinations of differentially targeted GES and GPPS were transiently expressed in *N. benthamiana* leaves to determine how local boosting of GPP production affects geraniol production in the same or neighboring compartments. The headspace and leaf extracts were analyzed for geraniol and geraniol-derived products by GC-MS and LC-MS, respectively. I could show that the plastids have the highest potential for geraniol production and that plastids can exchange a small fraction of GPP to the cytosol and mitochondria. There was no indication of GPP exchange from the cytosol to the plastids. Moreover, isoprenyl diphosphate (IPP) use by cytosolic PaGPPS1 reduced plastidial monoterpene production, suggesting a preferred drain of IPP from the plastid when cytosolic IPP levels are low. Finally, I showed that GPP produced in the mitochondria can boost GES activity in the cytosol and plastids. Because GPP produced in the cytosol seems not to be available to the plastids, this suggests a direct exchange of GPP from the mitochondria to the plastids.

During the course of this project, *C. roseus* mRNA sequencing and proteomics data became available to the EU-project consortium. An integrated transcriptomics and
proteomics approach was used by our partners for gene discovery of the missing steps in the pathway. Candidate genes were subsequently characterized in \textit{in vitro} biochemical assays and by me in \textit{in planta} transient expression, resulting in the identification of the last four missing genes, from the strictosidine biosynthesis pathway (Chapter 5). \textit{In situ} hybridization showed a similar cellular localization for 8-hydroxygeraniol oxidoreductase (8-HGO), iridoid oxidase (IO), 7-deoxyloganetic acid glucosyl transferase (7-DLGT), 7-deoxyloganic acid hydroxylase (7-DLH) and geraniol 8-oxidase (G8O), which possibly form the first transcriptional regulon for the production of loganic acid in internal phloem associated parenchyma cells. The subcellular localization of IO, 7-DLH, 8-HGO and 7-DLGT was determined by tagging proteins with GFP. Expression of the fusion proteins in Catharanthus cells indicated that IO and 7-DLH are (as predicted) ER-associated, whereas 8-HGO and 7-DLGT were shown to be soluble proteins found in the cytosol and nucleus. Finally, I expressed the entire pathway transiently in \textit{N. benthamiana}. I could demonstrate that the entire pathway was functional but because of the many side reaction by endogenous \textit{N. benthamiana} enzymes it did require introduction of intermediates into the leaves to accumulate sufficient end product for detection. Thus, for future applications, elimination of competing side reactions (for example through glycosylation) may be needed to channel the carbon flux into the desired end-products.

With this accomplishment we have not only paved the way for the biotechnological production of pharmaceutical valuable strictosidine-derived terpene indole alkaloids, but have also provided novel insight into the flux capacity for terpene precursor biosynthesis, which has broad implications for the production of terpenoid products in plants.
Samenvatting

Terpen indolalkaloiden (TIAs) zijn secundaire metabolieën van planten die belangrijk zijn voor mensen omdat veel van deze stoffen een medicinale werking hebben (bv de antikanker stoffen vinblastine en vincristine). Echter, voor de productie van TIAs wordt nog grotendeels gebruik gemaakt van extractie uit medicinale planten. Eén van de belangrijkste bronnen voor vinblastine en vincristine is Catharanthus roseus, die echter maar heel lage concentraties van deze stoffen bevat. Volledige chemische synthese van deze stoffen is moeilijk en duur vanwege de complexe moleculaire structuur en dus is de productie van deze medicijnen nog steeds gebaseerd op de eelt van Catharanthus. Strictosidine is the algemene precursor van alle TIAs en het doel van mijn onderzoek was de productie van deze stof in een andere plant mogelijk te maken. Om de gehele biosynthese route van strictosidine tot expressie te brengen in een andere plant moeten eerst de genen geïdentificeerd en gekloneerd worden. Bij het begin van het project waren maar 6 van de 12 genen die nodig zijn voor de productie van strictosidine in C. roseus bekend. Dit proefschrift beschrijft de isolatie van de missende genen en de reconstructie van de hele biosynthese route in een heterologe host.

