Towards the engineering of the monoterpene secoiridoid pathway in transgenic tobacco plants

Bruna Miralpeix Anglada

Dipòsit Legal: L.954-2013
http://hdl.handle.net/10803/123289

TOWARDS THE ENGINEERING OF THE MONOTERPENE SECOIRIDOID PATHWAY IN TRANSGENIC TOBACCO PLANTS

by

Bruna Miralpeix Anglada

University of Lleida

Departament of Plant Production and Forest Science

Applied Plant Biotechnology

Doctoral Dissertation

2013

To my parents, who love me above all else

To my sister, wherever she is, she takes care of me

Supervisor: Prof. Paul Christou

Universitat de Lleida

Institució Catalana de Recerca i Estudis Avançats

_____ _____

Signature Date

Co-supervisor: Prof. Teresa Capell

Departement de Producció Vegetal i Ciència Forestal

Universitat de Lleida

_____ _____

Signature Date

Table of contents

CHAPTER 3: METABOLIC ANALYSIS OF TOBACCO PLANTS EXPRESSING THE TWO EARLY GENES OF THE MONOTERPENE SECOIRIDOID BIOSYNTHETIC PATHWAY

Abstract

The reconstruction of biosynthetic pathways by genetic engineering in heterologous organisms is one of the current aims of metabolic engineering and synthetic biology. Substantial advances have been reported in this context in microbes. However, many challenges still remain particularly in plant systems, due to their inherent complexity and organization at the subcellular and organelle levels. The creation of an exogenous metabolic pathway most often will influence the metabolite flow in plant cells. Monoterpene secoiridoid biosynthesis contributes the terpene component of terpenoid indole alkaloids produced in the medicinal plant *Catharanthus roseus*. *C. roseus* is a rich source of alkaloids; more than 130 different compounds have so far been reported, of which vincristine is highly important. Vincristine is mainly used to treat acute leukaemia, Hodgkin's disease, neuroblastoma, breast carcinoma, among other lymphomas.

My PhD thesis focuses on the development of fundamental knowledge and methodology to reconstruct the monoterpene secoiridoid pathway in tobacco plants using metabolic engineering. The thrust of the thesis concentrated on remaining challenges to engineer complex secondary metabolic pathways in plants. I investigated the early part of the pathway and I recovered a population of transgenic tobacco plants expressing different combinations of transgenes (2 to 4) through the use of a novel multigene transfer process.

Expression and metabolite analysis of a transgenic plant population generated in this study provided novel insights into the complex regulation and the high toxicity of the monoterpene secoiridoid molecules, demonstrating that the engineering of the early steps of the pathway has direct effects on non-target pathways.

The experimental part of the thesis is complemented by the development of a systematic patent search to create an IP database with all relevant patents related with the engineering of terpenoid indole alkaloid pathway in plants.

Resum

Mitjançant la ruta biosintètica dels monoterpens secoiridoïdes es genera el compost terpè, que s'englova dins dels terpens índol alcaloids produïts per la planta medicinal *Catharanthus roseus.* La meva tesi doctoral es centra en el desenvolupament de coneixements fonamentals i metodologia per reconstruir la ruta metabòlica dels monoterpens secoiridoïdes en plantes de tabac mitjançant enginyeria metabòlica.

Per tal de resoldre i donar respostes als reptes pendents de l'enginyeria de rutes metabòliques complexes en plantes, he enfocat la meva recerca en quatre parts independents i complementàries. Primer, centrat-me en la primera part de la ruta metabòlica dels monoterpens, he regenerat una població de plantes de tabac transgèniques mitjançant transferència directa de DNA. Aquestes plantes expressen de forma constitutiva diferents combinacions de transgens, des de 2 a 4. En una segona i tercera fase he fet un anàlisi a nivell metabolòmic i transcriptòmic del material vegetal que he generat, que m'ha proporcionat nous coneixements en la regulació de la ruta metabòlica dels monoterpens i en l'interacció entre rutes endògenes de metabolits secundaris. Aquests resultats són d'aplicació directa i poden ajudar a resoldre les dificultats que encara hi ha en l'enginyeria efectiva de les rutes metabòliques secundàries a les plantes.

Finalment, la part experimental de la meva tesi es complementa amb el desenvolupament d'una recerca sistemàtica de patents, per tal de crear una base de dades sobre propietat intelectual, amb totes les patents pertinents relacionades amb l'enginyeria de la ruta metabòlica dels terpens índole alcaloides en plantes. Aquesta base de dades és amplament utilitzada pels investigadors relacionat en aquest tema.

Resumen

La biosíntesis de monoterpenos secoiridoides aporta el compuesto terpeno de los terpenos índole alcaloides producidos por la planta medicinal *Catharanthus roseus.* Mi tesis doctoral se centra en el desarrollo de los conocimientos fundamentales i la metodología para reconstruir la ruta metabólica de los monoterpenos secoiridoides en plantas de tabaco mediante ingeniería metabólica.

Para resolver i aportar respuestas a los retos pendientes de la ingeniería de rutas metabólicas complejas en plantas, he enfocado mi investigación en cuatro partes independientes y complementarias. Pimero, centrándome en la primera parte de la ruta metabólica de los monoterpenos, he regenerado una población de plantas de tabaco transgénicas mediante transferencia directa de DNA. Estas plantas expresan de manera constitutiva diferentes combinaciones de transgenes, desde 2 a 4. En una segunda y tercera fase he echo un análisis a nivel metabolómico i transcriptómico del material vegetal que he generado, que me ha proporcionado nuevos conocimientos en la regulación de la ruta metabolómica de los monoterpenos y en la interacción entre rutas endógenas de metabolitos secundarios. Estos resultados son de aplicación directa y pueden ayudar a resolver las dificultades que todavía hay en la ingeniería efectiva de las rutas metabólicas secundarias en las plantas.

Finalmente, la parte experimental de mi tesis se complementa con el desarrollo de una búsqueda sistemática de patentes, para crear una base de datos sobre propiedad intelectual, con todas las patentes pertinentes relacionadas con la ingeniería de la ruta metabólica de los terpenos índole alcaloides en plantas. Ésta base de datos es ampliamente utilizada por los investigadores relacionados con éste tema.

Acknowledgements

My sincere gratitude goes to my supervisors Paul Christou and Teresa Capell for their support, encouragement and motivation. Special thanks for Paul for his time, advice and guidelines. Many thanks to Teresa for her help and optimism. I learn more things than science from you.

I also benefited greatly from the Smartcell collaborators, especially Kirsi-Marja Oksman and Anneli Ritala for their hospitality when I visited their laboratory and their constant support, encouragement and friendship. In addition, I am very grateful to Young Hae Choi and Antoine Champagne for his excellent assistance and help to understand my last experiments.

It is with immense gratitude that I acknowledge the support and help of Koreen Ramessar. With her teaching and advices I arrived here. She was more than a supervisor.

I am very grateful to work all this years with my best friend Maite Sabalza. I was very lucky to have her support in all the frustrations and also to all my good results. I shared memorable moments with her. In addition I would like to thank Gemma Farré and Georgina Sanahuja for their friendship, support and good times. I have a lot of good memories, which I will never forget.

In my daily work, I had blessed with my 'old' lab mates Shaista Naqvi, Ariadna Peremarti, Svetlana Dashevskaya, Sonia Gómez, Ludovic Bassie, Sol Maiam, Dawei Yuan and Changfu Zhu. They taught me how to work in a lab and they became a small family. Also, to my 'new' labmates Chao Bai, Raviraj Banakar, Eduard Perez, Evangelia Vamvaka, Uxue Zorrila, Gemma Masip, Judit Berman and Gemma Arjó for their support and friendship.

I would like to make a special mention and a great recognition to Jaume Capell for his general advice, care of the plants and the good conversations, and to Núria Gabernet who always smiling and helped me.

Last but not least, I would like to thanks my parents, Pere Miralpeix and Dolors Anglada, for their unending love and support. They give me always the light when I

see everything dark. I also owe sincere gratitude to my brother Anrau Miralpeix and my sister-in-law Carla Mateu, for always being by my side. Finally, I would also like to express my special thanks to Jordi Gallifa for his moral support, patience and love.

Due to the page limitations of this acknowledgement, my apologies go out to those whom I could not mention personally by name but a big thank you goes out to you, and to everyone who was important to the successful completion of this thesis.

Index of figures

CHAPTER 1: GENERAL INTRODUCTION

CHAPTER 2: GENERATION OF A TRANSGENIC PLANT POPULATION ENGINEERED WITH EARLY MONOTERPENE SECOIRIDOID PATHWAY GENES

CHAPTER 6: ENGINEERING THE MONOTERPENE SECOIRIDOID PATHWAY BY COMBINATORIAL MULITGENE TRANSFER

CHAPTER 7: INTELLECTUAL PROPERTY DATABASE FOR THE ENGINEERING OF THE TERPENOID INDOLE ALKALOID PATHWAY IN PLANTS

Index of tables

CHAPTER 1: GENERAL INTRODUCTION

CHAPTER 2: GENERATION OF A TRANSGENIC PLANT POPULATION ENGINEERED WITH EARLY MONOTERPENE SECOIRIDOID PATHWAY GENES

CHAPTER 3: METABOLIC ANALYSIS OF TOBACCO PLANTS EXPRESSING THE TWO EARLY GENES OF THE MONOTERPENE SECOIRIDOID BIOSYNTHETIC PATHWAY

CHAPTER 4: PROTEOMIC ANALYSIS OF TRANSGENIC TOBACCO PLANTS EXPRESSING THE EARLY PART OF THE MONOTERPENE SECOIRIDOID ALKALOID PATHWAY GENES

CHAPTER 5: MODULATION OF THE MONOTERPENE SECOIRIDOID PATHWAY IN TRANSGENIC TOBACCO PLANTS BY THE TRANSIENT EXPRESSION OF DOWNSTREAM PATHWAY GENES

CHAPTER 6: ENGINEERING THE MONOTERPENE SECOIRIDOID PATHWAY BY COMBINATORIAL MULITGENE TRANSFER

CHAPTER 7: INTELLECTUAL PROPERTY DATABASE FOR THE ENGINEERING OF THE TERPENOID INDOLE ALKALOID PATHWAY IN PLANTS

Abbreviations

A622	Isoflavone reductase
AACT	Acetoacetly CoA thiolase
AACT	Acetoacetyl-CoA acetyltransferase or thiolase
AACT	Acetyl-CoA acetyltransferase
ADCS	Aminodeoxychorismate synthase
ADH	Alcohol dehydrogenase
ADS	Amorphadiene synthase
AppFT	Patent application full text and image database
AS	Anthranilate synthase
AS	Anthranilate synthase
BAP	6-benzylaminopurine
BN-PAGE	Blue native polyacrylamide gel electrophoresis
CDP-MEP	4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate
CdRP	1-(o-carboxyphenylamino)-1-deoxyribulose 5-phosphate
CHS	Antisense chalcone synthase
CMK	4-diphosphocytidyl-2C-methyl-D-erythritol kinase
CMS	4-diphosphocytidyl-2C-methyl-D-erythritol synthase
CPR	Cytochrome P450 reductase
CRT	Carotene desaturase
CS	Chorismate synthase
D4H	Desacetoxyvindoline 4- hydroxylase
D4H	Deacetoxyvindoline 4-hydroxylase
DAHP	3-deoxy-D-arabino-heptulosonate-7-phosphate
DAT	Deacetylvindoline acetyltransferase
DBR	Artemisinic aldehyde reductase
2DGE	Two-dimensional gel-electrophoresis

DHAR	Dehydroascorbate reductase
DHD	3-dehydroquinate dehydratase
DHQS	3-dehydroquinate synthase
DL7H	Deoxyloganin 7-hydroxylase
DMAPP	Dimethylallyl diphosphate
(E)-DMNT	(E)-4,8-dimethyl-1,3,7-nonatriene
DXP	1-deoxy-D-xylulose-5-phosphate
DXP	Pyruvate into 1-deoxy-D-xylulose 5-phosphate
DXPS	1-deoxy-D-xylulose-5-phosphate synthase
DXR	1-deoxy-D-xylulose-5-phosphate reductoisomerase
DXR	DXP reducto-isomerase
E4P	Erythrose 4-phosphate
EPC	European Patent Convention
EPO	European Patent Office
EPSP	5-endolpyruvylshikimate 3-phosphate
ER	Endoplasmatic Reticulum
ESI-MS/MS	Electrospray ionization tandem mass spectrometry
FI-ICR	Fourier transform ion cyclotron resonance
FPPS	Farnesyl diphosphate synthase
G10H	Geraniol 10-hydroxylase
G3P	Glyceraldehyde 3-phosphate
GA-3P	Glyceraldehyde 3-phosphate
GC-MS	Gas chromatography–mass spectrometry
GES	Geraniol synthase
GGPPS	Geranyl geranyl diphosphate synthase
GPPS	Geraniol diphosphate synthase
HDS	2-methyl-2-(E)-butenyl 4-diphosphate synthase
10HGO	10-hydroxygeraniol oxidoreductase
HMBPP	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate

HMG	3-hydroxy-3-methylglutaryl-coenzyme A reductase
HMGCoA	3-hydroxy-3-methylglutaryl coenzyme A
HMGR	3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase
HMGRt	truncated 3-hydroxy-3-methylglutaryl coenzyme A reductase
HMGS	3-hydroxy-3-methylglutaryl coenzyme A synthase
HMGS	3-hydroxy-3-methylglutaryl coenzyme A by HMG-CoA synthase
ICAT	Isotope coded affinity tagging
ICAT	Isotopically-distinct tags
ICS	Isochrosmate synthase
IGPS	Indole-3-glycerol phosphate synthase
IP	Intellectual property
IPAP	Internal phloem-associated parenchyma
IPDL	Property Digital Library
IPP	Isopentenyl diphosphate
IPR	Intellectual property rights
IT	Ion trap
iTRAQ	Isobaric tags for relative and absolute quantification
JPO	Japan Patent Office
LAMT	Loganinc acid methyltransferase
LAMT	Loganic acid methyltransferase
LC	Liquid chromatography
LC-QTOF-MS	Liquid chromatography coupled to quadrupole time-of-flight mass spectrometry
M/Z	Mass-to-charge ratio
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MAT	Minovincinine-19hydroxy-o- acetyltransferase
MDD	Mevalonate diphosphate decarboxylase
MDR	Multidrug resistance
ME-cPP	2-C-methyl-D-erythritol 2,4-cyclodiphosphate
MEP	Methyl-erythriol-phosphate

MEP	2-C-methyl-D-erythritol 4-phosphate
MMSR	Multivariate mass spectra reconstruction
MPO	Methyl putrescine oxidase
MS	Mass spectrometry
MTC	10-oxogeraniol/iridoid cyclase
MudPIT	Multidimensional protein identification technology
MVA	Mevalonic acid
MVAK	Mevalonate kinase
MVAPK	5-diphosphomevalonate kinase
MVK	Mevalonate kinase
NAA	Napthalene acetic acid
NES	Nerolidol Synthase 1 from F. ananassa
NMR	Nuclear magnetic resonance spectroscopy
NMT	16-methoxy-2,3-dihydrotabersonine N-methyltransferase
ODC	Ornithine decarboxylase
PAI	Phosphoribosylanthranilate isomerase
PAIR	Patent application information retrieval
PAL	Phenylalanine ammonia lyase
PAT	Phosphoribosylanthranilate transferase
PatFT	Patent full text and image database
PCA	Principal component analysis
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PGR	Plant growth regulators
PLP	Tryptophan using pyridoxal 5-phosphate
PMF	Peptide mass fingerprint
PMK	Phosphomevalonate kinase
PMT	Putrescine-N-methyl transferase
PQLs	Protein quantity loci

PSY	Phytoene synthase
PTGS	Posttranscriptional gene silencing
PTS	Patchoulol synthase
Q	Quadrupole
QPRT	Quinolinate phosphoribosyl transferase
qRT-PCR	Quantitative real-time PCR
QTLs	Quantitative trait loci
RT-PCR	Reverse transcriptase polymerase chain reaction
SDH	Shikimate dehydrogenase
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SGD	Strictosidine glycosidase
SILAC	Stable isotope labeling with amino acids
SK	Shikimate kinase
SLS	Secologanin synthase
STR	Strictosidine synthase
TDC	Tryptophan decarboxylase
TDC	Tryptophan decarboxylase
TGS	Transcriptional gene silencing
TIA	Terpenoid indole alkaloids
TMSP-d_4	Trimethylsilyl propionic acid sodium salt
TOF	Time-of-flight
TS	Tryptophan synthase α subunit
USPTO	United Sates Patent and Trademark Office
WIPO	World Intellectual Property Organization

CHAPTER 1

GENERAL INTRODUCTION

CHAPTER 1: GENERAL INTRODUCTION

1.1. Secondary metabolites: roles, uses and applications

Plants produce a wide variety and high diversity of secondary metabolites (2ry metabolites) in addition to primary metabolites, via complex pathways that are regulated in highly sophisticated ways (Yazaki et al., 2004; Schäfer et al., 2009). Secondary metabolites are organic compounds with no apparent essential role in growth and development; yet plants often expend much energy in their biosynthesis (Peters et al., 2004). This has been rationalised by the observation that these natural products play important ecological roles in plant defence against insects and other herbivores, also against bacterial, fungi, viruses, and even other competing plants (Schäfer et al., 2009). They also play important roles in allelopathy and tritrophic interactions, as well as acting as attractants for pollination by seed-dispersing animals or for UV protection (Schäfer et al., 2009, Peters et al., 2004). Nicotine and caffeine, for example are strong insecticides (Casanova et al., 2002, Willoughby et al., 2006). Some 2ry metabolites exhibit strong biological activity, e.g. inhibition of DNA and protein synthesis or cardenolide activity (Yazaki et al., 2004).

Throughout history, as a result of the properties of their active secondary metabolites, plants have been used in medicine since prehistoric times, and many are still actively prescribed in traditional Asian medicines as well as being used in folk medicine around the world (Yazaki et al., 2004, Gómez-Galera et al., 2007). Many 2ry metabolites exhibit significant biological activities and can interact with molecular targets in human cells or microorganisms. Consequently, many of the drugs used in contemporary medicine are derived directly from over 50,000 plant species (Gómez-Galera et al., 2007) or indirectly as components in the structures of bioactive 2ry metabolites used for chemical synthesis with improved activities (Schäfer et al., 2009). Over 25% of the new drugs approved in the last 30 years are based on a molecule of plant origin, and about 50% of the top selling drugs derive from knowledge on plant secondary metabolism (Gómez-Galera et al., 2007). Examples of isolated 2ry metabolites, which are being used for medicinal purposes are morphine as analgesic, codeine as antitussive, reserpine as antihypertensive, digoxin as cardiotonic, vinblastine, vincristine, paclitaxel (taxol), camptothecin, demecolvine, and podophyllotoxin as antineoplasic agents, or artemisinin and quinine as antimalarial

compounds. These are only a few of the several irreplaceable medicinal products derived from plants (Figure 1) (Schäfer et al., 2009; Gómez-Galera et al., 2007). Seven plant-derived compounds with anticancer activity have received FDA approval for clinical use: Taxol/Paclitaxel (*Taxus brevifolia*), vinblastine and vincristine (*Catharanthus rouses*), topotecan and irinotecan (*Camptotheca acuminata*), and etoposide and teniposide (*Podophyllum peltatum*) (Gómez-Galera et al., 2007).

These plants are usually grown in plantations. Since 2ry metabolites accumulation in plants is usually low, production costs are high and consequently these drugs are very expensive (Schäfer et al., 2009). Cost-imposed constraints limit their broader accessibility to poor people, particularly in developing countries.

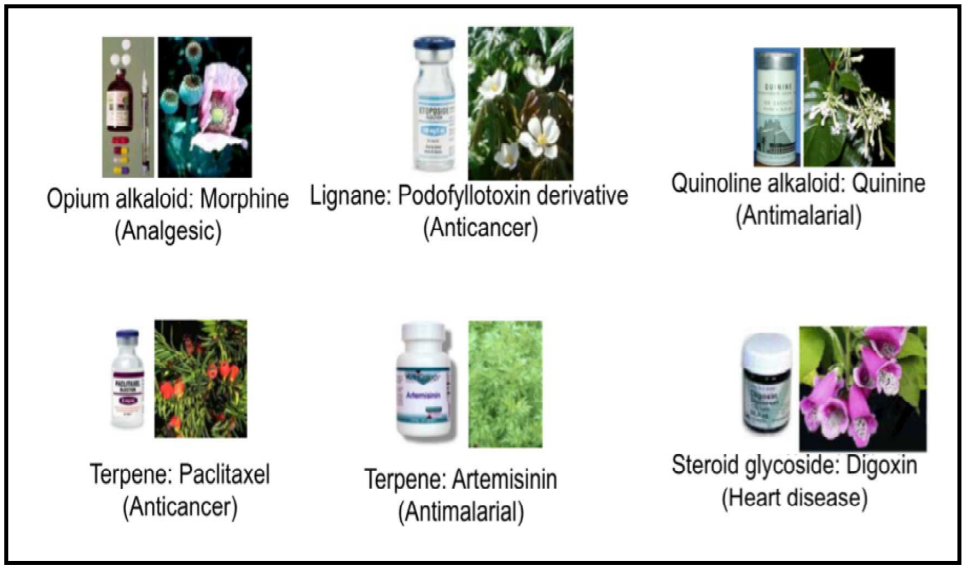

Figure 1. Some important medicinal products derived from plants

The uses of plant secondary metabolites are not just limited to the medical field, as they are also utilised in industry and agriculture as dyes, edible flavours, fragrances, fibres, fixatives, arrowhead poisons and tanning agents or as insecticides and herbicides (Yazaki et al., 2004). Others have found use as drugs, both medicinal and recreational, and many have important roles as phytonutrients for humans and livestock (Peters et al., 2004).

1.2. Major classes of secondary metabolites

Plants usually synthesize, transport and store 2ry metabolites in specific ways. The types of 2ry metabolites produced are sometimes but not always typical for particular systematic groups of plants. Among more than 1,000,000 structures of 2ry metabolites that have been identified thus far, one can distinguish between nitrogen-containing and nitrogen-free molecules. Among nitrogen-containing 2ry metabolites, alkaloids are the largest group with more than 20,000 structures, many with pronounced pharmacological and toxic properties. Also important are non-protein amino acids (7000 structures), amines (100 structures), cyanogenic glucosides (60 structures), glucosinolates (100 structures), alkamides (150 structures), as well as lectins and other peptides (2000 structures). In the class of nitrogen-free 2ry metabolites, even more structures have been determined. The largest class is terpenoids with more than 20,000 known compounds, including mono-, sesqui-, di-, and triterpenes with interesting biological activities. Another bioactive group, the polyphenols, is characterized by the presence of several phenolic hydroxyl groups, which can dissociate into O- ions under physiological conditions. Members of polyphenols are flavonoids, anthocyanins and tannins. In addition, phenylpropanoids, coumarins, lignans and anthraquinones often possess phenolic OH-groups (Schäfer et al., 2009; Yazaki et al., 2004).

1.1.1. Alkaloids

Alkaloids constitute the 2ry metabolite group with the largest number of biologically active compounds (Yazaki et al., 2004). They are low molecular weight nitrogen-containing basic compounds (Peter et al., 2004; Liu et al., 2007). There are ca: 20,000 alkaloids and they are found in about 20% of all flowering plant species where they serve eco-chemical roles, most often in plant defence. Many other uses have been discovered for alkaloids, particularly as pharmaceuticals, with a world market volume of 4 billion US$ (Julsing et al., 2006), as purgatives, antitussives, sedatives, analgesics such as morphine from the opium poppy, *Papaver somniferum* (Papaveraceae), which is the most effective analgesic known and also recreational drugs, including, the ubiquitous use of caffeine or the widely abused drug heroin (Peter et al., 2004; Yazaki

et al., 2004; Liu et al., 2007). The basic structures of many alkaloids originate from the amino residue of an amino acid (Yazaki et al., 2004) e.g. Phe, Tyr, Trp, Lys and Orn (Liu et al., 2007). Alkaloids can be classified in terms of their biological activity, their chemical structure, or according to their biosynthesis (Julsing et al., 2006). They are usually divided into five major groups depending on their amino acid origin (Julsing et al., 2006; Yazaki et al., 2004) (Table 1).

Table 1. Five major groups of alkaloids depending on their amino acid origin

Group		**Amino acid type**
I.	Tropane-, pyrrolidine- and pyrrolizide-alkaloids	Ornithine
II.	Benzylisoquinoline	Tyrosine
III.	Indolequinoline	Tryptophane
IV.	Pyridine	Lysine
V.	Quinolizidine- and piperidine-alkaloids	Lysine

Terpenoid indole alkaloids, as the name suggests, contain a terpenoid moiety derived from the iridoid monoterpenoid glycoside secologanin, in addition to the indole (alkaloid) core that is derived from tryptophan (Peter et al., 2004). Terpenoid indole alkaloids (TIAs) form a large group of structurally diverse molecules comprising over 2000 members (O'Connor et al., 2005). These complex natural products possess interesting pharmacological activities such as the antineoplastic agents vincristine and vinblastine from *Catharanthus roseus*, the antihypertensives reserpine and ajmalicine, and the anti-arrhythmic ajmaline. Some TIAs are believed to play a role in the defense of plants against pests and pathogens (Liu et al., 2007; Schäfer et al., 2009; O'Connor et al., 2005).

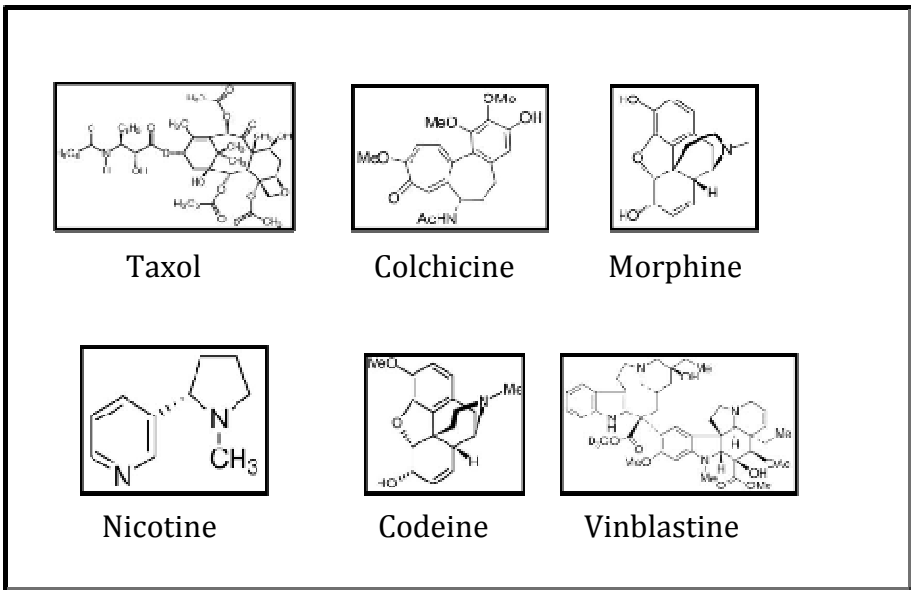

Figure 2. Examples of alkaloids structures

1.1.2. Terpenoids

Terpenoids (isoprenoids) represent a large and important class of natural products with more than 30,000 different structures (Julsing et al., 2006). They are organic compounds possessing branched chains constructed from the isoprene unit, 2-methylbutadiene, which is composed of five carbon atoms (Hegazy et al., 2006; Yazaki et al., 2004). There are more than 25,000 terpenoids with well characterized structures. Depending on the number of isoprene units they are classified as hemiterpenes (C_5), monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), sesterpenes (C_{25}) and triterpenes (C_{30}). Steroids are also biosynthetically classified as terpenoids because their basic structures originate from the common precursor of triterpenes, squalene. Terpenoids are distributed in microorganisms, plants and animals. On the basis of their chemical structure, terpenoids may possibly be the most diverse group of plant secondary metabolites (Yazaki et al., 2004).

Many monoterpenoids are fragrance components of the essential oils of flowers, herbs and spices, and they are extracted for use as flavours and perfumes. Sesqui- and diterpenoids have been extensively studied by plant pathologists and phytochemists, because the compounds of these subclasses function as phytoalexines to protect plants from herbivores or microbial pathogens (Yazaki et al., 2004). Furthermore diterpenoids (C_{20}) are of high interest and paclitaxel as a major representative of this group is a blockbuster drug. Carotenoids with a C_{40} backbone exhibit important functions in photosynthesis, pigmentation and as antioxidants. From a pharmaceutical point of view a very important class of terpenoids is the sterols which are derived from a triterpene (C_{30}) backbone and are used as starting material in the chemical synthesis of synthetic drugs including steroid hormones and contraceptives (Julsing et al., 2006).

1.1.3. Phenols

They are defined as low molecular weight (< 700 D) secondary metabolites, which contain at least one hydroxylated aromatic ring (Figure 3; Brodniewicz et al., 2012). It is estimated that ca 40% of organic carbon is bound in the form of polymeric structures containing phenolic functionality and constitute a large group of 2ry metabolites with ca 10,000 individual entities (Brodniewicz et al., 2012). Phenolic compounds of plant origin share some common properties. Many exhibit antioxidant activity, which is considered beneficial for human health; they are seldom pharmacologically inert and also, they are known to share certain metabolic pathways, by which they are conjugated and excreted from the human body (Brodniewicz et al., 2012). Phenolic compounds, such as flavonoids are responsible for flower colour, and phenylpropanoids, involved in lignin biosynthesis, are also very widely distributed in the plant kingdom (Yazaki et al., 2004).

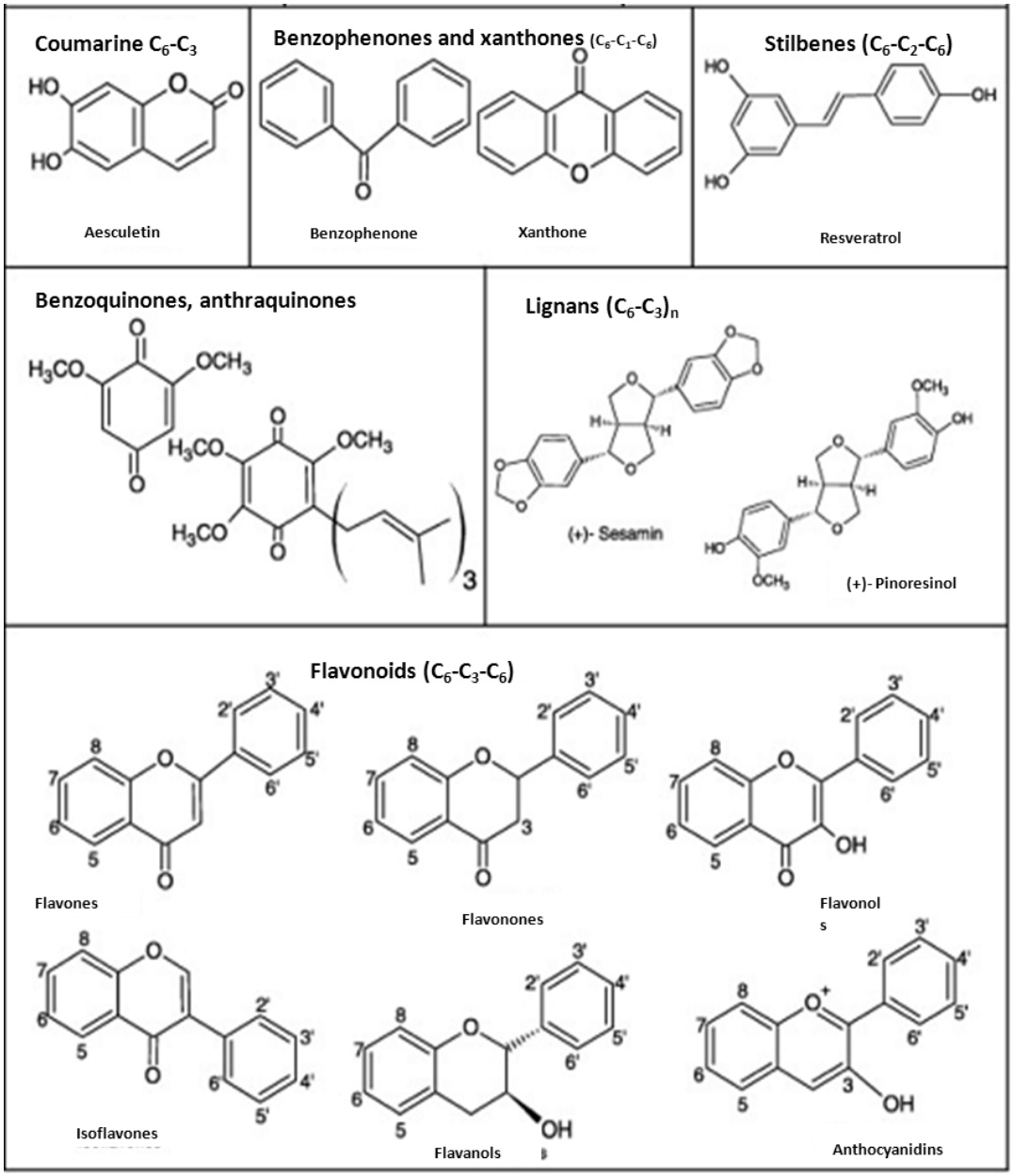

Figure 3. Examples of phenol structures (Soto et al., 2011)

1.3. Traditional sources of secondary metabolites

Some secondary products are only available as extracts from wild plants (Table 2), many of which grow very slowly and are seldom cultivated. If such natural compounds, many of which are difficult to synthesise chemically because of their complicated structures could be produced by cultured plant cells or tissues, this would generate a stable supply of the compounds, while also protecting rare wild plant species (Goklany et al., 2009; Yazaki et al., 2004). In the past, quantities needed to meet demand were relatively low; however, increasing commercial demand is fast outpacing supply. Currently between 4,000 and 10,000 medicinal plants are on the endangered species list and this number is expected to increase (Gómez-Galera et al., 2007). To counter over-exploitation of natural resources and consequent threats to biodiversity, sustainable practices have been recommended and several worldwide organizations have established guidelines for collection and sustainable cultivation of medicinal plants (Klingenstein et al., 2006). Cultivation of medicinal plants has conservation advantages; however, costs are frequently prohibitive because of their slow growth rate and the fact that many tropical plants are very difficult to cultivate in a commercial setting. Despite such difficulties, several compelling advantages call for serious efforts in the sustainable cultivation of endangered medicinal plants. Some of these advantages include authoritative and reliable identity of particular plant species and populations, reduction in genotypic and phenotypic variability, and control of the cultivation process to allow accurate prediction of the levels and purity of the principal ingredient(s). Breeding programs targeting selection of high yielding populations and certification programs to allow better quality control can also be implemented under cultivation (Gómez-Galera et al., 2007).

There are several pharmaceuticals based or derived from 2ry metabolites on the market that are very expensive, due to the fact that these compounds are only found in rare plants and often in extremely low amounts. To achieve a sustainable source of such compounds (Julsing et al., 2006) methods to produce them *in vitro* have been explored. Callus, cell suspension, and organ cultures (root and hairy root cultures) and even large-scale fermentation of suspended cells were successfully established over the last 40 years. The biotechnological production of secondary metabolites is

advantageous because of its production-rate stability, regardless of unpredictable climatic conditions (Yazaki et al., 2004). However, the employment of callus and suspension cultures of medicinal plants often encountered the problem of very low or insufficient product yields, since isolated plant cells quickly cease producing secondary metabolites when dedifferentiated although conversely, these dedifferentiated cells grow much more rapidly than differentiated cells (Yazaki et al., 2004). Apparently, the genes encoding the proteins necessary for biosynthesis, transport and storage of 2ry metabolites are not adequately expressed in most undifferentiated cell cultures. There are a few notable exceptions e.g. ginsenosides (*Panax ginseng*) (Mathur et al., 1999), shikonin (*Lithospermum erythrrhhizon*) (Hara et al., 1987), berberine (*Coptis japonica*) (Suzuki et al., 1993), rosmarinic acid (*Coleus blumei*) (Bauer et al., 2004), anthraquinones (*Morinda citrifolia*) (Ruffoni et al., 2010) and paclitaxel (*Taxus* sp.) (Sabater-Jara et al., 2010). More encouragingly, root and hairy root cultures, which are differentiated tissues, show excellent product yields for those 2ry metabolites that are produced in roots (which is unfortunately not the case for all 2ry metabolites). However, the large-scale fermentation of roots and hairy roots still is a challenge, although bioreactors have been developed for sustainable *in vitro* cultivation (Sevón et al. 2002; Schäfer et al., 2009).

1.4. *Catharanthus roseus*

1.4.1. General information (plant, source of medicinal compounds)

The genus *Catharanthus* belongs to the family Apocynaceae and is closely related to the genera *Vinca* and *Amsonia*. *Catharanthus* comprises eight species; *C. coriaceus* Markgr. *C. lanceus* (Bojer ex A.DC.) Pichon, *C. longifolius* (Pichon) Pichon, *C. ovalis* Markgr., *C. roseus* (L.) G.Don, *C. scitulus* (Pichon) Pichon, *C. trichophyllus* (Bak.) Pichon are endemic in Madagaskar, *C. pusillus* (Murray) G. Don originates from India (Van der Heijden et al, 2004). They are sub shrubs (30-90 cm tall), generally upright or decumbent, produce white latex and strong pungent smell when damaged (Aslam J. et al., 2009). *C. roseus* is also known as Madagascar periwinkle (Figure 4) (Dutta et al., 2007; Pan et al., 2009) and has a pantropic distribution, naturalized in continental Africa, America, Asia, Australia, South Europe and on some islands in the Pacific Ocean.

Table 2. Currently used Drugs in the United States that are obtained from flowering Plants (modified from Ethnoleaflets, 2010)

Used Drugs	**Plant Name**	**Family**	**Use**
Bromelain	*Ananas comosus* (pineapple)	Bromeliaceae	Antiinflammatory, proteolytic
Atropine	*Atropa belladonna, Duboisia myoporoides*	Solanaceae	Anticholinergic
Oatmeal Concentrate	*Avena sativa*	Gramineae	Skin Protectant
Capsicum Oleoresin	*Capsicum* species	Solanaceae	Topical analgesic
Papain	*Carica papaya*	Caricaceae	Proteolytic, mucolytic
Sennosides A + B	*Cassia acutifolia, Cassia angustifolia*	Leguminosae	Laxative
Vincristine, vinblastine	*Catharanthus roseus*	Apocynaceae	Antitumor, antileukemic agent
Quinine	*Cinchona* species	Rubiaceae	Antimalarial, antipyretic
Colchicine	*Colchicum autumnale*	Liliacae	Antitumor, antigout
Digoxin, Lanatoside C, Acetylidigitoxin	*Digitalis lanata, Digitalis purpurea*	Scrophulariaceae	Cardiotonic
Diosgenin	*Dioscorea* species	Dioscoreaceae	Contraceptive
Hyoscyamine, Scopolamine	*Duboisia myoporoides*	Solanaceae	Anticholinergic, sedative
Ephedrine, Pseudoephedrine	*Ephedra sinica*	Ephedraceae	Sympathomimetic, antihistamine
Sitosterols	*Glycine max*	Leguminosae	Lowering colesterol
Opium, Codeine, Morphine, Noscapine, Papaverine	*Papaver somniferum*	Papaveraceae	Analgesic, antitussive, antitussive
Pilocarpine	*Pilocarpus jaborandi*	Rutaceae	Parasympathomimetic
Psyllium husks	Plantago species	Plantaginaceae	Laxative

Table 2. Currently used Drugs in the United States that are obtained from flowering Plants (modified from Ethnoleaflets, 2010) **(cont.)**

Used Drugs	**Plant Name**	**Family**	**Use**
Podophyllin	*Podophyllum peltatum*	Berberidaceae	Genital warts
Prune Concentrate	*Prunus domestica*	Rosaceae	Laxative
Reserpine, Alseroxylon	*Rauvolfia serpentine*	Apocynaceae	Antihypertensive, tranquilizer
Deserpidine, Reserpine, Rescinnamine	*Rauvolfia vomitoria*	Apocynaceae	Antihypertensive, tranquilizer
Casanthranol	*Rhamnus purshiana*	Rhamnacee	Laxative
Rhubarb Root	*Rheum* species	Polygonaceae	Laxative
Ricinoleic Acid	*Ricinus communis*	Euphorbiaceae	Analgesic, anti-inflammatory
Veratrum viride, Cryptennamine	*Veratrum viride*	Liliaceae	Antihypertensive

C. roseus is a rich source of alkaloids; more than 130 different compounds have so far been reported of which bisindole alkaloids (about 40 compounds), most of those containing a plumeran (vindoline) and an ibogan (catharanthine) moiety, are very important medicinally active compounds (Van der Heijden et al., 2004). Some of the *C. roseus* alkaloids are marketed as pharmaceuticals as is the case of ajmalicine (Hydrosperan^{R}, Lamuran^{R}) which was introduced in 1957 for the treatment of hypertensia; vinblastine (vincaleukoblastine, Velbe^{R}), which was introduced in 1960 and is used in the treatment of Hodgkin's disease, non-Hodkin lymphomas, testiscarcionomas, and sometimes against breast cancer and chorio-carcinomas or its oxidized form vincristine (leurocristine, Oncovin^{R}), which was introduced in 1963 and is used against acute leukemia, Hodkin's disease, non-Hodgkin lymphomas, rhabdomysarcomas, Wilm's tumors in children and breast cancer. Anhydrovinblastine has also been used as an antieoplastic agent in the treatment of cervical and lung cancer (Van der Heijden et al., 2004; Aslam et al., 2009; Pereira et al., 2010). In fact, *C. roseus* is a remarkable factory of bioactive compounds and it is the leading single plant species reported to produce such a wide array of complex alkaloids (Perreira et al., 2010; Goklany et al., 2009). In addition alkaloids, other secondary metabolites have been isolated from *C. roseus*, including monoterpenoid glucosides (loganin, secologanin, sweroside, deoxyloganin, dehydrologanin), steroids (catasteron, brassinolides), phenolics, flavonoids and anthocyanins (Van der Heijden et al., 2004).

For commercial use, *C. roseus* plants are harvested 3 months after germination when they reach 45–55 cm and begin to flower. However, there are few reports on the effect of plant growth regulators (PGRs) on alkaloid production in *C. roseus* plants during the blooming period, the optimal time for harvest. Pre-harvest PGR application to *C. roseus* plants during flowering is carried out to collect plant material with high yields of alkaloids for commercial use and production (Pan et al., 2010). Traditionally, *C. roseus* is propagated through seeds, which leads to genetic segregation and decline in uniform yield of dimeric alkaloids (Aslam et al., 2009). Breeding programs have been set up not only for horticultural purposes, but also to modify alkaloid content (Van der Heijden et al, 2004).

The commercial importance of the dimeric alkaloids vinblastine and vincristine, which are formed by the condensation of catharanthine and vindoline, and also ajmalicine and serpentine, have prompted many attempts to produce them in plant

tissue culture systems over the past several decades. Most of these efforts, however, have not focused on genetically engineered cell cultures, tissue cultures, or plants but rather on non-engineered cell lines under elicitation, with precursor feeding, or in optimized media. While these classical methods have hastened characterization efforts, metabolic engineering offers the most promising method for improved product composition and increased alkaloid yield in plants and cultured cell/organ systems (Liu et al., 2007; Miralpeix et al. 2013; Rischer et al., 2013).

Figure 4. *Catharanthus roseus* (Lookfordiagnosis, 2012)

1.4.2. Vincristine/vinblastine

Vincristine and vinblastine are dimeric monoterpenoid indole alkaloids produced in *Catharanthus roseus*. They were discovered as a result of a drug-screening program in the late 1950's by Eli Lilly (Johnson et al, 1963; Van der Heijden et al, 2004). In early experiments mice implanted with the acute lymphocytic neoplasm, P-1534 leukemia, were treated with crude fractions of the plant. In those experiments certain alkaloidal fractions resulted in a reproducible extention of the life expectancy of the animals.. Further phytochemical investigations of this fraction in the isolation of vinblastine and vincristine, which were subsequently shown to be effective against P-1534 leukemia. Both alkaloids became highly effective anticancer agents and are currently in clinical use against leukaemia, Hodgkin's lymphoma and other cancers (Johnson et al, 1963; DeVita et al., 2008; O'Connor et al., 2005). During the last 40

years they have been used for the treatment and cure of thousands of patients because of their unique mode of action and their effectiveness (Van der Heijden et al, 2004).

The antineoplastic activity of the bisindole alkaloids is attributed to their ability to disrupt microtubules, causing dissolution of mitotic spindles and metaphase arrest in dividing cells (Van der Heijden et al, 2004). Although the alkaloids are classified as mitotic inhibitors, their antineoplastic activity in the clinical treatment of cancer most likely results from perturbation of a variety of microtubule-dependent processes, as well as from distribution of the cell cycle and induction of programmed cell death (Van der Heijden et al, 2004).

Vinblastine and vincristine are present only as minor constituents of the complex mixture of about 130 alkaloids produced by *C. roseus* (Van der Heijden et al, 2004; Julsing et al., 2006). In fact, vincristine, which is present in approximately 0.0003% content, represents the lowest level of any medicinally useful alkaloid produced on a commercial basis. For the production of 3 kg of *Vinca* alkaloids, which is the annual demand worldwide, around 300 tones of plant material has to be processed (Shukla et al., 2010).

As discussed above bisindole alkaloids occur in extremely low quantities *in planta* due to their highly mitotoxic nature, which poses the problem of cellular containment (Shukla et al., 2010). Attempts to produce *Vinca* alkaloids in plant cell cultures did not lead to a significant improvement and biotechnological approaches in plant cell cultures may not provide an instant solution to this problem (Julsing et al., 2006). One of the limitations for biotechnological production in plant cell cultures is that the condensation of catharanthine and vindoline is an absolute requirement for the formation of vinblastine and later, vincristine. Since vindoline is not present in the roots of *C. roseus* or in dedifferentiated suspension-cultured cells (it is only produced in the green parts of the plant) neither vincristine nor vinblastine have thus far been produced in hairy roots or dedifferentiated cultures (Perreira et al., 2010).

Structurally vincristine and vinblastine are bisindole alkaloids with vindoline attached to a tetracyclic indole derivative, carbomethoxyvelnamine (Figure 5) (Aslam et al., 2009). Both vinblastine and vincristine possess an identical velbanamine upper subunit and nearly identical vindoline-derived lower subunits differing only in the dihydroindole *N*-substituent (Ishikawa et al., 2009). Despide this small structural

difference, vinblastine and vincristine differ in their antitumor properties and dose-limiting toxicities (Ishikawa et al., 2009). Vincristine is used in combination therapy to treat acute leukemias and lymphomas and constitutes an important component of the regime that has been so successful in treating childhood leukemias. Vinblastine is often used in combination to treat bladder and breast cancers and is an integral part of the curative treatment regime for Hodgkin's disease (Ishikawa et al, 2009). The major limitation to the continued use of the vinca alkaloids is the emergence of drug resistance arising principally from overexpression of phosphoglycoprotein (Pgp), an efflux pump that transports many of the major drugs out of the cell. In fact, vinblastine represents one of the most studied prototypical substrates for Pgp efflux responsible for multidrug resistance (MDR) (Ishikawa et al., 2009).

These alkaloids are frequently incorporated into combination chemotherapy because of their lack of cross-resistance to DNA-alkylating drugs and their different mechanism of action (Van der Heijden et al, 2004).

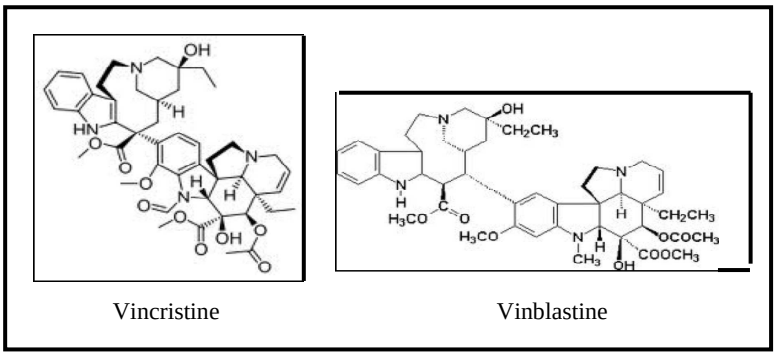

Figure 5. Vincristine and vinblastine structures

1.5. References

Almagro L., Perez A. J. L., Pedreño M. A. (2011) New method to enhance ajmalicine production in *Catharanthus roseus* cell cultures base don the use of cyclodextrins. *Biotechnology Letters*, 33: 381-385.

Aslam J., Mujib A., Nasim S. A. and Sharma M. P. (2009) Screening of vincristine yield in *ex vitro* somatic embryos derived plantlets of *Catharanthus roseus* L. (G) Don. *Scientia Horticulture,* 119: 325-329.

Bauer N., Leliak-Levanic D., Jelaska S. (2004) Ro2ry metabolites arinic acid synthesis in transformed callus culture of Coleus blumei Benth. *Journal of Bioscience*, 59: 554-560.

Brodniewicz T., Grynkiewicz G. (2012) Plant phenolics as drug leads- what is missing? *Acta Poloniea Pharmaceutica- Drug Research*, 69: 1203-1217.

Casanova H., Ortiz., Peláez C., Vallejo A., Moreno M.E., Acevedo M. (2002) Insecticide formulations based on nicotine oleate stabilized by sodium caseinate. *Journal of Agricultural and Food Chemistry*, 50:6389-6394.

DeVita V.T. and Chu E. (2008) A history of cancer chemotherapy. *American Association for Cancer Research,* 68:8643-8653.

Dutta A., Sen J. and deswal R. (2007) Downregulation of terpenid indole alkaloid biosynthetic pathway by low temperature and cloning of AP2 type C-repeat binding factor (CBF) from *Catharanthus roseus* (L). G. Don. *Plant Cell Rep*, 26:1869-1878.

Ethnoleaflets (2010) Origins of plant derived medicines. Accessed online 25 October 2010, http://www.ethnoleaflets.com/leaflets/maridass.htm

Food-info (2012) What is terpenes? Accessed online 21 December 2012, http://www.food-info.net/uk/qa/qa-fi69.htm

Goklany S., Loring R. H., Glick J. and lee-Parsons C. W. T. (2009) Assessing the Limitations to Terpenoid Indole Alkaloid Bosynthesis in *Catharanthus roseus* Hairy root Cultures Through Gene Expression Profiling and Precursor Feeding. *Biotechnol. Prog*, 25:1289-1296.

Gómez-Galera S., Pelacho A. M., Gené A., Capell T. and Christou P. (2007) The genetic manipulation of medicinal and aromatic plants. *Plant Cell Rep*, 26:1689-1715.

Hara Y., Morimoto T., Fujit Y. (1987) Production of shikonin derivatives by cell suspension cultures of *Lithospermum erythrorhizon*. *Plant cell Report*, 6:8-11.

Hegazy M. E. F., Kuwata C., Matsushima A., Ahmed A. A. and Hirata T. (2006) Biotransformation of sesquiterpenoids having α,β-unsaturated carbonyl groups with cultured plant cells of *Marchantia polymorpha*. *Journal of Molecular Catalysis B: Enzymatic,* 39: 13-17.

Hilliou F., Christou P., Leech M. J. (1998) development of an efficient transformation system for *Catharanthus roseus* cell cultures using particle bombardment. *Plant Science*, 140: 179-188.

Ishikawa H., Colby D. A., Seto S., Va P., Tam A., Kakei H., Rayl T. J., Hwang I., Boger D. L. (2009) Total synthesis of vinblastine, vincristine, related natural products, and key structural analogues. *Journal American Chemical Society,* 131: 4904-4916.

Johnson I.S., Armstrong J.G., Gorman M (1963) The vinca alkaloids: a new class of oncolytic agents. *American Association for Cancer Research,* 23:1390-1427.

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

Klingenstein F., Vogtmann H., Honner S., Leaman D. (2006) Sustainable wild collection of medicinal and aromatic plants: practice standards and performance

criteria. *Medicinal and aromatic plants*, 7:109–120.

Liu D. H., Jin H. B., Chen Y. H., Cui L. J., Ren W. W., Gong Y. F. and Tang K. X. (2007) Terpenoid Indole Alkaloids Biosynthesis and Metabolic Engineering in *Catharanthus roseus. Journal of Integrative Plant Biology*, 49: 961-974.

Lookfordiagnosis (2012) Catharanthus (Vinca de Madagascar; Vinca rosea), accessed online 26 December 2012, http://www.lookfordiagnosis.com/mesh_info.php?term=catharanthus&lang=2

Mathur A., Mathur A.K., Pal M. Unival G.C. (1999) Comparison of qualitative and quantitative in vitro ginsenoside production in callus cultures of three Panax species. *Planta Medica*, 65: 484-486.

Miralpeix B., Rischer H., Häkkinen S.T., Ritala A., Seppänen-Laakso T., Oksman-Caldentey K.M., Capell T., Christou P. (2013) Metabolic engineering of plant secondary products: which way forward? *Current Pharmaceutical Design* (in press).

Rischer H., Häkkinen S.T., Ritala A., Seppänen-Laakso T., Miralpeix B.,Capell T., Christou P., Oksman-Caldentey K.M. (2013) Plant cells as pharmaceutical factories. *Current Pharmaceutical Design* (in press).

Ruffoni B., Pistelli L., Bertoli A., Pisteli L. (2004) Plant cell cultures: Bioreactors for industrial production. *Advances in Experimental Medicine and Biology*, 698: 203-221.