Toen dit project begon was het geraniol synthase (GES) van C. roseus, dat de eerste stap in de strictosidine biosynthese route katalyseert, nog niet bekend. Tabak heeft zelf geen GES en daarom zijn voor de reconstructie van de strictosidine biosynthese route in tabak eerst twee andere GES genen getest (uit Valeriana officinalis en Lippia dulcis). Beide genen zijn eerst in de bacterie Escherichia coli tot expressie gebracht waardoor deze geraniol konden produceren als de precursor GPP werd gegeven (Hoofdstuk 2). Het GES is een zogenaamd monoterpeen synthase en die bevinden zich meestal in de plastiden omdat daar de precursor GPP aanwezig is. Echter, subcellulaire localisatie studies van GES met de fluorescente merker GFP lieten zien dat (zoals verwacht) VoGES weliswaar in de plastiden zit, maar LdGES in het cytoplasma. Vreemd genoeg was echter de activiteit van VoGES en LdGES in planta niet verschillend na transiente expressie in Nicotiana benthamiana, wat aangeeft dat GPP ook beschikbaar is in het cytoplasma. Bovendien konden we aantonen dat geraniol in N. benhtamiana ook omgezet wordt in een reeks andere producten door oxidatie en glycosylering (Hoofdstuk 2). Dezelfde omzettingen van geraniol werden ook waargenomen in stabel met VoGES of LdGES getransformeerde tabak. Omdat er verder geen verschil was tussen VoGES en LdGES is voor de rest van het project alleen met VoGES gewerkt.

De volgende stap na GES in de biosynthese route wordt gekatalyseerd door een cytochrom P450 enzyn, geraniol 8-hydroxylase (GBO). Voor deze stap was er een kandidaat gen bekend uit C. roseus dat we hebben vergeleken met twee mogelijke alternatieve cytochrom P450s uit Arabidopsis. Om te bepalen welk enzyn het meest effectief was zijn deze drie leden van de CYP76 family uitgebreid getest in gist en in
**Samenvatting**

*planta (Hoofdstuk 3).* In de gist expressie assay had AtCYP76C1 niet de juiste geraniol hydroxylase activiteit, maar AtCYP76C4 had wel geraniol 8- en 9-hydroxylase activiteit. Het *C. roseus* enzym, CrCYP76B6, werd geïdentificeerd als een regio-specifieke enzym dat twee opeenvolgende oxidaties van de C8 van geraniol kan uitvoeren, wat leidt tot de vorming van 8-oxygeraniol. Deze enzymactiviteit werd bevestigd in *in planta* bladpons assays. Als VoGES samen met een van de twee cytochrom P450s tot expressie werd gebracht in *N. benthamiana* werd geraniol omgezet naar hydroxygeraniol en derivaten hiervan (Hoofdstuk 3). AtCYP76C4 was slechts half zo efficient als CrCYP76B6 in de omzetting van geraniol. Daarom werd voor de reconstitutie van de strictosidine biosynthese route verder CrCYP76B6 gebruikt.

In *Hoofdstuk 4* hebben we geprobeerd een geranyldifosfaat synthase (GPPS) te identificeren dat gebruikt kan worden om de substraat input van de monoterpeen biosynthese route te stimuleren en zo de opbrengst te vergroten. Drie GPPS kandidaat genen (twee uit *Picea abies* en één uit Arabidopsis) werden getest voor stimulering van geraniol productie door transiente overexpressie in *N. benthamiana* in combinatie met VoGES. Alleen het GPPS1 van *P. abies* (PaGPPS1) was in staat om de productie te verhogen.

Om de GPP biosynthese capaciteit van de verschillende sub-cellulaire compartimenten te testen zijn GaGPPS1 en VoGES beide voorzien van plastiden-, cytoplasma- en mitochondriën-importsignalen. Al de combinaties van verschillend gelokaliseerde VoGES en GPPS zijn vervolgens tot expressie gebracht in *N. benthamiana* om te bepalen hoe lokale GPP synthese de productie in hetzelfde of nabijgelegen compartiment kan stimuleren. De headspace en bladextracten werden geanalyseerd voor geraniol en geraniol-afgeleide producten met behulp van respectievelijk GCMS en LCMS. Ik kon aantonen dat de plastiden de grootste capaciteit hebben voor geraniol productie en dat plastiden een geringe hoeveelheid GPP naar cytoplasma en mitochondriën kunnen uitwisselen. Er was geen aanwijzing voor GPP uitwisseling van het cytoplasma naar de plastiden. Het was zelfs zo dat gebruik van isopenetyl difosfaat (IPP) door cytosolisch PaGPPS1 de monoterpeen productie in de plastiden verlaagde, wat suggereert dat er een voorkeur is voor transplantisch IPP van de plastiden naar het cytoplasma als daar het IPP niveau laag is. Tenslotte heb ik laten zien dat GPP in de mitochondriën de productie van geraniol in het cytoplasma en in de plastiden kan stimuleren. Omdat GPP van het cytoplasma niet naar de plastiden gaat duidt dit er op dat er een directe uitwisseling moet zijn van GPP tussen de mitochondriën en de plastiden.

Gedurende dit project kwamen *C. roseus*mRNA sequenties en proteoom data beschikbaar voor het EU-project consortium. Geïntegreerde transcriptomics en proteomics werd door onze partners gebruikt voor de identificatie van kandidaten voor de ontbrekende genen in de biosynthese route. Deze kandidaat genen werden gekeurteriseerd in *in vitro* biochemische assays en door mij met behulp van transiente expressie *in planta*, wat resulteerde in de identificatie van de laatste vier missende genen (Hoofdstuk 5). *In
situ hybridisatie liet een zelfde cellulare lokalisatie zien voor geraniol 8-oxidase (G8O), 8-hydroxygeraniol oxidoreductase (8-HGO), iridoid oxidase (IO), 7-deoxyloganetic acid glucosyl transferase (7-DLGT) en 7-deoxyloganic acid hydroxylase (7-DLH). Samen lijken deze genen een transcriptie regulon te vormen voor de productie van loganinezuur in de “internal phloem associated parenchyma” (IPAP) cellen.

Voor verdere sub-cellulaire lokalisatie studies werden de IO, 7-DLH, 8-HGO en 7-DLGT eiwitten gefuseerd met GFP. Expressie van de fusie eiwitten in Catharanthus cellen gaf aan dat IO and 7-DLH (zoals verwacht) geassocieerd aan het ER zitten, terwijl 8-HGO en 7DLGT oplosbare eiwitten bleken te zijn. Uiteindelijk heb ik de hele biosynthese route tot expressie gebracht in N. benthamiana en kon ik aantonen dat de hele route functioneel was. Echter, door de vele zij-reacties door endogene enzymen van N. benthamiana was het nodig om een intermediair van de biosyntheseroute mee te injecteren in het blad om voldoende eindproduct te kunnen accumuleren. Voor toekomstige toepassingen is het dus eerst nodig om de ongewilde zij-reacties (bv glycosylering) eerst te elimineren om de flux naar het gewenste eindproduct te verbeteren.

Met de resultaten beschreven in mijn proefschrift hebben we de weg geopend voor de biotechnologische productie van het farmaceutisch belangrijke strictosidine en daarvan afgeleide TIA's. Daarnaast hebben we ook nieuw inzicht verkregen in de flux capaciteit voor terpeen precursor biosynthese, wat bredere implicaties heeft voor de (heterologe) productie van terpenoiden in planten.
Acknowledgements

Time flies so fast, and people go in and out of your life so quickly, so you should never miss the opportunity to tell these people how much they mean to you. Today, after 5 years working on my PhD thesis, it is the time for me to express my full gratitude to all of the people who were scientifically and spiritually involved in the accomplishment of this thesis.