O'Connor S. E. and Maresh J. J. (2005) Chemestry and biology of monoterpen indole alkaloid biosynthesis. *Natural Product Reports*, 23: 532-547.

Pan Q., Chen Y., Wang Q., Yuan F., Xing S., Tian Y., Zhao J., Sun X. and Tang K. (2010) Effect of plant growth regulators on the biosynthesis of vinblastine, vindoline and catharanthine in *Catharnathus roseus*. *Plant Growth Regul*, 60: 133-141.

Pereira D. M., Ferreres F., Oliveira J. M. A., Gaspar L., Faria J., Valentao P., Sottomayor M. and Andrade P. B. (2010) Pharmacological effects of *Catharanthus roseus* root alkaloids in acetylcholinesterase inhibition and cholinergic neurotransmission. *Phytomedicine,* 17: 646-652.

Peters R. J. and Croteau R. B. (2004) Metabolic Engineering of Plant Secondary Metabolism. Handbook of Plant Biotechnology. *John Wiley & Sons, Ltd* Christou P and Klee H. Chichester, England, 1: 609-629.

Sabater-Jara A.B., Tudela L.R., López-Pérez A.J.(2010) In vitro culture of Taxus sp.: strategies to increase cell growth and taxoid production. *Phytochemical Reviews*, 9:343–356.

Schäfer H. and Wink M. (2009) Medicinally important secondary metabolites in recombinant microorganisms or plants: Progress in alkaloid biosynthesis. *Biotechnology Journal*, 4: 1684-1703.

Sevón N. and Oksman-Caldentey K.M. (2002) Agrobacterium rhizogenes-Mediated Transformation: Root Cultures as a Source of Alkaloids. *Plant Med,* 68: 859 – 868.

Shukla A. K., Shasany A. K., Verma R. K., Gupta M., Mathur A. K. and Khanuja S. P. S. (2010) Influence of cellular differentiation and elicitation on intermediate and late steps of terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Protoplasma,* 242: 35-47.

Soto M. L., Moure A., Domínguez H., Parajó J. C. (2011) Recovery, concentration and purification of phenolic compounds by adsorption: A review. *Journal of Food Engineering*, 105: 1-27.

Suzuki H., Ozaki Y., Suga C., Morimoto T., Satake M., Harada M. (1993) Dissolution test of Coptis rhizome and cultured cells of Coptis japonica. *Japanese Journal of Pharmacognosy*, 47: 311-315.

Van der Heijden R., Jacobs D. I., Snoeijer W., Hallard D. and Verpoorte R. (2004) The *Catharanthus* Alkaloids: Pharmacognosy and Biotechnology. *Current Medicinal Chemestry*. 11: 607-628.

Willoughby L., Chung H., Lumb C., Robin C., Batterham P., Daborn P.J. (2006)A comparison of Drosophila melanogaster detoxification gene induction responses for six insecticides, caffeine and phenobarbital. *Insect Biochemistry and Molecular Biology*, 36: 934–942.

Yazaki K. (2004) Natural products and metabolites. Handbook of Plant Biotechnology. *John Wiley & Sons, Ltd*. Christou P and Klee H. Chichester, England, 2: 811-861.

CHAPTER 2

GENERATION OF A TRANSGENIC PLANT POPULATION ENGINEERED WITH EARLY MONOTERPENE SECOIRIDOID PATHWAY GENES

CHAPTER 2: GENERATION OF A TRANSGENIC PLANT POPULATION ENGINEERED WITH EARLY MONOTERPENE SECOIRIDOID PATHWAY GENES

2.1. Abstract

The modulation of plant secondary metabolism is one of the key goals of plant biotechnology. Studies focusing on the biochemistry and enzymology of secondary metabolism and the isolation of genes encoding the corresponding enzymes led to the first attempts to engineer complex secondary metabolic pathways. The reconstruction of a complete pathway by combining relevant genes in a heterologous setting is one of the major aims of metabolic engineering programs. Several challenges remain to be addressed before the efficient engineering of complex metabolic pathways becomes routine. For example, because secondary metabolites are not involved in processes essential for survival, the pathways are not well characterized and must be inferred because some or all of the enzymes are unknown and the corresponding genes have not been cloned. Furthermore, the regulatory mechanisms controlling even the best-characterized pathways are often unknown, particularly those influencing transcription. The aim in this chapter was to optimize a transformation method for tobacco plants in order to generate a population of plants expressing genes involved in the early part of the monoterpene secoiridoid pathway, using four binary combinations of genes involved in the early part of the pathway: geraniol diphosphate synthase (*gpps*) and geraniol synthase (*ges*), the latter targeted to three different subcellular compartments (cytosol, mitochondria and chloroplasts). In separate experiments the geraniol synthase (*ges*) and geraniol 10-hydroxylase (*g10h*) genes were also used, yielding a population of transgenic plants containing and expressing these genes.

2.2. Introduction

The development of *in vitro* culture techniques and gene transfer methodology has allowed the improvement of agronomic traits in crops by genetic engineering. The extension of gene transfer technology beyond staple crops has allowed targets other than agronomic traits to be considered. One such target is secondary metabolism. The modulation of plant secondary metabolism was one of the early goals of classical phytochemistry (Hilliou et al., 1999). Studies focusing on the biochemistry and enzymology of secondary metabolism and the isolation of genes encoding the corresponding enzymes led to the first attempts to engineer complex secondary metabolic pathways. Genetic engineering in principle allows the generation of cells, organs, and/or plants with novel profiles of secondary products and the modulation of endogenous pathways to increase or reduce the yields of specific compounds (Hilliou et al., 1999). However, the genetic engineering of secondary pathways to enhance levels of commercially-important molecules has been restricted by the available systems for transformation and the recovery of whole plants from dedifferentiated transformed cells. Until recently, success was limited to single-step extensions of existing metabolic pathways, for example in *Atropa belladonna* (Zeef et al., 2000).

Metabolic engineering has been defined as the modification of metabolic output by introducing recombinant DNA (Peters et al., 2004). Metabolic engineering in plants may be used to achieve a number of different goals (Figure 1):

1) Introduce novel pathways or pathway steps
2) Increase the production of desirable compounds
3) Suppress the production of undesirable compounds
4) Combinations of the above

The first goal is always accomplished by the introduction of heterologous genes, but the other goals may also be achieved by the overexpression or suppression of endogenous genes. This may involve the expression of genes encoding appropriate enzymes controlled by strong promoters, the inhibition of endogenous genes using strategies such as antisense suppression or RNAi (Yazaki et al., 2004), and the targeting of regulatory components (Peters et al., 2004). A tissue-specific promoter

may be suitable in certain cases, e.g. if the intracellular substrate supply is limited to particular cells or the end product accumulates in cells of a special type.

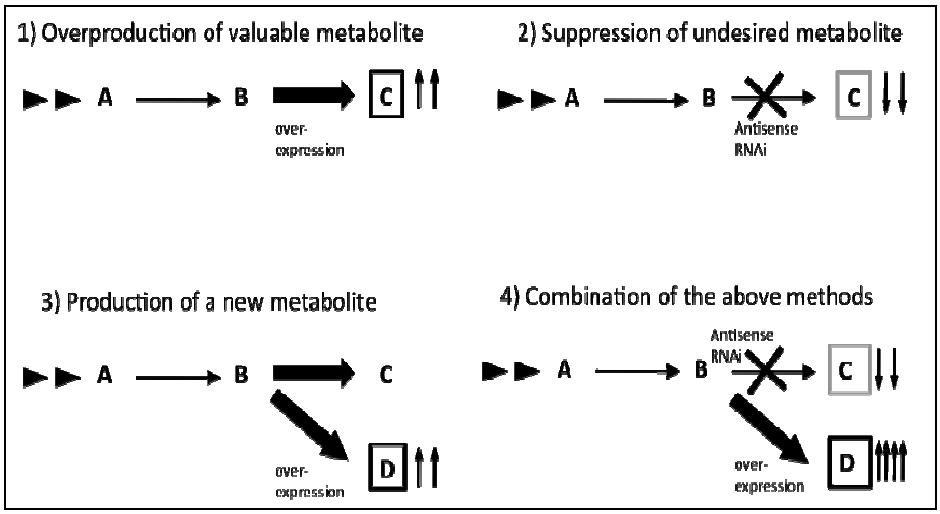

Figure 1. Goals of metabolic engineering (Peters et al., 2004).

The reconstruction of a complete biosynthesis pathway by combining the relevant genes in a host organism is one of the major aims of metabolic engineering. The carotenoid biosynthesis pathway in white maize provides a useful example, because the entire metabolic apparatus must be imported into the endosperm (Naqvi et al., 2009). Promising results have also been reported for the biosynthesis of artemisinin, taxol and strictosidine (Julsing et al., 2006). A number of artemisinin and taxol precursors have been produced in *Escherichia coli* (Martin et al., 2003; Picaud et al., 2005; Newman et al. 2006; Huang et al., 2011; Ajikumar et al., 2010). However, the most promising results include the production of 4.5% artemisinic acid as a proportion of dry weight in *Saccharomyces cerevisiae* (compared to 1.9% in *A. annua*) and the accumulation of 137–827 ng of amorpha-4,11-diene (the precursor of artemisinic acid) per gram fresh weight in transgenic tobacco leaves (compared to 5.2–10.8 mg/g fresh weight in *A. annua*), a useful starting point for artemisinin pathway engineering in plants (Ro et al. 2006; Farhi et al., 2011). Strictosidine has also been produced in different hosts including *E. coli* (Van der Heijden et al., 2004; Kutchan et al., 1999), *S. cerevisiae* (Geerlings et al., 2001) and tobacco cell

suspensions, which produced similar yields to *Catharanthus roseus* (Zarate et al., 2001).

Nevertheless, several challenges remain to be addressed before the efficient engineering of complex metabolic pathways becomes routine. For example, because secondary metabolites are not involved in processes that are essential for survival, the pathways tend not to be well characterized and must be inferred because some or all of the enzymes are unknown and the corresponding genes have not been cloned. Although many genes have been cloned from one or more species, there are few examples of completely-elucidated pathways for specific metabolites (Peters et al., 2004). For example, the terpenoid indole alkaloid (TIA) pathway comprises 30 or more enzymes required for the synthesis of the higher dimeric alkaloids (vinblastine and vincristine) but only 14 genes encoding enzymes (Table 1) and a few more encoding transcription factors have been cloned from *C. roseus* and related species (Liu et al., 2007; Turner et al. 2004; Tholl et al. 2004; Burke et al. 1999; Yang et al., 2011; Simkin et al. 2013; Collu et al., 2001; Ikeda et al. 1991; Murata et al. 2008; Irmler et al. 2000). Furthermore, the regulatory mechanisms controlling even the best-characterized pathways are often unknown, particularly those influencing the level of transcription (Peters et al., 2004).

The search for genes associated with plant secondary metabolism has been slow and challenging, because although genes in a common pathway are often clustered in bacteria, they tend to be dispersed throughout the genome in plants. The corresponding mutants are often unavailable therefore limiting the use of genetic approaches to characterize the steps in a complex pathway. Until the recent advent of large-scale approaches (genomics, transcriptomics, proteomics and metabolomics) each enzyme had to be isolated and sequenced to generate suitable degenerate primers for PCR or cDNA synthesis. Once the genes were cloned in one species it became relatively straightforward to identify and clone orthologs in other plants. Therefore, many secondary metabolism genes have been isolated over the last 10 years, thanks mostly to state-of-the-art “omics” technologies and reverse genetics (see Section 1.4.1.).

Many of these genes have been expressed in recombinant microorganisms and transgenic plants, and in some cases it has been possible to achieve the functional expression of two or even more genes in a pathway simultaneously. Metabolic

engineering has therefore made it possible to produce a few selected benzylisoquinoline alkaloids and key intermediates in the artemisinin and taxane biosynthesis pathways in recombinant *E. coli* and *S. cerevisiae*. Although these processes are still incomplete and not yet commercialized, they represent important steps towards the economic *in vitro* production of secondary metabolites (Schäfer et al., 2009).

Many metabolic reactions are compartmentalized in specific tissues and subcellular organelles, which adds complexity to the process of metabolic engineering. Furthermore, individual pathways may take place in different subcellular compartments and sometimes different tissues, requiring the shuttling of intermediates (Peters et al., 2004). The creation of an artificial metabolic pathway may also influence the overall metabolic flux in plant cells. A new metabolite may be produced, or the natural flow may be negatively regulated by the artificial pathway competing for intermediates. There have been several successful attempts to engineer metabolites that are regulated by biotic and abiotic environmental stresses or development e.g. different stages of cell differentiation specific to metabolite accumulation (Yazaki et al., 2004).

Table 1. List of cloned genes relevant to terpenoid indole alkaloid metabolism.

Gene	Enzyme	Source	Reference
Dxp	1-deoxy-D-xylulose-5-phosphate	*Streptomyces* sp., *Vanda* Mimi Palmer, *Hevea brasiliensis, Coleus forskohlii, Mentha x piperita,*	Lange et al., 1999; Kuzuyama et al, 2000; Chan et al., 2009;; Seetang-Nun et al., 2008; Engprasert et al. , 2005
Dxps	DXP synthase	*Amomum villosum, Withania somnifera, Catharanthus roseus, Croton stellatopilosus, Rhizobium radiobacter, Agrobacterium tumefaciens, Ginkgo biloba, Elaeis guineensis, Morinda citrifolia, Glycine max*	Yang et al., 2012; Gupta et al., 2012 ; Seo et al., 2007; Lee et al., 2007; Gong et al., 2006; Khemvong et al., *2005*; Han et al., 2003; Chahed et al., 2000; Wungsintaweekul et al., 2008; Zhang et al., 2008
Dxr	DXP reductoisomerase	*Ginkgo biloba, Zymomonas mobilis, Catharanthus roseus, Withania somnifera, Vanda Mimi Palmer, Hevea brasiliensis, Camptotheca acuminata, Rauvolfia verticillata, Coleus forskohlii, Arabidopsis thaliana, Artemisia annua, Salvia miltiorrhia*	Lange et al., *2011*; Gupta et al., *2012*; Grolle et al., 2000; Yang et al., 2012; Chan et al., 2009; Seetang-Nun et al., 2008; Yao et al., 2008; Liao et al., 2007; Schwender et al., 2007 ; Gong et al., 2005
Gpps	geranyl diphosphate synthase	*Picea abies, Abies grandis, Antirrhinum majus, Mentha x piperita, Mentha spicata, Phalaenopsis bellina, Arabiopsis Thaliana, Citrus sinensis, C. unshiu, Quercus robur Cinnamomum tenuipilum, Ocimum basilicum, Perilla citriodora, Catharanthus roseus*	Schmidt et al., 2008; Orlova et al., 2009; Burke et al., 1999; Hsiao et al., 2008; Bouvier et al., 2000; Turner et al., 2004;
Ges	geraniol synthase		Yang et al., 2005; Ijima et al., 2004; Ito et al., 2007; Simikin et al., 2013
Cpr	Cytochrome P450 reductase	*Catharanthus roseus, Centaurium erythraea, Nothapodytes foetida*	Schwarz et al., 2009; Huang et al., 2012; Irmler et al., 2000

Table 1. List of cloned genes relevant to terpenoid indole alkaloid metabolism **(cont.)**

Gene	Enzyme	Source	Reference
g10h	Geraniol-10-hydroxylase	*Catharanthus roseus, Swertia mussotii Franch*	Meijer et al., 1993; Wang et al., 2010a
10hgo	10-hydroxygeraniol oxidoreductase	*Camptotheca acuminate*	Valletta et al., 2010
As	Anthranilate synthase	*Arabidopsis thaliana, Azospirillum brasilense, Ruta graveolens*	De Troch et al., 1997; Bohlmann et al., 2003
Tdc	Tryptophan decarboxylase	*Oryza sativa, Catharanthus roseus*	Kang et al., 2007; De Luca et al., 1989
Str	Strictosidine synthase	*Rauwolfia verticillata, Catharanthus roseus*	Kutchan et al., 1988; Geerling et al., 2000
Sgd	Strictosidine glycosidase	*Catharanthus roseus, Rauvolfia serpentine*	Gerasimenko et al., 2002; Geerling et al., 2000
D4h	Desacetoxyvindoline 4-hydroxylase	*Catharanthus roseus*	Zhou et al., 2010; Vazquez-Flota et al., 1997
Dat	Deacetylvindoline acetyltransferase	*Catharanthus roseus*	Wang et al., 2010b
Mat	Minovincinine-19hydroxy-o-Acetyltransferase	*Catharanthus roseus*	Laflamme et al., 2001

2.3. Aims and objectives

The overall aim in this chapter was to adapt a previously-developed transformation method to engineer the early part of the monoterpene secoiridoid pathway in tobacco.

The specific objectives were:

- Develop a multigene transformation system for tobacco plants using direct DNA transfer.
- Introduce the early genes leading to geraniol formation, together with a selectable marker gene for kanamycin resistance to generate transgenic tobacco plants containing and expressing these genes.
- Select and regenerate putative transgenic plants.
- Analyze mRNA expression levels in transgenic plants to validate the transformation system.
- Create a population of plants expressing the early genes of the pathway for further analysis.

2.4. Materials and methods

2.4.1. Donor plants and explant preparation

Tobacco (*Nicotiana tabacum* cv. Petite Havana SR1) was selected as the host species. Tobacco seeds were surface-sterilized with 70% (v/v) ethanol for 30–60 s and rinsed with sterile water. The sterile seeds were transferred to MS medium (Murashige and Skoog, 1963) plus vitamins, 3% (w/v) sucrose and 0.8% (w/v) agar. Seedlings were grown at 23ºC with a 16-h photoperiod for approximately 3 weeks until the germ buds were large enough for transfer to a glass jar containing MS medium. Plants were grown under aseptic conditions. Leaves from these plantlets were cut into 1 cm^2 segments, placed on MS medium and incubated as above for 24-h with a 12-h photoperiod before transformation.

2.4.2. Plant transformation

Leaf explants were transformed by direct DNA transfer (particle bombardment). To establish the transformation and regeneration protocol, the explants were initially transformed with vector pTrak-AH, containing a kanamycin resistance marker (*npt*II) under the control of the constitutive CaMV 35S promoter (Figure 2), and plants were regenerated on medium supplemented with kanamycin.

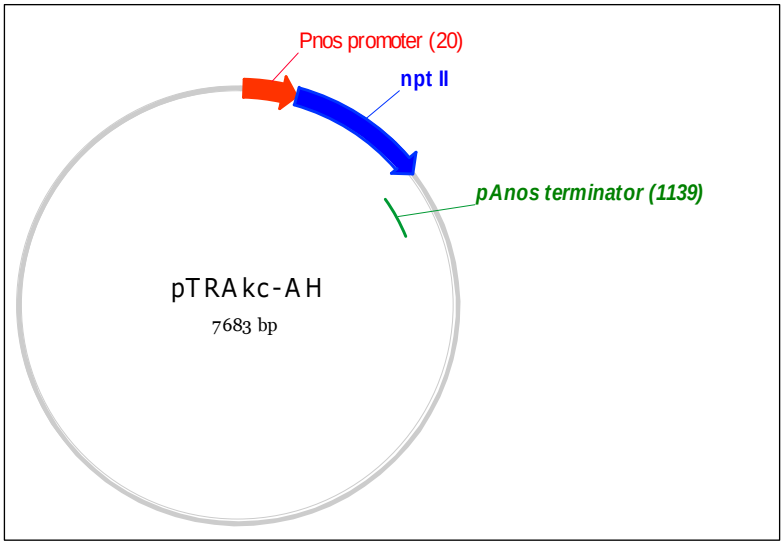

Figure 2. Transformation vector pTRAkc-AH, containing the kanamycin resistance marker *npt*II (provided by Prof. S. Schillberg, RWTH, Aachen, Germany).

Leaf explants were transformed with several combinations of the genes listed in Table 2. These comprised the first three genes in the monoterpene secoiridoid pathway: geranyl diphosphate synthase (*gpps*), geraniol synthase (*ges*) (each with appropriate signal peptides targeting the mitochondria and chloroplasts), and the subsequent gene geraniol 10-hydroxylase (*g10h* also known as *cyp76b6*). The constructs for *Arabidopsis thaliana gpps* and *Valeriana officialis ges* were provided by H. Bouwmeester (Wageningen University, Netherlands) (Figure 3). The construct for *Catharanthus roseus cyp76b6* in vector pMOG463 was provided by D. Werk (CNRS,

Strasbourg, France). All genes were controlled by the constitutive double CaMV 35S promoter. Transformation was carried out using a 3:3:1 molar ratio of plasmids carrying the two genes of interest and the selectable marker gene (Table 3).

Table 2. Combinations of constructs used for tobacco transformation experiments

Combination	*gpps*	*ges*	*g10h*
A	pIV2A2.1 *AtGpps* (cytosol)	pIV2B2.1*VoGes* (cytosol)	
B	pIV2A2.4 *AtGpps* (chloroplast)	pIV2B2.4 *VoGes* (chloroplast)	
C	pIV2A2.5 *AtGpps* (mitochondria)	pIV2B2.5 *VoGes* (mitochondria)	
D		pIV2B2.4 *VoGes* (chloroplast)	pMOG463-Cr*Cyp76b6*

Table 3. Plasmid DNA ratios for binary gene combinations plus the selectable marker gene

	Combination	**DNA amount** (µg)
A	pIV2A2.1-*AtGpps*:pIV2B2.1-*VoGes*:pTRAkc-AH	16.01:17.48:6.57
B	pIV2A2.4-*AtGpps*:pIV2B2.4-*VoGes*:pTRAkc-AH	16.13:17.55:6.32
C	pIV2A2.5-*AtGpps*:pIV2B2.5-*VoGes*:pTRAkc-AH	16.06:17.05:6.51
D	pIV2B2.4-*VoGes*: pMOG463-*Cyp76b6*	19.1:14.01:6.93

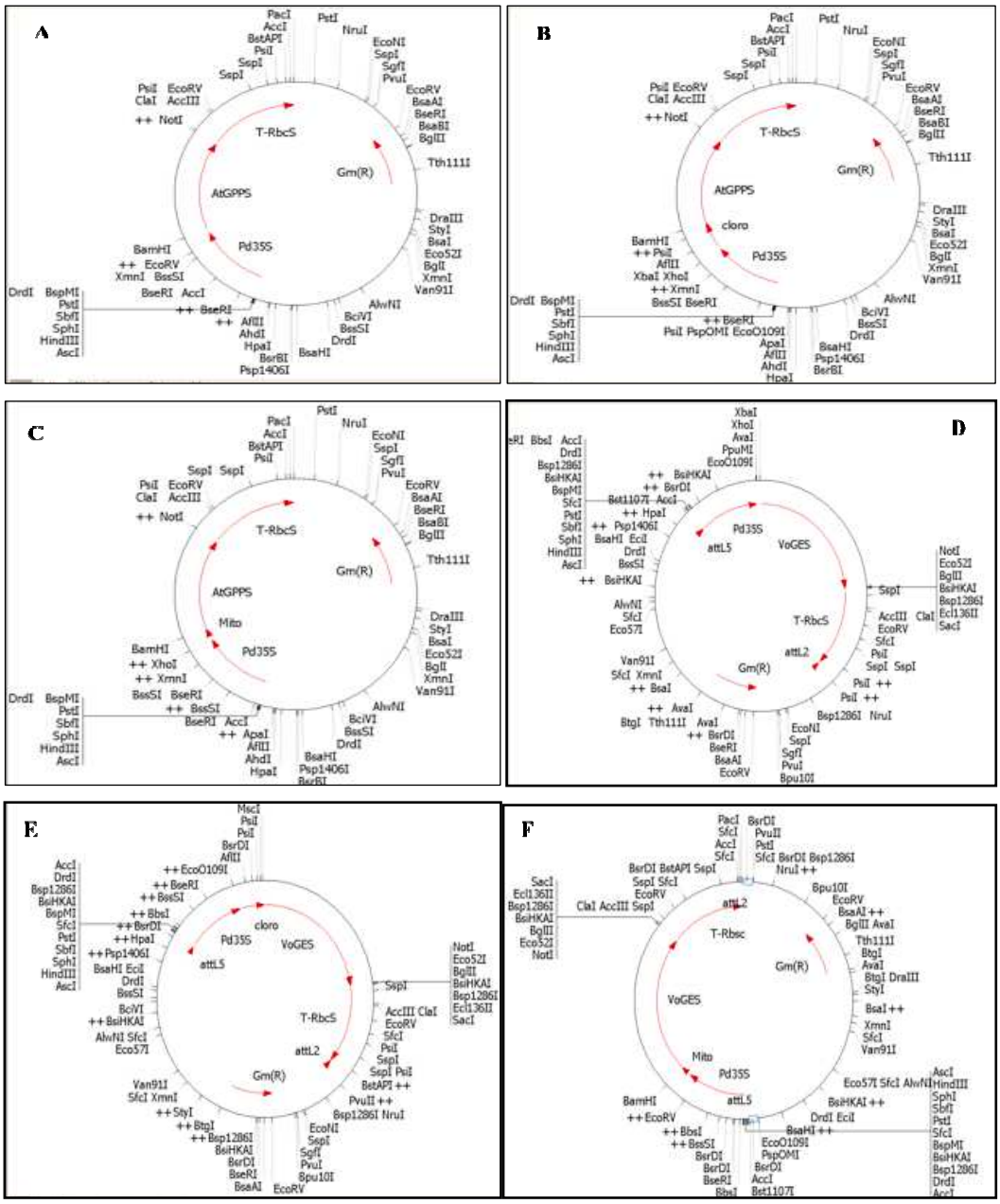

Figure 3. Transformation constructs for stable expression in tobacco plants. **A.** pVI2A2.1-AtGPPS containing *AtGpps* gene with no targeting signal. **B.** pVI2A2.4-AtGPPS containing *AtGpps* gene with a chloroplast targeting signal. **C.** pVI2A2.5-AtGPPS containing *AtGpps* gene with a mitochondrial targeting signal. **D.** pVI2B2.1-VoGES containing the *VoGes* gene with no targeting signal. **E.** B2.4-VoGES containing the *VoGes* gene with a chloroplast targeting signal. **F.** pVI2B2.5-VoGES containing the *VoGes* gene with a mitochondrial targeting signal.

2.4.3. Selection and regeneration of putative transgenic plants

Following transformation with the kanamycin resistance marker alone to validate the system, leaf discs were incubated overnight in the dark at 25ºC and then transferred to fresh MS1 medium, i.e. MS medium supplemented with 6-benzylaminopurine (BAP; 1 mg/l) and 1-napthalene acetic acid (NAA; 0.1 mg/ml) (Table 1 in annex 1). After 48 h, half of the bombarded leaf discs were transferred to MS2 selection medium, which is MS1 medium supplemented with 100 mg/l kanamycin sulfate (Table 1 in annex 1) as described by Leech et al. (1998). The other half was placed onto MS1 medium without kanamycin. The transformed leaf discs were compared to non-transformed leaf discs maintained under the same conditions but without transformation.

After 4–6 weeks, the non-transformed leaf discs on kanamycin-supplemented media did not develop shoots or callus, as expected (Figure 4a). The transformed leaf discs without selection developed a large number of shoots when bombarded at two different intensities (16 and 18 kV), although more shoots regenerated at the lower bombardment intensity. The transformed leaf discs under kanamycin selection produced few shoots and a small amount of callus (Figure 4b).

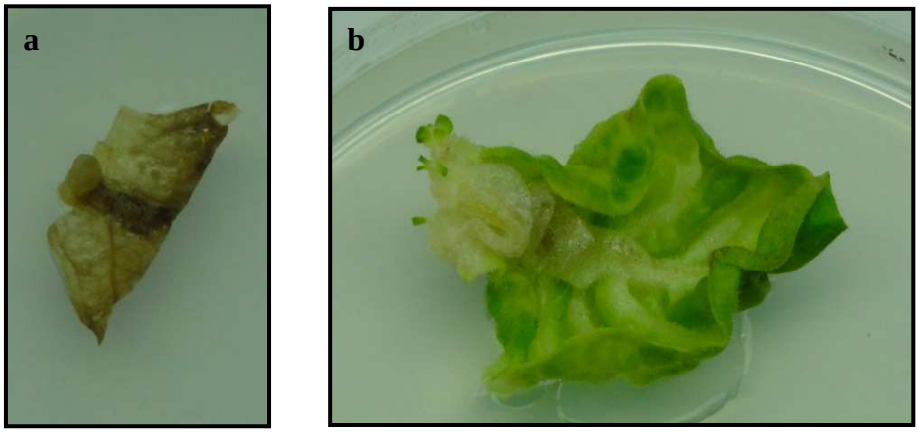

Figure 4. **A.** Non-bombarded leaf explant under kanamycin selection, showing no callus or shoot development. **B.** Bombarded leaf explant under kanamycin selection, showing initial callus and shoot development.

The transformation experiments using the different gene combinations with the selectable marker followed the procedure described in Section 2.5.4. After 4–6 weeks on MS2 medium, putative transgenic shoots were picked off and transferred to fresh MS3 medium, which is MS medium supplemented only with 100 mg/l kanamycin sulfate to promote root development (Table 1 in annex 1). After 8–12 weeks, developing plantlets with well formed shoots (>4–5 cm) and roots (>1–2 cm) were hardened off in soil (Figure 5). All experiments were carried out at 25ºC and all media constituents are listed in Table 1 in annex 1. The plants were self-pollinated and seeds from putative transgenic plants were germinated under kanamycin selection as described above.

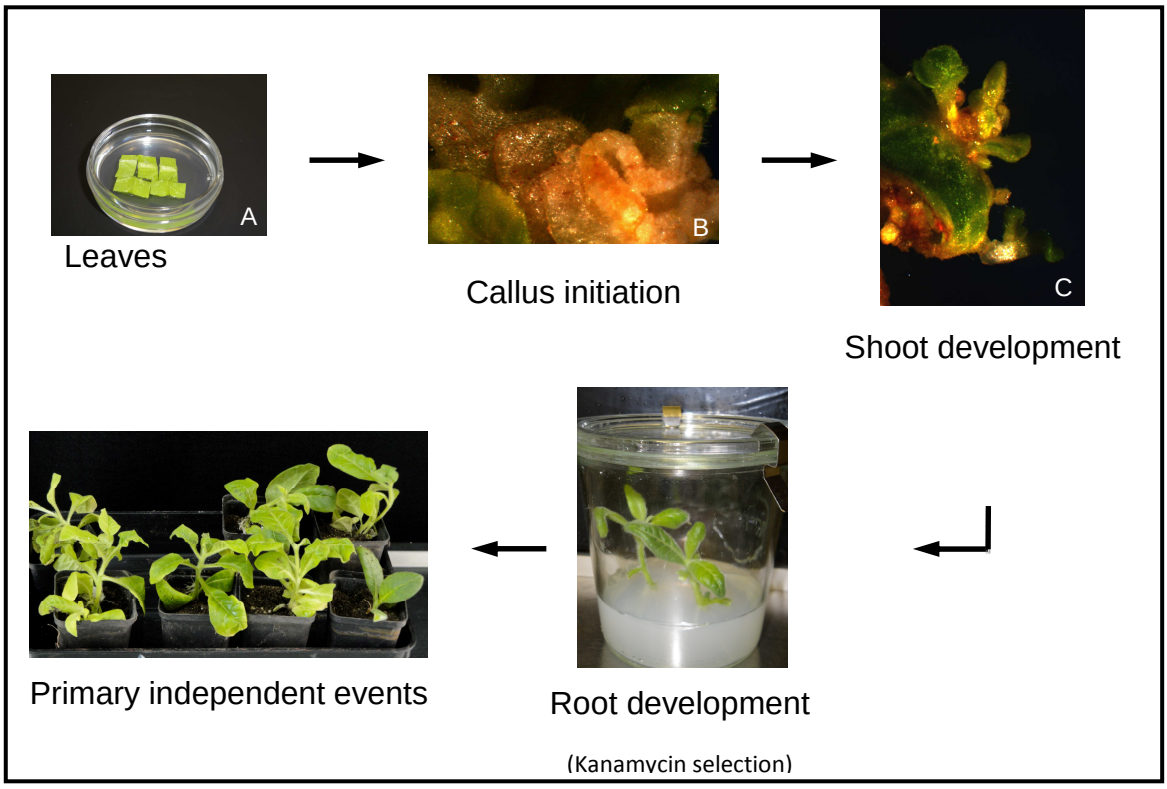

Figure 5. Selection and regeneration of transgenic tobacco plants

2.4.4. DNA analysis

2.4.4.1. Genomic DNA extraction

Genomic DNA was extracted from leaf material as described by Dellaporta et al. (1993), using 300 mg of frozen leaf tissue ground to a fine powder and mixed with 4 ml of extraction buffer (500 mM NaCl, 100 mM Tris-HCl, 50 mM EDTA, 20% SDS; pH 8). The mixture was shaken, incubated for 10 min at 65ºC, mixed with 4 ml 1:1 phenol:chloroform (Sigma, Steinheim, Germany) and centrifuged at 4500 rpm for 5 min at room temperature. The supernatant was removed, mixed with 150 μg RNase (Roche, Indianapolis, The USA) and incubated at 37ºC for 1 h. The phenol:chloroform extraction step was repeated and the mixture was centrifuged at 5000 rpm for 5 min at room temperature. The supernatant was then removed and genomic DNA was precipitated by adding 4 ml of isopropanol and incubating at room temperature for 1 h. The precipitated DNA was removed with a thin glass rod and washed in 1 ml 70% ethanol. Samples were centrifuged at 5000 rpm for 5 min at room temperature and the genomic DNA pellet was air-dried and dissolved in sterile distilled water. The DNA concentration was determined using a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and the quality was determined by analyzing 1 μl of isolated genomic DNA by 0.8% agarose gel electrophoresis.

2.4.4.2. Polymerase chain reaction (PCR)

The putative transgenic tobacco plants were initially characterized by polymerase chain reaction (PCR) using the forward and reverse primers listed in Table 4 to amplify 605-bp, 684-bp and 898-bp fragments within the *AtGpps*, *VoGes* and *Cyp76b6* genes, respectively. The PCR mixtures are listed in Table 5, and the thermal cycling conditions are summarized in Table 6. The pIV2A2.4-*AtGpps*, pIV2B2.4-*VoGes* and pMOG463-*Cyp76b6* plasmids (150 ng) were used as a positive controls and 150 ng of genomic DNA was used per sample. The PCR products, together with a DNA molecular weight marker (BenchTop 1-kb DNA Ladder, Promega, Madrid, Spain) were separated by 0.8% (w/v) agarose gel electrophoresis and visualized by ethidium bromide staining under UV light.

Table 4: Forward and reverse primers used to design probes for DNA and RNA blots.

Gene	Forward primer	Reverse primer
AtGpps	5'- GGGTGTGCAAGGAAAACAGT-3'	5'- TTTCCGAGAGAGGCAGATGT-3'
VoGes	5'-GGCTGATTCTTCCTTGCTTG-3'	5'-AAATCGTCCCACAGTCGAAGAA-3'
Cyp76b6	5'-TCCCCTGCTTGAAAAAGTTG-3'	5'-CACCCTTGTGAATTCGTCCT-3'

Table 5: PCR reagent mix

Component	PCR mix	Final concentration
5X GoTaq buffer (Promega)	5 µl	1 X
10 mM dNTPs	1 µl	0.2 mM
Forward primer (20 mM)	2.5 µl	1 mM
Reverse primer (20 mM)	2.5 µl	1 mM
Taq polymerase (Promega)	0.25 µl	1.25 U
Distilled water	36.75 µl	
DNA sample	2 µl	150 ng
Total PCR volume per reaction	50 µl	

Table 6: PCR thermal cycling conditions

	Temperature	Time	
Initial denaturation	95ºC	3 min	
Denaturation	94ºC	45 s	**34 cycles**
Annealing	60ºC	45 s	
Elongation	72ºC	2 min	
Final elongation	72ºC	10 min	

2.4.5. mRNA analysis

2.4.5.1. RNA extraction

Total RNA was extracted from 120 mg tobacco leaf tissue using Trizol® reagent (Invitrogen, Paisley, UK). Frozen leaf tissue was ground to a fine powder and incubated with 1.2 ml Trizol® for 5 min at room temperature, mixed with 240 µl of chloroform and centrifuged at 13,000 rpm for 15 min at 4ºC. The supernatant was transferred to a fresh tube, incubated with 600 µl isopropanol for 10 min at room temperature, centrifuged at 13,000 rpm for 10 min at 4ºC and washed with 1 ml 70% ethanol for 1 h. The samples were centrifuged at 13,000 rpm for 15 min at 4ºC and dissolved in sterile double distilled water that had been autoclaved with 1 µl of RNase inhibitor (Roche Diagnostics GmbH, Mannheim, Germany). RNA concentrations were determined using a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and quality was determined by analyzing 2 µl of isolated RNA by 1.2% agarose gel electrophoresis.

2.4.5.2. Reverse transcriptase polymerase chain reaction (RT-PCR)

Transgene expression was monitored by reverse transcriptase polymerase chain reaction (RT-PCR). DNA was removed from the total RNA using the RQ1 RNase-free DNase kit (Promega, Wisconsin, USA). Semi-quantitative RT-PCR was carried out by using 1 µg RNA as the template for first-strand cDNA synthesis with Omniscript® Reverse Transcription Kit (QIAGEN, Hilden, Germany). Then, 1–5 ng of cDNA was used for amplification as described in Section 2.4.5.2. The pIV2A2.4-*AtGpps*, pIV2B2.4-*VoGes* and pMOG463-CYP76B6 plasmids (150 ng) were used as positive controls and 10 µl of cDNA was used per sample. RT-PCR products, together with a DNA molecular weight marker (BenchTop 1-kb DNA Ladder, Promega, Madrid, Spain) were separated by 0.8% (w/v) agarose gel and visualized by ethidium bromide staining under UV light.

2.4.5.3. mRNA (northern) blot analysis

RNA aliquots (30 μg) were denatured and fractionated by 1.2% agarose-formaldehyde gel electrophoresis in 1x MOPS buffer (Sambrook et al., 1989). The RNA was transferred to a positively-charged nylon membrane (Roche Diagnostics GmbH, Mannheim, Germany) by standard capillary transfer (Sambrook et al. 1989). DIG-labeling was carried out as described by Capell et al. (2004) and each probe was purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The probes were denatured at 95ºC for 10 min and hybridized overnight at 50ºC in DIG Easy Hyb buffer (Roche Diagnostics GmbH, Mannheim, Germany). Membranes were washed twice for 5 min in 2x SSC + 0.1% SDS at room temperature, twice (25 min) in 0.5x SSC + 0.1% SDS, once (15 min) in 0.2x SSC + 0.1% SDS, and once (10 min) in 0.1 % SDS at 68ºC. After washing and immunological detection with anti-DIG-AP (Fab-Fragments, Diagnostics GmbH, Mannheim, Germany), chemiluminescence was detected by CSPD (Roche Diagnostics GmbH, Mannheim, Germany) on Kodak BioMax light film (Sigma-Aldrich, USA).

2.5. Results

2.5.1. Recovery of transgenic plants

In the early experiments, 13 kanamycin-resistant putative transgenic plantlets were recovered following bombardment with the *AtGpps* and *VoGes* combination plus the kanamycin resistance marker, but only 10 survived the 3–5 rounds of selection (Table 7). Another 38 kanamycin-resistant putative transgenic plantlets were recovered following transformation with *VoGes, Cyp76b6* and the kanamycin resistance marker.

Non-transgenic shoots appearing on leaf segments after bombardment died during the subsequent subculture steps, whereas transgenic shoots proliferated and regenerated into intact plants. All putative transgenic plants were self-pollinated and showed normal morphological and developmental characteristics.

Table 7. Regenerated plantlets from GPPS/GES transformation experiments

Number of regenerated plantlets	Genes	Plantlets surviving after selection
3	pIV2A2.1- *AtGpps* + pIV2B2.1- *VoGes* (cytosol)	3
5	pIV2A2.4 - *AtGpps* + pIV2B2.4- *VoGes* (chloroplast)	4
5	pIV2A2.5- *AtGpps* + pIV2B2.5- *VoGes* (mitochondria)	3

2.5.2. DNA analysis

Genomic DNA extracted from 10 independent GPPS/GES putative transgenic plants was analyzed by PCR (Figures 6 and 7) revealing the presence of the expected 605-bp *AtGpps* DNA fragment in six of the 10 transgenic plants, and the expected 684-bp *VoGes* DNA fragment in seven of them (Table 8). The positive controls also produced the 605-bp and 684-bp DNA fragments whereas no amplification products were observed in the negative control lane (with no template DNA) or in the DNA from wild-type plants. The *AtGpps* and *VoGes* genes were integrated in 60% and 70% of the lines respectively, and 50% of the lines contained both genes (Table 9).

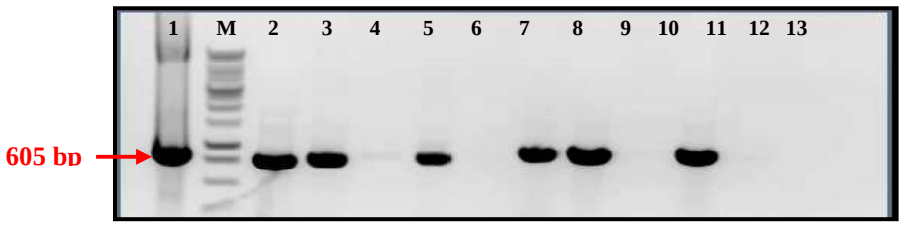

Figure 6. PCR amplification of *AtGpps* in leaf genomic DNA from putative GPPS/GES transgenic tobacco plants. Lane 1: positive control plasmid pIV2A2.4-*AtGpps*; Lane M: DNA molecular weight marker (1-kb ladder). Lanes 2–11: genomic DNA (150 ng) from transgenic events 1–10. Lane 12: negative control wild-type genomic DNA. Lane 13: no DNA template control. Positive events are identified by the presence of the expected 605-bp fragment.

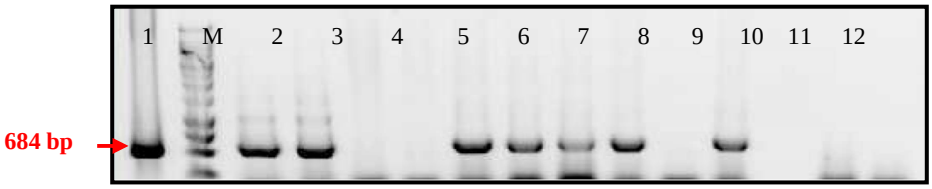

Figure 7. PCR amplification of *VoGes* in leaf genomic DNA from putative GPPS/GES transgenic tobacco plants. Lane 1: positive control plasmid pIV2A2.4-*VoGes*; Lane M: DNA molecular weight marker (1-kb ladder). Lanes 2–11: genomic DNA (150 ng) from transgenic events 1–10 (events 10 and 9 are loaded out of order, in lanes 10 and 11 respectively). Lane 12: no sample. Lane 13: negative control wild-type genomic DNA. Lane 14: no DNA template control. Positive events are identified by the presence of the expected 684-bp fragment.

Table 8. PCR results from the GPPS/GES transformation experiment. The input genes were *AtGpps* and *VoGes*, and the selectable marker gene was *nph*II. (+) and (-) indicate the presence and absence of a PCR amplification product, respectively.

Plant	**Target**	***AtGpps***	***VoGes***
1	Cytosol	+	+
2	Cytosol	+	+
3	Cytosol	+	-
4	Chloroplast	+	-
5	Chloroplast	-	+
6	Mitochondria	+	+
7	Mitochondria	+	+
8	Chloroplast	-	+
9	Mitochondria	+	+
10	Chloroplast	-	-

Table 9. Frequency of transgene presence detected by PCR.

Gene	Frequency of transgene presence
AtGpps	60%
VoGes	70%
AtGpps + VoGes	50%

The GES/CYP76B6 transformation experiment yielded 38 putative transgenic plants. PCR results (Figures 8 and 9) revealed the presence of the expected 684-bp *VoGes* DNA fragment in 30 plants and the expected 898-bp *cyp76b6* DNA fragment in 33 plants (Table 10). The positive controls also produced the expected 684-bp and 898-bp DNA fragments whereas no amplification products were observed in the negative control lane (no template DNA) or the wild-type DNA. The *VoGes* and *cyp76b6* genes were integrated in 64.3% and 62.5% of the lines respectively, and 58.9% of the lines contained both genes (Table 11).

Figure 8. PCR amplification of *VoGes* in leaf genomic DNA from putative GES/CYP76B6 transgenic tobacco plants. Lane 1: positive control plasmid pIV2A2.4-*VoGes*; Lane M: DNA molecular weight marker (1-kb ladder). Lanes 1–10: genomic DNA (150 ng) from putative transgenic plants. Lane WT: negative control wild type DNA. Lane N: negative control no DNA template. Positive events are identified by the presence of the expected 684-bp fragment.

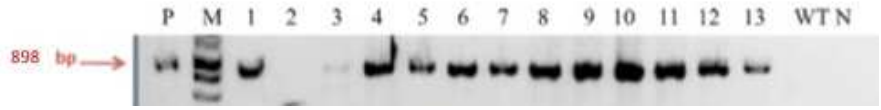

Figure 9. PCR amplification of *cyp76b6* in leaf genomic DNA from putative GES/CYP76B6 transgenic tobacco plants. Lane P: positive control plasmid pIV2B2.4-*Cyp76b6*. Lane M: DNA molecular weight marker (1-kb ladder). Lanes 1–13: genomic DNA (150 ng) from putative transgenic plants. Lane WT: negative control wild type DNA. Lane N: negative control no DNA template. Positive events are identified by the presence of the expected 898-bp fragment.

Table 10. PCR results from the GES/CYP76B6 transformation experiment. The input genes were *VoGes* and *cyp76b6*, and the selectable marker gene was *npt*II. (+) and (-) indicate the presence and absence of a PCR amplification product, respectively.

Plant	***VoGes***	***cyp76b6***	**Plant**	***VoGes***	***Cyp76b6***
1	-	-	20	+	+
2	+	+	21	-	+
3	+	+	22	-	+
4	-	+	23	-	+
5	+	-	24	+	+
6	-	+	25	+	-
7	-	-	26	+	+
8	+	+	27	+	+
9	+	+	28	+	+
10	+	+	29	+	+
11	+	+	30	+	+
12	+	+	31	+	+
13	+	+	32	+	+
14	+	+	33	+	+
15	-	+	34	+	+
16	+	+	35	+	+
17	+	-	36	+	+
18	+	+	37	+	+
19	+	+	38	+	+

Table 11. Frequency of transgene presence detected by PCR.

Gene	**Frequency of transgene presence**
VoGes	64.3%
cyp76b6	62.5%
VoGes + cyp76b6	58.9%

2.5.3. mRNA analysis

After screening for the presence of each input transgene, expression analysis revealed that three plants from the GPPS/GES experiment expressed *AtGpps* and *VoGes*, and 27 plants from the GES/CYP76B6 experiment expressed *VoGes* and *cyp76b6*.

These expression profiles were confirmed by mRNA blot analysis. The *AtGpps* and *VoGes* transcripts were detected in events 1, 7 and 9 (lanes 1, 4 and 6 in Figure 10A and lanes 2, 4, and 7 in Figure 10B) whereas the other events expressed neither transgene. Expression of the selectable marker was confirmed by the germination of T_1 seeds from the putative transgenic tobacco plants under kanamycin selection, with a Mendelian segregation ratio of 3:1 (Figure 11). Nineteen representative transgenic plants from the GES/CYP76B6 experiment were also analyzed by mRNA blot to confirm the expression of the *VoGes* and *cyp76b6* transgenes, 13 of which are shown in Figure 12.

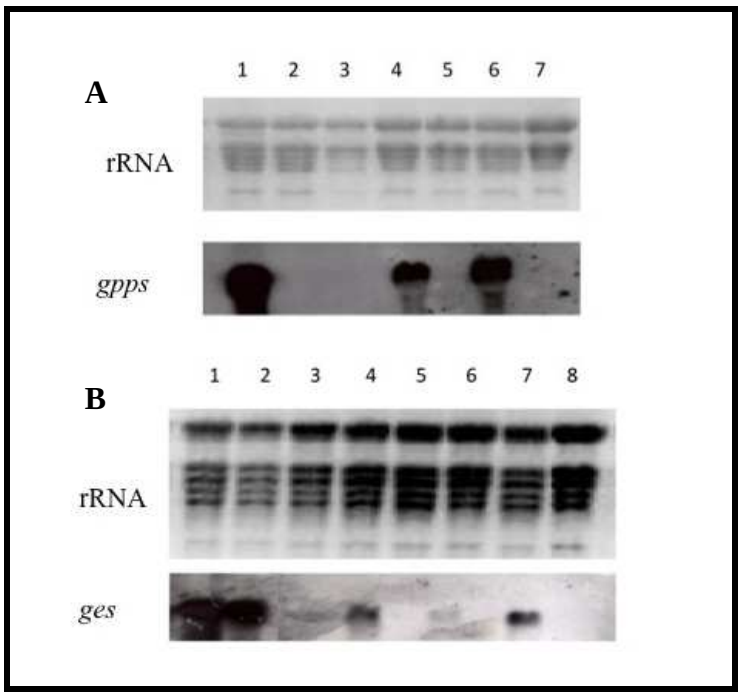

Figure 10. Gel blot analysis of 30 µg total leaf RNA from the GPPS/GES experiment probed with **(A)** *AtGpps* (lanes 1–6 represent transgenic events 1, 2, 4, 7, 8 and 9; lane 7 is the wild-type negative control) and **(B)** *VoGes* (lane 1 is a positive control plant expressing *VoGes* in the cytosol, provided by Wageningen University; lanes 2–7 represent transgenic events 1, 2, 7, 5, 6, 8 and 9; lane 8 is the wild-type negative control).

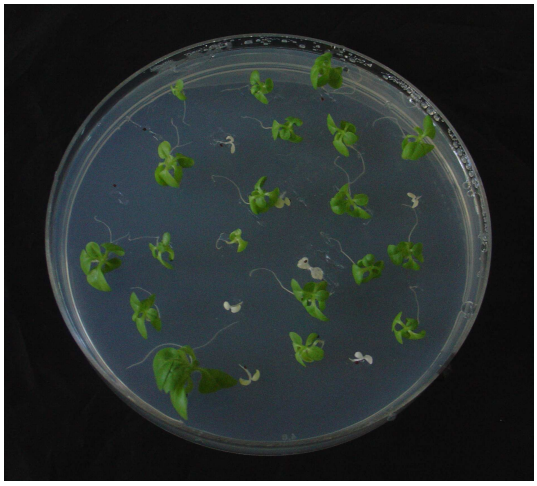

Figure 11. T_1 seeds germinating under kanamycin selection (19 of 25 seeds, representing a Mendelian segregation ratio of 3:1).

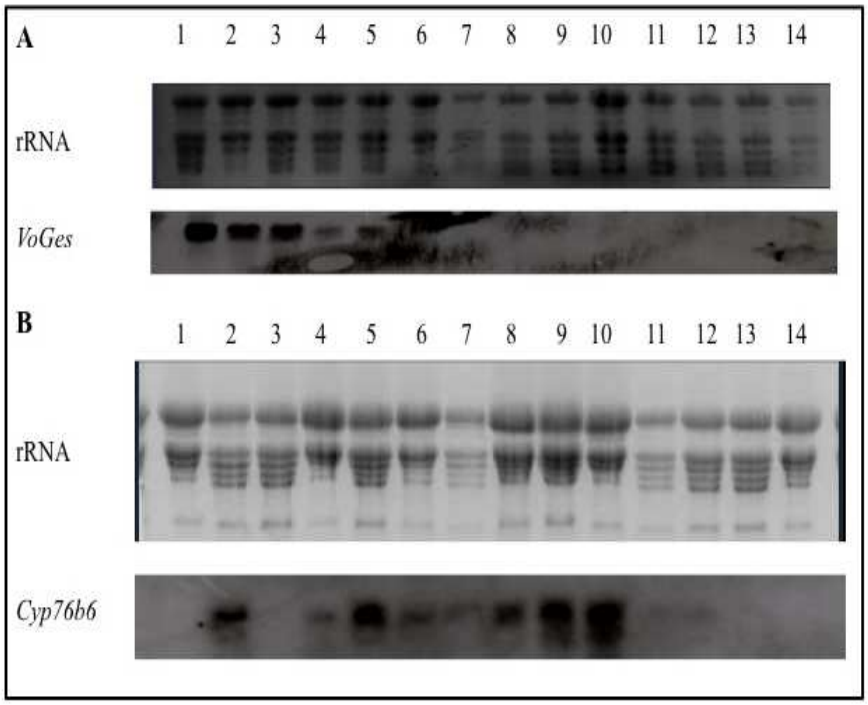

Figure 12. Gel blot analysis of 30 µg total leaf RNA from the GES/CYP76B6 experiment probed with **(A)** *VoGes* and **(B)** *cyp76b6*. In each panel, lanes 1–13 represent different transgenic events and lane 14 is the wild-type negative control.

2.6. Discussion

The universal precursor of monoterpenes is geranyl diphosphate (GPP) produced by the action of geranyl diphosphate synthase (GPPS), which catalyzes the condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) into a C_{10} product. This enzyme is similar to farnesyl diphosphate synthase (FPPS), which condenses two molecules of IPP with one molecule of DMAPP to form the C_{15} precursor of the sesquiterpenes and triterpenes, and to geranylgeranyl diphosphate synthase (GGPPS), which condenses three molecules of IPP with DMAPP to form the C_{20} precursor of diterpenes and tetraterpenes (Burke et al., 2002). These enzymes, collectively termed short-chain prenyltransferases, are homologous and function at the branch points of isoprenoid metabolism where they regulate the flux distribution of IPP into the various terpenoid families (Burke et al., 2002; Burke et al., 1999; Bouvier et al., 2000; Tholl, 2006).