I would like to thank my daily supervisor and co-promoter, Dr. Sander van der Krol. I would say, scientifically, I spend most of my time with you. Your great research vision and scientific thinking really impressed me and helped me to change my scientific concept. Your detailed instructions on my papers and presentations paved my way as an independent researcher. Your responsible attitude and scientific input for the Smartcell project helped our project to go smoothly and successfully. I am really glad to had you as my daily supervisor; all I learn from you is far beyond just this PhD project and it will have a huge influence on my future research.

I would like to thank my promoter, Prof. Richard Immink whose help was essential for the successful completion of my PhD. I am very grateful for your quick response to the scientific questions I had and to the problems I encountered during my PhD. You are not only a great scientist with a lot of knowledge, but also a great person with a kind heart.

I would like to thank Prof. Harro Bouwmeester who gave me the opportunity to start my project. Your broad scientific network and flexibility helped the project to go easy and successfully. Your patience and approachability gave me a lot encouragement and confidence during this scientific adventure.

Many thanks to Dr. Maarten Jongsma for his valuable comments on my experiments and papers and to Dr. Ric de Vos for his suggestions and instructions on the LC-MS analysis.

Geert Stoopen, Francel Verstappen, Miriam Goedbloed, Marielle Schreuder and Bert Schipper who help to make the lab work so much fun. Without your support, my experiments would have not been so successful. Rina Anthonijsz thanks for your arrangement for all the paper works which made my PhD defence become possible.

I would like to express my gratitude to my MSc thesis supervisor Yujun Liu and Jianhe Wei. Your encouragement and valuable advice will always be a source of inspiration in my scientific life.

Furthermore, I would like to extend my gratitude to my colleagues in the terpene projects. Ting yang (the first Chinese I met in Wageningen), Qing Liu, Liping Gao, Aldana Ramirez, Ting Hieng Ming (Jimmy), Jun He, Bo Wang, Thierry Delatte, Iris Kappers, Jules Beekwilder, Katja Cankar, Benyamin Houshyani, Mohammad majdi, Neli Prota and Yuanyuan Zhang. We started as a colleague, now I consider many of you as a very important friend.
Acknowledgements

I am also very grateful to the bachelor and master student who I had the opportunity to supervise. Kobus, Qian, Marina, Sonja and Kirsten thanks a lot for your great input to this project and I am very happy to have worked with you.

My friends in plant physiology Desaleng, Natalia, Julio, Anna, Wei Song, Esmer, Yanxia, Xi, Hanzi, Bing, Giovanni, Beatriz, Imran, Yunmeng, Rik, Deborah, Cecilia, Peter, Ralph and Manickam, I would like to thank you for your friendship and support throughout all these years. I apologize to my other colleagues in Plant Physiology whose names do not appear here but I can tell you that I will not forget you and I feel really grateful for the time we spent together.

My friends in Wageningen, Fang Xu, Lisong Ma, Hui Li, Jia Li, Xu Cheng, Yafen Lin, Yingwen Zhang, Lin ke, Ying Li, Chunxu Song, Wei Qin, Wei Liu, Na Li, Chenlei Hua, Lisha Zhang, Wei Gao, Yanru Song, Peng Zhu, Songlin Xie, Ke Peng, Tingting Xiao, Wang Yan, Suxian Zhu, Xianwen Ji, Huchen Li, Jimeng Li, Chunting Lang, Zhen Wei, Jianhua Zhang, Yanli Wang, and Shuhang Wang, and Joel Schroter, Sonja, Anouk, Lisette, Marthe, Didi, Le Deng, Xi Xiao, Loes, Kadi, Marjon, Ilse and Bart, I had a lot fun with you and I enjoyed all the dinners and so many other activities we did together.

My friends in China, Dingliang Zhang, Zhixin Li, Zhi Wang, Jun Cao, Lifeng Zhou, Haishen Chang, Li Jun Cai, Chao Shi, Lianlei Zhang, Yan Liu, Anan Chen, Ping Chen. I am very happy that after so many years away we still keep on contact, I am very grateful for your unconditional friendship, love and support. I miss you and love you all!

I would also like to thank my beloved parents and my sister, no words can fully express my gratitude to your unlimited support and love.

Special thanks for all the Smartcell partners for intense collaboration and inspiration and Chinese Scholarship Council for the financial support for my PhD.

Last but not the least, I would like to thank those who will be with me at my defence ceremony in the Aula and those who I could not mention in these acknowledgments, thanks for all of your company and support which make my life so beautiful.
Curriculum vitae

Lemeng Dong was Born on July 24, 1983 in Baoding, Hebei province, China. Her fascination to chemistry was first recognized in the High School which she graduated at 2001. Soon after that, she took the opportunity to go to Agricultural University of Hebei (China), and studied in Horticulture. Due to her excellent performance, she was offered an immediate enrolment in the Agricultural University of Hebei for a MSc without a need to do an additionally national exam. However, her great interests on chemistry made her to pursue another direction to work on the natural products, therefore, she took the major of medicinal plant in Beijing Forestry University. She used only two and half years to obtain the MSc degree which supposed to be three years. In 2009, she came to Wageningen University in the laboratory of plant physiology to continue her research on natural products. The work related to this thesis is aim to produce anticancer compounds in tobacco to lower the price for the consumers. After graduation, she will continue with the research on natural products.
List of publications


## Education Statement of the Graduate School
### Experimental Plant Sciences

**Issued to:** Lemeng Dong  
**Date:** 7 March 2014  
**Group:** Plant Physiology, Wageningen University & Research Centre

### 1) Start-up phase

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>First presentation of your project</td>
<td></td>
</tr>
<tr>
<td>Engineering of the terpene pathway in indole alkaloid biosynthesis</td>
<td>Jan 25, 2009</td>
</tr>
<tr>
<td>Writing or rewriting a project proposal</td>
<td></td>
</tr>
<tr>
<td>Engineering of the terpene pathway in indole alkaloid biosynthesis</td>
<td>2009</td>
</tr>
<tr>
<td>Writing a review or book chapter</td>
<td></td>
</tr>
<tr>
<td>MSc courses</td>
<td></td>
</tr>
<tr>
<td>Laboratory use of isotopes</td>
<td></td>
</tr>
</tbody>
</table>

**Subtotal Start-up Phase:** 7.5 credits*

### 2) Scientific Exposure

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPS PhD student days</td>
<td></td>
</tr>
<tr>
<td>2nd Joint International Retreat of PhD Students in Plant Sciences, Cologne, Germany</td>
<td>Apr 15-17, 2010</td>
</tr>
<tr>
<td>EPS PhD Students Day, Utrecht University</td>
<td>Jun 01, 2010</td>
</tr>
<tr>
<td>ExPectationS Day</td>
<td>Nov 19, 2010</td>
</tr>
<tr>
<td>EPS PhD Day, Wageningen University</td>
<td>May 20, 2011</td>
</tr>
<tr>
<td>3nd Joint International Retreat of PhD Students in Plant Sciences, Orsay, France</td>
<td>Jul 05-08, 2011</td>
</tr>
<tr>
<td>4th European Retreat of PhD students in Plant Sciences, Norwich, UK</td>
<td>Aug 14-17, 2012</td>
</tr>
<tr>
<td>EPS theme symposia</td>
<td></td>
</tr>
<tr>
<td>EPS theme 3 'Metabolism and Adaptation', Wageningen University</td>
<td>Feb 10, 2011</td>
</tr>
<tr>
<td>EPS theme 3 'Metabolism and Adaptation', University of Utrecht</td>
<td>Apr 26, 2012</td>
</tr>
<tr>
<td>NWO Lunteren days and other National Platforms</td>
<td></td>
</tr>
<tr>
<td>ALW meeting 'Experimental Plant Sciences', Lunteren</td>
<td>Apr 19-20, 2010</td>
</tr>
<tr>
<td>ALW meeting 'Experimental Plant Sciences', Lunteren</td>
<td>Apr 04-05, 2011</td>
</tr>
<tr>
<td>ALW meeting 'Experimental Plant Sciences', Lunteren</td>
<td>Apr 02-03, 2012</td>
</tr>
<tr>
<td>Seminars (series), workshops and symposia</td>
<td></td>
</tr>
<tr>
<td>Seminar series: Plant Physiology (4 X)</td>
<td>2009-2012</td>
</tr>
</tbody>
</table>
### Education statement