Monoterpenoid biosynthesis in plants occurs in different subcellular compartments, including the endoplasmic reticulum, cytosol, plastids and mitochondria (Orlova et al., 2009; Bouvier et al., 2000). Until recently, monoterpene synthesis was thought to occur exclusively in non-green plastids, using IPP and DMAPP formed via the plastidial methyl-erythriol-phosphate (MEP) pathway (Orlova et al., 2009). This compartmentalization is partly based on the toxicity of monoterpenes at concentrations as low as 100 ppm, which inhibits growth and photosynthetic electron transfer, and causes membrane disorganization (Bouvier et al., 2000; Orlova et al., 2009). To circumvent these effects, monoterpenes are sequestered into specialized secretory structures such as the glandular trichomes of catmint (*Nepeta racemosa*) which accumulate approximately 30 ng of the monoterpenoid nepetalactone. This mechanism allows monoterpenes to accumulate up to 1.5–20% of the dry weight of lemongrass (*Cymbopogon* spp.) and *Eucalyptus* spp. leaves. A characteristic feature of these secretory structures is the presence of non-green plastids that may form non-specialized structures (Bouvier et al., 2000).

The plastid localization of GPPS, the key branch-point enzyme leading to monoterpene biosynthesis, was anticipated because the GPPS pre-protein includes a predicted N-terminal plastid-targeting transit peptide and because the substrates IPP and DMAPP are provided by the plastidial MEP pathway (Turner et al., 2004; Burke

et al., 2002). Studies in *Citrofortunella mitis* and *Narcissus pseudonarcissus* demonstrated that GPPS is indirectly localized in leucoplasts and chloroplasts, as well as in the non-green plastids of secretory cells, which are widely recognized as the sites of monoteprene synthesis in plants (Turner et al., 2004; Bouvier et al., 2000). The dual expression of two isoforms targeted to different cellular compartments from a single gene is not unique (Bouvier et al., 2000). In the isoprenoid pathway for example, it has been shown that one Arabidopsis FPPS gene encodes both a cytosolic and a mitochondrial form of the enzyme (Bouvier et al., 2000). The potential dual targeting of GPPS, coupled to the possible exchange of prenyl diphosphate substrates between the plastid and cytosolic compartments, may reconcile the contradictory data from previous reports describing GPPS exclusively in the plastids or cytosol but not in both compartments (Bouvier et al., 2000). Exchange between the plastid and the cytosol could occur at the level of IPP or GPP (Bouvier et al., 2000). IPP and perhaps DMAPP can be exchanged between compartments, particularly the plastids to the cytosol. The trafficking of isoprenoid intermediates between organelles, probably mediated by specific metabolite transporters, depends on the plant species, tissue and physiological status of the plant (Orlova et al., 2009).

Despite the ubiquity of prenyltransferases, GPPS has been investigated in a limited number of species, where it may exist as either a homodimer or heterodimer (Orlova et al., 2009; Schmidt et al., 2008; Bouvier et al., 2000; Burke et al., 1999). The first heterodimeric GPPS was found in peppermint (*Mentha x piperita*), which is a model for the study of monoterpene biosynthesis (Turner et al., 2004; Burke et al., 1999). The enzyme was only functional as a heterodimer and the deduced amino acid sequences of the 28-kDa and 37-kDa subunits have only minimal similarity (Schmidt et al., 2008). Similar heterodimeric forms of GPPS have also been identified in *Antirrhinum majus* and *Clarkia breweri* (Tholl et al., 2004; Turner et al., 2004). The large subunit of these enzymes shares 62–75% identity with GGPPS pre-proteins but only approximately 25% identity with FPPS. The small subunit shows minimal similarity to GGPPS (Burke et al., 2002; Schmidt et al., 2008). The greater conservation between GPPS and GGPPS (compared to FPPS) suggests that both GPPS subunits are evolutionarily derived from plastidial GGPPS rather than cytosolic FPPS (Burke et al., 1999; Turner et al., 2004).

The heterodimeric subunit architecture of GPPS was unexpected because other short-chain prenyltransferases are homodimers. However, a precedent for heterodimeric prenyltransferases was established by microbial hexaprenyl and heptaprenyl diphosphate synthases (Burke et al., 1999).

The only homodimeric GPPS described thus far is found in Arabidopsis (Schmidt et al., 2008) but similar albeit uncharacterized sequences are present in GenBank e.g. from *Citrus sinensis, C. unshiu* and *Quercus robur* (Schmidt et al., 2008). The difference observed between peppermint and Arabidopsis GPPS could reflect the localization of heterodimeric peppermint GPPS in leucoplast trichomes, whereas the Arabidopsis homodimeric form (36–38 kDa) is compartmentalized in parenchymal tissues containing chloroplasts or non-green plastids despite the presence of predicted mitochondrial targeting sequences (Bouvier et al., 2000). This was confirmed by immunofluorescence localization in Arabidopsis, *Pinus* spp. and *Citrofortunella* spp., as well as strong labeling of secretory cell leucoplasts in *Pinus* and *Citrofortunella*. The abundant and constitutive labeling observed in the chlorenchyma of diverse species contrasts with the findings in peppermint, perhaps reflecting the reliance on crude polyclonal antisera for immunolocalization, which may encourage cross-reaction (Turner et al., 2004). Presumably, the subcellular distribution of GPPS is an important factor affecting substrate allocation for the production of monoterpenes (Turner et al., 2004).

Currently, there is almost no information available about endogenous GPPS enzymes and their corresponding genes in tobacco, which produces only small amounts of volatile terpenoids, although four candidate genes have been identified (Orlova et al., 2009). The small GPPS subunit from snapdragon (*AmGpps.ssu*) has been overexpressed in tobacco and can find endogenous partner(s), modifying the product specificity and shifting the metabolic flux towards GPP formation, thus affecting terpenoid biosynthesis *in planta* (Orlova et al., 2009). The expression of *AmGpps.ssu* into tobacco plants suggests that the protein has a strong affinity for tobacco GGPPS and can prevent or compete with GGPPS homodimerization, at least the expression levels detected in transgenic plants (Orlova et al., 2009). Because GGPPS and GPPS are both localized in plastids, this suggests that the enzyme subunits can interact at any time to modulate product specificity and produce GPP efficiently (Orlova et al., 2009). To prevent uncontrolled or undesirable heterodimer formation in plastids,

leading to potentially deleterious effects on GGPP-dependent metabolic pathways, the level of GPPS must be tightly regulated. Such regulation in plants appears to occur at the transcriptional level, as demonstrated in snapdragon and hop (Orlova et al., 2009).

To avoid potentially problematic interactions with endogenous prenyltransferases in our experiments, *AtGpps* was also targeted to the cytosol and mitochondria. Transgenic tobacco plants expressing *AtGpps* in the cytosol and mitochondria provide a better understanding of the localization and interactions with other prenyltransferases and the monoterpene pathway in tobacco plants. It was difficult to regenerate tobacco plants co-expressing *AtGpps* and *VoGes*. Only 10 kanamycin-resistant transgenic plantlets survived selection and none of the transgenic plants expressed *AtGpps* and *VoGes* in the plastids, the natural subcellular localization, perhaps reflecting the lack of an optimized tobacco transformation protocol. We improved the transformation protocol by increasing the subculture frequency to every 2 weeks instead of 4, lowering the intensity of bombardment and incident light during regeneration, and including thiamine in the medium. Orlova et al. (2009) suggested that tight regulation of *AmGpps.ssu* is crucial for plant survival: high levels of *AmGpps.ssu* expression were lethal as shown by the small number of transformants obtained from 1200 transformation events, and three lines expressing high levels of *AmGpps.ssu* were unable to mature under greenhouse conditions. Only plants with low to moderate levels of *AmGpps.ssu* were able to produce mature plants, although the latter still showed signs of delayed growth and chlorosis (Orlova et al., 2009). The suppression of *AmGpps.ssu* increased the amount of monoterpenes emitted by both leaves and flowers, and resulted in an uneven distribution of different monoterpene compounds (Orlova et al., 2009).

These results suggest that there is metabolic crosstalk between the plastidial MEP and cytosolic MVA pathways in tobacco and that the introduction of *AmGpps.ssu* increases the flux towards GPP in the plastids, thus reducing the IPP pool and its transport to the cytosol, ultimately limiting the synthesis of sesquiterpenes (Orlova et al., 2009). The inability to recover transgenic plants expressing *AtGpps* and *VoGes* in the chloroplast (the natural localization of the corresponding enzymes) is consistent with the observations of Orlova et al. (2009). The increased flux towards GPP in plastids and the subsequent reduction of the IPP pool, appear to be critical for plant regeneration. Compartmentalization of GPPS in the chloroplasts, mitochondria or

cytosol resulted in toxicity effects that limited the recovery of transgenic plants, but chloroplast localization was the most toxic. Bouvier et al. (2000) reported that Arabidopsis GPPS is predicted to accumulate in the mitochondria based on *in silico* analysis but this could not be confirmed *in planta*. More studies are required to understand the toxicity of *AtGpps* and the potential impact of avoidance strategies based on subcellular compartmentalization.

The expression of GES is also associated with toxic effects, e.g. severe damage to grapevine (*Vitis vinifera*) plants expressing *Ocimum basilicum* GES (*ObGes*) resulting in brown-colored embryogenic calli and somatic embryos that die (Fisher et al., 2013). GES has also been localized in plastids in *Ocimum basilicum* (Iijima et al., 2004), *Perilla frutescens* and *Perilla citriodora* (Ito et al., 2007). The recently cloned *Catharanthus roseus* GES (*CrGes*) is localized in the stroma and stromules (Simikin et al., 2013). Stromules are found in close association with the ER suggesting an exchange of metabolites between the two organelles *in vivo* (Simikin et al., 2013). This scenario fits well with the next step of monoterpene synthesis, i.e. the conversion of geraniol into 10-hyroxygeraniol, catalyzed by G10H. This enzyme is a cytochrome P450 which is anchored in the ER membrane with the catalytic domain probably exposed to the cytosol in the vicinity of stromules (Simkin et al., 2013; Guirimand et al., 2009). Stromules may therefore facilitate the export of geraniol into the cytosol and thereby its conversion into 10-hydroxygeraniol by ER-anchored G10H (Simkin et al., 2013).

G10H belongs to the CYP76B subfamily, and is designed CYP76B6 (Collu et al., 2001). It is not related to the CYP71 or CYP72 families even though both are known to possess G10H activity (Collu et al., 2001). However, it is related to Arabidopsis CYP76C1, which can also hydrolyze geraniol *in vitro*. There is 47% amino acid identity between G10H and CYP76C1 and there is a greater degree of identity in the C-terminal region (Collu et al., 2001).

The GES/B6 transformation experiments achieved a higher regeneration frequency (38 transgenic plants from ~10 transformed explants) than the GPPS/GES transformation experiments (10 transgenic plants from five transformed explants). The GES/CYP76B6 experiment achieved a higher transformation frequency than the GPPS/GES experiment because the protocol described by Leech et al. (1998) was

optimized and there was no associated *AtGpps* toxicity. *VoGes* toxicity was less severe than observed in the GPPS/GES experiments.

DNA blot analysis to confirm the presence of the input transgenes in the putative GPPS/GES and GES/CYP76B6 transgenic plants showed that 50% of the plants in the first experiment contained both input transgenes (*AtGpps* and *VoGes*) and 60% of the plants in the second experiment contained both input transgenes (*VoGes* and *cyp76b6*). RNA blot analysis confirmed the expression of *AtGpps* and *VoGes* in three transgenic plants and of *VoGes* and *Cyp76b6* in 19. There was a good correlation between transgene integration and expression, consistent with other transgene combinations reported in the literature (Zhu et al., 2009; Naqvi et al., 2009; Ramessar et al., 2007). Naqvi et al. (2009) regenerated 75 transgenic maize plants, 13% of which expressed all five input transgenes. Direct DNA transfer experiments occasionally generate transgene fragments that can be identified by PCR but fail to express a product, meaning that transgene content does not necessarily predict expression profiles and metabolic characteristics (Naqvi et al., 2009). Germination under kanamycin selection indicated the predicted Mendelian segregation of the transgenes suggesting that all genes had integrated into a single locus and that the tobacco transformation protocol is efficient. These results provide proof of concept for the experiments described in Chapter 5, which involve engineering the later steps of the monoterpene secoiridoid pathway.

2.7. Conclusions

Monoterpene compounds such as geraniol, geraniol diphosphate and 10-hydroxygeraniol are toxic to plant cells. They are sequestered within subcellular organelles to avoid systemic toxicity, which adds a layer of complexity to transformation experiments involving the metabolic engineering of the terpenoid pathway. Among 10 regenerated plants transformed with *AtGpps* and *VoGes*, three plants expressed both transgenes at levels detectable by mRNA blot analysis. Similarly *Cyp76b6* and *VoGes* were co-expressed in 19 of 38 transgenic plants. This population of transgenic plants confirms the utility of a tobacco transformation method using direct DNA transfer by particle bombardment and will facilitate further studies of the monoterpene pathway as described in subsequent chapters.

2.8. References

Ajikumar P. K., Xiao W-H., Tyo K. E. J., Wang Y., Simeon F., Leonard E., Mucha O., Phon T. H., Pfeifer B., Stephanopoulos G. (2010) Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli Science*, 330: 70-74.

Bohlmann J., DeLuca V., Eilert U., Martin W. (1995) Purification and cDNA cloning of anthranilate synthase from *Ruta graveolens*: modes of expression and properties of native and recombinant enzymes. *The Plant Journal*, 7: 491-501.

Bouvier F., Suire C., d'Harlingue A., Backhaus R.A., Camara B. (2000) Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *The Plant Journal* 24: 241-252.

Burke C., Croteau R. (2002) Interaction with the Small Subunit of Geranyl Diphosphate Synthase Modifies the Chain Length Specificity of Geranylgeranyl Diphosphate Synthase to Produce Geranyl Diphosphate. *The Journal of Biological Chemistry*, 277: 3141–3149.

Burke C.C., Wildung M.R., Croteau R. (1999) Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. *PNAS,* 96: 13062–13067.

Capell T., Bassie L., Christou P. (2004) Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. *Proc. Natl. Acad. Sci USA,* 101: 9909-9914.

Carretero-Paulet L., Ahumada I., Cunillera N., Rodríguez-Concepción M., Ferrer A., Boronat A., Campos N. (2002) expression and molecular analysis of the arabidopsis *dxr* gene encoding 1-deoxy-d-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-*c*-methyl-d-erythritol 4-phosphate pathway. *American Society of Plant Physiologists*, 129: 1581-1591.

Chahed K., Oudin A., Guivarc'h N., Hamdi S., Chénieux J-C., Rideau M., Clastre M. (2000) 1-Deoxy-D-xylulose 5-phosphate synthase from periwinkle: cDNA identification and induced gene expression in terpenoid indole alkaloid-producing cells. *Plant Physiology and Biochemistry*, 38-559-566.

Chan W-S., Abdullah J.O., Namasivayam P., Mahmood M. (2009) Molecular characterization of a new 1-deoxy-d-xylulose 5-phosphate reductoisomerase (DXR) transcript from *Vanda Mimi Palmer*. *Scientia Horticulturae*, 121: 378-382.

Collu G., Unver N., Peltenburg-Looman A.M.G., van der Heijden R., Verpoorter R., Memelink R. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme envolved in terpenoid indole alcaloide biosíntesis. *FEBS Letters* 508: 215-220.

De Troch P., Dosselaere F., Keijers V., de Wilde P., Vanderleyden J. (1997) Isolation and Characterization of the *Azospirillum brasilense trpE(G)* Gene, Encoding Anthranilate Synthase. *Current Microbiology*, 34: 27-32.

De Waal, Verpoorte R. (1993) Purification of the cytochrome P-450 enzyme geraniol 10-hydroxylase from cell cultures of Catharanthus roseus. *Journal of Chromatography A*, 635: 237-249.

Dellaporta S. L., Wood J., Hicks J. B. (1993) A plant DNA minipreparation: Version II. *Plant Mol Bio Rep*, 1: 19-21.

DeLuca V., Marineau C., Brisson N. (1997) Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopadecarboxylases. *National Acad. Science*, 86: 2582–2586.

Engpraset S., Taura F., Shoyama Y. (2005) Molecular cloning, expression and characterization of recombinant 1-deoxy-d-xylulose-5-phosphate reductoisomerase from Coleus forskohlii Briq. *Plant Science*, 169: 287-294.

Fisher M.J.C., Meyer S., Claudel P., Perrin M., ginglinger J.F., Gertz C., Masson J. E., Werck-Reinhardt D., Hugueney P., Karst F. (2013) Specificity of *Ocimum basilicum* geraniol synthase modified by its expresión in different heterologous Systems. *Journal of Biotechnology*, 163: 24-29.

Geerlings A., Martinez-Lozano Ibañeez M., Memelink J., van der Heijden R., Verpoorte R (2000) Molecular Cloning and Analysis of Strictosidine □-D-Glucosidase, an Enzyme in Terpenoid Indole Alkaloid Biosynthesis in Catharanthus roseus. *The Journal of Biological Chemistry*, 275: 3051–3056.

Geerlings A., Redondo F. J., Contin A., Memelink J., van der H. R., Verpoorte R. (2001) Biotransformation of tryptamine and secologanin into plant terpenoid indole alkaloids by transgenic yeast. *Appl. Microbiol. Biotechnol.*, 56: 420–424.

Gerasimenko I., Sheludko Y., Ma X., Stöckigt J. (2002) Heterologous expression of a *Rauvolfia* cDNA encoding strictosidine glucosidase, a biosynthetic key to over 2000 monoterpenoid indole alkaloids. *European Journal of Biochemistry*, 269: 2204–2213.

Gong Y., Liao Z., Guo B., Sun X., Tang K. (2006) Molecular Cloning and Expression Profile Analysis of *Ginkgo biloba* DXS Gene Encoding 1-Deoxy-D-xylulose 5-Phosphate Synthase, the First Committed Enzyme of the 2-C-Methyl-D-erythritol 4-Phosphate Pathway. *Planta Med*, 72: 329-335.

Grolle S., Bringer-Meyer S., Sahm H. (2000) Isolation of the *dxr* gene of *Zymomonas mobilis* and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *FEMS Microbiology Letters*, 191: 131-137.

Guirimand G., Burlat V., Oudin A., Lanoue A., St-Pierre B., Courdavault V. (2009) Optimization of the transient transformation of *Cahtaranthus roseus* cells by particle bombardment and its applicaiton to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase. *Plant Cell Rep*, 28:1215-1234.

Gupta P., Agarwal A. V., Akhtar N., Sangwan R. S., Singh S. P., Trivedi P. K. (2013) Cloning and characterization of 2-C-methyl-d-erythritol-4-phosphate pathway genes for isoprenoid biosynthesis from Indian ginseng, *Withania somnifera*. *Protoplasma*, 250: 285-295.

Han Y., Roytrakul S., Verberne M. C., van der Heijden R., Linthorst H. J. M., Verpoorte R. (2003) Cloning of a cDNA encoding 1-deoxy-d-xylulose 5-phosphate synthase from Morinda citrifolia and analysis of its expression in relation to anthraquinone accumulation. *Plant Science*, 164: 911–917.

Hilliou F., Christou P., Leech M. J. (1998) development of an efficient transformation system for *Catharanthus roseus* cell cultures using particle bombardment. *Plant Science*, 140: 179-188.

Hsiao Y.Y., Jeng M.F., Tsai W.C., Chuang Y.C., Li C.Y., Wu T.S., Kuoh C.S., Chen W.H., Chen H.H. (2008) A novel homodimeric geranyl diphosphate synthase from the orchid *Phalaenopsis bellina* lacking a DD(X)$_{2–4}$D motif. *The Plant Journal*, 55: 719–733.

Huang F.C., Sung P.H., Do Y.Y. (2012) Differential expression and functional characterization of the NADPH cytochrome P450 reductase genes from *Nothapodytes foetida*. *Plant Science,* 190: 16-23.

Huang Q., Roessner C. A., Croteau R., Scott A. (2001) Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol, *Bioorganic & Medicinal* Chemistry, 9: 2237–2242.

Iijima Y., Gang D.R., Fridman E., Lewinsohn E., Pichersky E. (2004) Characterization of Geraniol Synthase from the Peltate Glands of Sweet Basil. *Plant Physiology,* January, 134: 370–379.

Ikeda H, Esaki N, Nakai S et al (1991) Acyclic monoterpene primary alcohol: NADP super(+)oxidoreductase of Rauwolfia serpentina cells: the key enzyme in biosynthesis of monoterpene alcohols. *J Biochem,* 109:341–347.

Irmler S, Schroder G, St-Pierre B, Crouch NP, Hotze M, Schmidt J, Strak D, Matern U, Schroder J (2000) Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J* , 24:797–804.

Ito M., Honda G. (2007) Geraniol synthases from perilla and their taxonomical significance. *Phytochemistry* 68: 446–453.

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

Kang S., Kang K., Lee K., Back K. (2007) Characterization of rice tryptophan decarboxylases and their direct involvement in serotonin biosynthesis in transgenic rice. *Planta,* 227: 263-272.

Khemvong S., Suvachittanont W. (2005) Molecular cloning and expression of a cDNA encoding 1-deoxy-d-xylulose-5-phosphate synthase from oil palm Elaeis guineensis Jacq. *Plant Science,* 169: 571-578.

Kutchan T. M., Bock A., Dittrich H. (1994) Heterologous expression of the plant proteins strictosidine synthase and berberine bridge enzyme in insect cell culture. *Phytochemistr*y, 35: 353–360.

Kutchan T.M., Hampp N., Lottspeich F., Beyreuther K., Zenk M.H. (1988) The

cDNA clone for strictosidine synthase from *Rauvolfia serpentina* DNA sequence determination and expression in *Escherichia coli*. *FEBS letters*, 237: 40-44.

Kuzuyama T., Takagi M., Takahashi S., Seto H. (2000) Cloning and Characterization of 1-Deoxy-D-Xylulose 5-Phosphate Synthase from *Streptomyces* sp. Strain CL190, Which Uses both the Mevalonate and Nonmevalonate Pathways for Isopentenyl Diphosphate Biosynthesis. *J. Bacteriol*, 182: 891-897.

Laflamme P., St-Pierre B., De Luca V. (2001) Molecular and Biochemical Analysis of a Madagascar Periwinkle Root-Specific Minovincinine-19-Hydroxy-*O*-Acetyltransferase. *American Society of Plant Physiologists*, 125: 189-198.

Lange B.M., Croteau R. (1999) Mevalonate-Independent Pathway in Plants: Cloning and Heterologous Expression of 1-Deoxy-d-xylulose-5-phosphate Reductoisomerase from Peppermint. *Archives of Biochemistry and Biophysics*, 365: 170-174.

Lee J.K., Oh D.K., Kim S.Y. (2007) Cloning and characterization of the dxs gene, encoding 1-deoxy-d-xylulose 5-phosphate synthase from Agrobacterium tumefaciens, and its overexpression in Agrobacterium tumefaciens. *Journal of Biotechnology*, 128: 555-566.

Leech M. J., May K., Halard D., Verpoorte R., De Luca V., Christou P. (1998) Expression of two consecutive genes of a secondary metabolic pathway in transgenic tobacco: molecular diversity influences levels of expression and product accumulation. *Plant Mol Bio*, 38:765-774.

Liu D. H., Jin H. B., Chen Y. H., Cui L. J., Ren W. W., Gong Y. F., Tang K. X. (2007) Terpenoid Indole Alkaloids Biosynthesis and Metabolic Engineering in *Catharanthus roseus. Journal of Integrative Plant Biology*, 49: 961-974.

Martin V. J. J., Pitera D. J., Withers S. T., Newman J. D., Keasling J. D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology* 21: 796 – 802.

Murashige T., Skoog F. (1963) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol Plantarum*, 15: 473-497.

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

Naqvi S., Zhu C., Farré G., Ramessar K., Bassie L., Breitenbach J., Perez Conesac D., Sandmann G., Capell T., Christou P. (2009) Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proc Natl Acad Sci USA*, 106: 7762–7767.

Newman J. D., Marshall J., Chang M., Nowroozi F., Paradise E., Pitera D., Newman K. L., Keasling J. D. (2006) High-level production of amorpha-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. *Biotechnology and Bioengineering*, 95: 684–691.

Orlova I. Nagegowda D., Kish C., Gutensohn M., Maeda H., Varbanova M., Fridman E., Yamaguchi S., Hanada A., Kamiya Y., Krichevsky A., Citovsky V., Pichersky E., Dudareva N. (2009) The small subunit of Snapdragon geranyl diphosphate synthase modifies the chain length specificity of tobacco geranylgeranyl diphosphate synthase in planta. *Plant Cell*, 21: 4002-4017.

Peters R. J., Croteau R. B. (2004) Handbook of Plant Biotechnology. *John Wiley & Sons, Ltd* Christou P and Klee H. England, 1: 609-629.

Picaud S., Brodelius M., Brodelius P. E. (2005) Expression, purification and characterization of recombinant (E)-β-farnesene synthase from Artemisia annu. *Phytochemistry* 66: 961-967.

Ramessar K., Rademcher T., Sack M., Stadlmann J., Platis D., Stiegler G., Labrou N., Altman F., Ma J., Stöger E., Capell T., Christou P. (2007) Cost-effective production of a vaginal protein microbicide to prevent HIV transmission. *PNAS*, 105: 3727-3732.

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

Sambrook J., Fritsch E. F., Maniatis T (1989) Molecular cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Lanoratory Press.

Schäfer H., Wink M. (2009) Medicinally important secondary metabolites in recombinant microorganisms or plants: Progress in alkaloid biosynthesis. *Biotechnology Journal*, 4: 1684-1703.

Schmidt A., Gershenzon J. (2008) Cloning and characterization of two different types of

geranyl diphosphate synthases from Norway spruce (Picea abies). *Phytochemistry*, 69: 49–57.

Schwarz H., Liu B., Peters S., Barillas W., Beerhues L. (2008) Purification, cDNA cloning and functional expression of NADPH-cytochrome P450 reductase from *Centaurium erythraea* cell cultures. *Plant Biology*, 11: 300–306.

Schwender J., Muller C., Zeidler J., Lichtenthaler H.K. (1999) Cloning and heterologous expression of a cDNA encoding 1-deoxy-D-xylulose-5-phosphate reductoisomerase of Arabidopsis thaliana. *FEBS Letters*, 455: 140-144.

Seetang-Nun Y., Sharkey T., Suvachittanont W. (2008) Molecular cloning and characterization of two cDNAs encoding 1-deoxy-d-xylulose 5-phosphate reductoisomerase from Hevea brasiliensis. *Journal of plant Physiology*, 165: 991-1002.

Seo M.J., Im E.M., Nam J.Y., Kim S.O. (2007) Increase of CoQ_{10} production level by the coexpression of decaprenyl diphosphate synthase and 1-deoxy-D-xylulose 5-phosphate synthase isolated from *Rhizobium radiobacter* ATCC 4718 in recombinant *Escherichia coli*. *Journal of Biotechnology*, 17: 1045-1048.

Simkin A. J., Miettinen K., Claudel P., Burlat V., Guirimand G., Courdavault V., Papon N., Meyer S., Godet S., St-Pierre B., Giglioli-Guivarc'h N., Fischer M.J.C., Memelink J., Clastre M. (2013) Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. Phytochemistry 85: 36–43.

Tholl D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Current Opinion in Plant Biology*, 9:297–304.

Turner G.W., Croteau R. (2004) Organization of Monoterpene Biosynthesis in Mentha. Immunocytochemical Localizations of Geranyl Diphosphate Synthase, Limonene-6-Hydroxylase, Isopiperitenol Dehydrogenase, and Pulegone Reductase. *Plant Physiology*, 136: 4215–4227.

Valletta A., Trainotti L., Santamaria A.R. Pasqua G. (2010) Cell-specific expression of tryptophan decarboxylase and 10-hydroxygeraniol oxidoreductase, key genes involved in camptothecin biosynthesis in *Camptotheca acuminata* Decne (Nyssaceae)

Van der Heijden R., Jacobs D. I., Snoeijer W., Hallared D., Verpoorte R. (20049 The

Catharanthus alkaloids: pharmacognosy and biotechnology. *Curr. Med. Chem*, 11: 607-628.

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

Wang J., Liu Y., Cai Y., Zhang F., Xia G., Xiang F. (2010a) Cloning and Functional Analysis of Geraniol 10-Hydroxylase, a Cytochrome P450 from *Swertia mussotii*.

Wang Q., Yuan F., Pan Q., Li M., Wang G., Zhao J., Tang K. (2010b) Isolation and functional analysis of the *Catharanthus roseus* deacetylvindoline-4-*O*-acetyltransferase gene promoter. *Plant Cell Reports*, 29: 185-192.

Wu S.J., Shi M., Wu J.Y. (2010) Cloning and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene for diterpenoid tanshinone biosynthesis in *Salvia miltiorrhiza* (Chinese sage) hairy roots. *Biotechnology and Applied Biochemistry*, 52: 89-95.

Wungsintaweekul J., Sirisuntipong T., Kongduang D, Losuphanporn T., Ounaroon A., Tansakul P., De-eknamkul W. (2008) Transcription Profiles Analysis of Genes Encoding 1-Deoxy-D-xylulose 5- Phosphate Synthase and 2C-Methyl-D-erythritol 4-Phosphate Synthase in Plaunotol Biosynthesis from *Croton stellatopilosus*. *Pharm. Bull.*, 31: 852—856.

Yang J., Adhikari M.N., Liu H., Xu H., He G., Zhan R., Wei J., Chen W. (2012) Characterization and functional analysis of the genes encoding 1-deoxy-d-xylulose-5-phosphate reductoisomerase and 1-deoxy-d-xylulose-5-phosphate synthase, the two enzymes in the MEP pathway, from *Amomum villosum* Lour. *Molecular Biology Reports*, 39: 8287-8296.

Yang T, Stoopen G, Yalpani N, Vervoort J, de Vos R, Voster A, Verstappen FWA, Bouwmeester HJ, Jongsma MA (2011) Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. *Metabolic Engineering*,13: 414-425.

Yang T., Li J., Wang H.X., Zeng Y. (2005) A geraniol-synthase gene from Cinnamomum tenuipilum. *Phytochemistry*, 66: 285-293.

Yao H., Gong Y., Zuo K., Ling H., Qiu C., Zhang F., Wanf Y., Pi Y., Liu X., Molecular cloning, expression profiling and functional analysis of a DXR gene encoding 1-deoxy-d-xylulose 5-phosphate reductoisomerase from *Camptotheca acuminata*. *Journal of Plant Physiology*, 165: 203–213.

Yazaki K. (2004) Handbook of Plant Biotechnology. *John Wiley & Sons, Ltd*. Christou P and Klee H. England. 2: 811-861.

Yoonram K., Takahashi S., Rattanapittayaporn A., Koyama T., Wititsuwannakul D., Witisuwannakul R. (2008) cDNA, from Hevea brasiliensis latex, encoding 1-deoxy-d-xylulose-5-phosphate reductoisomerase. *Plant Science*, 175: 694–700.

Zarate R., Dirks C., van der H. R., Verpoorte R. (2001) Terpenoid indole alkaloid profile changes in Catharanthus pusillus during development. *Plant Sci*, 160: 971–977.

Zeef L. A. H., Christou P., Leech M. J. (2000) Transformation of the tropane alkaloid-producing medicinal plant *Hyoscyamus muticus* by particle bombardment. *Transgenic Research*, 9: 163-168.

Zhang M., Li K., Zhang C., Gai J., Yu D. (2009) Identification and characterization of class 1 *DXS* gene encoding 1-deoxy-d-xylulose-5-phosphate synthase, the first committed enzyme of the MEP pathway from soybean. *Molecular Biology Reports*, 36: 879-887.

Zhu C., Naqvi S., Breitenbach J., Sandmann G., Christou P., Capell T. (2009) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *PNAS*, 105: 18232-18237.

CHAPTER 3

METABOLIC ANALYSIS OF TOBACCO PLANTS EXPRESSING TWO EARLY GENES FROM THE MONOTERPENE SECOIRIDOID BIOSYNTHETIC PATHWAY

CHAPTER 3: METABOLIC ANALYSIS OF TOBACCO PLANTS EXPRESSING TWO EARLY GENES FROM THE MONOTERPENE SECOIRIDOID BIOSYNTHETIC PATHWAY

3.1. Abstract

Plants produce diverse and species-dependent secondary metabolites that enhance their survival in different environments. These compounds are produced via complex biosynthetic pathways that are regulated in a highly sophisticated manner. Metabolic engineering has been investigated as a strategy to produce such compounds sustainably, but only a few such attempts have been successful and many challenges remain, reflecting the complexity of the pathways and their organization at the subcellular level. We investigated the impact of importing the early monoterpene secoiridoid pathway into tobacco by carrying out a comparative metabolic profiling of wild-type and transgenic plants. This analysis revealed the complete suppression of the terpenoid pathway in the transgenic plants, as well as important interactions with non-target pathways. These results provide further insight into the development of effective metabolic engineering strategies in plants for the enhanced synthesis of secondary products.

3.2. Introduction

Plants produce diverse and species-dependent secondary metabolites that enhance their survival, and this requires the sophisticated regulation of complex metabolic pathways (Verpoorte et al. 2000; Schäfer et al., 2009). Many secondary metabolites possess useful biological activities, allowing them to interact with specific targets in mammals, insects and microorganisms. Therefore more than 40% of current drugs for human health are derived directly from plants (Verpoorte et al., 2002; Gómez-Galera et al., 2007) or are based indirectly on model structures or scaffolds derived from plants, then chemically modified to introduce novel activities (Schäfer et al., 2009). This is described in more detail in Chapter 1, Section 1.1.1.

Many *in vitro* strategies have been explored for the sustainable production of secondary metabolites, including callus, cell suspension and organ cultures (e.g. hairy roots) and even the large-scale fermentation of suspended cells (Miralpeix et al. 2013). However, dedifferentiated callus and cell suspension cultures from medicinal plants often produce only low yields of secondary products because the necessary genes are not expressed (Yazaki et al., 2004). The compartmentalization of the metabolic enzymes in the cytosol, endoplasmic reticulum, vacuoles and plastids also adds complexity to the synthesis of secondary compounds (O´Connor and Maresh, 2006).

3.2.1. Engineering of complex plant secondary metabolic pathways

There are few examples of successful metabolic engineering for the production of complex secondary products because the metabolic pathways are long and complex, the enzymes and substrates are compartmentalized at the tissue and subcellular levels and the entire pathway is under strict multilevel regulation (Yazaki et al., 2004). Metabolic engineering often influences the overall metabolic flux in plant cells, not only generating new products but also modulating the availability of existing products by competing for intermediates. One of the most efficient ways to induce the biosynthesis of metabolites is by applying biotic or abiotic stress. Many pathways are elicited by stress, as well as depending on the tissue, cell type and developmental stage (Yazaki et al., 2004).

The reconstruction of a complete metabolic pathway by combining the relevant genes in a heterologous background is one of the key aims of metabolic engineering and synthetic biology. The carotenoid pathway in corn provides a useful example of total heterologous biosynthesis (Naqvi et al., 2009). Promising results have also been achieved for the biosynthesis of artemisinin in tobacco (Fahri et al., 2011).

Naqvi et al. (2009) enigneered corn with four genes encoding enzymes representing three vitamin biosynthesis pathways to boost the accumulation of β-carotene, ascorbate and folate simultaneously in the seed endosperm. The levels of β-carotene

(pro-vitamin A) were increased 169-fold by the introduction of the corn phytoene synthase (*psy1*) cDNA and the *Pantoea ananatis crt*I gene (encoding carotene desaturase). To amount of ascorbate (vitamin C) was increased by 6-fold by expressing the rice dehydroascorbate reductase (*dhar*) cDNA, and the level of folate was doubled by expressing the *E. coli folE* gene, encoding GTP cyclohydrolase (GCH1). This was the first successful modification of three distinct metabolic pathways simultaneously in plants.

Farhi et al. (2011) reported the metabolic engineering of tobacco to produce the terpenoid artemisinin. Tobacco plants were transformed with five linked genes representing the mevalonate and artemisinin pathways: deregulated 3-hydroxy-3-methylglutaryl-coenzyme A reductase (*tHMG*) from yeast, and *A. annua* genes encoding two cytochrome P450 reductases (*CPR* and *CYP71AV1*), amorphadiene synthase (*ADS*) and artemisinic aldehyde reductase (*DBR2*). The simultaneous expression of all transgenes demonstrated that the artemisinin pathway could be reconstructed successfully in a heterologous background such as tobacco. Although the artemisinin levels in the transgenic tobacco plants were lower than those in *A. annua*, this experimental platform provided proof of concept for the reconstruction of secondary metabolism in heterologous plants (Farhi et al., 2011)

Despite these successful examples of terpenoid pathway reconstruction, the modulation of secondary metabolism in whole plants can disrupt other pathways with the increased metabolic load, inhibiting growth and inducing stress symptoms such as chlorosis (Miralpeix et al., 2013). The regulation of the terpenoid pathway is complex and the control mechanisms are unknown (Muntendam et al. 2009), in part reflecting the simultaneous participation of the MVA and MEP pathways (see Chapter 5, Section 5.2). The cytoplasmic MVA pathway requires acetyl-CoA units for the biosynthesis of isopentenyl diphoshpate (IPP) the universal C_5 building block for all isoprenoids. The methyl erythritol phosphate (MEP) pathway also produces IPP, but operates in the plastids and requires pyruvate and 3-phosphoglycerate derived from fixed CO_2 via the Calvin cycle (Wu et al. 2008; Verpoorte et al., 2002; Gómez-Galera et al. 2007). The coexistence of these pathways in higher plants may result in the intracellular trafficking of precursors, the extent of which depends on the species and

the environment. The regulation of this trafficking is not yet fully understood (Muntendam et al. 2009) as discussed further in Chapter 5, Section 5.2.4.

Terpenoid engineering may have a direct effect on other branches of the pathway (Sandman et al. 2006). Constitutive overexpression of tomato *Psy-1* in transgenic tomato or tobacco increased the carotenoid content, but reduced the chlorophyll content and generated a dwarf phenotype (Fray et al. 1995; Busch et al. 2002). These examples reveal metabolic interactions between the different isporenoid branches and their common precursors (Sandman et al. 2006). Terpenoid engineering must therefore include a consideration of multiple layers of regulation, including the compartmentalization of enzymes. Monoterpene biosynthesis takes place in the plastids, but the overexpression of monoterpene synthase in either the plastid or cytoplasm resulted in the accumulation of novel monoterpenes and in some cases novel derivatives such as glyco-conjugates (Wu et al., 2012). Re-directing sesquiterpene metabolism to the mitochondria had an equivalent impact on overall terpene metabolism. Kapper et al. (2005) generated transgenic *Arabidopsis thaliana* plants expressing Nerolidol Synthase 1 from *F. ananassa* (FaNES1), a strawberry linalool/nerolidol synthase, and directing it to accumulate in the mitochondria instead of the cytosol. Ubiquinone biosynthesis takes place in the mitochondria and Arabidopsis produces a farnesyl diphosphate (FPP) synthase isoform that is targeted to the mitochondria, so FPP should be available in this compartment. The transgenic plants not only produced (3S)-(*E*)-nerolidol, which confirmed the presence of FPP in the mitochondria, they also produced the unusual homoterpene (*E*)-4,8-dimethyl-1,3,7-nonatriene ((*E*)-DMNT) (Kapper et al., 2005). Arabidopsis therefore appears to possess enzymes that convert (3*S*)-(*E*)-nerolidol into (*E*)-DMNT. This was confirmed by feeding (*E*)-nerolidol to wild-type Arabidopsis leaves, resulting in the formation of (*E*)-DMNT. Mitochondria use FPP to produce ubiquinone and heme A, but the introduction of a sesquiterpene synthase into these organelles does not deplete the available FPP to the extent that growth is inhibited (Kapper et al., 2005). This example shows that it is possible to engineer terpene metabolism in plants, but the low levels of sesquiterpenes demonstrate the tight regulation of their biosynthesis (Kapper et al., 2005).

Wu et al. (2006) also demonstrated that cytoplasmic sesquiterpene metabolism could be engineered in the chloroplast to overcome endogenous regulation and accumulate a

complex mixture of industrially-valuable sesquiterpenes. They increased the levels of patchoulol by targeting the FPPS and patchoulol synthase (PTS) to the plastids, to overcome the bottleneck posed by the small and tightly-regulated pool of FPP. More recently, Kumar et al. (2012) extended this ectopic engineering strategy by integrating genes representing the endogenous cytoplasmic MVA pathway into the plastid genome to achieve high-level expression of the enzymes upstream of mevalonate. They expressed the six early genes of the cytoplasmic MVA pathway to increase the synthesis of isoprenoid metabolites, plus a selectable marker gene. The transgenes encoded phosphomevalonate kinase (PMK), mevalonate kinase (MVK), mevalonate diphosphate decarboxylase (MDD), acetoacetly CoA thiolase (AACT), 3-hydroxy-3-methylglutaryl coenzyme A (HMGCoA) synthase (HMGS), and truncated HMGCoA reductase (HMGRt) from yeast (Kumar et al., 2012). They reported higher levels of products normally associated with both the plastid and cytosolic pathways, including mevalonate, carotenoids, sterols and squalene, as well as triacylglycerides. This was the first report to describe the integration of six heterologous eukaryotic genes representing a cytoplasmic pathway into the plastid genome (Kumar et al., 2012).

3.3. Aims and Objectives

The aim of the experiments described in this chapter was to explore the metabolic changes produced by the expression of the early monoterpene secoiridoid pathway enzymes in transgenic tobacco plants. We evaluated the metabolic profile of transgenic tobacco plants expressing geranyl diphosphate synthase (*AtGpps*) and geraniol synthase (*VoGes*) in two non-native subcellular compartments, the cytosol and mitochondria. We carried out a metabolomic analysis of the transgenic plants by nuclear magnetic resonance (NMR) spectroscopy in order to determine whether the expression of *AtGpps* and *VoGes* affected the overall metabolic profile of the plants.

3.4. Materials and methods

3.4.1. Transformation vectors

All constructs used for tobacco transformation were based on vector pIV2 and are described in Section 2.4.2. The expression cassette comprised the double CaMV 35S promoter, the signal for mitochondrial targeting, and the coding region and the chrysanthemum terminator RbcS1 (pIV2). The coding region comprised either the Arabidopsis geranyl diphosphate (*Gpps*) gene (GenBank accession no. AT2G34630) yielding vector pIV2A2.4AtGPPS, or the *Valeriana officialis* geraniol synthase (*Ges*) gene (GenBank accession no. JX494702.1) yielding vector pIV2B2.4VoGES. Two further constructs were produced without the mitochondrial targeting signal (pIV2A2.1AtGPPS and pIV2B2.1VoGES). The fifth construct (pTRAuxAH) contained the *npt*II gene controlled by the constitutive Pnos promoter and Pnos terminator (Figure 2 in Chapter 2, Section 2.4.2).

3.4.2. Transformation, selection and regeneration of transgenic plants

Leaf sections were transformed by direct DNA transfer (particle bombardment) as described in Chapter 2, Section 2.4.2, using a 3:1 molecular weight ratio of the vector containing the gene of interest against the *npt*II selectable marker. Two combinations of genes were used, as listed in Table 1. Bombarded leaf pieces were incubated in the dark and transferred to fresh media every 2 weeks as described in Chapter 2, Section 2.4.3.

Table 1. Combinations of constructs used for tobacco transformation experiments

Combination	***gpps***	***ges***
1	pIV2A2.1 *AtGpps* (cytosol)	pIV2B2.1*VoGes* (cytosol)
2	pIV2A2.5 *AtGpps* (mitochondria)	pIV2B2.5 *VoGes* (mitochondria)

For practical reasons, we focused on the in-depth analysis of one independent transgenic plant line (L-1) transformed with combination 1 (Table1) as described in Chapter 2, Sections 2.5.3 and 2.4.5, and two independent transgenic plant lines (L-7 and L-9) transformed with combination 2.

3.4.3. Quantitative real-time PCR (qRT-PCR)

First-strand cDNA was synthesized from 2 µg of total RNA using the Omniscript Reverse Transcription kit (QIAGEN). Quantitative real-time PCR (qRT-PCR) was carried out using a BioRad CFX96TM system. Each 25 µl reaction comprised 5 ng of cDNA, 1x iQ SYBR green supermix (BioRad, Hercules, CA, USA) and 5 µM of the forward and reverse primers. The primers for *A622* (isoflavone reductase), *odc* (ornithine decarboxylase), *pmt* (putrescine-N-methyl transferase), *mpo* (methyl putrescine oxidase) and *qprt* (quinolinate phosphoribosyl transferase) were designed using sequences from GenBank (Table 2 in annex 1). Relative mRNA expression levels were calculated on the basis of serial dilutions of cDNA (100–0.16 ng), which were used to generate standard curves for each gene. Triplicate amplifications were carried out in 96 well optical reaction plates by first heating to 95ºC for 5 min followed by 40 cycles at 95ºC for 30 s, then 59.4ºC, 56.3ºC, 58.3ºC, 50.7ºC or 52.0ºC for 30 s (depending on the gene) and 72ºC for 30 s. Amplification specificity was confirmed by product melt curve analysis over the temperature range 50–90 ºC with fluorescence acquired after every 0.5ºC increase, and the fluorescence threshold value and gene expression data were calculated with BioRad CFX96TM software. Values represent the mean of three replicates ± standard error. Amplification efficiencies were compared by plotting the ΔC_t values of different primer combinations of serial dilutions against the log of starting template concentrations using the CFX96TM software. Relative expression levels were determined by adjusting C_t values of the samples to standard curves derived from serial dilutions of tobacco cDNA, which were generated for each gene in every run separately. The expression levels of each gene were normalized against actin.

3.4.4. Nuclear magnetic resonance (NMR) spectroscopy

Sample extraction and NMR spectroscopy were carried out as previously described (Kim et al., 2010). A dried 50 mg sample was transferred to a 2 ml microtube and topped up with 1.5 ml methanol-d_4 and D_2O in KH_2PO_4 buffer, pH 6.0, containing 0.01% (w/w) TMSP-d_4 (trimethylsilyl propionic acid sodium salt). The mixture was vortexed for 1 min, ultrasonicated for 30 min, and centrifuged for 20 min at 13,000 rpm at room temperature. A 800 μl aliquot of the supernatant was transferred to a fresh tube and centrifuged at 13,000 rpm for 5 min, then 750 μl of the supernatant was transferred to a 5 mm NMR tube for ^{1}H-NMR analysis. Deuterated methanol and water were purchased from Cambridge Isotope (Andover, MA, USA). We added 3.2 mM TMSP (tri-methylsilylpropionic sodium salt-2,2,3,3-d_4) to the NMR tubes as a reference for the calibration of NMR shifts, and also as an internal standard for the quantification of metabolites. The samples were analyzed using a 600 MHz NMR spectrometer, with 268 scans per sample.

3.4.5. Multivariate data analysis

The ^{1}H NMR spectra were automatically binned using AMIX software (v3.7, Biospin, Bruker). Spectral intensities were scaled to total intensity and the region of δ 0.30 – δ 10.00 was reduced to integrated regions of 0.04 ppm. The regions of δ 4.7 – δ 5.0 and δ 3.28 – δ 3.34 were excluded from the analysis because of the residual water and methanol signals, respectively. SIMCA-P$^+$ software (v13.0, Umetrics, Umeå, Sweden) was used for principal component analysis (PCA) with Pareto scaling. The X-data set from the PCA represented the ^{1}H NMR data (256 binned data) and the Y-data set was divided into two classes representing transgenic and wild type plants.

3.5. Results

3.5.1. Metabolic profiling

Six tobacco plant samples, namely L-1, L-7 and L-9 including three wild type negative controls were analyzed by ^{1}H NMR spectroscopy. Principal component analysis (PCA) revealed a correlation between the expression of the *AtGpps* and *VoGes* transgenes and the accumulating of metabolites (^{1}H NMR spectra). The ^{1}H NMR intensity data were used as the X-data set for the PCA and transgene expression was used as the Y-data set. The expression of *AtGpps* and *VoGes* (G) was clearly separated from the negative controls (wild type, W) in Y-data set (Figure 1a).

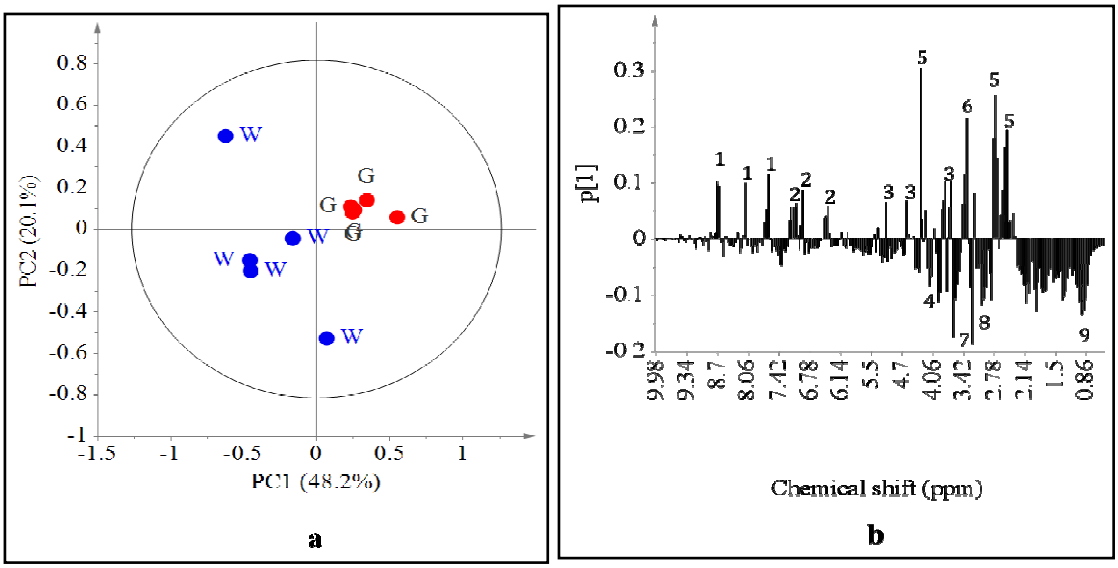

Figure 1. Score plot (a) and loading plot (b) from the principal component analysis (PCA) of plant samples. W: wild type, G: transgenic plants expressing *VoGes* and *AtGpps*. The circle represents the Hotelling T2 with 95% confidence in score plots. 1: nicotine, 2: 4-caffeoyl quinic acid, 3: glucose, 4: inositol, 5: malic acid, 6: betaine, 7: choline, 8: aspartic acid, 9: methyl signals of terpenoids.

The separation predominantly reflected (i) the suppression of terpenoids; (ii) the higher accumulation of nicotine, 4-caffeoyl quinic acid, glucose, malic acid and betaine; and (iii) lower levels of inositol, choline, aspartic acid and glutamic acid in the transgenic plant samples (Figure 1b and 2).

Two major pyridine alkaloids (nicotine and anabasine) were detected in the wild type plants. However, the intermediate anabasine was completely suppressed in transgenic plants (Figure 2).

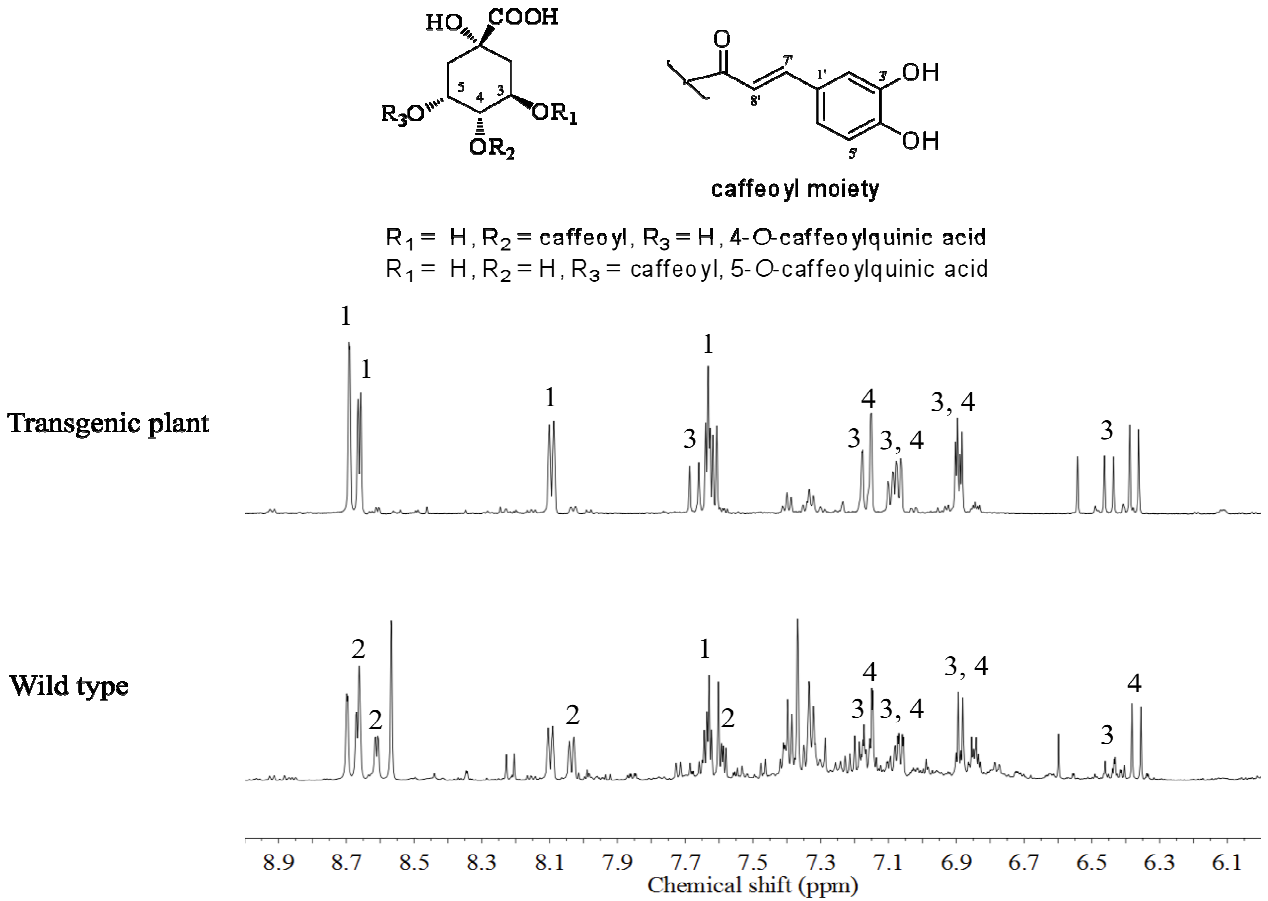

Figure 2. [1]H NMR spectra of wild type and transgenic plant samples in the range δ 6.0 – δ 9.0 (a). 1: nicotine, 2: anabasine, 3: 4-caffeoyl quinic acid, 4: 5-caffeoyl quinic acid (chlorogenic acid).

3.5.2. Quantitative real-time PCR

We used quantitative real-time PCR with SYBR Green dye detection to identify differences in the expression and regulation of the nicotine and phenylalanine pathway genes, thus providing insight into the modulation of endogenous metabolic profiles. The expression of five different endogenous pyridine alkaloid genes representing the nicotine pathway (A622, *odc*, *pmt*, *qprt*, *moc*) and one phenylalanine ammonia-lyase gene (*pal1*) involved in the lignin pathway, were compared in transgenic lines L-1, L-7 and L-9, as well as wild-type controls (Figure 3 and 4). We detected transcripts representing all genes in all samples. Relative quantification by quantitative real-time PCR demonstrated that the transgenic plants expressing *AtGpps* and *VoGes* in the cytosol (L-1) or mitochondria (L-7 and L-9) accumulated lower levels of the *qprt*, *odc*, *pmt* and *A622* transcripts ($P < 0.05$). In contrast, *mpo* transcripts accumulated to significantly higher levels in the transgenic plants ($P < 0.05$). These results confirm that the nicotine pathway genes are differentially regulated in the transgenic and wild type plants (Figure 3). The abundance of phenylalanine ammonia lyase transcripts confirmed that the *pal1* gene was upregulated in all transgenic plants (Figure 4).

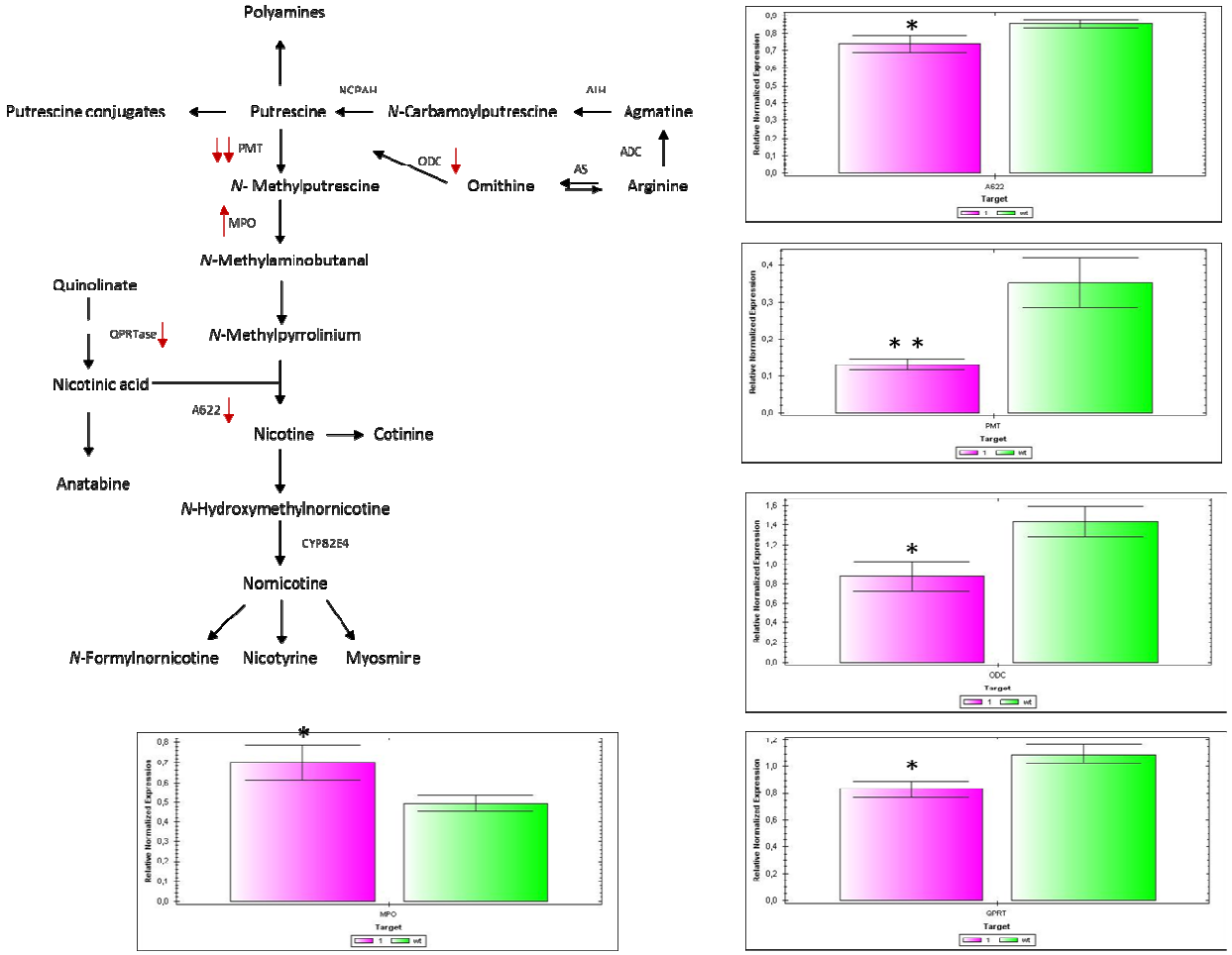

Figure 3. Transcript levels of endogenous nicotine pathway genes: *A622* (isoflavone reductase), *odc* (ornithine decarboxylase), *pmt* (putrescine-N-methyl transferase), *mpo* (methyl putrescine oxidase) and *qprt* (quinolinate phosphoribosyl transferase) in transgenic tobacco plants**.** Values represent the mean of two real time PCR experiments ± SD. Expression levels were normalized against actin. Pink: Transgenic plants with *AtGpps* and *VoGes* targeted to the mitochondria or cytosol. Green: Wild type plants ($P < 0.05$). ↑: up-regulation of the gene in transgenic plants. ↓: down-regulation of the gene in transgenic plants.

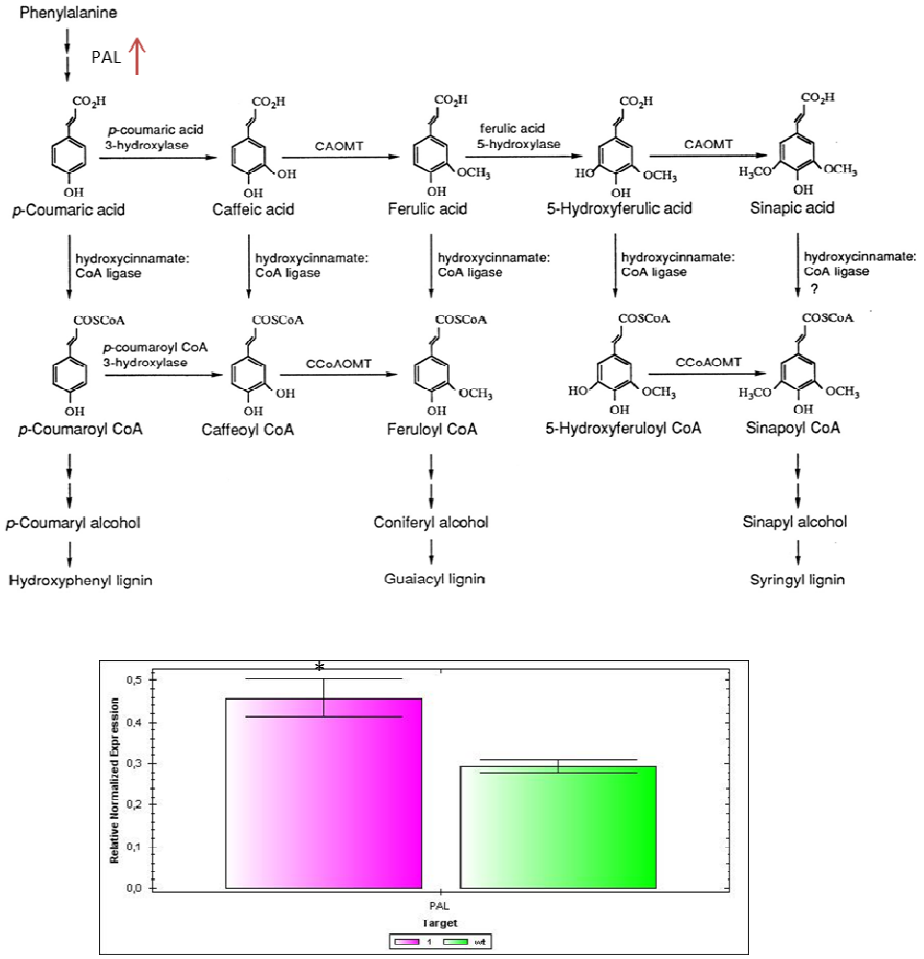

Figure 4. Transcript levels representing genes from the endogenous monolignol biosynthesis pathway (Zhong et al., 1998). PAL: phenylalanine ammonia lyase. Values represent the mean of two real time PCR experiments ± SD. Expression levels were normalized against actin. Pink: Transgenic plants with *AtGpps* and *VoGes* targeted to the mitochondria or cytosol. Green: Wild type plants ($P < 0.05$). ↑: up-regulation of the gene in transgenic plants.

3.6. Discussion

The Arabidopsis geranyl diphosphate (*AtGpps*) and *Valeriana officialis* geraniol synthase (*VoGes*) genes were expressed in tobacco plants and directed into two different subcellular compartments (the cytosol and mitochondria). After screening the initial population of transgenic events, one lead line (L-1) was identified with high levels of both proteins in the cytosol and two lead lines (L-7 and L-9) were identified with high levels of both proteins in the mitochondria (see Chapter 2).

Metabolic profiling of these transgenic plants revealed the complete suppression of terpenoid production compared to negative controls. This suggested that the high-level constitutive expression of *AtGpps* in tobacco leaves might be toxic (see Chapter 2) or might induce RNA silencing (sense-suppression) of transgenes and/or endogenous genes (Takizawa et al., 2007; Howles et al., 1996). Because GPP acts upstream of farnesyl diphosphate synthase (FPP, C_{15}), the first step in the synthesis of diterpenes, and geranylgeranyl diphosphate (GGPP, C_{20}), the first step in the synthesis of sesquiterpenes, the inhibition of GPP may supppress terpenoid biosynthesis by removing precursors. Orlova et al. (2009) reported that the small subunit (SSU) of snapdragon (*Antirrhinum majus*) GPPS in tobacco plants increased the total GPP activity and monoterpene emission from leaves and flowers, indicating that the catalytically inactive GPPS.SSU interacted with endogenous large subunit partner(s) and formed an active snapdragon/tobacco GPPS in plants. The formation of chimeric GPPS also resulted in leaf chlorosis, increased light sensitivity and dwarfism reflecting the loss of chlorophylls, carotenoids and gibberellins. The transgenic plants also had reduced levels of sesquiterpene emission.

Similar homology-dependent gene silencing has been observed in a number of transgenic plants at both the transcriptional and posttranscriptional levels (Matzke and Matzke, 1995). Van der Krol et al. (1988) expressed an antisense chalcone synthase (CHS) gene in transgenic plants to investigate the impact of endogenous genes expressed in ectopic tissues. In *Petunia hybrid*, chalcone synthase is encoded by a multigene family, one member of which is predominantly expressed in floral tissues. Antisense CHS gene driven by the CaMV 35S promoter could create a pool of antisense RNA before the onset of endogenous *CHS* gene expression, providing an instantaneous high ratio of antisense to sense *CHS* mRNA. Some of the transgenic

petunia plants lacked flower pigmentation due to the absence of flavonoids (van der Krol et al., 1988). The effectiveness of the antisense *CHS* suppression was demonstrated by the fact that CHS gene expression could be inhibited even in a heterologous plant system, e.g. tobacco plants expressing the petunia antisense *CHS* gene construct (van der Krol et al., 1988).

Other examples that suggest *AtGpps* induces gene silencing include the constitutive overexpression of a tomato phytoene synthase in tomato plants and the overexpression of French bean (*Phaseolus vulgaris*) phenylalanine ammonia lyase (*pal*) in tobacco plants (Fray et al., 1995; Howles et al., 1996). The transgenic tomato plants were stunted and produced lower levels of lycopene in the fruit (Fray et al., 1995), whereas the endogenous tobacco *pal* gene was severely repressed at both the RNA and enzyme activity levels, resulting in stunted growth and abnormal curled leaves (Howles et al., 1996). The overexpression of rate-limiting enzymes may therefore suppress the corresponding pathway by inducing feedback inhibition (Broun and Somerville, 2001).

Our metabolic profiling data also revealed changes in non-target pathways related to stress responses, e.g. nicotine and lignin biosynthesis. Nicotine is an insecticidal alkaloid produced exclusively in *Nicotiana* species from ornithine and/or arginine via putrescine (Takizaw et al., 2007). The first committed step in nicotine biosynthesis is the conversion of putrescine to N-methylpyrrolinium by the sequential action of putrescine N-methyltransferase (PMT) and N-methylputrescine oxidase (MPO) (Häkkinen et al., 2007) (Figure 3). Nicotinic acid is synthesized from quinolinic acid via the pyridine nucleotide cycle, and is controlled by quinolinate phosphoribosyltransferase (QPRT) (Wang et al. 2009). Anabasine is derived from anatabine and is synthesized from nicotinic acid (Wang et al., 2009) (Figure 3).

The metabolic profile of the transgenic tobacco plants revealed an increase in the nicotine content over wild type levels but no traces of the intermediate anabasine. To understand the changes in pyridine alkaloid biosynthesis, we evaluated the transcript levels of endogenous nicotine pathway genes. Interestingly, the *qprt*, *odc*, *pmt* and *A622* transcripts were lower in the transgenic plants than the wild type, whereas *mpo* transcripts were significantly more abundant. These results suggest that the upregulation of *mpo* increases the nicotine content. Shoji et al. (2008) reported the overexpression of *mpo* in tobacco BY-2 cells using the constitutive CaMV 35S

promoter, increasing the nicotine content by up to 24-fold compared with wild-type BY-2 cells (Shoji et al., 2008). Furthermore, cell lines overexpressing *mpo* showed a reduction (≥50%) in the accumulation of anatabine, anabasine and anatalline compared with wild type cells. This suggested that the overexpression of *mpo* in the transgenic plants boosted the nicotine levels as the expense of anabasine. DeBoer et al. (2012) showed that the overexpression of *qprt* and *A622* boosted the accumulation of anabasine in the upper leaves of tobacco plants. The low levels of the *qprt* and *A622* transcripts in our transgenic plants may therefore explain the lack of anabasine.

Our transgenic plants also accumulated more 4-caffeoyl quinic acid than wild type plants. This phenylpropanoid is an intermediate in the biosynthesis of lignin via phenylalanine (Figure 4) (Zhong et al., 1998). The signaling and defense-related gene *pal1* encodes phenylalanine ammonia lyase, which catalyzes the first step in the phenylpropanoid biosynthesis pathway, converting L-phenylalanine into *trans*-cinnamic acid (Pan et al., 2013; Howles et al., 1996). The *pal1* gene was up-regulated in the transgenic plants, and this may be responsible for the accumulation of 4-caffeoyl quinic acid. PAL activity is tightly regulated during development to control the synthesis of flavonoid pigments and lignin during xylogenesis, and also by environmental cues to control the synthesis of protective compounds (Elkind et al., 1994). The induction of *pal1* in our tobacco plants suggests that PAL transcription is a major control point for the initiation of phenylpropanoid biosynthesis in response to developmental and environmental cues (Elkind et al., 1994).

Another non-target metabolite that accumulated in the transgenic plants was betaine, one of two major organic osmolytes that accumulate in response to environmental stresses such as drought, salinity, extreme temperatures, UV radiation and heavy metals (Ashraf et al., 2007). Although their precise roles in osmotolerance remain controversial, these compounds are thought to be involved in the protection of cellular macrocomponents, such as protein complexes and membranes, mediating osmotic adjustments in plants grown under stress conditions (Ashraf et al., 2007; Sakamoto et al., 2000).

Our data suggest that the metabolite changes in non-target pathways reflect a stress response induced by the expression of *AtGpps* and *VoGes*, suppressing terpenoid biosynthesis and modulating normal physiological processes. Sesquiterpenoids and diterpenoids have been studied extensively by plant pathologists and phytochemists,

because they function as phytoalexines to protect plants from herbivores or microbial pathogens (Yazaki et al., 2004). The modulation of this pathway induces changes in the levels of nicotine and 4-caffeoyl quinic acid, the latter giving rise to caffeine. Both nicotine and caffeine are potent insecticides (Casanova et al., 2002; Willoughby et al., 2006).

The low levels of aspartic acid and glutamic acid in the transgenic plants reveal changes in primary metabolism. These amino acids may be associated with the γ-aminobutyrate (GABA)-shunt, a central stress-related pathway (Fait et al., 2007). GABA metabolism through the GABA-shunt may confer protection against oxidative stress, pH regulation, redox regulation, energy production and the maintenance of physiological carbon/nitrogen balance (Fait et al., 2007; Lancien et al., 2006). GABA may function as a buffer for glutamate and other amino acids, and as a potential compatible osmolyte to protect against osmotic stress (Lancien et al., 2006). The accumulation of so-called “compatible” osmolytes, such as glucose, fructose, inositol, proline and GABA appears to be a common metabolic response to salinity in order to maintain osmotic balance and protect protein structures (Zhang et al., 2011). In addition to glutamic acid, aspartic acid, choline and inositol also are related to osmotic stress but these were found at lower levels in the transgenic plants (Zhang et al., 2011). As discussed above, betaine was found at high levels and also is related to salinity stress. The low levels of aspartic acid in contrast of the high levels of malic acid suggest a relationship with the malic-aspartate shuttle, which regulates osmosis. The loss of choline is probably related to salinity-induced alterations in membrane synthesis because choline is an intermediate in the biosynthesis of cell membrane components (Zhang et al., 2011). The low level of choline suggests that the suppression of terpenoid biosynthesis may inhibit membrane synthesis or enhance membrane degradation. Inositol may be used to manage osmotic stress under salinity stress, and this also plays an important role in membrane biosynthesis, membrane protection by scavenging free radicals, signaling and the biosynthesis of cell wall components (Klages et al., 1999).

Most of the non-target metabolic changes we observed are directly or indirectly related to osmotic stress, suggesting that the suppression of terpenoid biosynthesis due to the expression of *AtGpps* and *VoGes* induces an osmotic effect in the cell thus triggering an osmotic stress response.

3.7. Conclusion

The metabolomic analysis of transgenic tobacco plants expressing *AtGpps* and *VoGes* provided insight into the interaction between heterologous and endogenous pathways. The sense suppression of terpenoid biosynthesis in the transgenic plants highlighted the importance of understanding the multilevel regulation of endogenous metabolic pathways, so that effective metabolic engineering strategies can be designed for secondary metabolism. The unexpected impact on non-target metabolites suggests that the expression of *AtGpps* and *VoGes* in transgenic tobacco plants triggers a stress response, resulting in the accumulation of protective metabolites such as nicotine and betaine, and the depletion of others such as choline or inositol.

3.8. References

Ashraf M., Foolad M. R. (2007) Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environmental and Experimental Botany*, 59: 206-216.

Busch M., Seuter A., Hain R. (2002) Functional analysis of early steps of carotenoid biosynthesis in tobacco. *Plant Physiology*, 12: 439 – 453.

DeBoer D. K., Dalton H. L. Edward F. J., Ryan S. M., Hamill J. D. (2012) RNAi-mediated down-regulation of ornithine decarboxylase (ODC) impedes wound-stress stimulation of anabasine synthesis in Nicotiana glauca. *Phytochemistry*, 86: 21-28.

Farhi M., MArhevka E., Ben-Ari J., Algamas-Dimantov A., Liang Z., Zeevi V., Edelbaum O., Spitzer-Rimon B., Abeliovich H., Scwartz B., Tzfira T., Vainstein A. (2011) Generation of the potente anti-malaria drug artemisinin in tobacco. *Nature Biotechnology*, 29: 1072-1074.

Fray R.G., Wallace A., fraser P. D., Valero D., Hedden P., Bramley P. M. (1995) Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. *Plant Journal*, 8: 693-701.

Gómez-Galera S., Pelacho A. M., Gené A., Capell T., Christou P. (2007) The genetic manipulation of medicinal and aromatic plants. *Plant Cell Rep*, 26: 1689-1715.

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

Kapper I. F., Aharoni A., van Herpen T. W. J. M., Luckerhoff L. L .P., Dicke M., Bouwmeester H. J. (2005) Genetic engineering of terpenoid metabolismo attracts bodyguards to *Arabidopsis*. *Science*, 309: 2070-2073.

Klages K., Boldingh H., Smith G. S. (1999) Accumulation of *myo*-inositol in *Actinidia* seedlings subjected to salt stress. *Ann Bot*, 84: 521-527.

Kim H. K., Choi Y. H., Verpoorte R. (2010) NMR-based metabolomic analysis of plants. *Nature Protocols* 5: 536 – 549.

Kumar S., Hahn F. M., Baidoo E., Kahlon T., Wood D. F., maman C. M., Cornish K., Keasling J. D., Daniell H., Whalen M. C. (2012) Remodeling the isoprenoid pathway in tobacco by expressinf the cytoplasmic mevalonate pathway in chloroplasts. *Metabolic engineering*, 14: 19-28.

Miralpeix B., Rischer H., Häkkinen S.T., Ritala A., Seppänen-Laakso T., Oksman-Caldentey K-M., Capell T., Christou P. (2013) Metabolic engineering of plant secondary products: which way forward? *Current Pharmaceutical Design*. *In press*

Muntendam R., Melillo E., Ryden A., Kayser O. (2009) Perspective and limits of engineering the isoprenoid metabolism in heterologous hosts. *Appl Microbiol Biotechnology*, 84: 1003-1019.

Naqvi S., Zhu C., Farre G., Ramessar K., Bassie L., Breitenbach J., Conesa D. P., Ros G., Sandmann G., Capell T., Christou P. (2009) Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proc Natl Acad Sci USA*, 106: 7762-7767.

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

Pan X., Li Y., Zhang H., Huang R., Liu W., Ming J., Liu S. Li X. (2013) Expression of signalling and defence-related genes mediated by over-expression of JERF1, and increased resistance to sheath blight in rice. *Plant Pathology*, *in press*. DOI: 10.1111/ppa.12064.

Sandman G., Römer S., Fraser P. D. (2006) Understanding carotenoid metabolism as a necessity for genetic engineering of crop plants. *Metabolic Engineering*, 8: 291-302.

Schäfer H., Wink M. (2009) Medicinally important secondary metabolites in recombinant microorganisms or plants: Progress in alkaloid biosynthesis. *Biotechnology Journal*, 4: 1684-1703.

Shoji T., Hashimoto T. (2008) Why does Anatabine, But not Nicotine, Accumulate in Jasmonate-Elicited Cultured Tobacco BY-2 Cells? *Plant Cell Physiology*, 49: 1209-1216.

Shoji T., Kajikawa M., Hashimoto T. (2010) Clustered Transcription Factor Genes Regulate Nicotine Biosynthesis in Tobacco. *The Plant Cell*, 22: 3390–3409.

Takizawa M., Hori K., Inai K., Takase H., Hashimoto T., Watanabe Y. (2007) A virus-induced gene silencing approach for the suppression of nicotine content in *Nicotiana benthamiana*. *Plant Biotechnology*, 24: 295-300.

Verpoorte R., van der Heijden R., Memelink J. (2000) Engineering the plant cell factory for secondary metabolite production. *Transgenic Research*, 9: 323–343.

Verpoorte R., Memelink J. (2002) Engineering secondary metabolite production in plants. *Current Opinion in Biotechnology*, 13: 181–187.

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

Wu S., Chappell J. (2008) Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. *Current opinion in Biotechnology*. 19: 145–152.

Wu S., Jiang Z., Kempinski C., Nybo S. E., Husodo S., Williams R., Chappell J. (2012) Engineering triterpene metabolismo in tobacco. *Planta*, 236: 867-877.

Yazaki K. (2004) Handbook of Plant Biotechnology. *John Wiley & Sons, Ltd*. Christou P and Klee H. England. 2: 811-861.

Yazaki K., Sasaki K., Tsurumaru Y. (2009) Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry*, 70: 1739-1745.

Zhang J., Zhang Y., Dy Y., Chen S., Tang H. (2011). Dynamic metabonomic responose of tobacco (*Nicotiana tabacum*) plants to salt stress. *Journal of proteome research*, 10: 1904-1914.

Zhong R., Morrison W.H., Negrel J., Yea Z.H. (1998) Dual Methylation Pathways in Lignin Biosynthesis. *The Plant Cell*, 10: 2033–2045.

CHAPTER 4

PROTEOMIC ANALYSIS OF TRANSGENIC TOBACCO PLANTS EXPRESSING EARLY GENES OF THE MONOTERPENE SECOIRIDOID ALKALOID PATHWAY

CHAPTER 4: PROTEOMIC ANALYSIS OF TRANSGENIC TOBACCO PLANTS EXPRESSING EARLY GENES OF THE MONOTERPENE SECOIRIDOID ALKALOID PATHWAY

4.1. Abstract

The early part of the monoterpene secoiridoid pathway was reconstructed in tobacco plants by coexpressing the genes *VoGes* and *Cyp76b6*. We analyzed the flower proteome from two transgenic tobacco lines expressing both transgenes at high levels and compared them with wild-type plants. The microsomal fractions were isolated, purified and analyzed by mass spectrometry after labeling with isobaric tags for relative and absolute quantification (iTRAQ). We identified and quantified 1398 proteins but there were no significant changes in abundance between the transgenic and wild type samples. The GES and CYP76B6 proteins were not detected.

4.2. Introduction

4.2.1. Proteomic tools

Proteomics is the systematic analysis of the protein complement of the genome. It is a rapidly developing and innovative field under the umbrella of large-scale biology, and promises to bridge the divide between DNA sequence information and biological systems (Twyman, 2012).

The term is commonly and broadly used to encompass almost any aspect of protein structure, expression and function, but it often refers to the identification (and quantitation) of all the protein components in a cell or tissue. Proteomic tools are diverse, reflecting the many different structural and functional properties of proteins, e.g. polypeptide synthesis, post-translational modification, protein-protein interactions, degradation, activity and the structures of protein complexes (Twyman, 2012; Bantscheff et al., 2012). Proteomics also provides the means to validate predicted gene functions (Twyman, 2012).

Over the last two decades, two-dimensional gel-electrophoresis (2DGE) has been the method of choice for the separation of complex protein mixtures. In this technique, protein mixtures are first separated by isoelectric focusing according to their net

charge using an immobilized pH gradient (IPG) gel. Orthogonal second-dimension separation is usually achieved by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions, which separates proteins according to their molecular weight (Thelen et al., 2007; Lilley et al. 2006). This technique is powerful enough to resolve thousands of the major proteins in a tissue or subcellular fraction (Lilley et al. 2006). Other orthogonal combinations have been described such as replacing the strong denaturing agent SDS with mild detergents and the dye Coomassie Brilliant Blue, a technique known as blue native polyacrylamide gel electrophoresis (BN-PAGE). In combination with second-dimension SDS-PAGE, this allows the separation in two dimensions of strongly hydrophobic proteins and provides information on native interactions by comparing subunit clusters on blue native (BN) and SDS gels (Eubel et al. 2005).

Gel-based methods have several limitations and often fail to detect proteins of low abundance (e.g. transcription factors), or low solubility (e.g. membrane-spanning receptors). Therefore, non-gel-based proteomics approaches have been developed based on liquid chromatography (LC) coupled to mass spectrometry (MS), e.g. multidimensional protein identification technology (MudPIT) (Bantscheff et al., 2012; Hennig et al. 2007; Di Palma et al., 2012) (Figure 1). LC-based methods offer flexibility because a wide range of stationary and mobile phases can be used to resolve complex biological samples at the protein or peptide levels. In multidimensional LC, proteins are usually digested into peptides prior to separation first by cation exchange chromatography and then C_{18} reversed-phase column chromatography. The advantage of such an approach is that the resolved peptides from the C_{18} column can be directly introduced into a mass spectrometer. Chemical tagging (e.g. stable isotope labeling) allows the relative quantification of protein samples by LC-MS (Bantscheff et al., 2012). MudPIT uses multiple liquid chromatographic separations interfaced with electrospray ionization tandem mass spectrometry (ESI-MS/MS). The sample peptides are separated by cation exchange and reversed-phase chromatography before ESI-MS, resulting in the acquisition of many hundreds of tandem spectra with 0.5% variance between two separate analytical runs and a dynamic range over five orders of magnitude (Figure 1). The system is fully automated in that the spectra can be acquired automatically with minimal pre-fractionation, e.g. using a 2DGE gel system (Twyman, 2012).

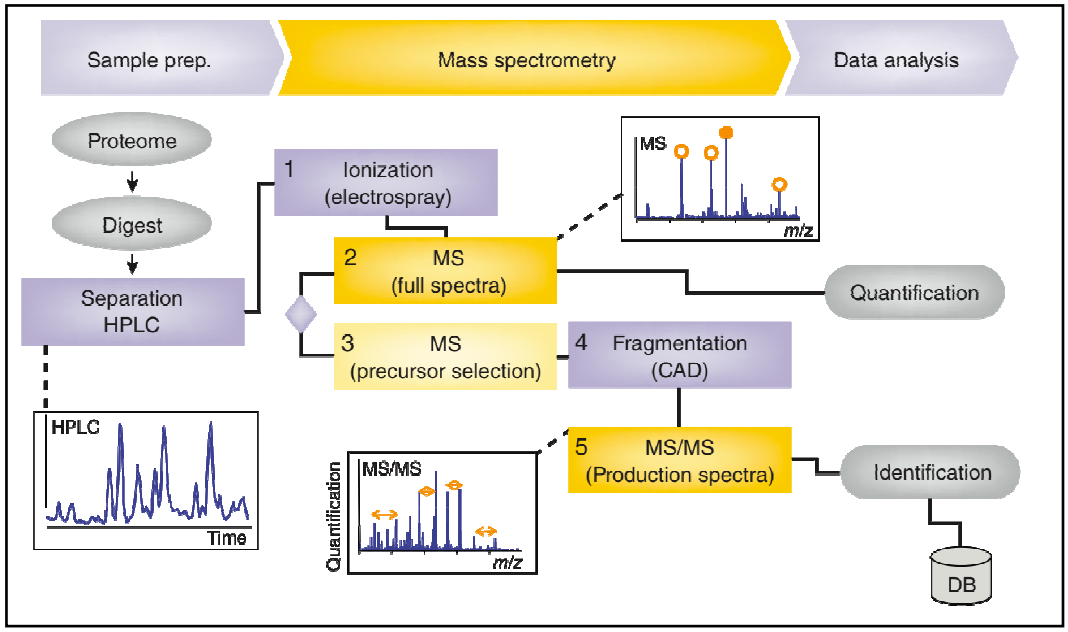

Figure 1. Workflow of a typical proteomic experiment. Proteins are digested to produce a omplex mixture of peptides, which are separated by HPLC before analysis by MS (Domon et al., 2010)

More recent methods for protein quantification include ICAT (isotope coded affinity tagging), SILAC (stable isotope labeling with amino acids), iTRAQ and metabolic labeling. Protein samples are labeled with isotopically-distinct tags (ICAT), or isobaric tags that yield intense MS/MS signature ions (iTRAQ), or isotopically-distinct amino acids (SILAC), or a stable isotope-labeled compound as the sole source of one element (metabolic labeling). Protein quantification can then be achieved by comparing the MS peak areas of the peptides derived from the two samples (Hennig et al. 2007; Twyman, 2012; Bantscheff et al., 2012). However, the quantitative reproducibility and the number of replicate experiments required for statistical significance are yet to be completely resolved and ICAT technology is still under development (Twyman, 2012).

The limitations of 2DGE in comparative studies have been addressed by the development of difference in-gel electrophoresis (DIGE) which involves the labeling of comparative samples with one of three spectrally-resolvable fluorescent CyDyes (Cy2, Cy3 and Cy5) allowing the multiplex separation of different samples on the same gel to eliminate reproducibility errors (Thelen et al., 2007; Rose et al., 2006; Twyman, 2012; Lilley et al., 2006). Following electrophoresis, the gel is scanned

using a variable wavelength laser-based imaging system and the distinct excitation and emission spectra of Cy3 and Cy5 allow the rapid quantitation of proteins in either of the original two extracts and the direct comparison of protein abundance between samples. The incorporation of Cy2 allows the comparison of three protein samples provided an internal standard is available for sample normalization. This allows the accurate statistical analysis of protein expression across multiple DIGE gels. The DIGE approach not only eliminates inter-gel variation, but also allows rapid and accurate comparative analysis (Rose et al., 2006; Twyman, 2012; Lilley et al. 2006).

Separated proteins can be identified using a number of different methods, but mass spectrometry after digestion into peptides with a site-specific protease (usually trypsin) is the method of choice for both protein identification and the characterization of post-translational modifications (Rose et al., 2006; Twyman, 2012; Lilley et al. 2006). Two MS platforms are widely used in proteomics. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS is generally used to measure the masses of peptides derived from a parent protein, generating a 'peptide mass fingerprint' (PMF). Several software packages are available that can compare the peptide mass list with a predicted 'theoretical' list of tryptic peptide fragments for every protein in the public databases, together with equivalent translated genomic and EST datasets (Rose et al., 2006). A common alternative to the PMF approach is *de novo* sequencing by electrospray ionization tandem mass spectrometry (ESI-MS/MS) which yields the amino acid sequences of selected tryptic peptides (Rose et al., 2006).

4.2.2. Limitations of proteomic methods

Plant samples can be challenging to analyze using proteomic methods because the protein concentrations are low and the cells contain abundant endogenous proteases that interfere with downstream separation and analysis. Other molecules that interfere with extraction and separation include cell wall and storage polysaccharides, lipids, phenolic compounds and many secondary metabolites (Rose et al., 2006). Indeed, the prevalence of these compounds possibly represents the most significant problem associated with plant proteomics (Rose et al., 2006).

Protein extraction and sample preparation are critical steps in proteomics, and ideally the extraction protocol should reproducibly capture and solubilize the full

complement of proteins in a given sample while minimizing post-extraction artifacts and non-protein contaminants (Rose et al., 2006). However, given the diverse size, charge, hydrophobicity, post-translational modification, complex formation and cellular distribution properties of proteins, no single extraction protocol or solvent system can capture the entire proteome (Rose et al., 2006). A range of different extraction protocols have therefore been developed, involving many permutations of physical treatments, solvents and buffers (Rose et al., 2006). Sample preparation has to be optimized for each sample type, always considering the purpose behind each experiment (Bantscheff et al., 2010). The analysis of large proteomes consisting of thousands of proteins often suffers from a limited dynamic range, such that only the most abundant proteins are detected (Figure 2). The dynamic resolution can be improved by fractionating the proteome into smaller sub-proteomes (Twyman, 2012).

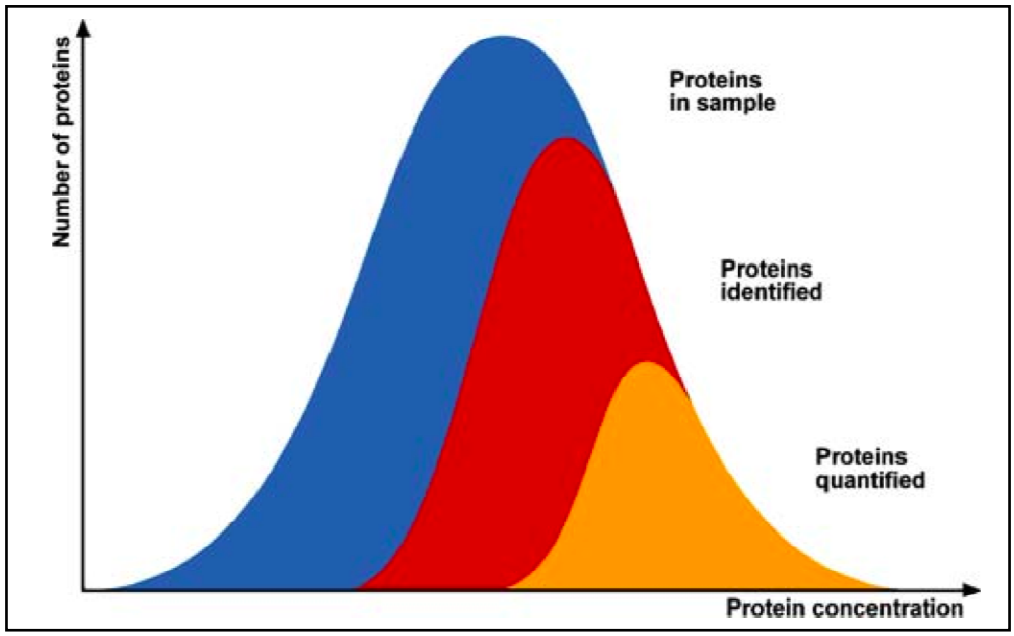

Figure 2. Schematic representation of the fraction of a proteome that can by identified or quantified by mass-spectrometry-based proteomics (Bantscheff et al., 2012).

As discussed above, gel-based methods for protein separation are incompatible with certain classes of proteins, including hydrophobic proteins, membrane-bound proteins, glycoproteins, and scarce proteins such as transcription factors and signaling molecules (Rose et al., 2006; Eubel et al. 2005; Hennig et al. 2007; Whasburn et al 2001; Twyman, 2012). It is also difficult to automate 2DGE techniques, limiting the throughput and introducing greater experimental variability through manual intervention (Rose et al., 2006). LC-based approaches such as MudPIT can generate an exhaustive catalogue of proteins but the data are not quantitative (Twyman, 2012; Rose et al., 2006). ICAT, ITRAQ and similar methods have been developed to allow the direct quantitative comparison of proteins by mass spectrometry (Rose et al., 2006).

The acquisition of quantitative data from 2D gels requires accurate methods for image capturing, spot editing and quantification, annotation, comparisons and the generation of web-formatted datasets. This requires substantial manual intervention even when dedicated software packages are used, due to the inherent variability in the preparation of multiple samples and subtle differences in the physical nature of gel matrices. These factors ensure that duplicate protein samples never generate identical 2DGE profiles (Twyman, 2012; Rose et al., 2006). Spot processing software can cost between $10,000 and $100,000 depending on the sophistication, but all packages require considerable effort to familiarize the user (Roses et al. 2006). Comparative 2DGE therefore need to be verified manually and many experimental replicates are needed for statistical validation, making the approach error-prone and laborious (Twyman, 2012).

Another bottleneck lies in the development of bioinformatics tools and robust prediction algorithms based on plant gene and protein sequences in order to facilitate data organization and the identification of proteins. Although 14 plant genomes are now sequenced and more than 70 have been mapped, this nevertheless is a small amount of data compared to equivalent resources microbes, insects and mammals (Bantscheff et al., 2012). However, the data gap has been partially bridged by large-scale EST sequencing projects that have generated more than 40,000 ESTs representing hundreds of plant species. These sequences are available to the public via the NCBI database dbEST.

4.2.3. Applications of proteomics in plants

One of the most common applications of proteomics is to compare contrasting samples and identify the proteins that are differentially expressed. The 2DGE-based comparative proteomics approach has been used in this way to investigate developmental processes such as germination, secondary cell wall and wood formation, somatic embryogenesis and leaf senescence, and to contrast global expression in wild plants with that in mutants and transgenic lines (Twyman, 2012; Bantscheff et al., 2012). Similarly, comparative proteomics has been used to study plant interactions with symbiotic organisms and pathogens, and responses to stimuli such as jasmonic acid, UV radiation, toxic metals, salinity, osmotic stress, anoxia, ozone stress and drought (Twyman, 2012; Bantscheff et al., 2012). Changes in the patterns of protein phosphorylation can also be monitored to study signal transduction cascades (Twyman, 2012). Comparative proteomics has also been used to investigate genetic diversity among cultivars of maize, rice, wheat and pepper, ecotypes of maritime pine, and genetic variability in a relative of sweet potato (Twyman, 2012).

Proteomics can also be used to associate patterns of protein expression with complex phenotypes much as DNA-based genetic markers are used to map quantitative trait loci (QTLs). Quantitative variations in the proteome, known as protein quantity loci (PQLs), can be associated with such phenotypes and candidate proteins can be identified for subsequent analysis. This approach has been used, for example, to identify PQLs associated with drought stress in maize (Twyman, 2012; Bantscheff et al., 2012).

Another major application of proteomics is as a platform for functional genomics. Next-generation sequencing methods provide deeper genome (DNA-Seq) and transcriptome (RNA-Seq) coverage by assembling millions of short reads, and have considerably increased the rate of data acquisition while lowering the costs. Much larger genomic contigs can be generated and full-length transcripts are easier to generate than with standard cDNA libraries thus improving the quality of proteomic annotation by peptide mass fingerprinting and de novo sequencing (Champagne et al., 2013). RNA-Seq data are particularly useful for protein annotation because they are less complex than DNA-Seq data (absence of introns and intergenic regions), thus

facilitating the identification and annotation of open reading frames. In addition, transcriptome sequencing allows the identification of alternative splice variants and multiple polyadenylation sites, an overlooked aspect of plant gene expression (Champagne et al., 2013).

Proteomics can also be applied to subcellular fractions, allowing protein localization to be integrated into models based on other aspects of protein behavior (Twyman, 2012; Lilley et al., 2006). Organelle-specific proteomic analysis can also reveal the dynamic nature of the protein population in specific subcellular compartments, e.g. in response to developmental and environmental signals. The isolation of proteins from specific subcellular compartments or organelles can also enrich for proteins that are comparatively rare and difficult to detect in total protein extracts (Twyman, 2012).

4.2.4. Proteomic analysis of secondary metabolism

Plant secondary metabolites are often produced by specialized structures such as glandular trichomes, which secrete particular metabolites in response to stress. Some of these compounds are commercially valuable, especially those produced by the *Lamiaceae* (mint, lavender), *Solanaceae* (tobacco), *Asteraceae* (sweet wormwood) and *Cannabaceae* (hop, cannabis) (Champagne et al., 2011). Trichome secondary metabolic pathways are therefore interesting and potentially valuable, e.g. those leading to the synthesis of terpenes, phenylpropanoids, alkaloids, and derivatives of sugars and lipids, so it is useful to consider the proteome of such structures to gain insight into the corresponding enzymes. Accordingly, Cutsem et al. (2011) analyzed the tobacco trichome proteome and identified many enzymes involved in secondary metabolic pathways and stress responses. Because tobacco sequence data are still limited, the identification of the trichome proteins relied predominantly on sequence data from other species (Cutsem et al., 2011). Many of the enzymes identified in this study were found to be involved in the synthesis of terpene precursors, e.g. geranylgeranyldiphosphate synthases and terpene synthases, which might be expected given that terpenes account for up to 73% of the trichome exudate in tobacco (Cutsem et al., 2011). They also identified eight P450 enzymes, including a member of the CYP71 family that modified the terpene profile of the trichome exudate and conferred increased resistance to aphids, a member of the CYP73 family homologous to

cinnamic acid 4-hydroxylase, and others involved in the formation of flavonoid and phenylpropanoid precursors. This study also led to the identification of 165 membrane proteins containing at least one transmembrane region, including eight ABC transporters, which couple active transport with ATP hydrolysis (Cutsem et al., 2011). Three of these have been characterized and are known to translocate ditrepenes (NtPDR1), fatty acids (AtABCD1) and play a role in cuticular transport (AtABCG11), but little is known about the others (Cutsem et al., 2011).

Secondary metabolism in plants is diverse, so there is no universal model system that is appropriate. For example, to study the terpenoid indole alkaloid compounds produced by the medicinal plant *Catharanthus roseus* it is necessary to work directly with this non-model species because there is no alternative (Champagne et al., 2013).

Jacob et al. (2005) carried out an in-depth proteomic analysis of *Catharanthus roseus* to demonstrate the feasibility of the approach, but most of the proteins identified were housekeeping proteins involved in common cellular processes such as glycolysis, oxidative phosphorylation, protein synthesis and folding (Jacob et al., 2005). It was assumed that most proteins responsible for secondary metabolism were too scarce to be detected, although the study identified two isoforms of strictosidine synthase which condenses tryptamine and secologanin to form strictosidine, the common precursor of all terpenoid indole alkaloids (Jacob et al., 2005). Among 88 protein spots selected for analysis, 56 were identified by peptide mass fingerprinting and the remainder could not be identified, the latter representing interesting targets for further investigation because the absence of homologous sequences in the databases suggests a potentially novel function (Jacob et al., 2005).

A more recent proteomic analysis of the same species (Champagne et al., 2011) identified 1663 proteins, most of which were housekeeping proteins involved in primary metabolism. However, 63 enzymes potentially involved in secondary metabolism were also identified, including 22 involved in terpenoid biosynthesis and 16 putative transporters that could translocate the metabolites. Even so, the function of 30% of the proteins identified in this study remains unclear or unknown, revealing a persistent gap in our knowledge of secondary metabolism which must be bridged by functional analysis of the corresponding proteins (Champagne et al., 2011).

4.3. Aims and objectives

The aim of the experiments described in this chapter was to characterize, at the protein level, transgenic tobacco lines expressing two early genes of the monoterpene secoiridoid pathway, i.e. geraniol synthase (*Ges*) and geraniol 10-hydroxylase (*Cyp76b6*). The objective was to carry out a proteomic analysis of the transgenic tobacco plants in order to determine whether expression of *Ges* and *Cyp76b6* affected the proteome of these transgenic plants.

4.4. Materials and methods

4.4.1. Plant materials

Tobacco leaf pieces were transformed by particle bombardment as described in Chapter 2, Section 2.4.2., using the primary vector and the selectable marker construct n a 3:1 molecular weight ratio. Transgenic lines H and L were selected as described in Chapter 2, Section 2.6.1., based on their accumulation of the mRNAs for plastid-targeted geraniol synthase (*VoGes*) and cytoplasmic geraniol 10-hydroxilase (*G10h*) as detected by mRNA blot (Chapter 2, Section 2.6.3). Transgenic T_1 seeds from both lines were germinated under kanamycin selection (100 mg/L) and three replicate plants from each line were transferred to soil at the same day. Flowers from each plant were harvested at maturity.

4.4.2. Sample processing

Frozen flowers from the three replicate plants of lines L and H (and three wild type plants as negative controls) were processed for proteomic analysis by Antoine Champagne (Louvain University, Belgium). Corollas from the flowers were removed from the green calyx and ground in a mortar and pestle. A 500 mg sample of frozen powder was homogenized in 50 mM Tris-HCl, pH 8.0, 2 mM EDTA, 5 mM DTT, protease inhibitor mix (1 mM PMSF, 2 μg/ml each of leupeptin, aprotinin, antipain, pepstatin and chymostatin), 0.6% w/v polyvinylpolypyrrolidone, 30 mM spermine. The homogenate was centrifuged for 5 min at 5300 x *g* and 4ºC, and the supernatant was then centrifuged for 30 min at 72,000 x *g* and 2ºC. The final supernatant was

precipitated with chloroform/methanol to recover microsomal fractions (Van Custem et al., 2011).

4.4.3. Reduction, alkylation, digestion and iTRAQ-labeling

Proteomic analysis was carried out by Antoine Champagne (Louvain University, Belgium). Proteins were suspended by vortexing in 50 mM NH_4HCO_3, 0.1% RapiGest (Waters) for 30 min at room temperature. Disulfide bond were reduced in 25 mM Tris(2-carboxyethyl) phosphine for 1 h at 60°C, then cysteine residues were blocked in 200 mM methyl methanethiosulfonate for 15 min at room temperature in the dark. The proteins were then digested with sequencing grade modified trypsin (Promega, Leiden, Netherland) at a protease/protein ratio of 1/20 w/w at 37°C for 16 h. RapiGest lysis was carried out by adding trifluoroacetic acid (1% final concentration), incubating for 1 h at 37°C, and centrifuging at 130,000 × g for 45 min at 4°C. The supernatant was vacuum dried (SpeedVac SC 200, Savant) and the peptides were labeled with isobaric tags for relative and absolute quantification (iTRAQ) according to the manufacturer's protocol (Applied Biosystems, Foster City, California). The samples were labeled with tags 113, 114 and 115 for the wild-type plants and tags 116–121 for the transgenic plants.

4.4.4. On-line MudPIT separation

On-line MudPIT separation was carried out by Antoine Champagne (Louvain University, Belgium). Three hundred µg of peptides were solubilized in loading buffer (2% ACN, 1% TFA) and separated on an Ultimate 3000 LC Packings system. The digested proteins were separated by ten-step LC using multidimensional protein identification (MudPIT) adapted from the method described by Motoyama et al. (2007). Briefly, the peptide mix (300 µl) was loaded (30 µl/min) onto a weak anion (Synchropack WAX) and strong cation (Saphir SCX) mixed-bed ion exchanger column (2:1, w/w) (Dr. Maisch GmbH, custom made, 1 mm × 150 mm, 6 µm) and was eluted in ten steps of 25, 50, 75, 100, 125, 150, 200, 250, 375 and 500 mM ammonium acetate in eluent A (pH 6.8). Peptides were eluted using a step gradient as follows: initial conditions 100% A; 0–5 min to 85% A/15% B; 5–65 min to 45% A/55% B; 65–75 min to 75% A/25% B, 75–80 min at 75% A/25% B. Each fraction

was recovered on a C_{18} trapping column (PepMap 100, 300 m × 5 mm, 5 µm, 100 Å pore size) before peptide separation on a reversed-phase analytical C_{18} nano-column (PepMap 100, 75 µm × 15 cm, 3 µm, 100 Å pore size). The mobile phases were water/ACN/TFA (96:4:0.1, v/v/v) as eluent A and water/ACN/TFA (20:80:0.1, v/v/v) as eluent B. The injection volume was 6.4 µl, the flow rate was 200 nl/min, and the column temperature was 25°C. Peptides were eluted with a step gradient as follows: initial conditions 95% A/5% B; 0–265 min to 60% A/40% B; 265–295 min to 40% A/60% B; 295–310 min to 10% A/90% B, 310–320 min at 10% A/90% B. The eluted fractions (10 x 360) were spotted onto two MALDI plates together with ionization matrix (4 mg/ml α-cyano-4-hydroxycinnamic acid, 70% ACN, 0.1% TFA, 10 mM $NH_4H_2PO_4$) using a Probot (LC Packings).

4.4.5. MALDI-MS/MS and database analysis

MALDI-MS/MS was carried out by Antoine Champagne (Louvain University, Belgium) as previously described (Szopinska et al., 2011). Briefly, the spotted plates were analyzed on an Applied Biosystems 4800 MALDI time-of-flight (TOF)/TOF Analyzer using a 200-Hz solid state laser operating at 355 nm. MS/MS spectra were obtained by automatic selection of the 15 most intense precursor ions per spot. Collision-induced dissociation was implemented with energy of 1 kV with air as the collision gas at a pressure of 1 x 10^6 Torr. Data were collected using Applied Biosystems 4000 Series Explorer™ software. LC-MS/MS data were processed using ProteinPilot software and the Paragon search algorithm (Shilov et al., 2007; Applied Biosystems/MDS SCIEX/4800 v4.0.8085). The MS/MS data were used to search the *Viridiplantae* part of the NCBInr database (downloaded on 3/1/2013). This algorithm in ProteinPilot was used with the “iTRAQ 8plex peptide labeled” sample type and a “biological modification ID focus” selected in the analysis method. “Quantification” as the sample type, “MMTS” as cysteine modification, “4800 TOF/TOF” as the instrument, and the “Thorough” and “Correction bias” preset search setting. In each experiment, bias was corrected for unequal mixing during the combination of the different labeled samples, based on the assumption that most proteins do not change in expression. Thus, if samples from each experimental condition are not combined in exactly equal amounts, bias correction fixes this systematic error. All reported proteins were identified with 95% or greater confidence, as determined by

ProteinPilot unused scores (>1.3) using the most stringent threshold of false positive discovery rate, i.e. below 10%. Protein grouping with ProteinPilot removed redundant hits.

4.5. Results

4.5.1. Protein identification

Samples from our transgenic plant lines and wild type controls were analyzed using iTRAQ technology for quantitative comparison, as recently reported in transgenic plants with a silenced H+-ATPase gene (Szopinska et al., 2011). The peptides were detected by MS, and the corresponding proteins were identified when at least one peptide exceeded the threshold score required to match a protein sequence in NCBInr database (downloaded on 3/1/2013).

We identified 1216 proteins (99% confidence) and 1398 proteins (95% confidence) with at least one peptide, and approximately 600 proteins (99% confidence) with at least two peptides. This is an unusually large number of proteins for the iTRAQ method, which typically identifies a few hundred proteins (Szopinska et al., 2011).

Twenty-eight proteins related to secondary metabolism were identified by manually screening the 1398 proteins (Table 1) although only two proteins (8-hydroxygeraniol dehydrogenase and geraniol dehydrogenase) were related to the monoterpene secoiridoid pathway. Many of the remaining proteins were related to the MEP, phenylpropanoid, flavonoid and shikimate pathways, as well as some proteins related to the synthesis of terpenoids, sesquiterpenoids, carotenoids, anthocyanins, alkaloids, isoquinoline alkaloids and phenylalanine.

Table 1. Proteins related to secondary metabolism found in the flowers of transgenic plants expressing *VoGes* and *Cyp76b6*.

N	Unused score	% Cov	Peptides (95%)	Name	Subcellular localization	Origin Species
28	19,33	61,40	14	Isoflavone reductase homolog		*Arabidopsis thaliana; Solanum tuberosum; Nicotiana tabacum; Zea mays; Medicago sativa*
75	13,63	35,85	7	Cinnamoyl-CoA reductase; Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase		*Arabidopsis thaliana; Malus domestica; Pyrus communis*
135	10,48	32,17	6	Caffeic acid 3-O-methyltransferase		*Capsicum annuum; Capsicum chinense; Catharanthus roseus; Populus tremuloides; Coffea canephora*
159	9,74	34,77	5	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	chloroplastic	*Arabidopsis thaliana; Oryza sativa subsp. Japonica*
203	8,09	34,13	7	Isopentenyl-diphosphate Delta-isomerase	chloroplastic	*Camptotheca acuminata; Arabidopsis thaliana; Clarkia breweri*
221	7,84	34,07	4	8-hydroxygeraniol dehydrogenase; Geraniol dehydrogenase; Mannitol dehydrogenase		*Catharanthus roseus; Ocimum basilicum; Fragaria ananassa; Mesembryanthemum crystallinum; Apium graveolens*
225	7,77	32,18	4	Chalcone--flavonone isomerase		*Nicotiana tabacum; Petunia hybrida; Camellia sinensis*
327	6,03	36,49	3	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	chloroplastic	*Arabidopsis thaliana; Oryza sativa subsp. japonica*
393	5,41	35,62	3	Perakine reductase; Probable aldo-keto reductase		*Rauvolfia serpentina; Glycine max; Arabidopsis thaliana*
521	4,21	29,85	2	1-deoxy-D-xylulose 5-phosphate reductoisomerase	chloroplastic	*Mentha piperita; Arabidopsis thaliana; Oryza sativa subsp. japonica*
557	4,02	31,38	2	Farnesyl pyrophosphate synthase; Farnesyl diphosphate synthase		*Lupinus albus; Parthenium argentatum; Artemisia annua; Artemisia spiciformis*

Table 1 continued

N	Unused score	% Cov	Peptides (95%)	Name	Subcellular localization	Species
570	4,00	26,52	6	Polyphenol oxidase; Catechol oxidase	chloroplastic	*Solanum lycopersicum; Solanum tuberosum*
627	3,77	21,23	2	Phenylalanine ammonia-lyase		*Nicotiana tabacum; Solanum lycopersicum; Solanum tuberosum*
661	3,62	30,94	2	Acetylajmalan esterase; GDSL esterase/lipase At1g28590; GDSL esterase/lipase At5g03980; GDSL esterase/lipase At1g28600; GDSL esterase/lipase At1g31550		*Rauvolfia serpentina; Arabidopsis thaliana*
663	3,60	30,58	2	Caffeoyl-CoA O-methyltransferase; Probable caffeoyl-CoA O-methyltransferase At4g34050		Nicotiana tabacum; Eucalyptus globulus; Arabidopsis thaliana; Solanum tuberosum
706	3,28	26,63	2	Dihydroflavonol-4-reductase; Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase		Callistephus chinensis; Arabidopsis thaliana; Gerbera hybrida; Malus domestica; Pyrus communis
728	3,15	22,67	2	15-cis-phytoene desaturase; Phytoene dehydrogenase; 15-cis-phytoene desaturase	Chloroplastic chromoplastic	Capsicum annuum; Solanum lycopersicum; Glycine max; Arabidopsis thaliana; Narcissus pseudonarcissus
816	2,69	25,67	1	Anthocyanin 5-aromatic acyltransferase; Malonyl-coenzyme:anthocyanin 5-O-glucoside-6'''-O-malonyltransferase; Malonyl-coenzyme A:anthocyanin 3-O-glucoside-6''-O-malonyltransferase; Agmatine coumaroyltransferase; BAHD acyltransferase DCR		Gentiana triflora; Salvia splendens; Dahlia pinnata; Arabidopsis thaliana
846	2,54	32,72	2	4-coumarate--CoA ligase		Nicotiana tabacum; Solanum tuberosum; Vanilla planifolia
884	2,35	36,93	5	Isopentenyl-diphosphate Delta-isomeraseOS		Camptotheca acuminata; Clarkia breweri; Clarkia xantiana
899	2,26	26,92	1	Shikimate O-hydroxycinnamoyltransferase; Anthranilate N-benzoyltransferase protein		Nicotiana tabacum; Arabidopsis thaliana; Dianthus caryophyllus

Table 1 continued

N	Unused score	% Cov	Peptides (95%)	Name	Subcellular localization	Species
901	2,26	23,96	2	Salutaridinol 7-O-acetyltransferase; Vinorine synthase; Deacetylvindoline O-acetyltransferase; BAHD acyltransferase At5g47980; Shikimate O-hydroxycinnamoyltransferase		Papaver somniferum; Rauvolfia serpentina; Catharanthus roseus; Arabidopsis thaliana
952	2,08	24,22	2	Cinnamoyl-CoA reductase; Tetraketide alpha-pyrone reductase 1; Dihydroflavonol-4-reductase; Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase		Arabidopsis thaliana; Dianthus caryophyllus; Pyrus communis
1137	2,00	26,05	1	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	chloroplastic/ chromoplastic	Solanum lycopersicum; Mentha piperita; Oryza sativa subsp. japonica; Arabidopsis thaliana
1169	2,00	25,32	1	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	chloroplastic	Catharanthus roseus; Oryza sativa subsp. japonica; Arabidopsis thaliana
1188	2,00	17,80	1	Xanthoxin dehydrogenase; Zerumbone synthase; Momilactone A synthase; Sex determination protein tasselseed-2		Arabidopsis thaliana; Zingiber zerumbet; Oryza sativa subsp. japonica; Zea mays
1219	1,91	28,11	3	Caffeoyl-CoA O-methyltransferase		Nicotiana tabacum; Populus tremuloides; Populus trichocarpa
1260	1,74	25,61	1	Cinnamoyl-CoA reductase; Tetraketide alpha-pyrone reductase 1; Dihydroflavonol-4-reductase; Dihydroflavonol-4-reductase		Arabidopsis thaliana; Vitis vinifera; Callistephus chinensis

N: rank of the specified protein relative to all other proteins in the list of detected proteins

Unused score: measurement of protein identification confidence taking into account peptides from spectra that have not already been “used” by higher-scoring proteins.

% Cov: percentage of matching amino acids from identified peptides having confidence greater than 0 divided by the total number of amino acids in the sequence

Peptides: the minimal number of high-confidence (95%) peptides used for identification.

4.5.2. Protein quantitation

We quantified the identified proteins by comparing the wild type replicate samples thus yielding a baseline for comparison, with the transgenic lines samples. There was no significant difference among the wild type samples, or between the wild type and transgenic plants. Therefore it was not possible to associate the observed differences in the metabolite profiles with any specific protein.

4.5.3. Geraniol synthase and geraniol 10-hydroxylase proteins

Although several proteins involved in secondary metabolism were identified (Table 1), the list did not include the products of the *VoGes* and *Cyp76b6* transgenes (geraniol synthase and geraniol 10-hydroxylase).

4.6. Discussion

Proteomics has been used to study many aspects of plant biology, including secondary metabolism. Jacobs et al. (2005) used 2DGE to demonstrate the feasibility of a proteomic approach for the identification of proteins involved in alkaloid biosynthesis, and to find novel sequences in *Catharanthus roseus* cell suspension culture. They analyzed 88 spots and identified 56 proteins by peptide mass fingerprinting, but relied almost exclusively on sequences from other plants reflecting the lack of *C. roseus* genome and protein sequence data. Several proteins were identified by MALDI-MS/MS including a novel strictosidine synthase, confirming that 2DGE coupled to MS can be used to identify proteins associated with secondary metabolism in plants (Jacobs et al., 2005).

Champagne et al. (2012) also used proteomics to identify candidate proteins involved in the synthesis and transport of secondary metabolites in cultured *C. roseus* cells. The comparison of two independent cell lines with different terpenoid indole alkaloid metabolic profiles by DIGE revealed 358 proteins with at least a two-fold difference in abundance between samples, 172 of which were identified by MS although most corresponded to housekeeping proteins (Champagne et al., 2012). DIGE analysis is not optimal for large-scale protein identification, especially because membrane-bound

enzymes and transporters are not detected using this approach (Champagne et al., 2012). Indeed, 1663 proteins were identified in one of the cell lines by LC-MS, but this may overestimate the number of genes involved because the lack of *C. roseus* sequence data makes it more difficult to identify isoforms and splice variants. The 1663 proteins represented only 1464 protein groups (Champagne et al., 2012). Furthermore, protein identification relied on an annotated *C. roseus* EST database rather than the entire NCBInr database, so the probability that different groups actually represent the same protein was low (Champagne et al., 2012). The 22 identified proteins that were directly relevant to TIA metabolism included tryptophan decarboxylase, anthranilae synthase, three isoforms of strictosidine synthase, potential homologs of polyneuridine-aldehyde esterase and vinorine synthase, a tabersonine 10-hydroxylase (CYP71D12), 16-methoxy-2,3-dihydrotabersonine *N*-methyltransferase (NMT), deacetoxyvindoline 4-hydroxylase (D4H) and DAT (Champagne et al., 2012). Although the last three enzymes in the TIA pathway were identified (tabersonine 16-hydroxylase, D4H and DAT) there was no trace of vindoline and vinblastine in the suspension cells. Further enzymatic steps might be required to produce these metabolites in suspension cells, or the enzymes were present in an inactive or barely-active form so that the concentration of the metabolites is below the detection threshold (Champagne et al., 2012). Some enzymatic steps were proposed by analogy with other species, but were not identified directly in *C. roseus*. The MS results therefore led to the identification of several candidates, such as 10HGO, DLGT and 7DLH (Champagne et al., 2012). However, the presence of a conserved sequence in two different plant species does not necessarily mean the corresponding proteins are orthologs. Classical biochemical approaches will be necessary to confirm the roles of enzymes acting at particular steps in the TIA pathway (Champagne et al., 2012). This proteomic investigation nevertheless represents an important data resource the identification of enzymes potentially involved in metabolic steps lacking a candidate enzyme in *C. roseus*.

Although 2DGE is the classical proteomic approach, it is not convenient for the analysis of proteins that are scarce, small (<50 kDa), hydrophobic/membrane-bound or that have extreme pI values (Van Custem et al., 2011; Champagne et al., 2012). The gel-free proteomic analysis was far more useful and led to the identification of a large number of proteins as described in this chapter.

The analysis reported in this chapter identified 600 proteins with at least two peptides, and 1398 with at least one peptide (with 95% of confidence). However, these numbers might overestimate the true number of proteins because we cannot predict the abundance of isoforms or the extent of gene families. In this context, MS analysis can identify multiple isoforms within a group or even several groups (Champagne et al., 2012). We were able to identify more proteins than in previous studies, where typically about 100 proteins can be identified. For example, Szopinska et al. (2011) compared plasma membrane proteome modifications in yeast cells exposed to mild (0.4 M NaCl) or high (1 M NaCl) salinity stress for three different time points. Using a well-established procedure to purify yeast plasma membrane proteins, a LC-iTRAQ revealed 113 plasma membrane proteins, compared to 29 identified by 2DGE (Szopinska et al., 2011; Insenser et al., 2006). Among the 113 plasma membrane proteins, 68% were integral membrane proteins, 18% were lipid-anchored and 14% were tightly associated with the plasma membrane (Szopinska et al., 2011).

Champagne et al. (2012) reported that ~30% of the proteins they identified had unclear or unknown functions, revealing important gaps in our knowledge of plant metabolism. This highlights the limitations of proteomic approaches that rely on databases containing non-annotated proteins, particularly in the context of secondary metabolism because the pathways are not present in well characterized model plants that benefit from the availability of complete genome sequences (Champagne et al., 2012). Similarly, the manual screening of 1398 proteins allowed us to identify only 28 associated with secondary metabolism, underscoring the relative lack of data relating to secondary metabolic pathways. We identified five of the eight proteins representing the MEP pathway, seven representing the flavonoid pathway and five representing the phenylpropanoid pathway, as well as other representing the shikimate, alkaloid, carotenoid, anthocyanin, isoquinoline alkaloid, sesquiterpenoid, and terpenoid pathways (Table 2). Although we identified proteins representing pathways that are unrelated to monoterpene secoiridoid biosynthesis, there were no quantitative changes in transgenic plants compared to wild type, suggesting the transgenes had no effect on these pathways at the proteome level.

Table 2. Classification of the identified proteins related to secondary metabolism

2ry metabolite pathway	N	Name
Flavonoid biosynthesis	28	Isoflavone reductase homolog
	75	Cinnamoyl-CoA reductase; Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase
	225	Chalcone-flavonone isomerase
	706	Dihydroflavonol-4-reductase; Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase
	816	Anthocyanin 5-aromatic acyltransferase; Malonyl-coenzyme:anthocyanin 5-O-glucoside-6’’’-O-malonyltransferase; Malonyl-coenzyme A:Anthocyanin 3-O-glucoside-6”-O-malonyltransferase; Agmatine coumaroyltransferase; BAHD acyltransferase DCR
	952	Cinnamoyl-CoA reductase; Tetraketide alpha-pyrone reductase 1; Dihydroflavonol-4-reductase; Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase
	1260	Cinnamoyl-CoA reductase; Tetraketide alpha-pyrone reductase 1; Dihydroflavonol-4-reductase; Dihydroflavonol-4-reductase
Phenylpropanoid biosynthesis	135	Caffeic acid 3-O-methyltransferase
	570	Polyphenol oxidase; Catechol oxidase
	627	Phenylalanine ammonia-lyase
	663	Caffeoyl-CoA O-methyltransferase; Probable caffeoyl-coA O-methyltransferase At4g34050
	1219	Caffeoyl-CoA O-methyltransferase
MEP pathway	159	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
	327	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
	521	1-deoxy-D-xylulose 5-phosphate reductoisomerase
	1137	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
	1169	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
Terpenoid biosynthesis	203	Isopentenyl-diphosphate Delta-isomerase
	221	8-hydroxygeraniol dehydrogenase; Geraniol dehydrogenase; Mannitol dehydrogenase
	393	Perakine reductase; Probable aldo-keto reductase
	557	Farnesyl pyrophosphate synthase; Farnesyl diphosphate synthase
	661	Acetylajmalan esterase; GDSL esterase/lipase At1g28590; GDSL esterase/lipase At5g03980; GDSL esterase/lipase At1g28600; GDSL esterase/lipase At1g31550
	728	15-cis-phytoene desaturase; Phytoene dehydrogenase; 15-cis-phytoene desaturase
	846	4-coumarate--CoA ligase
	884	Isopentenyl-diphosphate Delta-isomeraseOS
	1188	Xanthoxin dehydrogenase; Zerumbone synthase; Momilactone A synthase; Sex determination protein tasselseed-2

Table 2. Classification of the identified proteins related to secondary metabolism **(cont.)**

2ry metabolite pathway	N	Name
Shikimate pathway	899	Shikimate O-hydroxycinnamoyltransferase; Anthranilate N-benzoyltransferase protein
	901	Salutaridinol 7-O-acetyltransferase; Vinorine synthase; Deacetylvindoline O-acetyltransferase; BAHD acyltransferase At5g47980; Shikimate O-hydroxycinnamoyltransferase

As discussed in Chapter 2, Section 2.6.3, the accumulation of *VoGes* and *Cyp76b6* mRNA was confirmed in transgenic tobacco lines C and H. However, we did not detect the corresponding proteins in the proteome dataset. This may reflect the low sensitivity of the 2DGE approach, which only detects abundant proteins. It is easier to detect scarce transcripts because they can be amplified, but if the proteins are also scarce then they are likely to fall below the detection threshold of the methods we used (Jacobs et al., 2005; Goossens et al., 2007). The CYP76B6 protein was difficult to detect using non-gel technique in *C. roseus*, suggesting it is either not present in flowers, or is transported to other subcellular compartments or tissues to avoid toxicity (Champagne et al., 2012).

The first steps of the monterpene secoiridoid pathway (prior to 10-hydroxylation) in *C. roseus* take place in the internal (adaxial) phloem parenchyma of young leaves and flowers (Burlat et al. 2004; Mahroug et al., 2007). The genes encoding the relevant enzymes are expressed at higher levels in the young cells at the revolute base of the leaves and gene expression then declines as the cells mature towards the top of the leaves (Mahroug et al., 2007). This basipetal expression gradient suggests that the mRNA half-life is shorter than the corresponding protein half-life for all these genes, so that only small amounts of protein are synthesized depending on the developmental stage (Mahroug et al., 2007). Therefore, one potential explanation for our ability to detect transgene mRNA but not protein is that we analyzed the corollas of mature flowers, which contain few internal phloem parenchyma cells.

VoGes and *Cyp76b6* were expressed constitutively, and the proteins were targeted to the plastids and endoplasmic reticulum membrane, respectively. Plastid gene expression is controlled predominantly at the post-translational level by the regulation of protein complex assembly and proteolysis (Apel et al., 2010). Although our

understanding of plastid proteases has increased over the last decade, almost nothing is known about stability or instability determinants within the substrate proteins (Apel et al., 2010). Plastid proteases may therefore have degraded the GES protein thus preventing its detection.

4.7. Conclusions

MALDI-MS/MS combined with iTRAQ technology is a powerful approach that allows transgenic lines to be compared with wild type plants at the proteomic level. This avoids the limitations of traditional gel-based techniques and allows the identification and quantitation of many proteins. We identified 600 proteins with at least two peptides and 1398 with at least one peptide (with 95% confidence). The quantitative analysis of the flower proteome of two transgenic tobacco plants expressing *VoGes* and *Cyp76b6* at high levels revealed no significant changes in the proteome compared to wild-type tobacco plants. GES and CYP76B6 were not identified in the proteome of the transgenic plants despite the presence of the corresponding transcripts, suggesting these proteins accumulate to levels below the detection threshold of our protein analysis methods.

4.8. References

Apel W., Schulze W. X., Bock R. (2010) Identification of protein stability determinants in chloroplasts. *The Plant Journal,* 63: 636–650.

Bantscheff M., Lemeer S., Savitski M. M, Kuster B. (2010) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem,* 404: 939-965.

Burlat V., Oudin A., Courtois M., Rideau M., St-Pierre B. (2004) Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of Catharanthus roseus implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *The Plant Journal*, 38: 131-141.

Champagne A., Boutry M. (2013) Proteomics of non model plant species. *Proteomics,* 13: 663–673.

Champagne A., Rischer H., Oksman-Caldentey K-M., Boutry M. (2012) In-depth proteome mining of cultured Catharanthus roseus cells identifies candidate proteins involved in the synthesis and transport of secondary metabolites. *Proteomics,* 12: 3536–3547.

Collu G., Unver N., Peltenburg-Looman A. M. G., van der Heijden R., Verpoorter R., Memelink R. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme envolved in terpenoid indole alcaloide biosíntesis. *FEBS Letters,* 508: 215-220.

Cutsem, E., Simonart, G., Degand, H., Faber, A. M. (2011) Gel-based and gel-free proteomic analysis of *Nicotiana tabacum* trichomes identifies proteins involved in secondary metabolism and in the (a)biotic stress response. *Proteomics,* 11: 440–454.

Di Palma S., Hennrich M. L., Heck A. J., Mohammed S. (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J. Proteomics,* 75: 3791-3813.

Domon B., Aebersol R. (2010) Options and cosiderations when selecting a quantitative proteomic strategy. *Nature Biotechnology,* 28: 710-721.

Eubel H., Braun H. P., Millar A. H. (2005) Blue-native PAGE in plants: a tool in analysis of protein-protein interactions. *Plant Methods,* 1: 1-13.

Goossens A., Rischer H. (2007) Implementation of functional genomics for gene discovery in alkaloid producing plants. *Phytochem Rev*, 6: 35-49.

Hennig L. (2007) Patterns of beauty – omics meets plant development. *Trens in Plant Science*, 12: 1360-1385.

Inser M., Nombela C., Molero G., Gil C. (2006) Proteomic analysis of detergent-resistant membranes from *Candida albicans*. *Proteomics*, 6: 74-81.

Jacobs D. I., Gaspari M., van der Greef J., van der Heijden R., Verpoorte R. (2005) Proteome analysis of the medicinal plant *Catharanthus roseus*. *Planta*, 221: 690–704.

Lilley S. K., Dupree P. (2006) Methods of quantitative proteomics and their application to plant organelle characterization. *Journal of Experimental Botany*, 57: 1493–1499.

Mahroug S., Burlat V., St-Pierre B. (2007) Cellular and sub-cellular organisation of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus. Phytochem Rev*, 6: 363–381.

Mallick P., Kuster B (2010) Proteomics: a pragmatic perspective. *National biotechnology*, 28: 695-709.

NCBI (2013) Genome, accessed online 04 May, http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html#MAP

Piette A.S., Derua R., Waelkens E., Boutry M., Duby G. (2011) A phosphorylation in the C-terminal auto-inhibitory domain of the plant plasma membrane H□-ATPase activates the enzyme with no requirement for regulatory 14-3-3 Proteins. *The Journal of Biological Chemistry*, 286: 18474-18482.

Rose J. K. C., Bashir S., Giovannoni J. J., Jahn M. M., Saravanan R. S. (2006) Tackling the plant proteome: practical approaches, hurdles and experimental tools. *The Plant Journal*, 39: 715–733.

Szopinska, A., Degand, H., Hochstenbach, J. F., Nader, J. (2011) Rapid response of the yeast plasma membrane proteome to salt stress. *Mol. Cell Proteomics* 10.

Thelen J. J., Peck S. C. (2007) Quantitative Proteomics in Plants: Choices in Abundance. *The Plant Cell*, 19: 3339–3346.

Twyman R. (2012) Principles of proteomics. Advanced text. Catherine Jones. England, 2n Edition.

CHAPTER 5

MODULATION OF THE MONOTERPENE SECOIRIDOID PATHWAY IN TRANSGENIC TOBACCO PLANTS BY THE TRANSIENT EXPRESSION OF DOWNSTREAM PATHWAY GENES

CHAPTER 5: MODULATION OF THE MONOTERPENE SECOIRIDOID PATHWAY IN TRANSGENIC TOBACCO PLANTS BY THE TRANSIENT EXPRESSION OF DOWNSTREAM PATHWAY GENES

5.1. Abstract

Monoterpene secoiridoid biosynthesis contributes the terpene component of terpenoid indole alkaloids produced in the medicinal plant *Catharanthus rosues*. We generated transgenic tobacco plants expressing the early genes of the pathway, which encode the enzymes geraniol synthase and 10-hydroxygeraniol synthase. None of our transgenic plants accumulated detectable amounts of geraniol, presumably due to toxicity. However the transient expression of geraniol synthase in the transgenic plants resulted in high levels of geraniol and some of its derivatives. The transient expression of downstream pathway genes in the transgenic tobacco plants already expressing the two upstream genes did not achieve the production of 10-hydroxygeraniol, 10-hydroxygeranial, 10-oxogeranial or iridodial, as determined by LC-MS/TOF. However, all the agroinfiltrated samples produced unknown compounds that were not detected in the negative controls. These could not be assigned to particular structures due to the limited information available in structural databases. Our results suggest that the expression of downstream pathway genes produced novel metabolites that cannot presently be identified.

5.2. Introduction

Catharanthus roseus produces a wide range of terpenoid indole alkaloids (TIAs). Some of them are used to treat anxiety (serpentine), arterial hypertension (ajmalicine) or cancer (vinblastine, vincristine and their hemisynthetic derivatives, such as vinorelbine and vinflumine). The economic potential of *C. roseus* TIAs has ensured that the physiological, cellular, biochemical and molecular aspects of their biosynthesis have been studied for the past 30 years. TIAs are derived from strictosidine, which is produced by the condensation of the monoterpene precursor secologanin with the indole precursor tryptamine as shown in Figure 1 (Hedhili et al., 2007).

Secologanin is produced from isopentenyl diphosphate (IPP) via a number of steps including the formation of the monoterpene geraniol. In higher plants, two pathways are involved in the biosynthesis of the active isoprene unit (IPP and its isomer DMAPP). Plants have evolved to maintain the well-known eukaryotic mevalonic acid (MVA) pathway in the cytosol, also known as the classical pathway, but have also acquired the more recently discovered prokaryotic 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway, also known as the alternative pathway, from the endosymbiotic ancestor of plastids (Hemmerlin et al., 2003; Hedhili et al., 2007; Liu et al., 2007; Oudin et al., 2007). Under normal physiological conditions, cytoplasmic isoprenoids (such as sterols, sesquiterpenes, ubiquinones, or the side chain of mitochondrial ubiquinone) are synthesized from IPP via MVA, whereas plastidial isoprenoids (such as hemiterpenes, monoterpenes, diterpenes and carotenoids) originate from DMAPP and IPP, produced simultaneously via the MEP pathway (Hemmerlin et al., 2003; Hedhili et al., 2007; Liu et al., 2007; Oudin et al., 2007). In *C. roseus*, the MEP pathway is the major source of terpenoids (Hedhili et al., 2007).

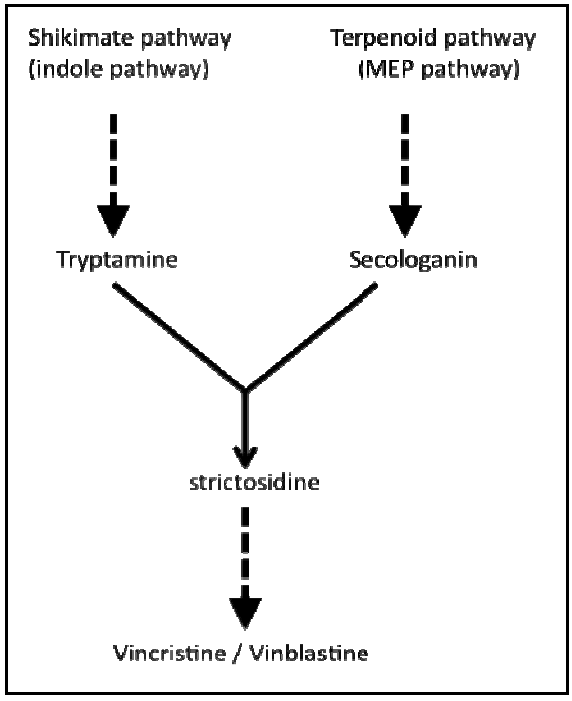

Figure 1. The terpenoid indole alkaloid pathway for the production of Vinca alkaloids

5.2.1. The shikimate and tryptophan pathways

The shikimate pathway links carbohydrate metabolism to the synthesis of aromatic amino acids (phenylalanine, tyrosine and tryptophan), which can in turn act as precursors for various primary and secondary metabolites. The pathway begins with two intermediates of carbohydrate metabolism, phosphoenolpyruvate (PEP) (from glycolysis) and erythrose 4-phosphate (E4P) (from the pentose phosphate pathway), which are converted to chorismate in a sequence of seven reactions as shown in Figure 2 (Goddijn et al., 1992; Guirimand et al., 2009; Dutta et al., 2007; Maeda et al., 2012; Tzin et al., 2010). All the shikimate pathway enzymes have been biochemically characterized, and the corresponding genes have been identified in both microbes and plants (Table 1) (Maeda et al., 2012). Based on experimental evidence and predictions of subcellular localization, all the shikimate pathway enzymes are found in the plastids (Maeda et al., 2012).

The first committed step in the shikimate pathway is catalyzed by 3-deoxy-D-*arabino*-heptulosonate-7-phosphate (DAHP) synthase, which condenses PEP and E4P to produce DAHP as shown in Figure 2 (Tzin et al., 2010; Maeda et al., 2012). Two DAHP synthase genes have been isolated from several plant species, *DAHPS1* and *DAHPS2*. The genes are differentially expressed in *Arabidopsis thaliana*, *Solanum lycopersicum* (tomato) and *Solanum tuberosum* (potato) (Table 1). *DAHPS2* (*shkB* in potato) is constitutively expressed, whereas *DAHPS1* (*shkA* in potato) is strongly induced in response to wounding and pathogen infection (Tzin et al., 2010; Maeda et al., 2012). The subsequent step is catalyzed by 3-dehydroquinate synthase (DHQS), which converts DAHP to 3-dehydroquinate. A *DHQS* gene has been isolated from tomato by complementing an *E. coli* mutant. This single-copy gene is strongly expressed in tomato roots, and its expression is induced by elicitor treatment in cell suspension cultures (Tzin et al., 2010; Maeda et al., 2012).

The third and fourth enzymatic reactions in the shikimate pathway include (*a*) the dehydration of 3-dehydroshikimate to introduce the first double bond in the ring and (*b*) the reversible reduction of 3-dehydroshikimate into shikimate using NADPH. These reactions are catalyzed by 3-dehydroquinate dehydratase (DHD) and shikimate dehydrogenase (SDH), the latter also known as shikimate:$NADP^+$ oxidoreductase (Figure 2). In plants, these enzymes are fused to form a bifunctional DHD-SDH

complex (Maeda et al., 2012). A single gene encoding DHD-SDH, which carries a putative plastid targeting peptide, has been identified in several plant species, including Arabidopsis (Table 1). The *Nicotiana tabacum* (tobacco) genome exceptionally contains two genes encoding plastid and cytosolic DHD-SDH efficiently (Tzin et al., 2010; Maeda et al., 2012). Suppression of the plastid *DHD-SDH* in tobacco leaves by RNA interference (RNAi) resulted in the accumulation of both 3-dehydroquinate and shikimate (the substrate and product of the enzyme, respectively) and reduced the levels of phenylalanine, tyrosine, lignin and chlorogenic acid (Tzin et al., 2010; Maeda et al., 2012). Shikimate kinase (SK), the fifth enzyme in the shikimate pathway, catalyzes the phosphorylation of the one C3 hydroxyl groups of shikimate using ATP as a cosubstrate to yield shikimate 3-phosphate. Different plant species have different numbers of isozymes, e.g. one in tomato, two in Arabiopsis, and three in *Oryza sativa* (rice) (Tzin et al., 2010; Maeda et al., 2012). The penultimate step in the shikimate pathway is catalyzed by 5-endolpyruvylshikimate 3-phosphate (EPSP) synthase (also known as 3-phosphoshikimate 1-carboxyvinyltransferase). EPSP synthase transfers the enolpyruvyl moiety of PEP to the 5-hydroxyl position of shikimate 3-phosphate, forming EPSP (Figure 2) (Tzin et al., 2010; Maeda et al., 2012). Genes encoding EPSP synthases have been isolated from a number of plants (Table 1). Two isozymes with very similar kinetic properties have been purified from *Zea mays* (maize) and the Arabidopsis genome contains two EPSP synthase genes, one of which is known to be functional. EPSP synthase is generally expressed constitutively at low levels but tissue-specific and developmentally regulated isozymes are found in *Petunia hybrida* (petunia) flowers, probably support the production of volatiles derived from phenylalanine (Maeda et al., 2012). Chorismate synthase (CS) catalyzes the final step of the shikimate pathway, the 1,4-*anti*-elimination of the 3-phosphate and C6-*pro*-R hydrogen from EPSP, introducing the second double bond into the ring to produce chorismate (Figure 2) (Tzin et al., 2010; Maeda et al., 2012). Thus far, plant genes encoding CS have been isolated from *Corydalis sempervirens* and tomato, the latter producing two differentially expressed isozymes (Maeda et al., 2012).

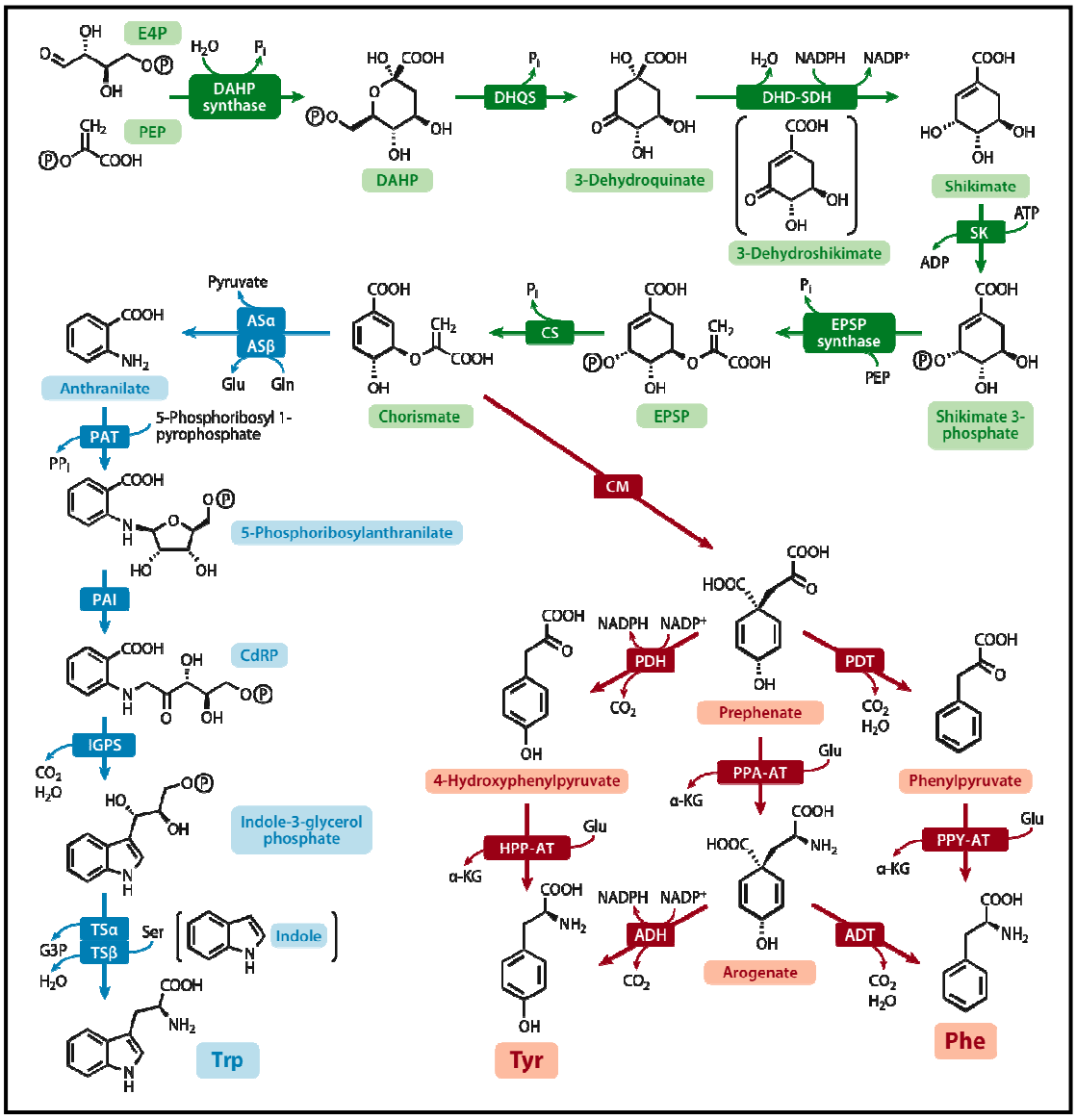

Figure 2. The shikimate pathway in plants (shown in green), the tryptophan (Trp) pathway (blue) and the synthesis of phenylalanine (Phe) and tyrosine (Tyr) via arogenate or phenylpyruvate/ 4-hydroxyphenylpyruvate (red). E4P: D-erythrose 4-phosphate; PEP: phosphoenolpyruvate; CdRP: 1-(o-carboxyphenylamino)-1-deoxy-ribulose 5-phosphate; DAHP: 3-deoxy-D-arabino-heptulosonate 7-phosphate; EPSP: 5-enolpyruvylshikimate 3-phosphate; G3P, glyceraldehyde 3-phosphate; Gln: glutamine; Glu: glutamate; α-KG: α-ketoglutarate; Pi: inorganic phosphate; PPi: inorganic diphosphate; Ser: serine. Enzyme abbreviations: ADH: arogenate dehydrogenase; ADT: arogenate dehydratase; ASα: anthranilate synthase α subunit; ASβ: anthranilate synthase β subunit; CM: chorismate mutase; CS: chorismate synthase; DHD: 3-dehydroquinate dehydratase; DHQS: 3-dehydroquinate synthase; HPP-AT: 4-hydroxyphenylpyruvate aminotransferase; IGPS: indole-3-glycerol phosphate synthase; PAI: phosphoribosylanthranilate isomerase; PAT: phosphoribosylanthranilate transferase; PDH: prephenate dehydrogenase; PDT: prephenate dehydratase; PPA-AT: prephenate aminotransferase; PPY-AT: phenylpyruvate aminotransferase; SDH: shikimate dehydrogenase; SK: shikimate kinase; TSα: tryptophan synthase α subunit; TSβ: tryptophan synthase β subunit (Maeda et al., 2012).

In plants, chorismate is a common precursor for at least four metabolic pathways leading to the formation of tryptophan, phenylalanine/tyrosine, salicylate/phylloquinone, and folate. Four enzymes – CM, anthranilate synthase (AS), isochrosmate synthase (ICS) and aminodeoxychorismate synthase (ADCS) – catalyze the committed steps of these pathways and compete for chorismate (Tzin et al., 2010; Maeda et al., 2012). The tryptophan pathway converts chorismate into tryptophan via six enzymatic reactions (Figure 2). All enzymes involved in the pathway are localized in the plastids (Table 1) (Maeda et al., 2012).

Anthranilate synthase (AS) is an amino-accepting chorismate-pyruvate lyase that catalyzes the formation of anthranilate (Figure 2). AS consists of large α and small β subunits (ASα and ASβ, respectively), which form an α/β heterodimer or an α_2/β_2 tetramer (Tzin et al., 2010; Maeda et al., 2012). ASα binds to chorismate and catalyzes the amination and pyruvate elimination reactions, whereas ASβ hydrolyzes glutamine and provides ammonia to ASα. The binding of chorismate to ASα triggers a conformational change to an active state and creates an intermolecular channel for ammonia transfer from ASβ to ASα. The AS enzyme is allosterically inhibited by tryptophan, which binds to ASα and restricts its conformational change, suggesting that ASα is responsible for feedback inhibition by tryptophan (Tzin et al., 2010; Maeda et al., 2012). Most ASα subunits in plants are feedback-sensitive to tryptophan, with the known exceptions of feedback-insensitive tobacco and *Ruta graveolens* ASα. Higher plants contain at least two genes encoding ASα and one gene encoding ASβ (Table 1). One of the ASα genes is constitutively expressed, whereas the other is regulated developmentally and induced in response to wounding and pathogens, suggesting its involvement in the production of natural products derived from the tryptophan pathway as a form of defense (Goddijn et al., 1992; Guirimand et al., 2009; Dutta et al., 2007; Maeda et al., 2012).

Phosphoribosylanthranilate transferase (PAT) transfers the phosphoribosyl moiety from phosphoribosyl pyrophosphate to anthranilate and produces 5-phosphoribosylanthranilate. Phosphoribosylanthranilate isomerase (PAI) catalyzes the irreversible rearrangement of 5-phosphoribosylanthranilate to 1-(o-carboxyphenylamino)-1-deoxyribulose 5-phosphate (CdRP), a reaction that can also occur non-enzymatically. Indole-3-glycerol phosphate synthase (IGPS) catalyzes the irreversible conversion of CdRP to indole-3-glycerol phosphate (Tzin et al., 2010;

Maeda et al., 2012). The final two reactions of the tryptophan pathway are catalyzed by the tryptophan synthase α subunit (TSα) and β subunit (TSβ), respectively (Figure 2) (Maeda et al., 2012). TSα catalyzes the reversible retroaldol cleavage of indole-3-glycerol phosphate to indole and glyceraldehyde 3-phosphate (G3P), and TSβ subsequently condenses indole and serine to produce tryptophan using pyridoxal 5-phosphate (PLP) as a cofactor (Tzin et al., 2010; Maeda et al., 2012).

Antisense suppression of IGPS1 reduces the levels of both tryptophan and auxin, whereas Arabidopsis mutants defective in tryptophan synthase produce less tryptophan but accumulate more auxin, suggesting that indole-3-glycerol phosphate serves as a key branch-point intermediate in tryptophan-independent auxin biosynthesis. The last step in the shikimate pathway involves the conversion of tryptophan to tryptamine catalyzed by tryptophan decarboxylase (TDC). The *Tdc* gene is both inducible and developmentally regulated (Goddijn et al., 1992; Guirimand et al., 2009; Dutta et al., 2007). Furthermore, *Tdc* gene expression is induced by external stress signals, such as fungal elicitor and UV light (Liu et al., 2007).

Table 1. Isolated enzymes and genes involved in the shikimate and tryptophan-biosynthesis pathways in plants (adapted from Maeda et al., 2012).

Enzyme name	**Origin**
Shikimate pathway	
3-Deaoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAHP synthase)	*Solanum tuberosum, Arabidopsis thaliana, Lycopersicon esculentum*
3-Dehydroquinate synthase (DHQS)	*Lycopersicon esculentum*
3-Dehydroquinate synthase dehydratase (DHD)	*Arabidopsis thaliana, Lycopersicon esculentum, Nicotiana tabacum*
Shikimate dehydrogenase (SDH)	
Shikimate kinase (SK)	*Arabidopsis thaliana, Lycopersicon esculentum, Oryza sativa*
5-*Enol*pyruvylshikimate-3-phosphate synthase (EPSP synthase)	*Arabidopsis thaliana, Solamun lycopericum, Petunia hybrida*
Chorismate synthase (CS)	*Corydalis sempervirens, Solamun lycopericum*
Tryptophan (Trp) pathway	
Anthranilate synthase α subunit (ASα)	*Arabidopsis thaliana, Ruta graveolens*
Anthranilate synthase β subunit (ASβ)	*Arabidopsis thaliana*
Phosphoribosylanthranilate transferase (PAT)	*Arabidopsis thaliana*
Phosphoribosylanthranilate isomerase (PAI)	*Arabidopsis thaliana*
Indole-3-glycerol phosphate synthase (IGPS)	*Arabidopsis thaliana*
Tryptophan synthase α subunit (TSα)	*Arabidopsis thaliana, Zea mays*
Tryptophan synthase β subunit (TSβ)	*Arabidopsis thaliana, Zea mays*

5.2.2. The mevalonate (MVA) pathway

The mevalonate pathway starts with the coupling of two molecules of acetyl-CoA to form acetoacetyl-CoA (Figure 3). This step is catalyzed by the enzyme acetoacetyl-CoA acetyltransferase or thiolase (AACT) (El-Sayed et al., 2007), which can be either degradative thiolases or biosynthetic thiolases. The degradative thiolases are involved in the β-oxidation of fatty acids, whereas the biosynthetic thiolases catalyze the condensation of two acetyl-CoA molecules to form acetoacetyl-CoA (Vishwakarma et al., 2012). Thiolase genes have been identified in *Clostridium*, *Rhizobium*, *Dictyostelium*, radish, Arabidopsis, sunflower and the rubber tree *Hevea brasiliensis*. However, little is known about the role of this enzyme in isoprenoid biosynthesis in plants (Vishwakarma et al., 2012).

The subsequent step is the condensation of one acetyl-CoA molecule with acetoacetyl-CoA to form 3-hydroxy-3-methylglutaryl coenzyme A by HMG-CoA synthase (HMGS) (Figure 3) (El-Sayed et al., 2007; Hemmerlin et al., 2003). The corresponding gene has been cloned from several plant species including *Camptotheca acuminata* (Kai et al., 2013), *Salvia miltiorrhiza* (Zhang et al., 2011) and *Taxus media* (Kai et al., 2006).

The key enzyme in the classical mevalonate pathway in plants was found to be 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase (HMGR). This catalyzes the formation of MVA by two successive reductions of HMG-CoA, using two molecules of NADPH as cofactors (Figure 3) (Hemmerlin et al., 2003; El-Sayed et al., 2007). Whereas animals have a single copy of HMGR, plants produce multiple isozymes with developmentally regulated and organ-specific expression profiles. Furthermore, HMGR activity responds to a variety of environmental and physiological signals including light, plant growth regulators, inhibitors, phosphorylation, metabolic feedback, wounding and plant pathogens (Dudey et al., 2003). The HMGR isoforms are thought to be associated with different branches of the cytoplasmic MVA pathway leading to different end products (Hemmerlin et al., 2003). Overexpression of *HMGR* cDNA in tobacco plants increased the apparent enzyme activity and the quantity of sterols and sterol pathway intermediates in the form of fatty acetyl esters but did not affect the accumulation of carotenoids or

chlorophylls. This confirmed the hypothesis that HMGR regulates phytosterol biosynthesis in plants (Hemmerlin et al., 2003).

Finally, mevalonate is phosphorylated to produce 5-diphosphomevalonate in a reaction catalyzed by mevalonate kinase (MVAK) and 5-diphosphomevalonate kinase (MVAPK). The 5-diphosphomevalonate is decarboxylated by 5-diphosphomevalonate decarboxylase to produce isopentenyl diphosphate (IPP) (Figure 3) (El-Sayed et al., 2007; Wang et al. 2012). In higher plants, the mevalonate pathway is found in the cytoplasm and mitochondria and is used to produce sterols, sesquiterpenes and ubiquinones (Eisenreich et al., 2001; Adam et al., 1999).

5.2.3. The 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway

The mevalonate pathway was initially considered to be the only source of DMAPP and IPP, but the alternative 2C-methyl-D-erythritol 4-phosphate (MEP) pathway was discovered more recently (Rohdich et al., 2001). The non-mevalonate or MEP pathway is involved in the formation of plastid-derived terpenoids such as isoprene, monoterpenes, diterpenes and tetraterpenes (Adam et al., 1999).

The first step in the MEP pathway is the condensation of glyceraldehyde 3-phosphate (GA-3P) and pyruvate into 1-deoxy-D-xylulose 5-phosphate (DXP) by DXP synthase (Figure 3). The 1-deoxy-D-xylulose is a common precursor of thiamine and pyridoxal, but can be diverted efficiently to produce terpenoids (Rohdich et al., 2001; Hemmerlin et al., 2012). Detailed studies have showed that DXS requires TPP and divalent cations, such as Mg^{2+} or Mn^{2+} for activity (Dudey et al., 2003). A *dxs* gene encoding DXP synthase (DXPS) was initially isolated from *Escherichia coli.* Homologous genes were subsequently identified in other bacteria as well as in higher plants such as *Mentha piperita, Capsicum annuum* and *Catharanthus roseus*, where they are found in the nuclear genome and are preceded by putative N-terminal plastid targeting sequences (Chahed et al., 2000; Veau et al., 2000; Liu et al., 2007; Yazaki et al., 2004; Guirimand et al., 2009; Lange et al., 1998; Bouvier et al., 1998; Dudey et al., 2003).

In the second step, DXP is converted into 2C-methyl-D-erythritol 4-phosphate (MEP) (Figure 3). The reaction is catalyzed by DXP reducto-isomerase (DXR), encoded by

the *dxr* gene, which has been characterized in species such as, bacteria, Arabidopsis and *Mentha x piperita* (see Table 1 in Chapter 2, section 2.2) (Kuzuyama et al., 2000; Schwender et al., 1999; Lange et al., 1999). The precise reaction mechanism of this enzyme is still unclear. It catalyzes the transformation of DXP into MEP in two steps. An intramolecular rearrangement leads thereby to the formation of the enzyme-bond intermediate MEP, which is concomitantly reduced in a NADPH-dependent manner (Hemmerlin et al. 2012; Dudey et al., 2003). The subsequent steps in the pathway have been elucidated in *E. coli*, Arabidopsis and other higher plants: 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (CMS) and 4-diphosphocytidyl-2C-methyl-D-erythritol kinase (CMK) convert MEP into 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP) by phosphorylating the hydroxyl group at the 2-position (Figure 3). The latter compound is then transformed into 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-cPP) by ME-cPP synthase (MECS). ME-cPP is reduced to 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate (HMBPP) by 2-methyl-2-(*E*)-butenyl 4-diphosphate synthase (HDS). In the final step, this compound is converted by a single enzyme, HMBPP reductase, into a 5–6:1 ratio of IPP and DMAPP. This ratio is subsequently adjusted to 7:3 by IPP/DMAPP synthase (IDS) (Liu et al., 2007; Yazaki et al., 2004; Guirimand et al., 2009; Chahed et al., 2000;Veau et al., 2000; Hemmerlin et al., 2012; Dudey et al., 2003).

It appears that the MEP pathway is duplicated in many plant species, with genes such as *dxs* organized in differentially expressed multigene families (Hemmerlin et al., 2012). Indeed, at least two genes were found for CMK and HDR in *Ginkgo biloba*, for MCT and MDS in *Hevea brasiliensis*, for HDR in *Pinus* species, for MDS and HDS in *Populus trichocarpa* and for HDR in *Oryza sativa* (Hemmerlin et al., 2012). Based on the analysis of sub-plastidial proteomes, this alternative MEP pathway thought to reside in the plastidial stroma (Hemmerlin et al., 2012).

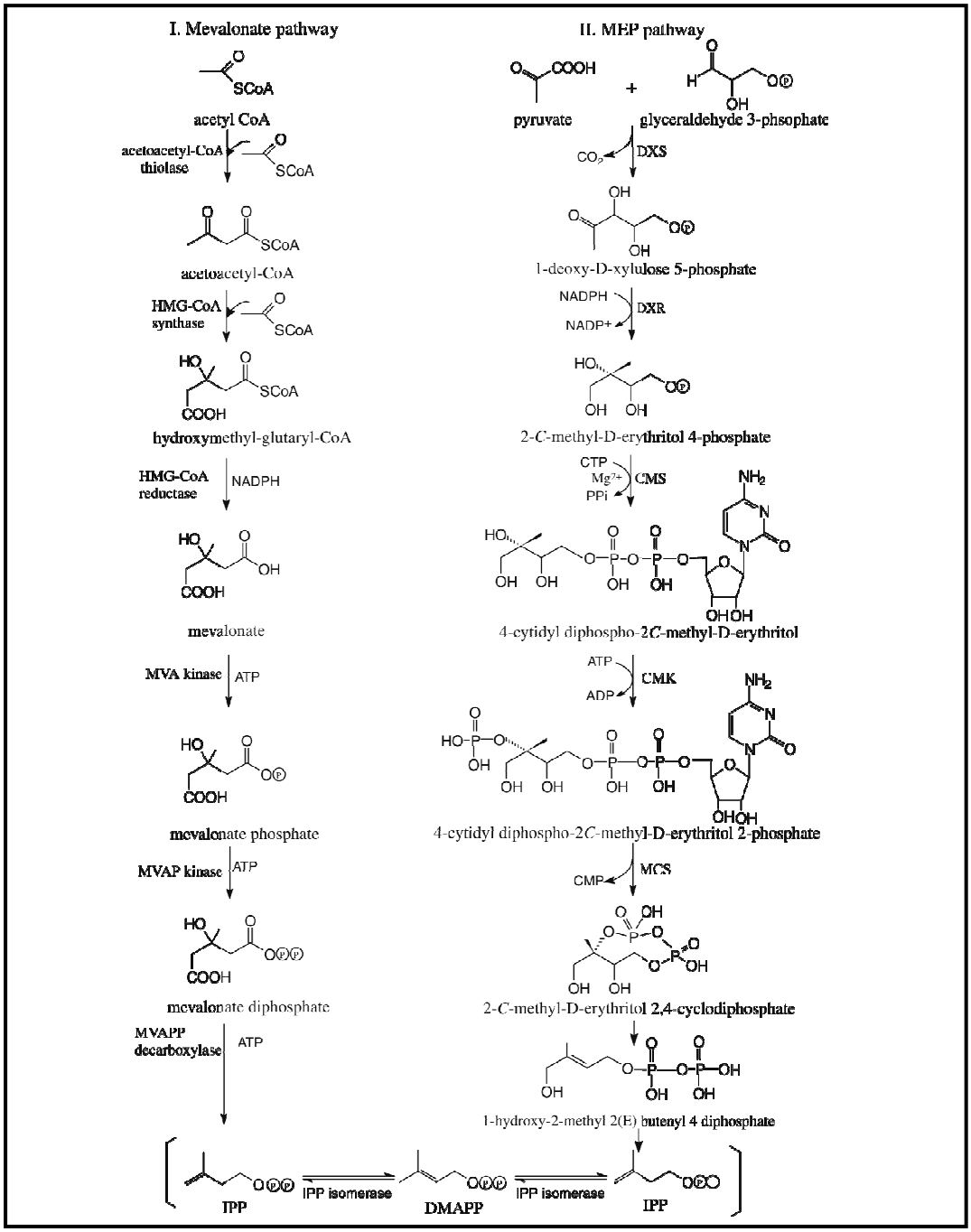

Figure 3. The biosynthesis of IPP via the mevalonate and MEP pathways, showing the corresponding enzymes (El-Sayed et al., 2007).

5.2.4. Cross-talk between the MVA and MEP pathways

The MVA and MEP pathways are localized in the cytosolic and plastid compartments, respectively, which allows them to operate independently. But there is evidence in plants that they cooperate in the biosynthesis of certain metabolites (Laule et al., 2003; Chahed et al., 2000; Oudin et al., 2007). Adam et al. (1999) fed chamomile plants with 1-[1-^{13}C] deoxy-D-xylulose to study the origin of the IPP incorporated into sesquiterpenes. Labeling the chamomile flowers showed that the three IPP units were of mixed origin, with site-specific overlays of both pathways. The biogenetic first two units are formed predominantly via the MEP pathway (~10% MVA pathway) whereas the terminal unit can be derived from either pathway with an equal probability (Adam et al., 1999).

The plastid C_{10} building block geranyl diphosphate (GPP) is formed from two IPP units derived from the MEP pathway, condensed by the prenyltransferase GPP synthase. Then, plastidial GPP and IPP are transferred to the cytosol or another cellular compartment that has access to MVA-derived cytoplasmic IPP. In this compartment, plastidial GPP and IPP and cytoplasmic IPP are condensed by another prenyltransferase (farnesyl diphosphate (FPP) synthase) to form FPP. Because FPPS can assemble FPP from DMAPP and IPP as well as GPP and IPP, chimeric FPP molecules are formed that include isoprene units from both the MEP and MVA pathways (Adam et al., 1999).

The possibility of cross talk between the MEP and MVA pathways cannot be completely ruled out during IPP synthesis, and may also occur during the synthesis of β-carotene, phytol and lutein in *C. roseus*, and monoterpene and sesquiterpene volatiles in lima beans (Laule et al., 2003; Verma et al., 2012; Oudin et al., 2007). Metabolic cross-talk between the MVA and MEP pathways may occur during the biosynthesis of several terpenoid compounds in plants (Eisenreich et al., 2001). The extent of this crosstalk depends on the species and the presence and concentration of exogenous precursors. Generally, crosstalk is minimal (<1%) in intact plants under physiological conditions (Eisenreich et al., 2001; Hemmerlin et al. 2003).

The existence of two pathways for IPP raises fundamental questions about the origin of monoterpenes in *C. roseus*. Veau et al. (2000) could not find a correlation between

hmgr gene expression and the production of monoterpenes in *C. roseus*, although they could not exclude a minor contribution, but they clearly demonstrated a correlation between MEP pathway gene expression and monoterpene accumulation, supporting other studies showing that the monoterpenes are formed predominantly via the MEP pathway (Veau et al., 2000; Hedhili et al., 2007). Hedhili et al., 2007 suggested that derivatives from the MVA pathway could be involved in the regulation of the MEP pathway, after suppressing monoterpene synthesis by inhibiting *hmgr* in *C. roseus* cell suspension cultures (Hedhili et al., 2007). The inhibition could be overcome by supplying the cells with exogenous mevalonic acid, and if ^{14}C-radiolabelled mevalonic acid was used then none of the labeled carbon was incorporated into the newly synthesized alkaloids but it was incorporated into sterols and prenylated proteins. These data suggested that the IPP produced via the MVA pathway is not incorporated into secologanin and that compounds originating from the MVA pathway may be involved in the regulation of secologanin biosynthesis (Hedhili et al., 2007). Oudin et al., 2007 reported that CaaX-prenyltransferases catalyzing the prenylation of proteins are essential for the expression of MEP genes including *dxs, dxr* and also the secoiridoid monotperne gene *g10h*. The nature of the metabolites exchanged between the compartments and how this process is regulated remain to be established (Eisenreich et al., 2001).

5.2.5. The monoterpene secoiridoid pathway

Secologanin is synthesized via the methylerythritol phosphate (MEP) pathway, and the resulting IPP and DMAPP are condensed head-to-tail in the presence of GPPS synthase to yield geranyl diphosphate (GPP), the universal precursor for all monoterpenes (Verma et al., 2012; Ginis et al., 2011). GPPS has been characterized in plants such as *Mentha piperita* (pepermint) (Turner et al. 2004; Burke et al. 1999) and Arabidopsis (Schmidt et al. 2008), but not *C. roseus* (Liu et al., 2007; Guirimand et al., 2009) (Figure 2).

The hydroxylation of geraniol at the C-10 position to form 10-hydroxygeraniol is catalyzed by geraniol 10-hydroxylase (G10H) and constitutes the first committed step in secologanin biosynthesis (Ginis et al., 2011; Wang et al. 2010). This membrane-associated monooxygenase from the CYP76B subfamily is dependent on NADPH and

oxygen, and is reversibly inhibited by light. G10H also requires a cytochrome P450 reductase (CPR) for activity because it transfers electrons to cytochrome P450 (Liu et al., 2007; O'Connor et al., 2005; Ginis et al., 2011). G10H expression is strongly induced along with other early monoterpenoid pathway genes in *C. roseus* by cytokines, ethylene and methyl jasmonate (Ginis et al., 2011) but it is feedback inhibited by terpenoid indole alkaloids such as catharanthine, vinblastine and vindoline. 10-Hydroxygeraniol is oxidized further into 10-oxogeraniol in the presence of NADP-oxidoreductase, and the resulting 10-oxogeranial is converted into iridodial by a cyclization reaction catalyzed by 10-oxogeraniol/ iridodial cyclase (Ginis et al., 2011; Verma et al., 2012).

At the end of the pathway, 7-deoxyloganin is converted into secologanin by sequential activity of deoxyloganin 7-hydroxylase (DL7H), which converts 7-deoxyloganin into loganin, then secologanin synthase (SLS) converts loganin into secologanin by the oxidative rupture of the methylcyclopentane ring (Verma et al., 2012; Liu et al., 2007; O'Connor et al., 2005). It has been shown that the synthesis of secologanin requires NADPH and molecular oxygen, and is inhibited by carbon monoxide and several other cytochrome P450 inhibitors (Liu et al., 2007; O'Connor et al., 2005). The multi-step conversion of GPP to secologanin via geraniol is still not completely understood. Only four cDNAs encoding P450-dependent geraniol 10-hydroxylase (G10H), acyclic monoterpene primary alcohol dehydroxygenase, loganinc acid methyltransferase (LAMT) and P450-dependent secologanin synthase (SLS) have been identified. At least 11 enzymatic steps may be required to convert geraniol into secologanin via loganic acid and loganin (Verma et al., 2012) (Figure 4).

It is now thought that regulatory mechanisms controlling the production of terpenoid indole alkaloids in *C. roseus* target the terpenoid branch, and that the accumulation of secologanin is a major bottleneck in TIA biosynthesis (Chahed et al., 2000).

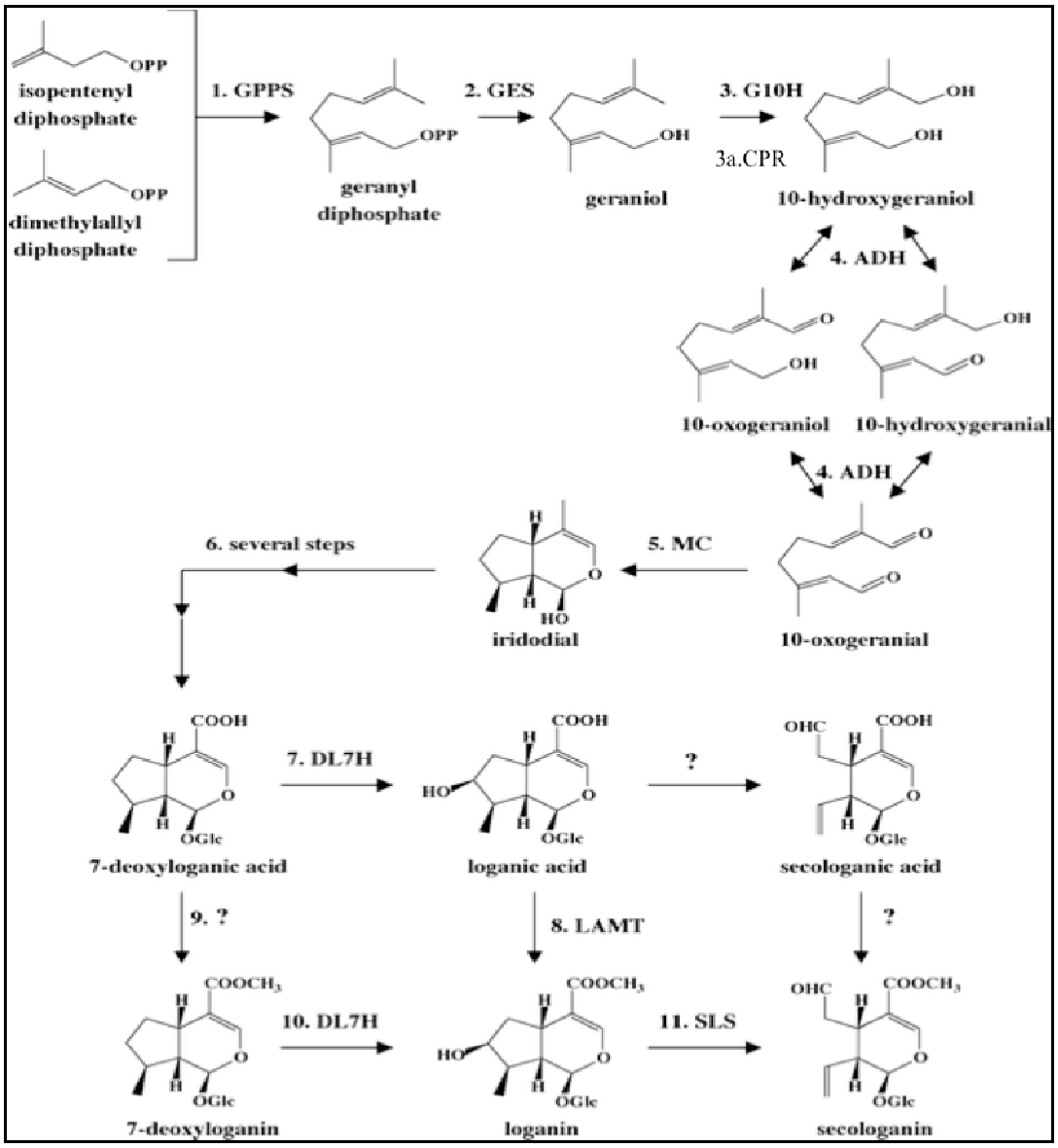

Figure 4. Monoterpene secoiridoid pathway for the production of secologanin. GPPS: Geraniol diphosphate synthase, GES: Geraniol synthase: G10H: Geraniol 10-hydroxylase; CPR: Cutocropme P450 reductase; ADH: Alcohol dehydrogenase; MC: Monoterpene cyclase; DL7H: Deoxyloganin 7-hydroxylase; LAMT: loganic acid methyltransferease; SLS: Secologanin synthase. (SmartCell, annual report 2010).

5.2.6. Multicellular organization of the monoterpenoid pathway

The enzymes in plant secondary metabolic pathways are often differentially expressed and localized in multiple subcellular compartments. The monoterpene secoiridoid biosynthetic pathway is no exception and the complex multicellular organization has been extensively characterized in *C. roseus*, which is now cited as an example of the extraordinary spatial complexity of secondary metabolism in plants (O'Connor et al., 2005; Guirimand et al., 2009).

A different model for the organization of the terpenoid biosynthesis pathway has emerged in the aerial organs of *C. roseus*. Sequentially, the MEP pathway appears to be localized within the internal phloem-associated parenchyma (IPAP) cells at the periphery of the stem and on the upper part of the vascular bundles (DXS, DXR, MECS and HDS) along the early monoterpene secoiridoid biosynthesis pathway (G10H), whereas SLS appears to accumulate in the epidermis (Guirimand et al., 2009; O'Connor et al., 2005; Verma et al., 2012; Oudin et al., 2007). The existence of this multicellular model indicates that unknown intermediates must be transferred from IPAP cells to the epidermis and others must be translocated from the epidermis to the idioblast and laticifer cells of the leaf (Verma et al., 2012).

Monoterpenoid enzymes and intermediates are also distributed among various subcellular compartments in *C. roseus*, requiring a mechanism for inter-organelle exchange (Guirimand et al., 2009; Verma et al., 2012). For secologanin biosynthesis, the geraniol derived from the plastidial MEP pathway is guided towards the vacuole following hydroxylation by G10H, which is localized in the endoplasmic reticulum (ER). This is also consistent with the subcellular localization of cytochorme P450 reductase in the ER for electron transfer to cytochrome P450 enzymes (Verma et al., 2012; O'Connor et al., 2005; Yazaki et al., 2004). Secologanin synthase (SLS) is also anchored to ER via an N-terminal helix and is thought to be localized in the tonoplast based on the presence of its substrate, loganin, and product, secologanin, in the vacuole (Verma et al., 2012).

5.3. Aims and objectives

The aim of the experiments described this chapter was evaluate the function of transgenes of the early monoterpene secoiridoid enzymes, geraniol diphosphate synthase (*AtGpps*) and geranyl synthase (*VoGes*), after boosting the pathway by transient expression of downstream genes. We also explore the hypothesis that the early enzymes of the monoterpene secoiridoid pathway might be activated by the presence of downstream enzymes of the pathway.

The objectives were:

- Transient expression of the upstream gene geraniol diphosphate synthase (*pineGpps*) and the downstream genes monoterpenoid cyclase (*Cr96*) and 10-hydroxygeraniol oxidoreductase (*Cr68*) in transgenic tobacco plants expressing geraniol synthase (*VoGes*) and geraniol 10-hydroxylase (*Cyp76b6*).
- Transient expression of upstream genes geraniol diphosphate synthase (*pineGpps*) and geraniol synthase (*VoGes*), and downstream genes monoterpenoid cyclase (*Cr96*) and 10-hydroxygeraniol oxidoreductase (*Cr68*) in transgenic tobacco plants expressing geraniol 10-hydroxylase (*Cyp76b6*) and cytocrome P450 reductase (*Cpr*).
- Preliminary characterization of the metabolites found in the agroinfiltrated leaf tissues.

5.4. Materials and methods

5.4.1. Transformation constructs

Two groups of transformation vectors were used: one for stable constitutive expression in tobacco plants and one for transient expression by agroinfiltration. The stable transformation construct for *Ges* was provided by collaborators from Wageningen University (The Netherlands). The *Ges* gene was cloned from *Valeriana officialis* and transferred to vector pIV2B2.4, which includes a plastid targeting signal sequence (Figure 3E in Chapter 2, Section 2.4.2). The stable transformation constructs for *Cyp76b6* and *Cpr* were provided by collaborators from CNRS (Strasbourg, France). Both genes were previously cloned from *Catharanthus roseus* and transferred to vector pMOG463. All genes were controlled by the constitutive double CaMV 35S promoter. The selectable marker *nptII* was used in the stable transformation experiments as described in detail in Figure 2, Chapter 2, Section 2.4.2. The transient expression constructs for *pineGPPS*, 10-*hgo* and *Mtc* were based on vector pCambia1300, whereas the *VoGes* transient expression construct was based on vector pBIN (Figure 5).

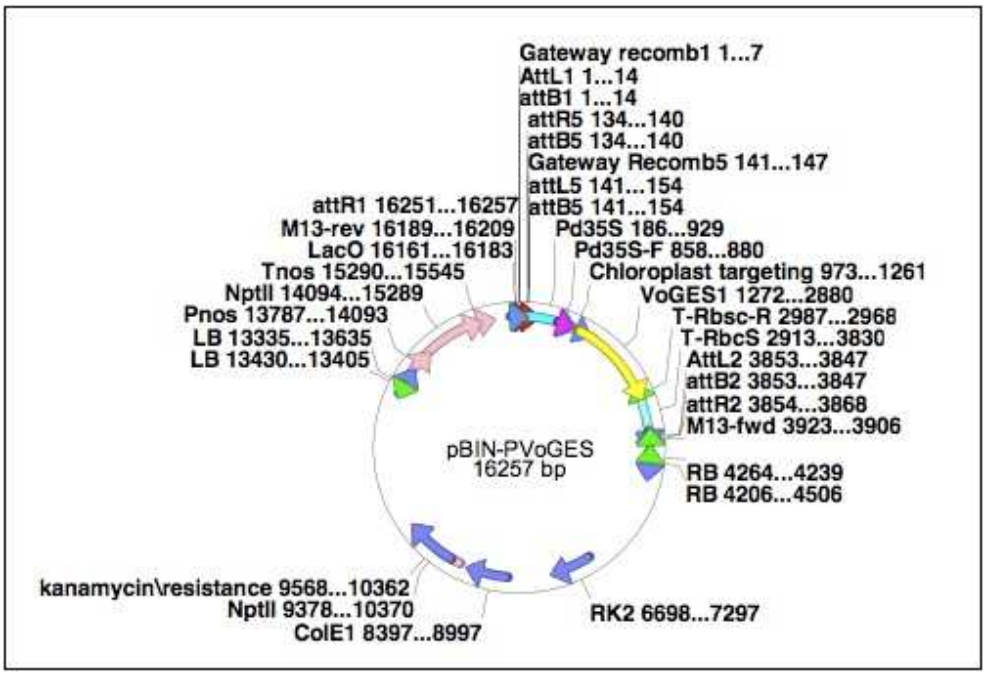

Figure 5. *VoGes* vector for transient expression in tobacco.

5.4.2. Transformation, selection and regeneration of transgenic plants

Leaf pieces were transformed by direct DNA transfer using particle bombardment as described in Chapter 2, section 2.4.2, using a 3:1 molecular weight ratio of the vector containing the gene of interest against the *npt*II selectable marker. Two combinations of genes were used, as listed in Table 2. Bombarded leaf pieces were incubated in the dark and transferred to fresh media every 2 weeks as described in Chapter 2, section 2.4.3.

Table 2. Combinations of constructs used for tobacco transformation experiments

Combination	*ges*	*g10h*	*cpr*
1	pIV2B2.4 *VoGes* (chloroplast)	pMOG463-Cr*Cyp76b6*	
2		pMOG463-Cr*Cyp76b6*	pMOG463-Cr*Cpr*

We recovered 27 transgenic tobacco plants from six independent lines transformed with combination 1 (Table2) as described in Chapter 2, Sections 2.5.3 and 2.4.5 and eight transgenic plants transformed with combination 2. Transgenic lines H and C were identified by mRNA blot as the lead lines for combinations 1 and 2, respectively. Both lines were used for the transient expression experiments described below.

5.4.3. Transient expression of monoterpene secoiridoid genes in transgenic tobacco leaves

Transient expression experiments were carried out at the Fraunhofer IME, Aachen, Germany. *Agrobacterium tumefaciens* cultures (strain LBA4404) individually expressing *VoGes, pineGpps, 10-hgo* (*Cr68*) and *Mtc* (*Cr96*) were grown overnight at 25–28ºC in 20 mL YEB medium supplemented with rifampicin (25 mg/ml), streptomycin (30 mg/ml) and kanamycin (25 mg/ml). The main cultures were inoculated 1:20 in YEB-medium containing streptomycin (30 mg/ml) and kanamycin (25 mg/ml) and were grown at 25–28°C. After 2 days, acetosyringone (20 μM), glucose (10 mM) and MES (10 mM) were added to the media and the pH was adjusted to 6.0 with NaOH. Bacterial clones were maintained in liquid culture media until the OD_{600} reached ~ 5.0. The OD_{600} was then adjusted to ~1.0 with *2x* infiltration medium (sucrose, 100 g/L; glucose 3.98 g/L; MS salts 8.6 g/L) and water. The agroinfiltration mixture was prepared with 50 ml of each culture following the combinations listed in Table 3 and it was diluted four times with 2x agroinfiltration medium and water. The suspensions were injected into the leaves of two different transgenic tobacco plants generated as described in chapter 2, using a 1-mL syringe. Three plants each from transgenic line H, expressing plastidial *VoGes* and *Cyp76b6*, and transgenic line C, expressing *Cyp76b6* and *Cpr* as described in Table 3, were agroinfiltrated and grown at 23°C. Three leaves were harvested per agroinfiltrated plant at 5 days and 7 days after agroinfiltration. The collected samples were ground in liquid nitrogen and stored at –80 ºC. Non-agroinfiltrated leaves from the same plants were collected as negative controls.

Table 3. Combinations of genes used in the agroinfiltration experiments

Transgenic plant line	**Integrated transgenes**	**Transient expression constructs**
H (H2, H5, H13)	*VoGes* (plastid) and *Cyp76b6*	*pineGpps*, *10-hgo* (*Cr68*) and *Mtc* (*Cr96*)
C (C5, C6, C7)	*Cyp76b6* and *Cpr*	*pineGpps*, *VoGes*, *10-hgo* (*Cr68*) and *Mtc* (*Cr96*)

5.4.4. Gas chromatography–mass spectrometry (GC-MS)

The agroinfiltrated samples were deglycosidated and analyzed by GC-MS at the Fraunhofer IME, Aachen, Germany. We dissolved 500 mg of ground sample in 1 ml of citrate-phosphate buffer (pH 5.4), sonicated for 15 min and added 200 µl of Viscozyme L (Sigma, Steinheim, Germany) followed by a 500 µl layer of heptane with 10 µg/ml *Z*-nerolidol as an internal standard. After incubation at 37ºC overnight, we added 1 ml of heptane, centrifuged for 10 and filtered the supernatant through anhydrous sodium sulfate in a glass Pasteur pipette to remove water. Another 0.5 ml of heptane was added and removed after centrifugation as above.

The different heptane fractions were combined before GC-MS analysis using a Shimadzu QP2010SE quadrupole mass spectrometer after separation with a Phenomenex Zebron ZB-5 ms column 30 m in length x 0.25 mm internal diameter containing 0.25 µm stationary phase, with a Guardian precolumn (5 m). One µl of sample was injected to the injection chamber at 250°C. The injection was split (1:10) and the ZB5 column was maintained at 45°C for 1 min before beginning a gradient of 10°C per min up to 300°C and maintaining 300°C for 7 min. The helium inlet pressure was checked by electronic pressure control to achieve a constant column flow of 1.0 ml/min. Mass detection, with electric ionization at 1 keV using the SIM/scan mode with diagnostic ion monitoring, was carried out for characteristic mass fragments of geraniol: 69, 84, 93, 111, 123 and 154 m/z. Products were identified by comparing the mass spectra to the compounds in spectral libraries such as NIST (NIST 11 Software, National Institute of Standards and Technology, Gaithersburg, Montgomery, Maryland). In addition, the Kovats Index, a concept to convert retention times into system-independent constants, were deduced

by comparison with the retention times of alkane standards and compared with available data to facilitate product identification.

5.4.5. Liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC-QTOF-MS)

Non-volatile compounds in tobacco leaves (Section 5.4.1) were analyzed at the Laboratory of Plant Physiology, Wageningen University, The Netherlands, by liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC-QTOF-MS), as described by De Vos et al. (2007).

Aliquots of 200 mg frozen, powdered material were extracted with 0.6 ml 99.9% methanol supplemented with 0.133% formic acid in 1.5 ml Eppendorf tubes. After vortexing briefly and sonicating for 15 min, the extracts were centrifuged and passed through 0.45 µm filters (SRP4, Sartorius, Germany), and 5 µl of the filtered 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 and leucine encephalin ([M – H]– = 554.2620) was used as a lock mass for online mass calibration.

LC-MS data were acquired, processed and visualized using MassLynx 4.0 (Waters). Mass data were processed using metAlign version 1.0 (www.metalign.nl). Baseline and noise calculations were carried out 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 ($[M - H]^-$ = 554.2620), was divided by the intensity range observed for this row throughout all samples. Each row was then mean centered. Finally, the normalized and log-transformed data matrix was used for principal component analysis implemented in GeneMath XT version 2.1.

Significant differences in the intensity of each aligned mass signal among samples were determined using Student's t-test (level of significance set at 0.05). Masses showing significant differences were manually checked in MassLynx. Metabolites were putatively identified 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. Geraniol glycosides were assigned names as previously described (Dong et al., 2013).

5.5. Results

5.5.1. Transgenic plants expressing early genes in the monoterpene secoiridoid biosynthesis pathway do not accumulate geraniol following agroinfiltration with downstream pathway genes

We analyzed samples of agroinfiltrated and non-agroinfiltrated leaves from the transgenic lines by GC-MS to check for the presence of anticipated metabolites. Geraniol was not detected in agroinfiltrated and non-agroinfiltrated leaves of transgenic line H, stably expressing *VoGes* and *Cyp76b6* and transiently expressing *pineGpps*, *10-hgo* (*Cr68*) and *Mtc* (*Cr96*) (Figure 6A,B). Trace amounts of geraniol (retention time 10.720 ÷ 10.833 min) were detected in agroinfiltrated leaf samples of line C, stably expressing *Cyp76b6* and *Cpr* and transiently expressing *pineGpps*, *VoGes*, *10-hgo* (*Cr68*) and *Mtc* (*Cr96*) (Figure 7A) but none was detected in non-agroinfiltrated leaves (Figure 7B).

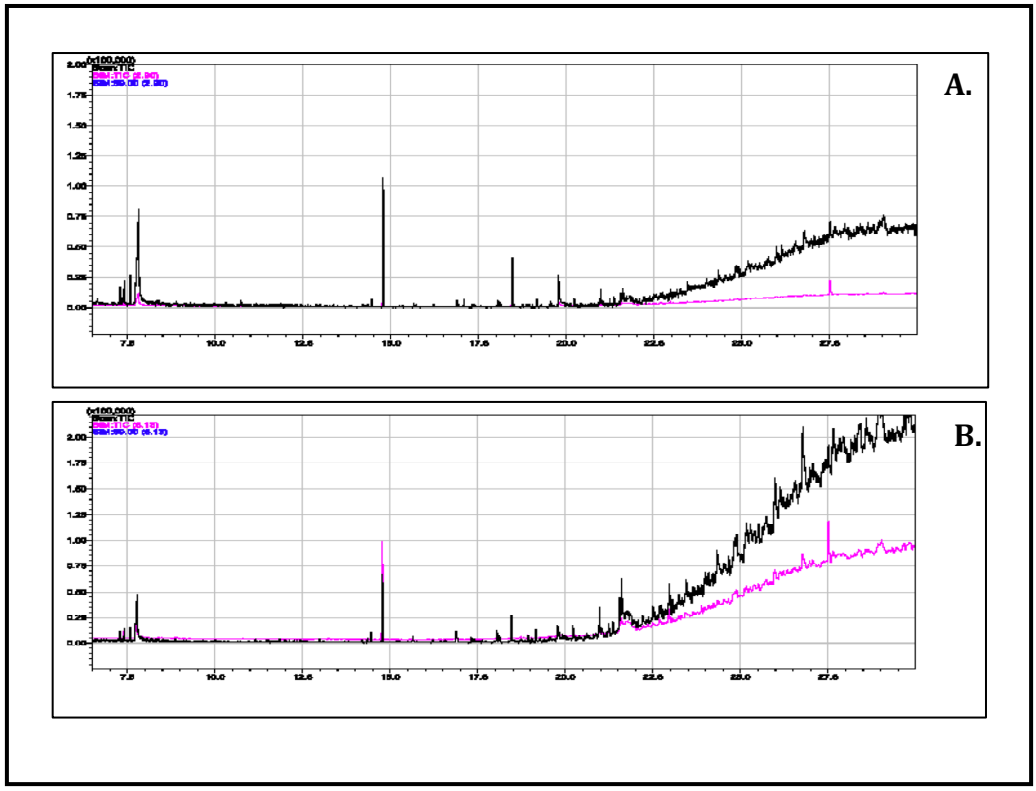

Figure 6. GC-MS spectra of leaf samples from transgenic line H stably expressing *VoGes*(chloro) and *Cyp76b6*). **A.** Leaves agroinfiltrated with *pineGpps*, *10-hgo* (*Cr68*) and *Mtc* (*Cr96*). **B.** Non-agroinfiltrated leaves as a negative control.

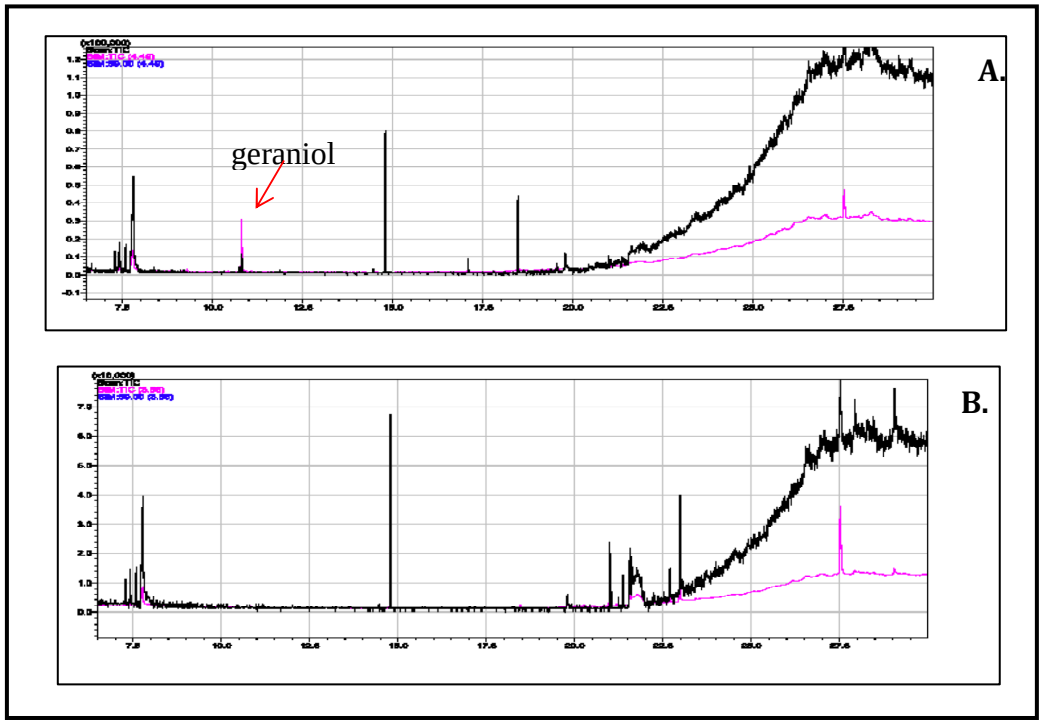

Figure 7. GC-MS spectra from transgenic line C, stably expressing *Cyp76b6* and *Cpr*. **A.** Leaves agroinfiltrated with *pineGpps*, *10-hgo* (*Cr68*) and *Mtc* (*Cr96*). **B.** Non-agroinfiltrated leaves as a negative control.

5.5.2. Transgenic plants expressing early genes in the monoterpene secoiridoid biosynthesis pathway do not accumulate downstream metabolites following agroinfiltration with downstream pathway genes

Having failed to detect significant amounts of geraniol in the agroinfiltrated plants, we next sought some of the downstream compounds in the pathway, i.e. 10-hydroxygeraniol, 10-hydroxygeranial, 10-oxogeranial and iridodial. None of these compounds could be detected by LC-MS/TOF in any of the agroinfiltrated samples. However, the agroinfiltrated samples of both transgenic lines revealed a multitude of new peaks with *m/z* values representing unknown compounds, which were not present in the negative controls. These included 60 non-related monoterpene secoiridoid compounds that could not be identified based on their *m/z* values (Table 3 in the annex 1). These differences, which are indicated in Figures 8 and 9, were confirmed by principal component analysis (Figure 10). PC1 and PC2 describe 35.9% and 21.2% of the total metabolic variation between non-infiltrated and agro-infiltrated leaves, respectively. The heat map representing all masses in Table 3 of the annex 1 (Figure 11) indicates that the major difference between lines H and C, which appears after agroinfiltration, is the presence of geraniol glycosides in line C. The other differences revealed by the heat map could not be identified due to the lack of controls and the anonymous nature of the peaks. The specific geraniol glycosides found in agroinfiltrated line C included acetyl-geraniol dihexose, dipentosyl-hexosyl-geraniol, malonyl-hexosyl geraniol (dimer), dihexosyl geraniol and some other unknown derivatives (Figure 12).

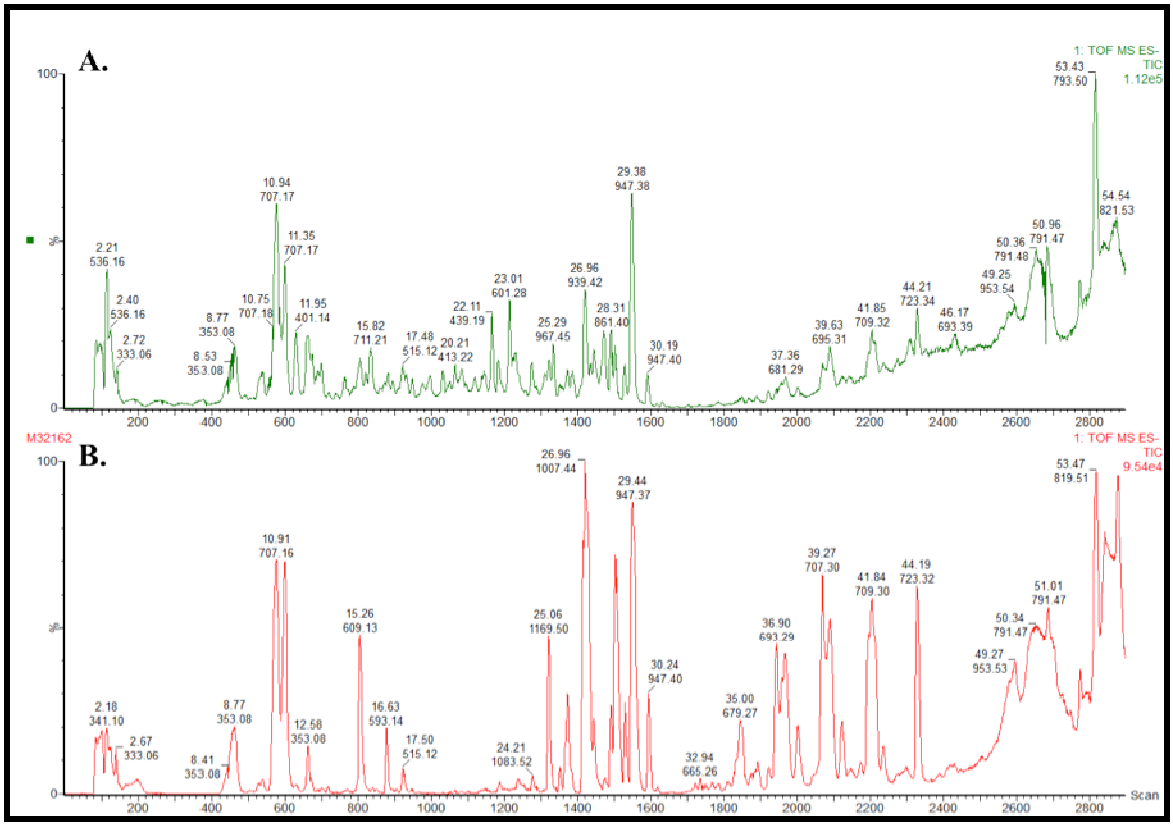

Figure 8. LC-MS/TOF spectra. The numbers above the peaks are *m/z* values. **A.** Mass spectrum representing line H (stably expressing *VoGes* and *Cyp76b6*) agroinfiltrated with *pineGpps*, *Cr68* and *Cr94*. **B.** Mass spectrum representing a non-agroinfiltrated sample of line H.

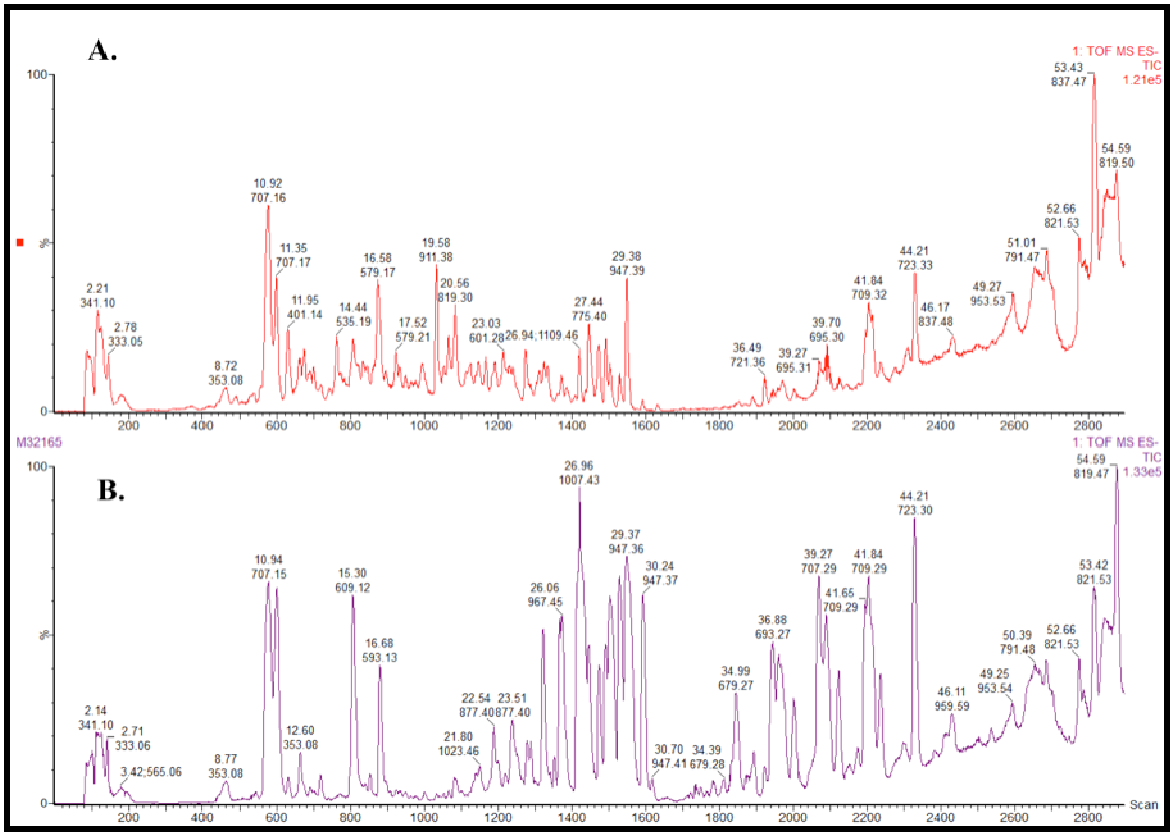

Figure 9. LC-MS/TOF spectra. The numbers above the peaks are *m/z* values. **A.** Mass spectrum representing line C (stably expressing *Cyp76b6* and *Cpr*) agroinfiltrated with *pineGpps*, *VoGes*, *Cr68* and *Cr94*. **B.** Mass spectrum representing a non-agroinfiltrated sample of line C.

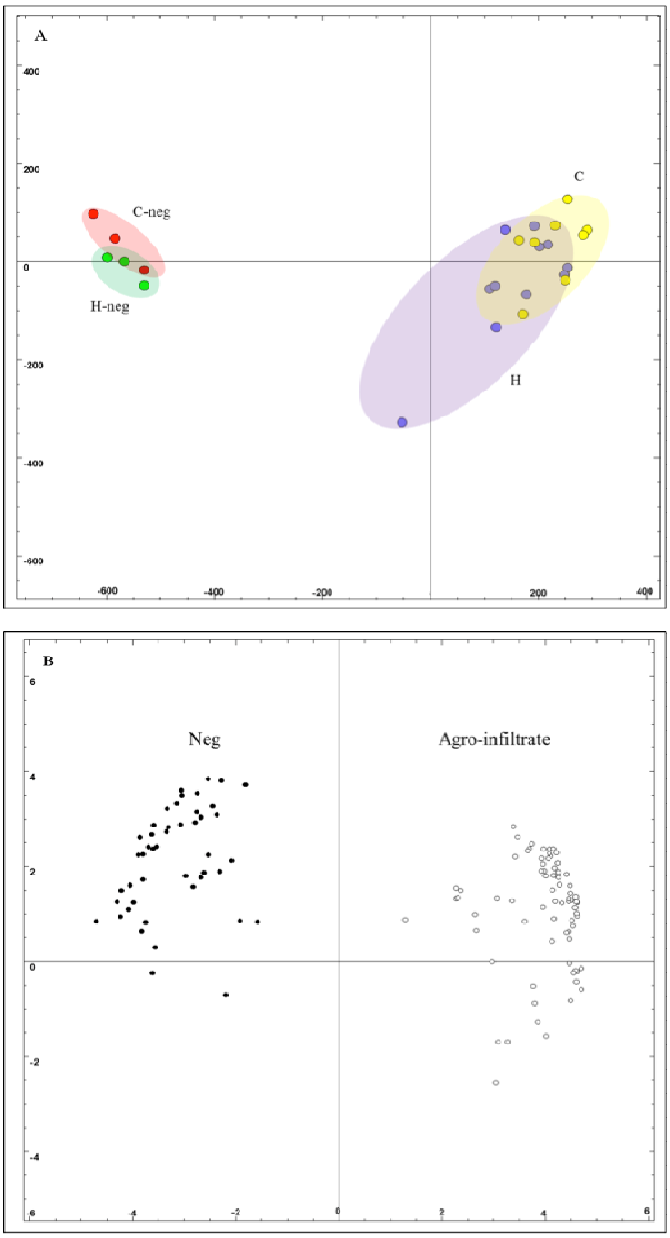

Figure 10. Principal component analysis of the LC-MS data. **A**. Clustering of four groups of samples along the first two principal components (PCs). H_neg: non-agroinfiltrated samples from three different plants of line H, which expressed *VoGes* and *Cyp76b6*. Negative control. C_neg: non-agroinfiltrated samples from three different plants of line C, which expressed *Cyp76b6* and *Cpr*. Negative control. H: three different agroinfiltrated samples with *VoGes*, *Cr68* and *Cr94* from the different transgenic plants of line H, which expressed *VoGes* and *Cyp76b6*. C: three different agroinfiltrated samples with *pineGpps*, *VoGes*, *Cr68* and *Cr94* from three different transgenic plants of line C, which expressed *Cyp76b6* and *Cpr*. **B.** Clustering of non-volatile compounds determining the non-infiltrated and agro-infiltrated leaves.

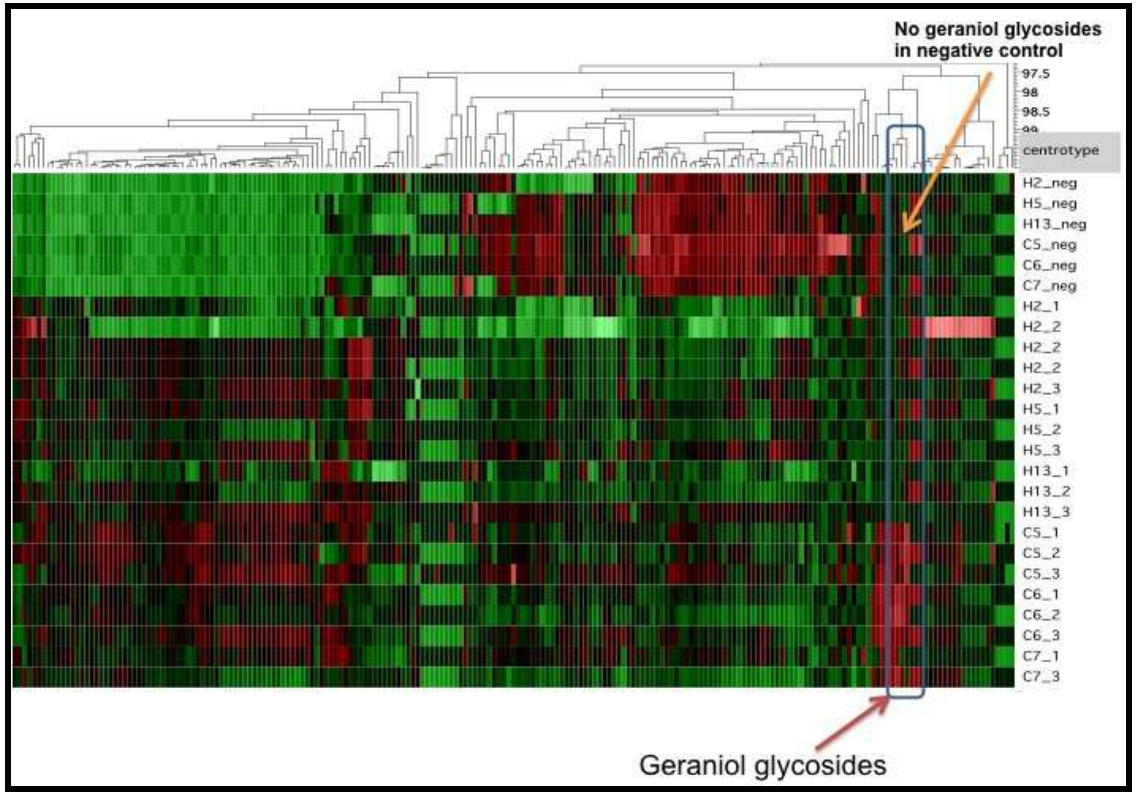

Figure 11. Heat map representing the LC-MS/TOF analysis. H2_neg, H5_neg and H13_neg: non-agroinfiltrated samples from three different plants of line H, which expressed *VoGes* and *Cyp76b6*. Negative controls. H2_1, 2, 3; H5_1, 2, 3 and H13_1, 2, 3: three different agroinfiltrated samples with *VoGes*, *Cr68* and *Cr94* from three different transgenic plants of line H, which expressed *VoGes* and *Cyp76b6*. C5_1, 2, 3; C6_1, 2, 3 and C7_1, 2, 3: three different agroinfiltrated samples with *pineGpps*, *VoGes*, *Cr68* and *Cr94* from three different transgenic plants of line C, which expressed *Cyp76b6* and *Cpr*.

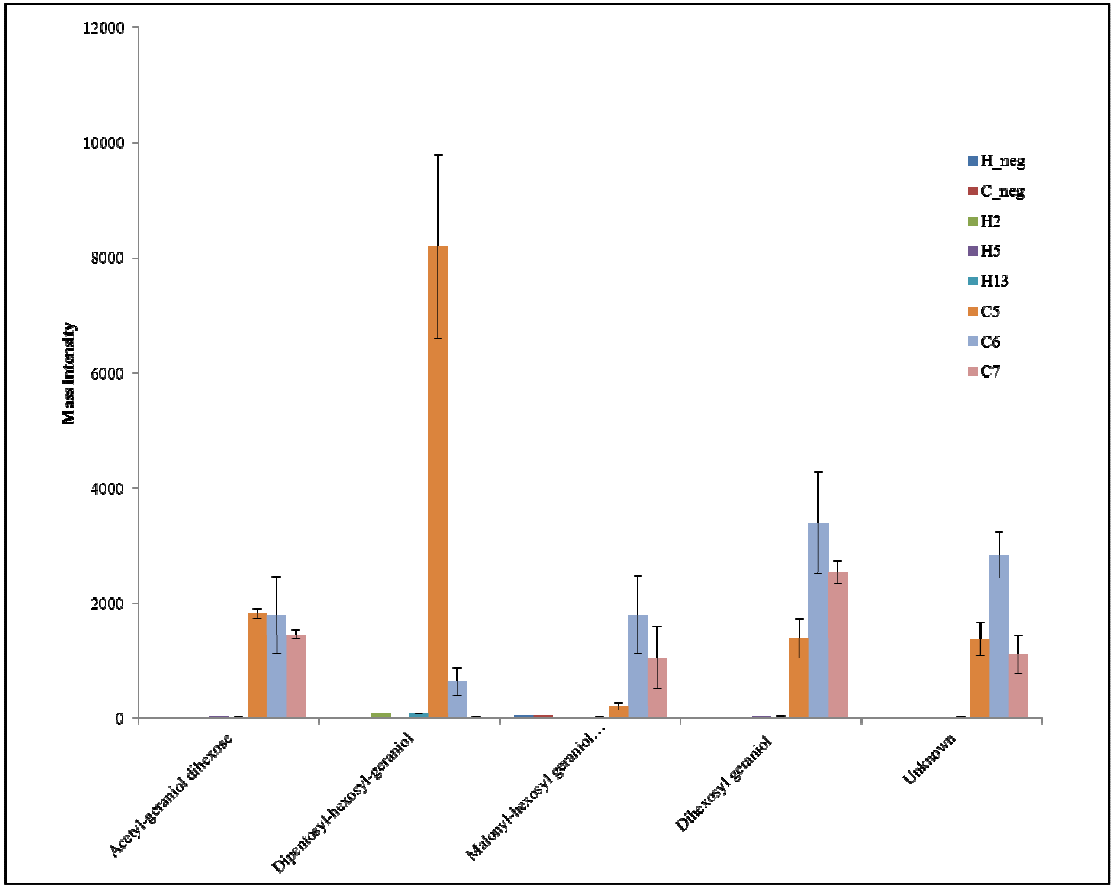

Figure 12. Geraniol derivatives detected by LC-MS/TOF. H_neg: Negative control. Non-agroinfiltrated sample from line H (expressing *VoGes* and *Cyp76b6*). C_neg: Negative control. Non-agroinfiltrated sample from line C (expressing *Cyp76b6* and *Cpr*). H2, H5 and H13: Samples of line H agroinfiltrated with *pineGpps*, *Cr68* and *Cr94*. C5, C6 and C7: Samples of line C agroinfiltrated with *pineGpps*, *VoGes*, *Cr68* and *Cr94*.

5.6. Discussion

Catharanthus roseus (Madagascar periwinkle) is an important natural source of monoterpenoid indole alkaloids, a diverse group of pharmaceutically-important molecules. However, these metabolites accumulate at extremely low levels within the plant, and chemical synthesis is not feasible because of their complex structures. The biosynthesis of monoterpenoid indole alkaloids has been studied in detail in order to find alternative production strategies for specific secondary metabolites. Production has been investigated in a number of different systems including plant cell/organ cultures and intact plants.

Only six cDNAs involved in secologanin biosynthesis have been isolated from the monoterpene secoiridoid pathway, namely geraniol diphosphate (*Gpps*) (Turner et al. 2004; Tholl et al. 2004; Burke et al. 1999), geraniol synthase (*Ges*) (Yang et al., 2011; Simkin et al. 2013), P450-dependent geraniol 10-hydroxylase (*Cyp76b6*) (Collu et al., 2001), acyclic monoterpene primary alcohol dehydrogenase (*Adh*) (Ikeda et al. 1991), loganic acid methyltransferase (*Lamt*) (Murata et al. 2008), and P450-dependent secologanin synthase (*Cyp72a1*) (Irmler et al. 2000). High-throughput RNA sequencing of *Catharanthus roseus* using the Illumina HiSeq2000 platform has produced an RNA-Seq dataset to facilitate the cloning of full-length open reading frames (http://www.cathacyc.org), allowing the comparative mining of gene expression and the selection of candidate genes to fill the current gaps in monoterpene secoiridoid biosynthesis pathway (Moerkercke et al., 2013). Two candidate genes were selected using this technology, namely *10-hgo* (*Cr68* tag) and *mtc* (*Cr94* tag). The candidates were selected based on preliminary enzymatic studies carried out in microbial systems by colleagues from Leiden University (The Netherlands) and CNRS (Strasbourg, France).

Transient expression in plants by agroinfiltration is often used as a rapid method to screen for candidate genes (Wroblewski et al. 2005, Faizal et al. 2012). However, transient expression assays based on agroinfiltration are restricted to species and tissues that are biologically compatible and physically accessible to *Agrobacterium tumefaciens.* The efficiency of transient expression assays is influenced by experimental variables that affect the virulence of *A. tumefaciens* and the physiological conditions of the plant (Wroblewski et al. 2005). Furthermore, the

transgene is not integrated into the genome and expression is therefore not biased by position effects resulting in significant differences between transient and stable expression. Transgene expression in transgenic plants often varies significantly due to the position of integration and other effects, thus complicating data analysis (Yang et al. 2000).

Geraniol, produced by geraniol synthase (*Ges*), was only detected in plants transiently expressing the *VoGes* gene. The expression of *Ges* is also associated with toxic effects, e.g. severe damage to grapevine (*Vitis vinifera*) plants expressing *Ocimum basilicum* GES (*ObGes*) resulting in brown-colored embryogenic calli and somatic embryos that tend to die (Fisher et al., 2013). This may explain the lack of geraniol in stable transformants because the trace amounts of geraniol produced following transient expression suggest that the *VoGes* gene is functional. Some form of regulation may prevent the accumulation of geraniol in the stable transformants. Terpenoids generally accumulate in specific cell types due to their cytotoxicity and rightly role in defense responses (Facchini et al. 2008). For example, terpenoids are sequestered into isolated idioblasts and laticifers in *C. roseus*, into root endodermis and stem cortex/pith in *Thalictrum flavum*, and into laticifers in opium poppy. The cellular localization of terpenoid biosynthesis in opium poppy and *C. roseus* is remarkably diverse and complex, establishing new paradigms in the cell biology of secondary metabolism (Facchini et al. 2008).

A large number of monoterpenoid and sesquiterpenoid synthase genes from different plant species have been characterized recently, and many of them have been used for metabolic engineering in plants (Yang et al., 2011). In addition to the direct products of terpene synthases expressed in plants, volatile and non-volatile derivatives are often produced due to the activity of endogenous enzymes. Common modifications include the oxidation, hydroxylation and acetylation of primary terpenoid skeletons (Yang et al., 2011). The accumulation of geraniol as well as geranial, geranic acid and geranyl acetate has been observed following the overexpression of *LdGes* and *ObGes* genes in transgenic maize and tomato (Yang et al., 2011; Davidovich-Rikanati et al., 2007). Seven different geraniol derivatives were found in transgenic maize, dominated by geranyl-6-*O*-malonyl-D-gluopyranoside with the others as minor components. In transgenic tomato plants, the most common geraniol derivatives were the oxidized forms geranic acid and geranial, due to the high levels of endogenous

alcohol dehydrogenases. Low levels of geranyl acetate were also detected in transgenic tomato plants (Davidovich-Rikanati et al., 2007). These data support the detection of geraniol derivatives including acetyl-geraniol dihexose, dipentosyl-hexosyl-geraniol, malonyl-hexosyl geraniol (dimer) and dihexosyl geraniol, in our agroinfiltrated samples (Figure 11 and 12).

Davidovih-Rikanati et al. (2007) refer to endogenous acetyl-CoA:alcohol acetyltransferase activity for the detection of traces amounts of geraniol acetylation forms such as geranyl acetate. Acetyl-CoA acetyltransferase (AACT) catalyzes the condensation of two-carbon acetyl CoA from the beginning of the mevalonate pathway into the 4C compound acetoacetyl-CoA, which is an important precursor for hormone biosynthesis. Recently it has been shown that this enzyme also regulates isoprenoid biosynthesis (Vishwakarma et al., 2012; Soto et al., 2012).

From a biochemical perspective, glycosyl conjugation reduces chemical reactivity whereas malonyl conjugations can facilitate the transport of glycosylated compounds into the vacuole, a process mediated by ATP-binding-cassette transporters (Yang et al., 2011). Malonylation may also increase the water-solubility of the glycosylated compound and prevent glycolysis by glycosidases (Yang et al., 2011).

The identification of the novel geraniol derivatives was made possible by applying a non-targeted LC-MS/TOF approach to analyze any changes in the non-volatile metabolite profile resulting from the transient expression of the first monoterpene secoiridoid genes. Five types of mass analyzer are commonly used, i.e., quadrupole (Q), ion trap (IT), time-of-flight (TOF), LTQ-Orbitrap and Fourier transform ion cyclotron resonance (FI-ICR) (Zhang et al., 2012). Advances in ion sources, mass calibration and analytical capabilities have led to greatly improved sensitivity, dynamic range, mass accuracy, and mass resolution. However, the identification of low-level compounds remains difficult in the presence of much more abundant compounds. The most abundant compounds interfere with the efficiency of the ionization process (Zhang et al., 2012; Guérard et al., 2011).

Many compounds were observed directly in the chromatograms from agroinfiltrated samples. Using LC-MS/TOF, we detected 60 new compounds in the agroinfiltrated samples, which clearly separated the agroinfiltrated and non-agroinfiltrated groups of samples during principal component analysis and are represented in the heat map

(Figure 11). The identification of target and non-target compounds was challenging, due to the absence of molecular formula for each candidate (10-hydroxygeraniol, 10-hydroxygeranial, 10-oxogeranial and iridodial) matching with our mass-to-charge ratio (*m/z*) and no database or literature hits recognizing the new compounds. The heat map shows the quantity of the masses detected variable color (listed in Table 3 in the annex 1), with green representing the detection and red representing full detection. The identification of compounds, especially secondary metabolites, by metabolomic profiling is also challenging. First, the number of commercially available standards representing secondary in particular plant species or tissues is low. Second, it is difficult to achieve optimized levels for all eluting compounds during automated online separation, PDA detection or MS analysis. The presence of overlapping compounds, low intensity mass signals, and confusion with mass signals for MS/MS fragmentation, the extraction of usable information for identification can be complex. Finally, the lack of dedicated software and databases that integrate spectral data and MS data limits the identification procedure to manual curation (Moco et al., 2006). It is clear that a major bottleneck in the large-scale identification of metabolites in complex plant tissues is the absence of a consolidated metabolite library and corresponding metabolite-specific data management system (Moco et al., 2006).

5.7. Conclusions

Transient expression of geranyl diphosphate synthase (*pineGpps*), geraniol synthase (*VoGes*), monoterpenoid cyclase (*Cr96*) and 10-hydroxygeraniol oxidoreductase (*Cr68*) in transgenic tobacco leaves expressing geraniol 10-hydroxylase (*Cyp76b6*) and a cytocrome P450 reductase (*Cpr*), showed that geraniol synthase is functional and can produce trace amounts of geraniol and its derivatives, as expected. Stable expression of *VoGes* did not promote the accumulation of geraniol because of toxicity. We cannot conclude that the expression of downstream genes boosts down the pathway activating the early monoterpene secoiridoid enzymes, since we were unable to detect the downstream metabolites 10-hydroxygeraniol, 10-hydroxygeranial, 10-oxogeranial and iridodial. However, LC-MS/TOF analysis detected 60 new compounds in both transgenic agroinfiltrated lines although they could not be identified. These preliminary results suggest that the expression of downstream genes led to the production of unknown metabolites that cannot currently be identified.

5.7. References

Adam K.P., Thiel R., Zapp J. (1999) Incorporation of 1-[1-(13)C]Deoxy-D-xylulose in chamomile sesquiterpenes. *Arch Biochem Biophys*, 369:127-132.

Bouvier F., Suire C., D'Harlingue A., Backhaus R.A., Camara B. (1998) Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *The Plant Journal*, 24:241–252.

Burke C.C., Wildung M.R., Croteau R. (1999) Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. *PNAS*, 96: 13062–13067.

Burlat V., Oudin A., Courtois M., Rideau M., St-Pierre B (2004) Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of Catharanthus roseus implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant Journal*, 38:131-141.

Cathacyc (2013) CathaCyc, a Metabolic Pathway Database Built from Catharanthus roseus RNA-Seq Data, accessed online 20 March 2013, http://www.cathacyc.org.

Chahed K., Oudin A., Guivarc'h N., Hamdi S., Chénieux J.C., Rideau M., Clastre M. (2000) 1-Deoxy-D-xylulose 5-phosphate synthase from periwinkle: cDNA identification and induced gene expression in terpenoid indole alkaloid-producing cells. *Plant Physiol. Biochem*, 38: 559−566.

Collu, G., Unver, N., Peltenburg-Looman, A.M.G., van der Heijden, R., Verpoorte, R., Memelink, J. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzymeinvolved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* 508: 215–220.

Davidovich-Rikanati R., Sitrit Y., Tadmor Y., Iijima Y., Bilenko N., Bar E., Carmona B., Fallik E., Dudai N., Simon J.E., Pichersky E., Lewinsohn E. (2007) Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nature Biotechnology*, 2: 899-901.

De Vos R.C.H., Moco S., Lommen A., Keurentjes J.J.B., Bino R.J., Hall R.D. (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols,* 2: 778-791.

Dong L., Goedbloed M., Verstappen F.W.A., Voster A., van der Krol [S.], Jongsma M.A., Bouwmeester H.J. (2013) Different subcellular location of Geraniol synthases with little effect on product level. *Metabolic engineering*. *In press*.

Dudey V.S., Bhalla R., Luthra R. (2003) An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *Journal of Biosciences,* 28: 637-646.

Dutta A., Sen J., Deswal R. (2007) Downregulation of terpenid indole alkaloid biosynthetic pathway by low temperature and cloning of AP2 type C-repeat binding factor (CBF) from *Catharanthus roseus* (L). G. Don. *Plant Cell Rep*. 26:1869-1878.

Eisenreich W., Rohdich F., Bacher A. (2001) Deoxyxylulose phosphate pathway to terpenoids. *Trends in Plant Science,* 6: 1360-1385.

El-Sayed M., Verpoorte R. (2007) Catharanthus terpenoid indole alkaloids: biosynthesis and regulation. *Phytochem Rev,* 6:277–305.

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

Faizal A., Geelen D. (2012) Agroinfiltration of intact leaves as a method for the transient and stable transformation of saponin producing Maesa lanceolata. *Plant Cell Rep*, 31:1517–1526.

Ginis O., Courdavault V., Melin C., Lanoue A., Giglioli-Guivarc'h N., St-Pierre B., Courtois M., Oudin A. (2012) Molecular cloning and functional characterization of *Catharanthus roseus* hydroxymethylbutenyl 4-diphosphate synthase gene promoter from the methyl erythritol phosphate pathway. *Mol. Bio. Rep.*, 39:5433-5447

Goddijn O.J.M., De Kam R.J., Zanetti A., Schilperoort, R.A., Hoge J.H.C. (1992) Auxin rapidly down-regulates transcription of the tryptophan decarboxylase gene from Catharanthus roseus. *Plant Molecular Biology,* 18:1113-1120.

Guérard F., Pétriacq P., Gakière B., Tcherkez G. (2011) Liquid chromatography/time-of-flight mass spectrometry for the analysis of plant samples: A method for simultaneous screening of common cofactors or nucleotides and application to an engineered plant line. *Plant Physiology and Biochemistry*, 49: 1117-1125.

Guirimand G., Burlat V., Oudin A., Lanoue A., St-Pierre B., Courdavault V. (2009) Optimization of the transient transformation of *Catharnathus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase. *Plant Cell Rep,* 28:1215-1234.

Hedhili S., Courdavault V., Giglioli-Guivarc'h N., Gantet P. (2007) Regulation of the terpene moiety biosynthesis of Catharanthus roseus terpene indole alkaloids. *Phytochem Rev,* 6:341–351.

Hemmerlin A., Hoeffler J-F., Meyer O., Tritsch D., Kagan I. A., Grosdemange-Billiard C., Rohmer M., Bach T.J. (2003) Cross-talk between the Cytosolic Mevalonate and the Plastidial Methylerythritol Phosphate Pathways in Tobacco Bright Yellow-2 Cells. *The Journal of Biological Chemistry*, 278: 26666–26676.

Ikeda H., Esaki N., Nakai S., Hashimoto K., Uesato S., Soda K., Fujita T. (1991) Acyclic monoterpene primary alcohol: NADP super(+)oxidoreductase of Rauwolfia serpentina cells: the key enzyme in biosynthesis of monoterpene alcohols. *J Biochem,* 109:341–347.

Irmler S., Schroder G., St-Pierre B., Crouch N.P., Hotze M., Schmidt J., Strak D., Matern U., Schroder J. (2000) Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J,* 24:797–804.

Kai G.Y., Miao Z.Q., Zhang L., Zhao D.L., Liao Z.H., Sun X.F., Zhao L.X., Tang K.X. (2006) Molecular cloning and expression analyses of a new gene encoding

3-hydroxy-3-methylglutaryl -CoA synthase from *Taxus* × *media*. *Biol Plant*, 50:359–366.

Kai G.Y., Li S.S., Wang W., Lu Y., Wang J., Liao P., Cui L.J. (2013) Molecular cloning and expression analysis of a gene encoding 3-hydroxy-3 methylglutaryl-CoA synthase from *Camptotheca acuminata*. *Russian Journal of Plant Physiology*, 60: 131-138.

Kuzuyama T., Takahashi S., Takagi M., Seto H. (2000) Characterization of 1-deoxy-d-xylulose 5-phosphate reductoisomerase, an enzyme involved in isopentenyl diphosphate biosynthesis, and identification of its catalytic amino acid residues. *The Journal Of Biological Chemistry* , 275: 19928–19932.

Lange B.M., Wildung M.R., McCaskill D., Croteau R. (1998) A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. *Proc Natl Acad Sci USA*, 95:2100-2104.

Lange B.M., Croteau R. (1999) Mevalonate-independent pathway in plants: cloning and heterologous expression of 1-deoxy-d-xylulose-5-phosphate reductoisomerase from peppermint. *Archives of Biochemistry and Biophysics*, 365: 170-174.

Laule O., Fürholz A., Chang H-S., Zhu T., Wang X., Heifetz P.B., Gruissem W., Lange B.M. (2003) Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in Arabidopsis thaliana. *PNAS*, 100: 6866 – 6871.

Liu D.H., Jin H.B., Chen Y.H., Cui L.J., Ren W.W., Gong Y.F., Tang K.X. (2007) Terpenoid Indole Alkaloids Biosynthesis and Metabolic Engineering in *Catharanthus roseus. Journal of Integrative Plant Biology*. 49: 961-974.

Maeda H., Dudareva N. (2012) The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annu. Rev. Plant Biol*, 63:73–105

Moco S., Bino R.J., Vorst O., Verhoeven H.A., de Groot J., van Beek T.A., Vervoort J., Ric de Vos C.H. (2006) A Liquid Chromatography-Mass Spectrometry-Based Metabolome Database for Tomato. *Plant Physiology*, 141: 1205–1218.

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

O'Connor S.E., Maresh J.J. (2005) Chemestry and biology of monoterpen indole alkaloid biosynthesis. *Natural Product Reports*, 23: 532-547.

Oudin A., Mahroug S., Courdavault V., Hervouet N., Zelwer C., Rodriguez-Concepcion M., St-Pierre B., Burlat V. (2007) Spatial distribution and hormonal regulation of gene products from methyl erythritol phosphate and monoterpene-secoiridoid pathways in Catharanthus roseus. *Plant Mol. Biol.*, 65:13–30.

Rohdich F., Kis K., BAcher A., Eisenreich W. (2001)The non-mevalonate pathway of isoprenoids: genes, enzymes and intermediates, *Current Opinion in Chemical Biology*, 5:535–540.

Schmidt M., Strack J., Matern D., Schröder J. (2000) Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J*, 24: 797–804.

Schwender J., Muller C., Zeidler J., Lichtenthaler H.K. (1999) Cloning and heterologous expression of a cDNA encoding 1-deoxy-D-xylulose-5-phosphate reductoisomerase of Arabidopsis thaliana. *FEBS Letters*, 455:140-144.

Simkin A.J, Miettinen K., Claudel P, Burlat V., Guirimand G, Courdavault V., Papon N., Meyer S., Godet S., St-Pierre B., Giglioli-Guivarc'h N., Fischer M.J.C., Memelink J., Clastre M. (2013) Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry,* 85: 36–43.

Soto G., Strizler M., Lisi C., Alleva K., Pagano M.E., Ardila F., Mozzicafreddo M., Cuccioloni M., Angeletti M., Ayub N.D. (2012) Acetoacetyl-CoA thiolase regulates the mevalonate pathway during abiotic stress adaption. *Journal of Experimental Botany*, 62: 5699–5711.

Tholl D. (2006) Terpene synthases and the regulation, diversity and biological roles of

terpene metabolism. *Current Opinion in Plant Biology*, 9:297–304.

Tikunov Y., Lommen A., De Vos C., Verhoeven H.A., Bino R.J., Hall R.D., Bovy A.G. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiology*, 139: 1125-1137.

Turner G.W., Croteau R. (2004) Organization of Monoterpene Biosynthesis in Mentha. Immunocytochemical Localizations of Geranyl Diphosphate Synthase, Limonene-6-Hydroxylase, Isopiperitenol Dehydrogenase, and Pulegone Reductase. *Plant Physiology,* 136: 4215–4227.

Tzin V., Galili G. (2010) New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Molecular Plant*, 6: 956-972.

Veau B., Courtois M., Oudin A., Chënieux J.C., Rideau M. (2000). Cloning and expression of cDNAs encoding two enzymes of the MEP pathway in Catharanthus roseus. *Marc Clastre Biochimica et Biophysica Acta,* 1517: 159-163.

Verma P., Mathur A.K., Srivastava A., Mathur A. (2012) Emergin trends in research on spatial and temporal organization of terpenoid indole alkaloid pathway in *Catharanthus roseus*: a literature update. *Protoplasma,* 249: 255-268.

Vishwakarma R.K., Ruby, Singh S., Sonawane P.D., Srivastava S., Kumari U., Santosh R.J., Kumar, Khan B.M. (2012) Molecular Cloning, Biochemical Characterization, and Differential Expression of an Acetyl-CoA C-Acetyltransferase Gene (*AACT*) of Brahmi (*Bacopa monniera*). *Plant Mol Biol Rep, in press.*

Wang H., Nagegowda D.A., Rawat R., Bouvier-Navé P., Guo D., Bach T.J., Chye M-L. (2012) Overexpression of Brassica juncea wild-type and mutant HMG-CoA synthase 1 in Arabidopsis up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance. *Plant Biotechnology Journal*, 10: 31–42.

Wang Q., Yuan F., Pan Q., Li M., Wang G., Zhao J and Tang K. (2010) Isolation and functional analysis of the *Catharanthus roseus* deacetylvindoline-4-*O*-acetyltransferase gene promoter. *Plant Cell Rep*. 29:185-192.

Wroblewski T., Tomczak A., Michelmore R. (2005) Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnology Journal*, 3: 259–273.

Yang T., Stoopen G., Yalpani N., Vervoort J., de Vos R., Voster A., Verstappen F.W.A., Bouwmeester H.J., Jongsma M.A. (2011) Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. *Metabolic Engineering*,13 414-425..

Yazaki K. (2004) Handbook of Plant Biotechnology. *John Wiley & Sons, Ltd*. Christou P and Klee H. England. 2: 811-861.

Zhang L., Yan X., Wang J., Li S., Liao P., Kai G. (2006) Molecular cloning and expression analysis of a new putative gene encoding 3-hydroxy-3-methylglutaryl-CoA synthase from *Salvia miltiorrhiza. Acta Physiologiae Plantarum*, 33: 953-961.

Zhang H., Wang S.Q., Liu Y., Luo L.P., Liu P., Qi L-W., Li P. (2011) Trace analysis in complex mixtures using a high-component filtering strategy with liquid chromatography–mass spectrometry. *Journal of Pharmaceuticaland Biomedical Analysis*, 70: 169– 177.

CHAPTER 6

ENGINEERING THE MONOTERPENE SECOIRIDOID PATHWAY BY COMBINATORIAL MULITGENE TRANSFER

CHAPTER 6: ENGINEERING THE MONOTERPENE SECOIRIDOID PATHWAY BY COMBINATORIAL MULITGENE TRANSFER

6.1. Abstract

A combinatorial transformation platform that has been established in our laboratory was used to generate a population of transgenic maize plants expressing different mixtures of input transgenes representing the carotenoid biosynthesis pathway. The simultaneous introduction of multiple transgenes by particle bombardment allowed the rapid recovery of diverse maize germplasm accumulating high levels of nutritionally important carotenoids. The library of plants expressing different transgene combinations also allowed the maize carotenoid pathway to be dissected, revealing previously unknown rate-limiting steps. A similar strategy was used to engineer the monoterpene secoiridoid pathway in tobacco plants. Two different combinations of genes were used to reconstruct the upstream pathway (five transgenes) and downstream pathway (four transgenes) in each case with an additional selectable marker. A diverse population of transgenic plants expressing different combinations of enzymes was recovered in preparation for biochemical and metabolomic analysis.

6.2. Introduction

6.2.1. Multigene engineering

The introduction of multiple transgenes into plants is necessary for sophisticated genetic engineering strategies, such as the stacking of transgenes specifying different agronomic traits, the expression of different polypeptide subunits making up a multimeric protein, the introduction of several enzymes acting sequentially in a metabolic pathway or the expression of a target protein and one or more enzymes required for specific types of post-translational modification (Agrawal et al., 2005; Alpeter et al., 2005; Naqvi et al. 2009a). Before specific multigene transfer methods were developed, multiple genes could be stacked in transgenic plants using multiple rounds of crosses among different transgenic lines or sequential transformation (the supertransformation of transgenic plants with additional transgenes). Both methods have two major drawbacks: the long and labor-intensive breeding program, and the

fact that the different transgenes are unlinked, leading to segregation in subsequent generations. The high level of heterozygosity in crops such as potato and cassava means that clonal propagation is favored and makes the abovementioned crossing strategies difficult to implement (Naqvi et al. 2009a; Agrawal et al., 2005; Alpeter et al., 2005).

The simultaneous transformation of plants with two or more transgenes (cotransformation) achieves the same goals in a single generation and often results in the cointegration of all transgenes at a single locus thus preventing segregation (Twyman et al., 2004; Alpeter et al., 2005). However, the process becomes less efficient as more transgenes are required, introducing a significant bottleneck in the progress of plant biotechnology (Naqvi et al. 2009a; Agrawal et al., 2005; Alpeter et al., 2005).

Transgenic plants are generated using two major strategies. One exploits the natural ability of *Agrobacterium tumefaciens* to transfer DNA from a resident plasmid into the plant genome. The other is collectively termed ‘direct DNA transfer’ and involves physical and chemical gene transfer methods such as microinjection, the transformation of protoplasts using polyethylene glycol or calcium phosphate, particle bombardment, electroporation and transformation using silicon carbide whiskers. The only similarity among the direct transfer techniques is that external physical or chemical factors mediate DNA delivery into the cell (Alpeter et al., 2005). Particle bombardment is the most convenient method for multiple gene transfer to plants because DNA mixtures comprising any number of different transformation constructs can be used without complex cloning strategies (Alpeter et al., 2005).

Particle bombardment as we know it today was developed by John Sanford and colleagues in the early 1980s (Sanford et al., 1987). It is a direct DNA transfer method in which high-velocity microprojectiles are used to introduce DNA into plant cells. The microprojectiles (tungsten or gold) are coated with a DNA mixture comprising any number of transformation constructs (Agrawal et al., 2005; Twyman et al., 2002; Twyman et al., 2004).

Particle bombardment is widely used because it has three key advantages over other direct transfer methods and *Agrobacterium*-mediated transformation. First, the delivery process is entirely physical and can therefore be applied to any species

without concern about the Agrobacterium 'host range'. Indeed, the range of species amenable to particle bombardment is restricted only by the competence of cells for regeneration, which means the technique is also genotype independent and thus useful for the transformation of elite cultivars as well as model varieties (Twyman et al., 2002; Twyman et al., 2004; Alpeter et al., 2005). Second, there is no dependence on a specific cell type. Unlike protoplast transformation, whisker-mediated transformation and electroporation, which rely on the availability of protoplasts or suspension cells, particle bombardment can be applied to any type of tissue, including explants of organized tissues (leaves, shoot tips and embryos), callus cultures, cell suspensions, whole plant organs *in situ* and even pollen grains (Twyman et al., 2002; 2004). Careful optimization is required to tailor the method for different species and cell types by selecting conditions that achieve the highest efficiency of transformation with the least cell damage. Important parameters include the acceleration method, particle velocity (controlled by the discharge voltage and/or gas pressure), particle size, and the use of different materials (tungsten or gold). Osmotic pretreatment prior to bombardment also increases the transformation efficiency in some species and/or tissues (Twyman et al., 2002; 2004).

Third, particle bombardment allows the transfer of many different genes at once using different plasmids, because these tend to concatenate so that a larger composite transgene integrates at a single locus resulting in stable inheritance (Twyman et al., 2002). In *Agrobacterium*-mediated transformation, the transgene must be placed between T-DNA repeats and further sequences such as *overdrive* and *transfer enhancer* may be required for efficient DNA delivery (Alpeter et al., 2005; Twyman et al., 2002; 2004). The introduction of multiple genes using *Agrobacterium tumefaciens* can be achieved by linking transgenes on a single vector or by using multiple strains carrying different vectors, but these approaches become progressively more laborious and less efficient as the number of different transgenes increases due to the instability of large T-DNAs, the lack of unique restriction sites during iterative cloning and the fact that larger input DNA sequences are more likely to fragment (Twyman et al., 2002; Naqvi et al, 2009a). An *Agrobacterium*-based multigene assembly vector system has been described, which goes some way towards addressing these limitations (Lin et al., 2003). However, assembly relies on a complex process involving the Cre/*lox*P site-specific recombination system and homing endonucleases

(Li et al., 2003). In contrast, there is no such biological constraint in particle bombardment and no vector DNA sequences are required for delivery. Cloning vectors are used in particle bombardment for convenience rather than necessity (Alpeter et al., 2005; Twyman et al., 2004). Particle bombardment offers a straightforward approach to multiple gene transfer in which the microprojectiles are coated with a mixture of transformation constructs (Zhu et al., 2008).

6.2.2. Genetic engineering of the monoterpene secoiridoid pathway

The monoterpene secoiridoid pathway, which is described in detail in Chapter 5, Section 5.2.5, produces secologanin via the methylerythritol phosphate (MEP) pathway (Verma et al., 2012; Ginis et al., 2011). The isoprene building block IPP and its isomer DMAPP are derived from the MEP pathway in the plastids, and these are condensed by the enzyme geranyl diphosphate synthase (GPPS) which is also located in this compartment (Turner et al., 2004; Burke et al., 2002). In *Citrofortunella mitis* and *Narcissus pseudonarcissus*, GPPS is indirectly localized in leucoplasts, chloroplasts and the non-green plastids of secretory cells, which are widely recognized as the sites of monoterpene synthesis in plants (see Chapter 2, Section 2.6) (Turner et al., 2004; Bouvier et al., 2000). GPPS may exist as a homodimer or heterodimer depending on the species (Orlova et al., 2009; Schimdt et al., 2008; Bouvier et al., 2000; Burke et al., 1999) although the only homodimeric GPP described thus far is found in *Arabidopsis thaliana* (Schmidt et al., 2008). Similar albeit uncharacterized sequences are present in GenBank e.g. from *Citrus sinensis, C. unshiu* and *Quercus robur* (Schmidt et al., 2008). GPPS cDNAs have also been isolated from several species including *Mentha x piperita* and *Antirrhinum majus* (Table 1; Burker et al., 1999; Tholl et al., 2004).

The small GPPS subunit from *Antirrhinum majus* (*AmGpps.ssu*) has been overexpressed in tobacco (Orlova et al., 2009). Both GPPS activity and monoterpene emission from leaves and flowers was greater in the transgenic plants than wild type controls, indicating that the catalytically-inactive GPPS.SSU interacted with endogenous large subunit partners and formed an active snapdragon/tobacco hybrid GPPS *in planta* (Orlova et al., 2009). However, this also caused the loss of chlorophylls, carotenoids and gibberellins, resulting in chlorosis, increased light

sensitivity and dwarfism. The transgenic plants also emitted lower amounts of sesquiterpenes, suggesting that the export of isoprenoid intermediates from the plastids into the cytosol was restricted (Orlova et al., 2009).

Geraniol diphosphate is converted to geraniol by the enzyme geraniol synthase (GES), and corresponding genes have been cloned from *Ocimum basilicum* (*ObGes*) and *Catharanthus roseus* (*CrGes*) as shown in Table 1 (Ijima et al., 2004; Simikin et al., 2013). *ObGes* has been expressed in *Arabidopsis thaliana* and *Vitis vinifera* (Fischer et al., 2013). The transgenic *A. thaliana* plants appeared normal, whereas the growth of the transgenic grapevine plants was severely impaired compared to wild type controls. This was noticeable at all stages of the transformation procedure, with a higher proportion of embryogenic calli exhibiting the typical brownish coloration which is ultimately lethal, reflecting the fact that monoterpenols are toxic to plants and tend to accumulate in glandular trichomes if they are present at high concentrations (Fischer et al., 2013). Metabolic analysis of the transgenic plants revealed the presence of geraniol in grapevine and in *A. thaliana* (Fischer et al., 2013).

The *CrGes* gene product is localized in the stroma and stromules in close association with the ER, suggesting an exchange of metabolites between the two organelles *in vivo* (Simikin et al., 2013). Such an exchange would be concordant with the next step in monoterpene synthesis, i.e. the hydroxylation of geraniol at the C-10 position to form 10-hydroxygeraniol, catalyzed by the cytochrome P450 geraniol 10-hydroxylase (G10H). This enzyme is anchored in the ER membrane with the catalytic domain probably exposed to the cytosol in the vicinity of stromules (Simkin et al., 2013; Guirimand et al., 2009). Stromules may therefore facilitate the export of geraniol into the cytosol hance its conversion into 10-hydroygeraniol by G10H (Simkin et al., 2013).

C. roseus G10H (CYP76B6) belongs to the CYP76B superfamily (Collu et al., 2001) and has been shown to synthesize 10-hydroxygeraniol both *in vitro* and *in vivo* (Wang et al., 2002c). *Arabidopsis thaliana* CYP76C1 can also hydroxylate geraniol *in vitro* (Wang et al., 2002c). The corresponding *CPR* genes have been characterized (Suttipanta et al. 2007; Wang et al 2010b). Hundreds of P450s have been identified in plants. Almost all eukaryotic P450s belong to the class II microsomal P450 system

which comprises two integral membrane proteins: P450 and NADPH-cytochrome P450 oxidoreductase (CPR), the latter acting as an electron donor (Jensen et al. 2010).

The estimated ratio between CPR and P450 in the microsomal membrane is 1:15, suggesting some form of regulation is required to ensure efficient electron transfer from CPR to the P450 enzymes. Substrate binding by P450 is thought to induce conformational changes that enhance the affinity of P450 for CPR, increasing the efficiency of electron donation from CPR to P450s ready to proceed into the catalytic monooxygenation cycle by binding a substrate molecule (Jensen et al. 2010). One potential issue associated with the heterologous expression of P450s is the supply of electrons from CPR to optimize P450 activity (Jensen et al. 2010). The enhanced P450 activity observed in the presence of CPR probably reflects the more efficient CPR–P450 interaction. The heterologous expression of P450 is usually supported by the endogenous CPR, probably reflecting the conservation of CPR residues that mediate P450 interactions. However, the efficiency of CPR–P450 interactions may depend on the CPR homolog because interactions between CPR and P450s are thought to be based on the different electrostatic potentials of the two proteins. Small variations in electrostatic potentials between different CPR proteins could therefore reduce the interaction efficiency (Jesen et al. 2010).

At least 11 enzymatic steps may be required to convert geraniol to secologanin via loganic acid and loganin (Loyola-Vargas et al. 2007), but only four cDNAs have been isolated (Verma et al., 2012). These encode the P450-dependent geraniol 10-hydroxylase (G10H; Collu et al. 2001; Wang et al., 2010a), acyclic monoterpene primary alcohol dehydrogenase (10HGO; Ikeda et al. 1991; Hallahan et al., 1995; Valleta et al., 2010), loganic acid methyltransferase (LAMT; Murata et al. 2008) and P450-dependent secologanin synthase (SLS; Irmler et al. 2000).

G10H cDNAs have been cloned from *Catharanthus roseus* (*CrG10h*) and *Swertia mussotii* (*SmG10h*) (Collu et al., 2001; Wang et al., 2010a). *SmG10h* encodes the 496-residue protein CYP76B10, which shows 80.2% identity to *C. roseus* G10H (Wang et al., 2010a). The enzyme has been overexpressed in *C. roseus* hairy roots and whole plants (Peacock et al., 2005; Gong et al., 2005; Wang et al., 2010b, Pan et al., 2012). Hairy roots expressing G10H produced 9.51 mg/g DW of terpenoid indole alkaloids, representing a three-fold increase over untransformed root lines (Gong et al., 2005).

Pan et al. (2012) reported the accumulation of vindoline, catharanthine and ajmalicine in transgenic *C. roseus* plants overexpressing G10H. Vindoline levels increased from 0.70 mg/g DW in the control to a maximum of 3.00 mg/g DW in the transgenic roots, whereas catharanthine increased from 1.99 to a maximum of 5.36 mg/g DW and ajmalicine increased from 0.05 to a maximum of 0.49 mg/g DW. *Swertia mussotii* plants overexpressing *SmG10h* accumulated 0.053 mg/g DW 10-hydroxygeraniol compared to 0.021 mg/g in wild type plants and 2.23 mg/g of the iridoid monoterpenoid swertiamarin compared to 1.31 mg/g in the wild type plants (Wang et al., 2010a). These studies demonstrated that G10H catalyzes a rate-limiting step in the monoterpenoid iridoid biosynthetic pathway in both *C. roseus* and *S. mussotii*, and this represents one of the key targets the metabolic engineering of monoterpenoid iridoid production.

In the next step, 10-hydroxygeraniol is oxidized to 10-oxogeraniol in the presence of 10-hydroxygeraniol NADP-oxidoreductase (10HGO), and the 10-oxogeranial is cyclized into iridodial by 10-oxogeraniol/iridodial cyclase, a monoterpene cyclase (Ginis et al., 2011; Verma et al., 2012). The acyclic monoterpene primary alcohol:NADP+ oxidoreductase (10HGO) is unstable, so the protein is poorly characterized in *Rauwolfia serpentina*, *Nepeta racemosa* and *Camptotheca acuminata* (Ikeda et al., 1991; Hallahan et al., 1995; Valleta et al., 2010). Although several monoterpene cyclases have been isolated, the 10-oxogeraniol/iridodial cyclase from *Rauwolfia serpentina* has been only partially characterized (Uesato et al., 1987).

At the end of the pathway, 7-deoxyloganin is converted into secologanin by the sequential activity of deoxyloganin 7-hydroxylase (DL7H), which converts 7-deoxyloganin into loganin and 7-deoxyloganic acid into loganic acid. Loganic acid methyltransferase (LAMT) catalyzes the conversion of loganic acid to loganin. Murata et al. (2008) have been reported the functional characterization of LAMT from *Catharanthus roseus*.

The last step in secologanin biosynthesis from loganin involves the oxidative rupture of the loganin methylcyclopentane ring catalyzed by secologanin synthase (SLS). The synthesis of secologanin requires NADPH and molecular oxygen, and is inhibited by carbon monoxide and several other cytochrome P450 inhibitors (Liu et al., 2007; O'Connor et al., 2005). SLS has been characterized as the cytochrome P450 enzyme

CYP72A1 from *Catharanthus roseus* (Irmler et al., 2000). The conversion of loganin into secologanin is also catalyzed by a cytochrome P450-dependent monooxygenase in *Lonicera japonica* (Yamamoto et al., 2000).

Table 1. Functional gene characterization in monoterpene secoiridoid pathway.

Gene	Enzyme	Source	Reference
gpps	Geranyl diphosphate synthase	*Picea abies, Cinnamomum tenuipilum, Abies grandis, Antirrhinum majus, Mentha x piperita, Mentha spicata, Ocimum basilicum*	Schmidt et al., 2008; Yang et al., 2005; Orlova et al., 2009; Burke et al., 1999; Ijima et al., 2004
ges	Gerniol synthase	*Catharanthus roseus, Ocimum basilicum*	Iijima et al., 2004; Simikin et al. 2013
cpr	Cytochrome P450 reductase	*Catharanthus roseus, Arabdiopsis thaliana*	Schwarz et al., 2009
g10h	Geraniol-10-hydroxylase	*Catharanthus roseus, Swertia mussotii Franch*	Collu et al., 2001; Wang et al., 2010a
10-hgo	10-hydroxygeraniol oxidoreductase	*Rauwolfia serpentina, Nepeta racemosa, Camptotheca acuminata*	Ikeda et al., 1991; Hallahan et al., 1995; Valleta et al., 2010
mtc	10-oxogeraniol/ Iridoid cyclase	*Rauwolfia serpentina, Nepeta racemosa, Camptotheca acuminata*	Ikeda et al., 1991; Hallahan et al., 1995; Valleta et al., 2010
lamt	Loganic acid methyltransferase	*Catharanthus roseus*	Murata et al., 2008
cyp72a1	Secologanin synthase	*Catharanthus roseus; Lonicera japonica*	Yamamoto et al., 2000; Irmler et al., 2000

6.3. Aims and objectives

The aim of the experiments described in this chapter was to understand the factors involved in the regulation of the monoterpene secoiridoid pathway in plants and to use the resulting knowledge to reconstruct a heterologous monoterpene secoiridoid pathway in tobacco using a previously-described combinatorial transformation method.

The objectives of the experiments can be summarized as follows:

- Develop a multi-gene combinatorial transformation system for tobacco
- Introduce five upstream genes and four downstream genes from the pathway into tobacco plants by combinatorial transformation, together with the kanamycin resistance selectable marker gene
- Select and regenerate putative transgenic plants
- Analyze gene expression in the transgenic plants at the mRNA level

6.4. Materials and methods

6.4.1. Transformation constructs

All transformation constructs were provided by collaborators from Leiden University (Netherlands) and were based on vector pRT101 (Figure 1). The expression cassette comprised the CaMV 35S promoter and the polyadenylation signal from CaMV strain Cabb B-D. The coding region comprised one of the following genes: geranyl diphosphate synthease (*Gpps*) from pine (GenBank accession no. AT2G34630) yielding vector pRT101-pineGPPS, geraniol synthase (*Ges*) from *Valeriana officialis* (GenBank accession no. JX494702.1) yielding vector pRT101-VoGES, geraniol synthase (*Ges*) from *Catharanthus roseus* (Shikin et al., 2012) yielding vector pRT101-CrGES, or the *C. roseus* genes for geraniol 10-hydroxylase (*Cyp76b6*), 10-hydroxygeraniol oxidoreductase (*10-hgo*), monoterrpenoid cyclase (*Mtc*), cytochrome P450 (CYP76A26), glucosyl transferase (*Dlgt*), loganic acid-*O*-methyltransferase (*Lamt*) or secologanin synthase (*Sls*), yielding vectors pRT101-

CYP76B6, pRT101-10GHO, pRT101-MTC, pRT101-CYP76A26, pRT101-DLGT, pRT101-LAMT and pRT101-SLS, respectively (Table 2).

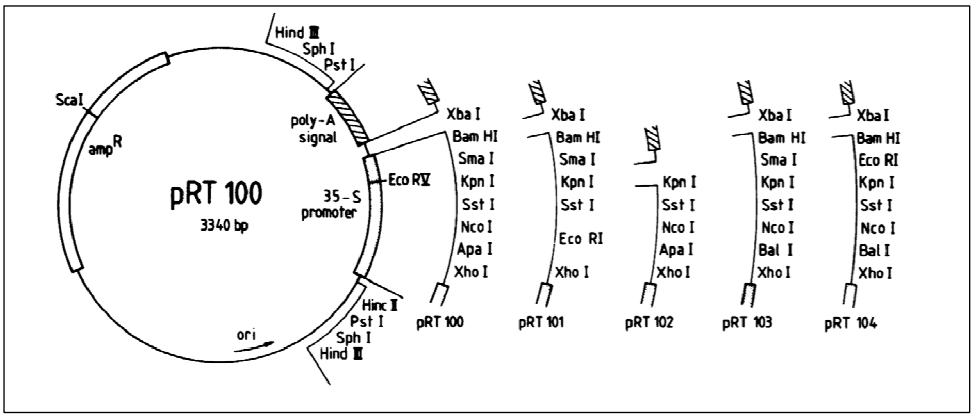

Figure 1. pRT100 family vectors used for tobacco transformation.

Table 2. Constructs used for tobacco transformation experiments.

Enzyme	Gene	Construct
Geranyl diphosphate synthase	*pineGpps*	pRT101-GPPS
Geraniol synthase	*VoGes*, *CrGes*	pRT101-VoGES, pRT101-CrGES
geraniol 10-hydroxylase	*CrCyp76b6*	pRT101-CYP76B6
10-hydroxygeraniol oxidoreductase	*Cr10-hgo*	pRT101-10HGO
Monoterpenoid cyclase	*CrMtc*	pRT101-MTC
Unknown	*CrCyp76a26*	pRT101-CYP76A26
Glucosyl transferase	*CrDlgt*	pRT101-DLGT
Loganic acid-*O*-methyltransferase	*CrLamt*	pRT101-LAMT
Secologanin synthase	*CrSls*	pRT101-SLS

The selectable marker gene *npt*II was used for stable transformation experiments as described in detail in Chapter 2, Figure 2, Section 2.4.2.

6.4.2. Transformation, selection and regeneration of putative transgenic plants

Leaf segments were transformed by particle bombardment in several experiments as described in Chapter 2, Section 2.4.2., using a 3:1 molecular weight ratio of the primary vectors and the selectable marker. Three combinations of genes were prepared as listed in Table 3. Bombarded leaves were incubated in the dark and placed on fresh medium every 2 weeks as described in Chapter 2, Section 2.4.3.

Table 3. Gene combinations used for tobacco transformation experiments.

Combination	Genes
Upstream 1	*PineGpps, VoGes, CrCyp76b6, Cr10hgo, Crmtc*
Upstream 2	*PineGpps, CrGes, CrCyp76b6, Cr10hgo, Crmtc*
Downstream	*CrCyp76a26, CrDlgt, CrLamt, CrSls*

6.4.3. Expression analysis

6.4.3.1. RNA extraction and mRNA blot analysis

Total RNA was extracted from 120 mg tobacco leaf tissue using Trizol reagent (Invitrogen, Paisley, UK) as described in Chapter 2, Section 2.4.5.1. We separated 30 μg RNA per lane by agarose gel electrophoresis and blotted onto nitrocellulose membranes as described in Section 2.4.5.3, then probed the membranes with sequences corresponding to *PineGpps, VoGes, CrGes, CrCyp76b6, Cr10hgo, Crmtc, CrCyp76a26, CrDlgt, CrLamt* and *CrSls* (Table 4).

Table 4: Forward and reverse primers used for designing probes for mRNA blots.

Gene	Forward primer	Reverse primer
PineGpps	5'- GGGTGTGCAAGGAAAACAG-3'	5'- TTTCCGAGAGAGGCAGATG-3'
VoGes	5'-GGCTGATTCTTCCTTGCTTG-3'	5'-ATCGTCCCACAGTCGAAGA-3'
CrGes	5'-GGTGGTGGAAACAATTAGG-3'	5'-TGCAAGAGGCATTTCTTCT-3'
CrCyp76b6	5'-TCCCCTGCTTGAAAAAGTTG-3'	5'-CACCCTTGTGAATTCGTCCT-3'
Cr10hgo	5'-GAGGGCAACTCTTGAGGA-3'	5'-TGTTATGCATTCCGGCAA-3'
Crmtc	5'-TTTGGCACCTCTGACAAGG-3'	5'-CCAAGGCTTTGATCCCAT-3'
CrCyp76a26	5'-GATCCAGAAGGAAGGGAATT-3'	5'-CCAAGATCCCAATCAAAAG-3'
CrDlgt	5'-GGGATCTTCCAACTTTCTTGA-3'	5'-GGCTTTCAACTTCATCCCTC-3'
CrLamt	5'-TGATTCCATTGAAATGCCTG-3'	5'-ATCACCATCAACCCTCCAA-3'
CrSls	5'-GGAAACAAGAAGGATGGTGG-3'	5'-GATGGAGCAACATCAAACT-3'

6.4.3.2. RT-PCR

Lamt and *Sls* transgene expression was monitored using the reverse transcriptase polymerase chain reaction (RT-PCR) as described in Chapter 2, Section 2.5.3.2. The pRT101-LAMT and pRT101-SLS plasmids (150 ng) were used as positive controls and 2 ng of cDNA per sample was used as the template as described in Chapter 2, Section 2.4.5.2. The primers listed in Table 3 amplified 657 bp and 567 bp fragments within the *Lamt* and *Sls* genes, respectively. The RT-PCR products, together with a DNA molecular weight marker (BenchTop 1-kb DNA Ladder, Promega, Madrid, Spain) were separated by 0.8% (w/v) agarose gel electrophoresis and visualized by ethidium bromide staining under UV light.

6.5. Results

6.5.1. Selection of transformants and plant regeneration

We regenerated 33 kanamycin-resistant plantlets following the bombardment of tobacco (*Nicotiana tabacum* cv. Petite Havana SR1) with the Upstream 1 combination of transgenes (under the control of the constitutive CaMV 35S promoter) and the *npt*II kanamycin resistance gene (Table 4). Leaf pieces transformed with the Upstream 2 combination of transgenes were severely impaired throughout the transformation

procedure, resulting in a brownish coloration and ultimately death in most cases (Figure 2). However, we recovered one kanamycin-resistant putative transgenic plantlet. Leaf pieces transformed with the Downstream combination of genes behaved in a similar manner although the phenotype was not so severe, yielding dwarf transgenic plantlets with close knots and small leaves (Figure 3). Twelve putative transgenic plants survived the selection and regeneration process following transformation with the Downstream combination of genes. All the transgenic plants were self-pollinated.

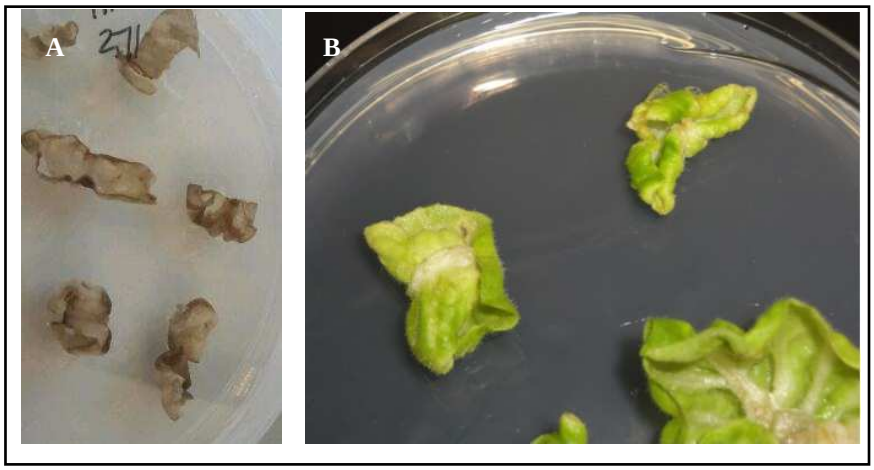

Figure 2. A. Dead leaf pieces transformed with the Upstream 2 combination of transgenes. **B.** Normal leaf pieces transformed with the Upstream 1 combination of transgenes.

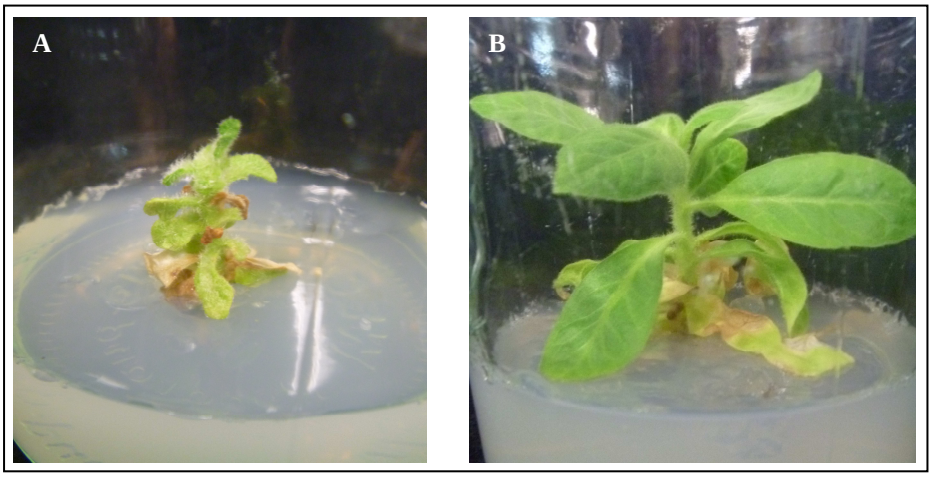

Figure 3. A. Putative transgenic tobacco plantlet transformed with the Downstream combination of transgenes. **B**. Normal putative transgenic tobacco plantlet transformed with the Upstream 1 combination of transgenes.

6.5.2. Expression analysis

The putative transgenic plants were screened for the expression of each input transgene by mRNA blot analysis. The transcripts representing each transgene from the Upstream 1 combination could be detected (Figure 4) generating the combinatorial population outlined in Table 5. Six of the 33 regenerated plants appeared to express only the *npt*II selectable marker gene because no other transcripts were detected.

Table 5. Expression analysis at the mRNA level based on the Upstream 1 transformation experiment. The input genes were *PineGpps*, *VoGes*, *Cyp76b6*, *10hgo* and *Mtc*, and the selectable marker gene was *npt*II. (+) and (-) indicate the presence or absence of a transcript, respectively.

Sample ID	*pineGpps*	*VoGes*	*Cyp76b6*	*10hgo*	*Mtc*
1	-	-	-	-	-
2	-	-	-	-	-
3	-	-	+	-	+
4	-	-	+	-	+
5	-	+	-	-	-
6	-	+	-	+	+
7	-	-	-	+	-
8	-	-	+	-	-
9	-	-	-	+	-
10	-	-	-	-	-
11	-	-	-	+	-
12	-	-	+	-	-
13	-	-	+	+	-
14	-	-	+	+	+
15	-	-	+	+	+
16	-	-	-	+	-
17	-	-	-	-	-
18	-	-	-	-	-
19	-	-	-	-	-
20	+	-	-	-	+
21	-	-	-	+	+
22	-	-	+	+	-
23	-	-	+	+	-

Table 5. Expression analysis at the mRNA level based on the Upstream 1 transformation experiment. The input genes were *PineGpps, VoGes, Cyp76b6, 10hgo* and *Mtc*, and the selectable marker gene was *npt*II. (+) and (-) indicate the presence or absence of a transcript, respectively **(cont.)**

Sample ID	*pineGpps*	*VoGes*	*Cyp76b6*	*10hgo*	*Mtc*
24	-	-	-	+	-
25	+	-	-	+	-
26	-	-	-	-	+
27	-	-	-	-	+
28	+	-	+	+	+
29	-	-	-	-	+
30	-	-	+	-	+
31	+	-	-	-	-
32	-	-	-	-	-
33		-	-	+	+

Only the *Cyp76a26* and *Dlgt* transcripts from the Downstream combination of genes could be detected by mRNA blot analysis (Figure 5). In order to confirm the absence of the *Lamt* and *Sls* transcripts, both were amplified by RT-PCR. The *Lamt* transcript was detected in three transgenic plants and the *Sls* transcript was detected in one, although the low expression levels made both transcripts difficult to detect (Figure 6). The Downstream transformation experiments yielded a combinatorial population as summarized in Table 6. Two of the 12 regenerated plants expressed only the *npt*II selectable marker because no other transcripts were detected.

All of the plants regenerated following transformation with the Upstream 2 combination of transgenes expressed only the *npt*II selectable marker because no other transcripts were detected.

Table 6. Expression results from the Downstream transformation experiment. The input genes were *Cyp76a26*, *Dlgt*, *Lamt*, *Sls* and the selectable marker gene was *npt*II. (+) and (-) indicate the presence or absence of a transcript, respectively.

Sample ID	*Cyp76a26*	*Dlgt*	*Lamt*	*Sls*
1	-	+	-	-
2	+	+	-	-
3	+	+	-	-
4	-	-	-	-
5	-	+	-	+
6	-	-	-	-
7	+	+	+	-
8	+	-	-	-
9	-	+	+	-
10	-	-	-	-
11	-	-	+	-
12	-	-	-	-

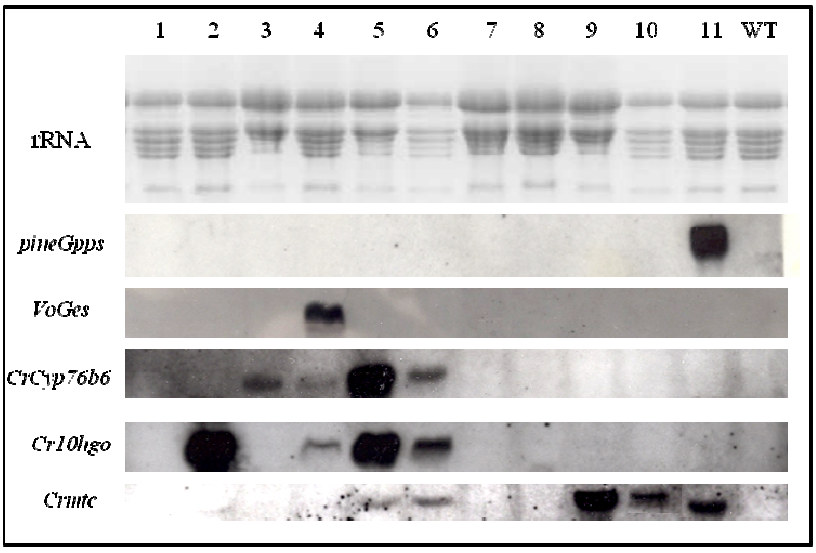

Figure 4. Gel blot analysis of 30 µg total leaf RNA from the Upstream 1 transformation experiment probed with *pineGpps*, *voGes*, *CrCyp76b6*, *Cr10hgo* and *Crmtc*. Lanes 1–11 are representative transgenic events; lane WT is the wild-type negative control.

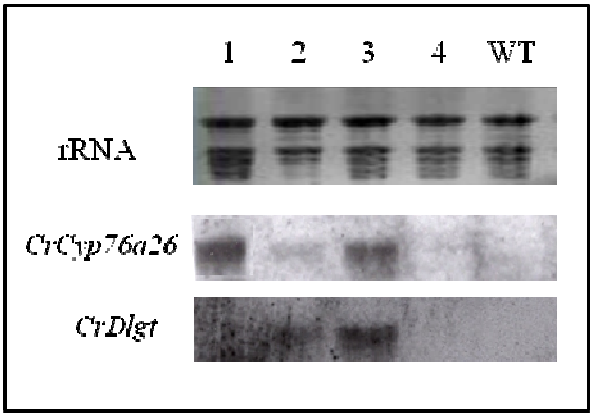

Figure 5. Gel blot analysis of 30 µg total leaf RNA from the Downstream transformation experiment probed with *CrCyp76a26* and *CrDlgt.* Lanes 1–4 are representative transgenic events; lane WT is the wild-type negative control.

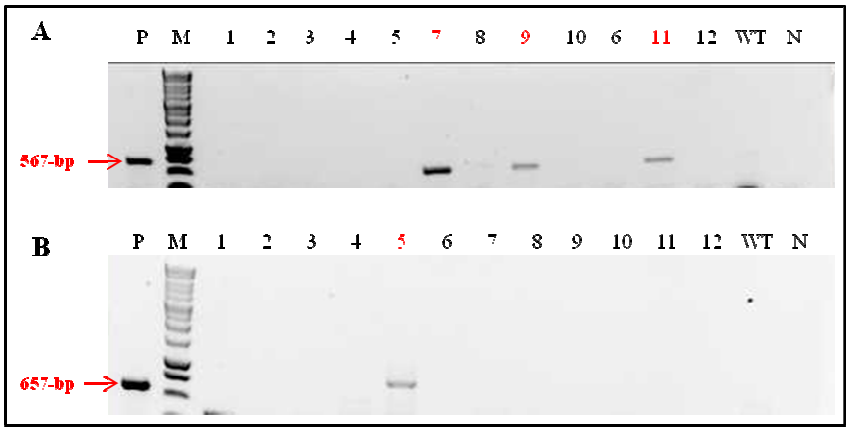

Figure 6. RT-PCR amplification of leaf cDNA from putative transgenic tobacco plants expressing the Downstream combination of genes (A) *Lamt* and (B) *Sls.* Lane P: positive control plasmid pRT101-*Lamt* and pRT101-*Sls* Lane M: DNA molecular weight marker (1-kb ladder). Lanes 1–12: cDNA (2ng) from putative transgenic plants. Lane WT: negative control wild-type cDNA. Lane N: negative control no cDNA template. Positive events are identified by the presence of the expected 567-bp and 657-bp fragments.

6.6. Discussion

Metabolic engineering in plants usually aims to generate novel compounds, improve the yields of extant desirable compounds and/or to reduce the amount of competing compounds in plants (Capell and Christou, 2004). Over the last few years, significant advances in metabolic engineering have been achieved by applying genomics and proteomics methods to characterize metabolic pathways and identify their components (Jacobs et al., 2000; Papin et al., 2003).

Gene transfer to plants provides an effective way to study and modify metabolic pathways precisely, and multigene engineering allows entire pathways to be reconstructed without endogenous regulation (Zhu et al., 2008). This in turn requires strategies to introduce multiple transgenes into plants and to ensure their coordinated expression over many generations (Capell and Christou, 2004) but this remains one of the most significant hurdles in plant molecular biology (Halpin 2005; Dafny et al., 2007). The probability that at least one transgene will fail to integrate (or will integrate but remain silent) increases with the number of genes introduced, such that large populations of transformants are required to ensure complete pathway reconstruction. Alternative approaches such as individual transformation followed by crossing to 'stack' the transgenes in one plant are unworkable for large numbers of transgenes because the development takes too long and the unlinked transgenes are likely to segregate in later generations. Zhu et al. (2008) addressed this challenge by developing a combinatorial transformation strategy in which multiple transgenes were introduced randomly into maize, resulting in a library of metabolic variants reflecting the expression of different combinations of transgenes. Some plants contained and expressed all the input genes and recapitulated the entire pathway under investigation, whereas others expressed subsets of transgenes and displayed corresponding metabolic profiles.

This chapter presents a similar strategy which was used to study the monoterpene secoiridoid pathway by expressing candidate genes encoding the relevant enzymes (Figure 7), since the pathway is not definitively understood (Loyola-Vargas et al., 2007). The monoterpene secoiridoid pathway starts with the condensation of IPP and DMAPP to form geraniol diphosphate, which is converted into secologanin via several steps, some of which are unknown.

One of the key challenges in metabolic engineering is that any targeted pathway must be understood in detail before interventions are made to avoid wasting resources on the development of futile transgenic lines. The longer and more complex the pathway, the more transgenic lines must be developed and tested independently before the most suitable intervention points are identified. Combinatorial transformation addresses this challenge and simultaneously changes the random nature of transgene integration during gene transfer from a drawback into an advantage. The approach is based on the creation of metabolic libraries comprising plants transformed with random selections of transgenes representing a given pathway (Zorrilla-López et al., 2013). The number of independent transgenic lines generated in a combinatorial transformation experiment depends on the expression level of the individual transgenes, particularly those acting at rate-limiting steps. The latter may depend on a number of factors, including the transgene insertion site, copy number, arrangement and integrity (Bock, 2013).

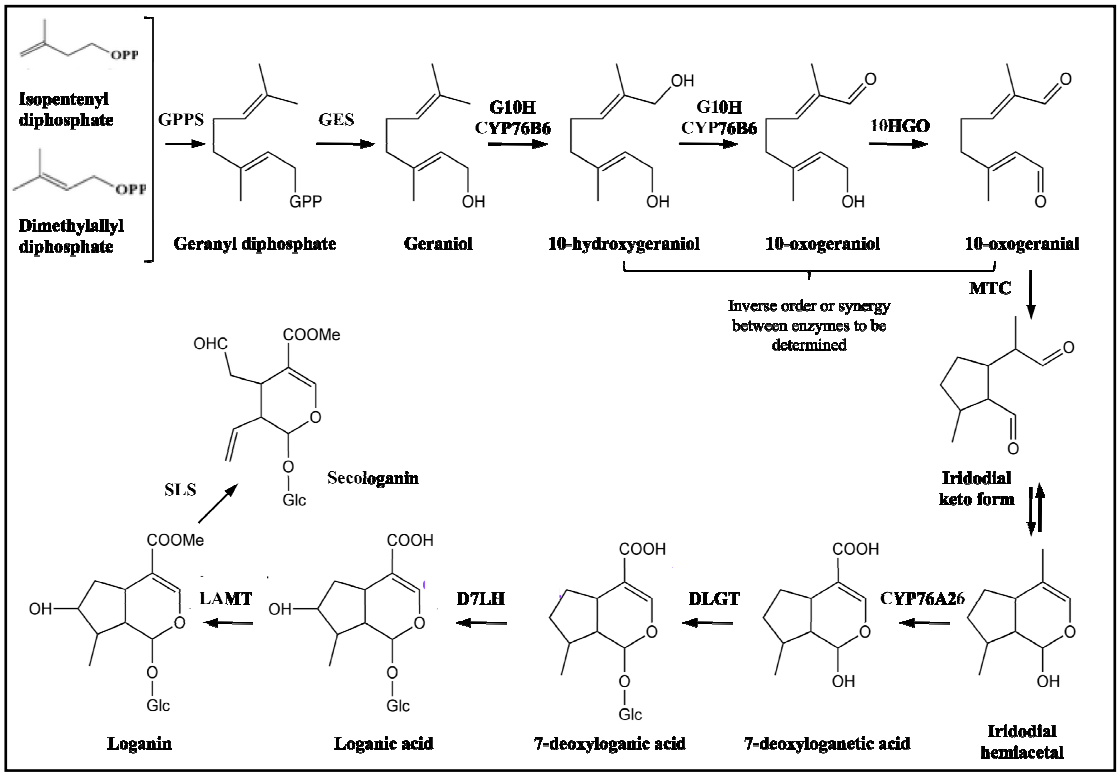

Figure 7. Putative monoterpene secoiridoid pathway leading to secologanin. GPPS: Geraniol diphosphate synthase, GES: Geraniol synthase: G10H: Geraniol 10-hydroxylase; CPR: Cytochrome P450 reductase; ADH: Alcohol dehydrogenase; MC: Monoterpene cyclase; DL7H: Deoxyloganin 7-hydroxylase; LAMT: SAM:loganic acid methyltransferease; SLS: Secologanin synthase.

The genes required to reconstruct the entire monoterpene secoiridoid pathway were split into Upstream and Downstream combinations to reduce the number of independent transgenic lines required to generate sufficient metabolic diversity and to facilitate the analysis of candidate genes proposed by our colleagues from CNRS (Strasburg, France) and Leiden University (Leiden, The Netherlands). The gene candidates *10-hgo, mtc* and *dlgt* were selected from an RNA-Seq dataset generated from the Illumina HiSeq2000-based RNA sequencing of *Catharanthus roseus* (www.cathacyc.org) in an attempt to fill the current gaps in the monoterpene secoiridoid biosynthesis pathway (Julsing et al., 2006).

The Upstream combination comprised the first five candidate genes in the pathway: *pineGpps, VoGes, CrCyp76b6, 10-hgo* and *mtc.* We regenerated 33 transgenic plants most of which expressed one (36.4%) or two transgenes (30.3%), but a small proportion of the plants expressed three (9%) and four (1%). None of them expressed all five transgenes. Wu et al. (2002) transformed rice with nine transgenes by particle bombardment and found that non-selected transgenes were present along with the selectable marker in approximately 70% of the plants, and that 56% carried seven or more genes. This percentage was much higher than expected based on independent integration frequencies, supporting a model suggesting that the integration of one transgene promotes the cointegration of more input DNA at the same locus (Kholi et al., 1998). All nine transgenes were expressed, and the expression of each gene was independent of the others (Wu et al., 2002; Zorrilla-López et al., 2013). The *Cyp76b6, 10-hgo* and *mtc* transgenes were the most frequently expressed (33.33%, 45.45% and 39.4%, respectively) whereas the *pineGpps* and *VoGes* transgenes were expressed with the lowest frequency (12.12% and 6.1%, respectively). These results suggest that transformants with high levels of GPPS and/or GES enzyme activity are selected against during the callus phase or during regeneration. Similarly, we were unable to recover transgenic plants from the Upstream 2 population expressing any gene other than the selectable marker. These results confirm the toxicity of *Gpps* and *Ges* as described in Chapter 2, Section 2.7.

We recovered 12 transgenic lines transformed with the Downstream combination of transgenes, but these plants were difficult to regenerate and displayed an atypical dwarf phenotype, with close knots and small leaves (Figure 4). Similar toxicity effects were reported when the biosynthesis pathway for the cyanogenic glucoside dhurrin

was reconstructed in *Arabidopsis thaliana, Lonicera japonicus, Nicotiana tabacum* and *Vitis vinifera* (Morant et al., 2007). The *A. thaliana* plants were stunted and showed signs of stress caused by the accumulation of toxic dhurrin intermediates and derivatives. It was impossible to regenerate transgenic *L. japonicus* plants because this species requires an extended callus phase (3-6 months), and unlike Arabidopsis it cannot detoxify the tyrosine-derived oxime by redirection into the glucosinolate pathway and also lacks the physiological machinery to handle the toxic compounds produced by the transgenes, so the tissues are probably exposed to cyanide intoxication (Morant et al., 2007). *N. tabacum* and *V. vinifera* were successfully transformed with the dhurrim pathway genes because these have shorter callus phases of 1 and 2-3 months, respectively (Morant et al., 2007).

It was also difficult to detect the *Lamt* and *Sls* transcripts by mRNA blot although this was possible using the more sensitive RT-PCR method (albeit with the risk of false positives) suggesting these two genes are expressed at minimal levels. Morant et al. (2007) produced 35 transgenic *N. tabacum* plants, but the transcript could only be detected in 10, resulting in significantly lower enzyme activity compared to *A. thaliana*. Likewise, only 2 out of 19 transgenic *V. vinifera* plants expressed the transgenes at detectable levels (Morant et al. 2007). The low gene expression could be due to transcriptional or posttranscriptional gene silencing (Morant et al., 2007). Transcriptional gene silencing (TGS) reduces transcriptional activity and is associated with homology in promoter regions. In contrast, posttranscriptional gene silencing (PTGS) involves sequence-specific degradation of the transcribed mRNA and is correlated with sequence homology in coding regions. All transgenes were controlled by the CaMV 35S promoter and there was no sequence homology in the coding regions suggesting that TGS is more likely to be the mechanism responsible for the low abundance of transcripts (Morant et al., 2007). Contrary to the prevailing view that the repetitious use of the same promoter may promote TGS, a number of transgenic plants have been generated containing five or more transgenes controlled by the same promoter without untoward effects (Naqvi et al., 2009). This study suggested that the TGS is not the reason of the low level of *Lamt* and *Sls* transcripts.

The *Cyp76a26* and *Dlgt* transgenes were expressed at higher frequencies (33.33% and 50%, respectively) compared to *Lamt* and *Sls* (25% and 8.3%, based on RT-PCR). Taken together with the abnormal phenotype of the *Sls* transgenic plants, these data

suggest that TGS may be caused by feedback inhibition induced by the natural toxicity of the new metabolic intermediates rather than the repetitious use of the same promoter (Naqvi et al., 2009b).

6.7. Conclusions

A combinatorial nuclear transformation approach developed by Zhu et al. (2008) was used to reconstruct the putative monoterpene secoiridoid biosynthetic pathway in tobacco (*Nicotiana tabacum*) plants in order to identify bottlenecks and gain insight into the pathway at the enzymatic, metabolic and regulatory levels. The inability to regenerate transgenic plants expressing all transgenes in the Upstream and Downstream combinations prompted us to investigate potential metabolomic perturbations resulting from the introduction of this heterologous pathway.

The transformation experiments with *CrGes* indicated the significant toxicity of this gene product, preventing the regeneration of transgenic plants. The Downstream combination of genes also had a negative impact, inducing a dwarf phenotype which may be associated with the limited accumulation of the *Lamt* and *Sls* transcripts.

The results presented in this chapter provided insight into the biosynthesis and regulation of the monoteprene secoiridoid pathway. However, further experiments based on targeted and non-targeted metabolomics are required to understand the perturbations in the metabolome resulting from the creation of this heterologous pathway in tobacco plants and to functionally characterize the transgenes and their products. Further experiments are also required to generate a larger population of transgenic plants for each combination of genes for further screening and evaluation. This will allow us to identify combinations of genes that do not induce negative effects during regeneration.

6.8. References

Agrawal P. K., Kohli A., Twyman R. M., Christou P. (2005) Transformation of plants with multiple cassettes generates simple transgene integration patterns and high expression levels. *Molelcular Breeding*, 16: 247–260.

Alpeter F., Baisakh N., Bock R., Capell T., Christou P., Daniell H., Datta K., Datta S., Dix P.J., Fauquet C-, huang N., Kholi A., Mooibroek H., Nicholson L., Nguyen T.T., Nugent G., Raemakers K., Romano A., Somers D.A., Stoger E., Taylor N., Visser R. (2005) Particle bombardment and the genetic enhancement of crops: myths and realities. *Molecular Breeding*, 15: 305-327.

Bock R. (2013) Strategies for metabolic pathway engineering with multiple transgenes. *Plant Mol Biol, in press.* DOI: 10.1007/s11103-013-0045-0

Capell T., Christou P. (2004) Progress in plant metabolic engineering. *Current Opinion Biotechnolgy*, 16: 167-173.

Cathacyc (2013) CathaCyc, a Metabolic Pathway Database Built from Catharanthus roseus RNA-Seq Data, accessed online 20 March 2013, http://www.cathacyc.org.

Dafny-Yelin M, Tzfira T (2007) Delivery of multiple transgenes to plant cells. *Plant Physiology*, 145:1118–1128.

Fischer M.J.C., Meyer S., Claudel P., Perrin M., Ginglinger J.F., Gertz C., Masson J.E., Werck-Reinhart D., Hugueney P., Karst F. (2013) Specificity of *Ocimum basilicum* geraniol synthase modified by its expression in different heterologous systems. *Journal of Biotechnology*, 163: 24-29.

Gong Y-F., Liao Z.-H., Pi. Y., Guo B-H., Miao Z-Q., Sun X-F., Tang K-Y. (2005) Engineering terpenoid indole alkaloids biosynthetic pathway in catharnathus roseus hairy root cultures by overexpressing the geraniol 10-hydroxylase gene. *Journal of Shanghai Jiatong University (Science)*, 10: 13-25.

Hallahan D.L., West J.M., Wallsgrove R.M., Smiley D.W.M., Dawson G.W., Pickett J.A., Hamilton J.G.C. (1995) *Archives of Biochemistry and Biophysics*, 318: 105-112.

Halpin C. (2005) Gene stacking in transgenic plants – the Challenger for 21st Century plant biotechnology. *Plant Biotechnology Journal*, 3:141-155.

Ikeda H., Esaki N., Nakai S., Hashimoto K., Uesato S., Soda S., Fujita T. (1991) Acyclic monoterpene primary alcohol:NADP+ oxidoreductase of Rauwolfia serpentina cells: the key enzyme in biosynthesis of monoterpene alcohols. *Journal of Biochemistry*, 109: 341-347.

Irmler S., Schröder G., St-Pierre B., Crouch N.P., Hotze M., Schmidt J., Strack D., Matern U., Schröder J. (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *The Plant Journal*, 24: 797-804.

Jacobs D.L., van der Heijden R., Verpoorte R. (2000) Proteomics in plant biotechnology and secondary metabolism research. *Phytochemical analysis*, 11. 277-287.

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

Kohli A., Christou P. (2008) Stable transgene bear fruit. *Nature Biotechnology*, 26: 653-654.

Kohli A., Leech M., Vain P., Laurie D.A., Christou P. (1998) Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot-spots. *Proc. Natl. Acad. Sci. USA*, 95: 7203–7208.

Lin L., Liu Y.G., Xu X., Li B. (2003) Efficient linking and transfer of multiple genes by a multigene assembly and transformation vector system. *Proc. Natl. Acad. Sci. USA*, 100: 5962–5967.

Loyola-Vargas V.M., Galaz-Ávalos R.M., Kú-cauich R. (2007) *Catharanthus* biosynthetic enzymes: the road ahead. *Phytochemisrtry Reveview*, 6: 307-339.

Morant A.V., Jorgensen K., Jorgensen B., Dam W., Olsen C. E., Moller B. L., Bak S. (2007) Lesson learned from metabolic engineering of cyanogenic glucosides. *Metabolomics*, 3:383-398.

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

Naqvi S., Farré G., Sanahuja G., Capell T., Zhu C., Christou P. (2009a) When more is better: multigene engineering in plants. *Trends in Plant Science*, 15: 48- 56.

Naqvi S., Zhu C., Farré G., Ramessar K., Bassie L., Breitenbach J., Perez Conesac D., Sandmann G., Capell T., Christou P. (2009b) Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proc Natl Acad Sci USA*, 106: 7762–7767.

Pan Q., Wang Q., Yuan F., Xing S., Zhao J., Choi Y.H., Verpoorte R., Tian Y., Wang G., Tang K (2012) Overexpression of ORCA3 and G10H in *Catharanthus roseus* plants regulated alkaloid biosynthesis and metabolism revealed by NMR-metabolomics. *PLoS ONE* 7, art. No. e43038.

Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A., Palsson, B. O. (2003) Metabolic pathways in the post-genome era. *Trends Biochem Sci*, 28: 250–258.

Sanford J., Klein J.C., Wolf T.M., Allen E.D. Nelson A. (1989) Delivery of substances into cells and tissues using a particle bombardment process. *Particulate Science and Technology*, 5: 27-37.

Twyman R.M., Christou P. (2004) Plant transformation technology - particle bombardment. In: Christou P, Klee H (eds) *Handbook of Plant Biotechnology*. John Wiley & Sons Inc., NY, pp 263-289.

Twyman R.M., Stoger E., Kohli A., Christou P. (2002) Foreign DNA: integration and expression in transgenic plnts. In: Setlow JK (ed) *Genetic Engineering: Principles and Practice, Volume 24*. Kluwer Academic/Plenum Publishers, NY, pp107-136.

Uesato S., Ikeda H., Fujita T., Inouye H., Zenk M. H. (1987) Elucidation of iridodial formation mechanism – partial purification and characterization of the novel monoterpene cyclase from *Rauwolfia serpentina* cell suspension cultures. *Tetrahedron Letters*, 38: 4431-4434.

Valletta A., Trainotti L., Santamaria A.R., Pasqua G. (2010) Cell-specific expression of tryptophan decarboxylase and 10-hydroxygeraniol oxidoreductase, key genes

involved in camptothecin biosynthesis in *Camtotheca acuminata* Decne (Nyssaceae). *BMC Plant Biology*, 10: 69-79.

Wang J., Liu Y., Cai Y., Zhang F., Xia G., Xiang F. (2010a) Cloning and functionl analysis of gernaiol 10-hydroxylase, a cytochrome P450 from *Swertia mussotii* Franch. *Biosci. Biotechnol. Biochem*, 74: 1583-1590.

Wang C-T., Liu H., Gao X-S., Zhang H-X. (2010b) Overexpression of *G10H* and *ORCA3* in the hairy roots of *Catharanthus roseus* improves catharanthine production. *Plant cell rep*, 29:887-894.

Wu L., Nandi S., Chen L., Rodriguez R.L., Huang N. (2002) Expression and inheritance of nine transgenes in rice. *Transgenic Research*, 11: 533-541.

Yamamoto H., Katano N., Ooi A., Inoue K. (2000) Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. *Phytochemistry*, 53: 7-12.

Zhu C., Naqvi S., Breitenbach J., Sandmann G., Christou P., Capell T. (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc Natl Acad Sci USA*, 105: 18232-18237.

Zorrilla-López U., Masip G., Arjó G., Bai C., Banakar R., Bassie L., Berman J., Farré G., Miralpeix B., Pérez-Massot E., Sabalza M., Sanahuja G., Vamvaka E., Twyman R.M., Christou P., Zhu C., Capell T. (2013) Engineering metabolic pathways in plants by multigene transformation. *The International Journal of developmental Biology. In press.*

CHAPTER 7

STRATEGIC PATENT ANALYSIS IN PLANT BIOTECHNOLOGY: TERPENOID METABOLIC ENGINEERING AS A CASE STUDY

CHAPTER 7: STRATEGIC PATENT ANALYSIS IN PLANT BIOTECHNOLOGY: TERPENOID METABOLIC ENGINEERING AS A CASE STUDY

7.1 Abstract

Intellectual property (IP) is the set of legal rights associated with intellectual activity and creativity, which is used to provide an economic incentive for inventors and thus promote innovation. Among the many different types of IP, the principal form that applies to scientific inventions is the patent. For an invention to be patentable it must be novel, non-obvious and commercially useful. The usefulness and obviousness of an invention is open to interpretation but novelty can be established by looking at related patents and the scientific literature. Scientists become adept at literature analysis and benefit from the availability of extensive literature databases such as PubMed and Web of Knowledge. In contrast, systematic patent analysis is often delegated to experts even though the major patent offices also have public-access databases that can be used in the same way as literature databases once an effective search strategy has been developed. The do-it-yourself patent search is a useful alternative to professional patent analysis particularly in the context of publicly-funded projects where funds for IP activities may be limited. As a case study, we analyzed patents related to the engineering of terpenoid indole alkaloid (TIA) metabolism in plants. We developed a focused search strategy to remove redundancy and reduce the workload without missing important and relevant patents. This resulted in the identification of approximately 50 key patents associated with TIA metabolic engineering in plants which could form the basis of a more detailed freedom-to-operate analysis. The structural elements of this search strategy could easily be transferred to other contexts making it a useful generic model for publicly-funded research projects.

7.2 Introduction

7.2.1 Intellectual property

Intellectual property rights (IPR) protect the creators of inventions, ideas and designs from exploitation without preventing the benefits filtering through to society (WIPO, 2008; Adcock, 2007). IP is therefore justified not only in terms of reward to the owner but also the broad societal gains (Adcock, 2007; Lei et al., 2009). IPR offers time-limited protection to the creators of intellectual property, aiming to promote creations that would probably not achieve commercial success without such protection (WIPO, 2008; Adcock, 2007).

The two main categories of IPR are “copyright and related rights items” and “industrial property items” (WIPO, 2008; Kannan, 2010). Copyright applies mainly to literary, artistic and musical works but may also apply to technical and architectural drawings, whereas industrial properties include inventions, trademarks and designs (Kannan, 2010). An invention has been defined as a process or a product that provides a new or better way of doing something or that provides a new solution to a problem. Inventions are protected by patents, which are defined by the World Intellectual Property Organization (WIPO) as “an exclusive right granted by the state for an invention that is new, involves an inventive step and is capable of industrial application” (WIPO, 2006). These requirements are often described as novelty, non-obviousness and utility, and they must be disclosed in a clear and complete manner in the patent application (Fleck et al., 2003; WIPO, 2006).

Once the patent is granted, other parties are prevented from benefiting from the invention by making, using, offering for sale or importing a patented product or process without permission from the inventor (WIPO, 2006). The patent is granted by national or regional patent offices for a limited time, generally 20 years from the date of filling, provided the required maintenance fees are paid on time. The period of protection is offered in return for the disclosure of the invention and ensures that the public will be able to practice the invention once the patent expires (Fleck et al. 2003; WIPO, 2006). Patents ensure economic value is generated from research and development, but because the protection is geographically limited it is important to make separate applications at different offices to protect the invention in multiple territories (Fleck et al., 2003; WIPO, 2006; Adcock, 2007).

The patent application sets out the legal scope of the patent, the nature of the invention, details of the inventor, the patent owner and other legal information (WIPO, 2006). It comprises a request, a description, claims, drawings (if necessary) and an abstract. The structure of the application is similar worldwide. The description must contain sufficient detail to be reconstructed without further inventive effort, and the claims must set out the limits of the invention and therefore the scope of the required protection (WIPO, 2006).

7.2.2. The prior art

When considering a patent application, the applicant must determine whether the invention meets the requirements of novelty, non-obviousness and utility. The definitions of non-obviousness and utility are subjective and can differ across jurisdictions, but novelty can be established by looking at two bodies of literature collectively termed the prior art. The first is the scientific literature, which is publically accessible and shows if previous inventions have been publically disclosed and are therefore considered to be in the public domain. Once disclosed in such a manner, an invention can no longer be patented. The second is the huge body of existing patents.

Most scientists become adept at searching the scientific literature, especially given the ubiquitous access to literature databases such as PubMed and Web of Knowledge. A researcher can therefore rapidly establish whether or not an invention is already in the public domain. However, because there are millions of patents both in force and expired, a search of the IPR prior art is time-consuming and complex, particularly when it comes to inventions that require intricate knowledge of both scientific concepts and patent law. Many large organizations employ specialist patent lawyers to oversee this process, but the expense involved is beyond the scope of small, publicly-funded research projects. An alternative approach is to use online facilities for patent analysis, which are hosted by most national patent offices. The patent offices in the three most important territories (Europe, USA and Japan) host the largest and most comprehensive databases, and two further databases (WIPO and CAMBIA) have global coverage (Figures 1 and 2). Therefore it is possible to screen these databases during early-stage research projects with potential commercial applications to ensure

that specific technologies and products will not encroach on existing IPR and are sufficiently novel to generate so-called foreground IP, i.e. novel IP generated during an ongoing project, in contrast to existing background IP brought into the project by the collaborating researchers.

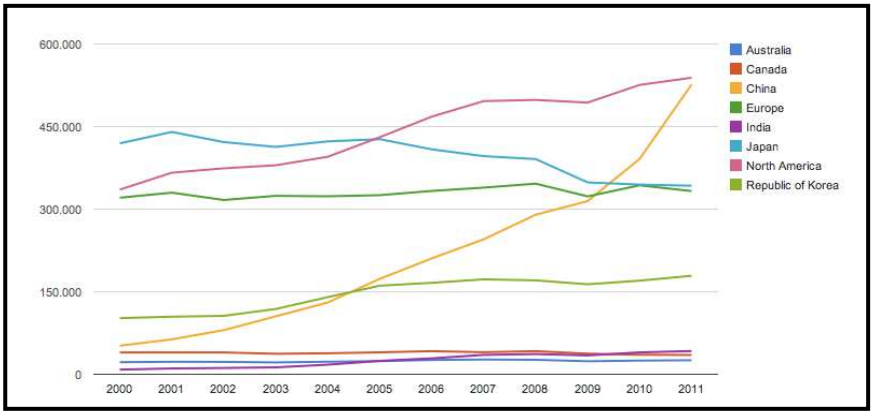

Figure 1. Total patent applications per year divided by national offices (WIPO, 2013).

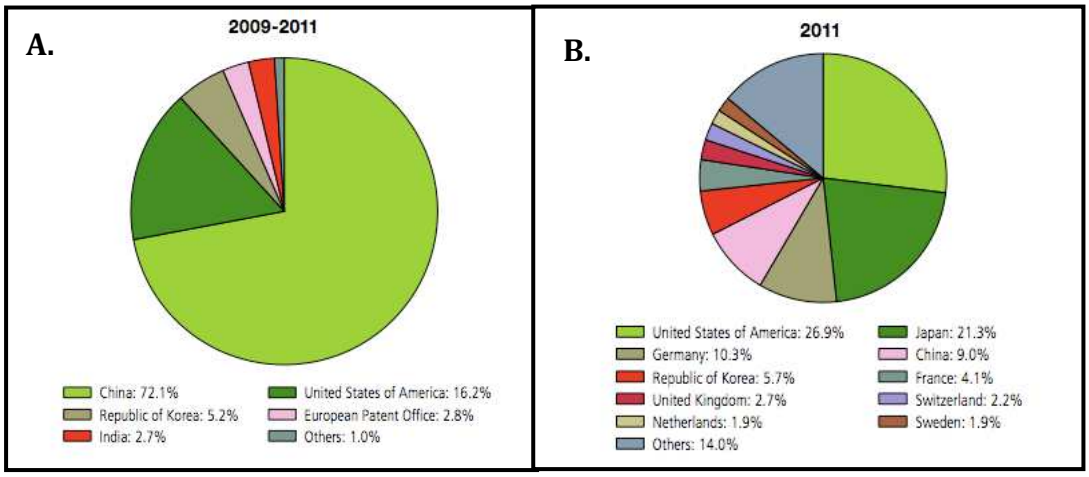

Figure 2 A. Contribution of national offices to the growth in patent applications worldwide to 2011 (WIPO, 2012) **B.** Country share in total patent applications by geographic regions in 2011 (WIPO, 2012).

7.3. Aims

The aims of the work described in this chapter were to develop a systematic strategy for the analysis of patents that relate to the metabolic engineering of the terpenoid indole alkaloid (TIA) pathway in plants in the context of the EU-funded project SmartCell (Seventh Framework Programme KBBE 222716).

7.4 Materials and methods

7.4.1 Data sources

We focused on the five major publicly-available patent databases offering freely-accessible search facilities over the internet.

The European Patent Organization is an intergovernmental organization established in 1973 on the basis of the European Patent Convention (EPO, 2013a). It has two bodies, the European Patent Office (EPO) and the Administrative Council, which supervises its activities (EPO, 2013a). Currently, 38 European countries (including all EU Member States) are members of the European Patent Convention (EPC). The EPC allows applicants to acquire patents that can be enforced in as many Member States as the applicant wishes, via a single application (Adcock, 2007). The EPO patent database (Espacenet) offers free access to more than 70 million patent documents published globally since 1836 (EPO, 2013b) and also provides access to patent family information (similar patents in other countries), legal status reports, other non-patent literature and links to the European Patent Register.

The United Sates Patent and Trademark Office (USPTO) is the federal agency responsible for granting US patents and registering trademarks. It has nine offices with different functions, the most relevant of which is the Office of the Commissioner for Patents (USPTO, 2013a). The USPTO website provides a range of resources including a patent full text and image database (PatFT), a patent application full text and image database (AppFT), patent application information retrieval (PAIR), a public search facility, and patent and trademark resource centers (USPTO, 2013b).

The Japan Patent Office (JPO) is ranked third in the world in terms of the proportion of applications received (WIPO, 2012). The JPO website contains the Industrial

Property Digital Library (IPDL) which provides public access to databases of patents and utility models, designs, or trademarks and figures (JPO, 2013).

The World Intellectual Property Organization (WIPO) is a United Nations specialized agency and the leading intergovernmental organization dedicated to the promotion and use of IP to stimulate innovation and creativity (WIPO, 2013a). The WIPO web provides a database of 18,649,424 patent documents including 2,171,684 applications representing all WIPO offices (WIPO, 2013b).

Finally, CAMBIA is an independent non-profit organization focusing on IP and information technology, which provides comprehensive patent databases as well as freedom to innovate and freedom to cooperate analysis (CAMBIA, 2013a,b).

7.4.2 General search methodology

Each database features a unique search query format so different search strings are required to access the information in each case, and the scope of that information is sometimes restricted by the permitted search terms.

The EPO patent database (Espacenet) offers smart search, advanced search and classification search options. Smart search allows the use of single or multiple words as search terms, or more complex Boolean search commands focusing on specific fields in the patent document such as title, abstract, year, inventor or patent number (Espacenet, 2013a). Advanced search can be used to combine search terms, e.g. patents from a particular year and country that have specific words in their title or abstract (Figure 3), and can restrict the search to different collections of publications such as worldwide (all patents from more than 90 countries), EP (full text of EU patents) and WIPO, a complete collection including the full text of published patents and applications (Espacenet 2013b). A group of patent publications in a particular technical area can be searched using the classification search, which is a powerful tool used by patent professionals (Espacenet, 2013a).

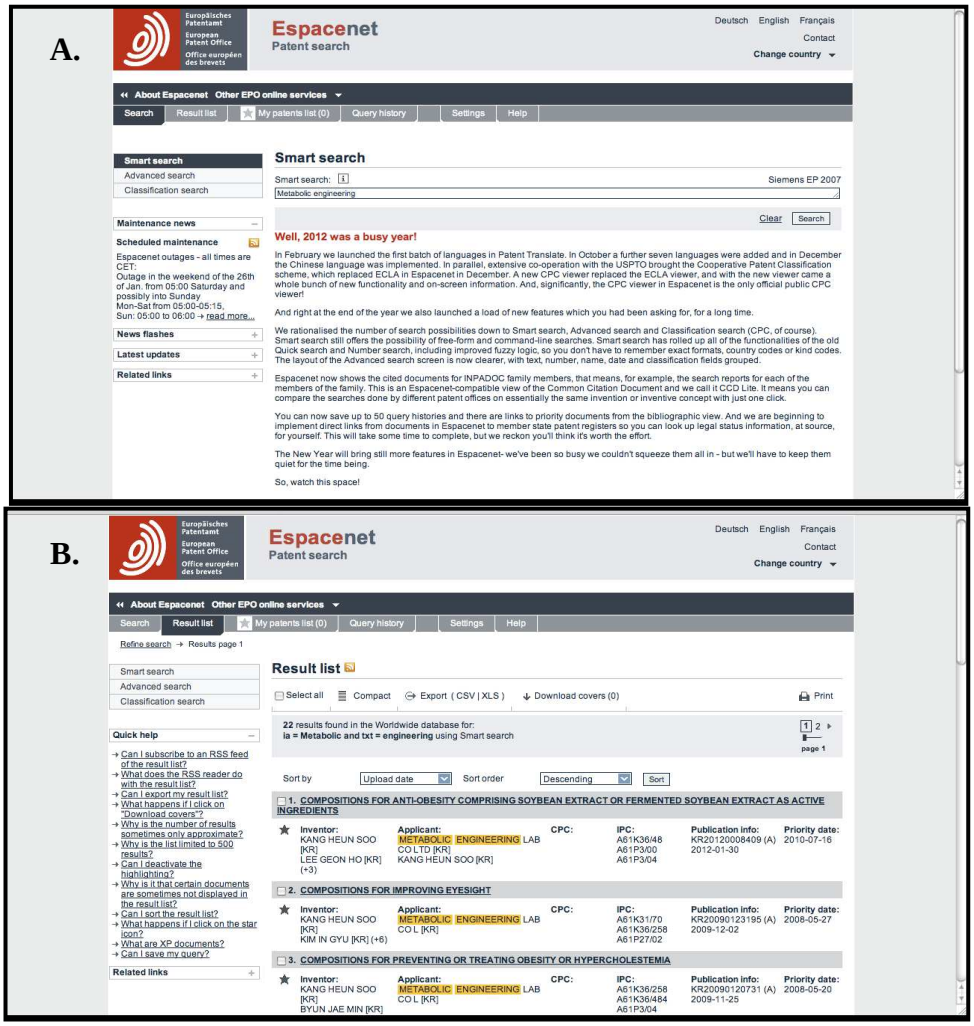

Figure 3. An 'advanced search' in the EPO patent database. A. 'Advanced search' using the term 'metabolic engineering'. B. Search results.

The USPTO patent full text and image database (PatFT) contains the full text of patents issued since 1976 and allows a simplified quick search, a customized advanced search and a patent number search option allowing individual patents to be retrieved (USPTO, 2013b). The quick search allows one or two terms to be used to search all fields or specific fields such as the abstract, title, patent number, assignee or inventor. More terms can be added with the advanced search option which also allows complex Boolean search strings (Figure 4).

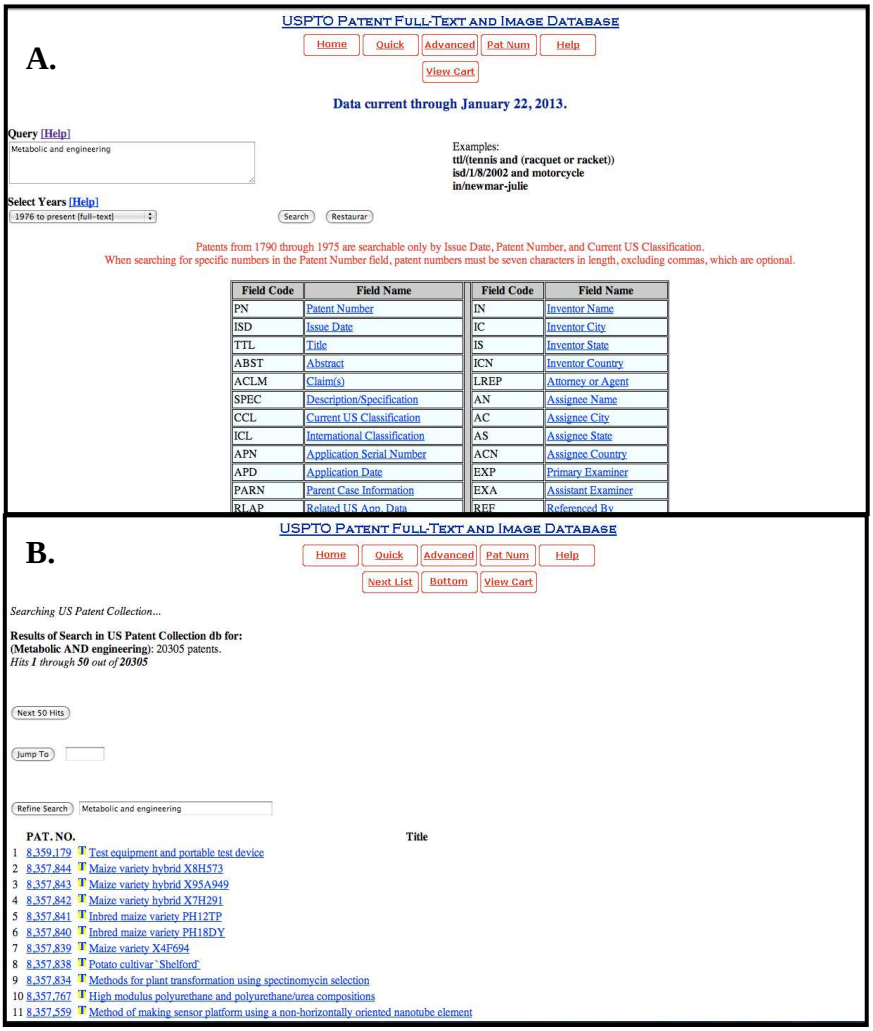

Figure 4. An 'advanced search' in the USPTO patent database. A. 'Advanced search' using the term 'metabolic engineering'. B. Search results.

The Japan Patent Office (JPO) search facility is more complex and more restricted than the equivalent EPO and USPTO resources. The "patent and utility model gazette database" and "patent and utility model concordance" facilities allow patents to be retrieved by code or application number. The FI/F-term search allows retrieval by F-term, a system for classifying patent documents based on the technical aspects, which overlaps with the IPC system overseen by WIPO. The most useful facility is the Patent Abstract of Japan search, where combinations of keywords can be used linked by Boolean operators, but this is limited to three terms and only searches the title, inventors and abstract (Figure 5). There currently appears to be no publicly-accessible system to carry out full-text searches of Japanese patent documents (IPDL, 2013).

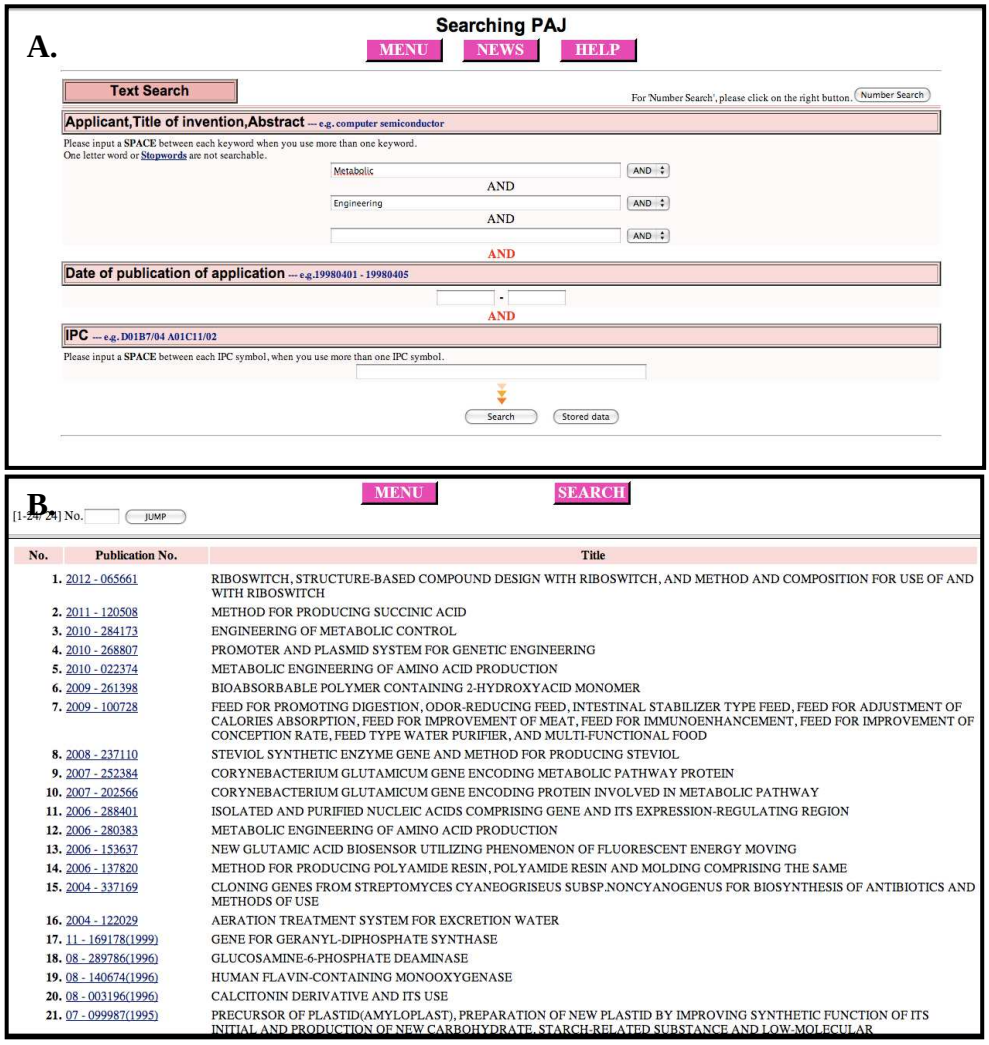

Figure 5. An 'advanced search' in the JPO patent database. A. 'Advanced search' using the term 'metabolic engineering'. B. Search results.

The WIPO database of 18,649,424 patent documents including 2,171,684 applications can be searched using Boolean operators by field, including the front page and full text. Although the WIPO databases can be searched via the EPO facility discussed above, the WIPO portal also allows searches to be restricted by individual WIPO offices as shown in Figure 6 (WIPO, 2013).

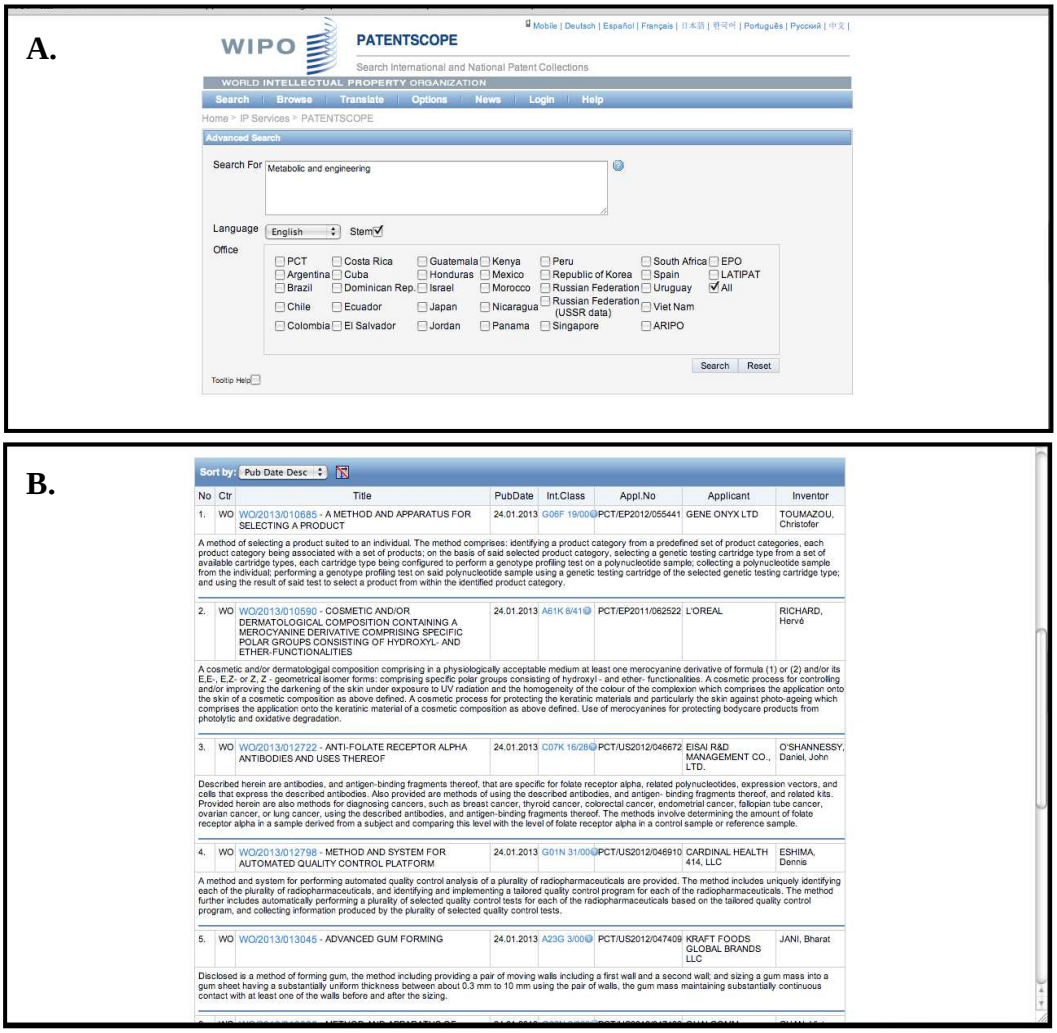

Figure 6. An 'advanced search' in the WIPO patent database. A. 'Advanced search' using the term 'metabolic engineering'. B. Search results.

The CAMBIA website provides a quick search option which allows keyword searches on the full text or one of the front pages sections (title, abstract, inventor and applicant) using the appropriate syntax (Figure 7) whereas the more complex 'structured search' allows the use of comprehensive Boolean search expressions over a larger range of fields. Finally, 'expert search' allows the full search query to be entered manually – this is more difficult to master but is a powerful way to carry out complex searches (Patentlens, 2013).

Figure 5. An 'advanced search' in the CAMBIA patent database. A. 'Advanced search' using the term 'metabolic engineering'. B. Search results.

7.5 Results

7.5.1 Development of a focused search strategy

Our task was to identify patents that might interfere with the development of a commercial process based on the engineering of TIA metabolism in plants. We therefore began by defining search terms related to the invention that would be sure to capture related patents. These terms fell into several categories (e.g. product related, process related and technology related). We initially carried out searches using broad

terms such as ‘metabolic engineering’ or ‘secondary product’ that were designed to capture many patents albeit at the expense of a large number of false positives. This was necessary to define the metabolic engineering patent space we wanted to focus on, and to ensure that key patents were not excluded in the first round of screening. For example, searching the USPTO database with ‘metabolic engineering’ resulted in more than 14,000 hits. We then intended to pair these broad terms with more specific tangible terms related to the TIA pathway using Boolean operators, e.g. ‘metabolic engineering AND strictosidine’. We drew up a list of genes/enzymes, intermediates and products relevant to the pathway and paired these with the four most fruitful general terms: ‘metabolic engineering’, ‘natural products’, ‘secondary metabolism’ and ‘secondary products’ (Table 1).

Table 1. List of terms to define the patent space “engineering the TIA metabolic pathway in plants”.

General concepts		
Cell factory Metabolic engineering Metabolic pathway	Metabolic pathway engineering Metabolic pathway extension Metabolic pathway regulation Metabolic regulation	Secondary metabolism Secondary product(s)
Tangibles (general terms)		
Alkaloid Binary vector Diterpenoid Indole alkaloid	MEP pathway Mevalonate pathway Monoterpenoid P450	Sesquiterpenoid Shikimate-pathway Terpenoid Transporters
Tangibles (genes/enzymes – primary)		
CPR – cytochrome P450 reductase G10H – geraniol 10-hydroxylase SLS – secologanin synthase AS (ASA2, ASA) – anthranilate synthase TDC – tryptophan decarboxylase 10-HGO – 10-hydroxygeraniol oxidoreductase DXPS – 1-deoxy-D-xylulose-5-phosphate synthase	STR1 – Strictosidine synthase SGD – strictosidine glycosidase T16H – tabersonine 16-hydroxylase D4H – desacetoxyvindoline 4-hydroxylase DAT – deacetylvindoline acetyltransferase DXPR – 1-deoxy-D-xylulose-5-phosphate reductoisomerase MAT – minovincinine-19-hydroxy-o-acetyltransferase	
Tangibles (genes/enzymes – secondary)		
Tropane/nicotine alkaloid pathway Isoquinoline alkaloid pathway	Monoterpenoid pathway Diterpenoid pathway	Sesquiterpenoid pathway Polyamine pathway

Table 1. List of terms to define the patent space "engineering the TIA metabolic pathway in plants"**(cont.)**

Tangibles (products/intermediates – primary)		
Geraniol 10-Hydroxygeraniol Ajmalicine Camtothecin	Catharanthin Secologanin Serpentine Strictosidine	Tryptamine Tryptophan Vinblastine Vincristine Vindoline
Tangibles (products/intermediates – secondary)		
Anabasine Anatabine Artemisinin Atropine Azadirachtin Bipyridine Cadaverine	Camtothecin Docetaxel Hyoscyamine Myosmine Nornicotine Paclitaxel	Shikonin Scopolamine Taxol Taxotere Tetrahydrocannabinol Tropinone
Tangibles (products/intermediates – tertiary)		
Berberine Buprerorphine Camptothecin Codeine Dephinidin Digoxin Dopamine Elaterin Geosmin Irinotecan	Limonene Macarpine Mandragorin Menthofuran Menthol Menthone Morphine Noscapine Oxycodone Papaverine Podophyllin	Podophyllotoxin Pulegone Reticuline Sanguinavine Sanquinarine Teniposide Tetrodotoxin Toposide Topotecan Tubocurarine
Tangibles (other resources)		
Promoters Terminators selectable and screenable marker genes Other functional DNA sequences		Plasmids and other vectors DNA libraries Databases Others..
Processes and methods		
Plant transformation Gene transfer to plants Genetic transformation (+plants) Genetic engineering (+plants) Transgenic plants	Transformation system (+plants) Transformation strategy (+plants) Agrobacterium Particle bombardment Biolistic	Microprojectile Gene gun Direct DNA transfer (+plants) Selection+regeneration+plants Cotransformation Multigene transformation

We found that this initial strategy required many repetitive searches that often yielded overlapping results because many patents discussing metabolic engineering list multiple genes, enzymes intermediates and products even if these are not directly related to the invention, and also because many patents contain more than one of the general terms we used to define the metabolic engineering patent space. We therefore developed a smart search strategy which used Boolean operators and wildcard characters to define an ideal term to capture the patent space we wanted to search.

The second round of searches was carried out by combining two different strategies, first by evaluating the cost of restricting the search to the title/abstract or title compared to the full text, and the second by refining the scope of the search with additional terms. To exclude metabolic engineering in microbes we added terms to ensure that the patent must contain the word plant or a derivative, and we also added terms identifying specific medicinal and model plants (e.g. tobacco, Arabidopsis, Catharanthus) linked by the Boolean operator OR to avoid ambiguity with the word plant, e.g. as a synonym for factory or heavy machinery.

We found the title/abstract search was too restrictive and generated too many false negatives. For example, searching USPTO with “P450 AND metabolic engineering AND plant AND (tobacco OR Arabidopsis OR Catharanthus)” generated 226 patents if a full-text search was used but only nine if this