**Smartcell meetings (8X)**  
2009-2012

**Host plant recognition by parasitic Orobanchaceae**  
Aug 26, 2010

**How to write a world class article**  
Oct 26, 2010

**New Biotrends in green chemistry - Dortmund**  
Dec 01-02, 2010

**Plant Breeding in the Genomics Era**  
Nov 25, 2011

**Workshop: New developments in plant endomembrane biology, University of Amsterdam**  
Jun 15, 2012

### International symposia and congresses

- **Terpnet, Kalmar, Sweden**  
  May 23-26, 2011

- **COST Action PlantEngine, Wageningen**  
  Feb 16-18, 2012

- **Terpnet 2013 (Grete, Greece)**  
  Jun 01-05, 2013

### Presentations

- **Poster: 2nd European of PhD Students in Plant Sciences, Cologne, DE**  
  Apr 15-17, 2010

- **Poster: AIW Meeting**  
  Apr 19-20, 2010

- **Poster: Biotrend**  
  May 25, 2010

- **Poster: EPS PhD Students Day**  
  Jun 01, 2010

- **Poster: Terpnet**  
  May 23-26, 2011

- **Poster: 3nd European Retreat of PhD Students in Plant Sciences, Orsay, FR**  
  Jul 05-08, 2011

- **Oral: WP1 in Aachen, Germany**  
  Feb 18, 2010

- **Oral: WP1 in Neuchatel, Switzerland**  
  Dec 08, 2010

- **Oral: Course Plant Metabolomics**  
  Apr 26, 2011

- **Oral: WP1 in Wädenswil, Switzerland**  
  Jun 15-17, 2011

- **Oral: Terpnet, Greece**  
  Jun 01-05, 2013

### IAB interview

**Meeting with a member of the International Advisory Board member**  
Feb 18, 2011

### Excursions

- **Rijkzwaan Excursion**  
  Sep 27, 2013

---

**Subtotal Scientific Exposure**  
19.9 credits*

### 3) In-Depth Studies

**EPS courses or other PhD courses**

- **Postgraduate course ‘Bioinformatics: A Users Approach’**  
  Aug 30-Sep 03, 2010

- **Postgraduate course ‘System Biology: statistical analysis of omics data’**  
  Dec 13-17, 2010

- **Postgraduate course ‘Plant metabolomics’**  
  Apr 26-28 2011

**Journal club**

- **Literature discussions at PRI and Plant Physiology**  
  2009-2012

**Individual research training**

---

**Subtotal In-Depth Studies**  
6.9 credits*
4) **Personal development**

<table>
<thead>
<tr>
<th>Skill training courses</th>
<th>date</th>
</tr>
</thead>
<tbody>
<tr>
<td>WGS course ‘Scientific Publishing’</td>
<td>Jan 21, 2011</td>
</tr>
<tr>
<td>WGS course ‘Techniques for Writing and Presenting a Scientific Paper’</td>
<td>Sep 06-09, 2011</td>
</tr>
<tr>
<td>WGS course ‘Career Orientation’</td>
<td>Mar 2012</td>
</tr>
<tr>
<td>R tutorial</td>
<td>Apr 20, 2012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organisation of PhD students day, course or conference</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Membership of Board, Committee or PhD council</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Subtotal Personal Development</th>
<th>3.3 credits*</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>TOTAL NUMBER OF CREDIT POINTS*</th>
<th>37.6</th>
</tr>
</thead>
</table>

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits.

* A credit represents a normative study load of 28 hours of study.
This research was conducted at the Laboratory of Plant Physiology of Wageningen University and was funded by the European Community’s Framework VII Program FP7/2007-2013 to the SMARTCELL project KBBE-2007-3-1-01. Lemeng Dong was supported by a PhD fellowship from China Scholarship Council.

Cover and layout design: Agilecolor Design Studio/Atelier (www.agilecolor.com)
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress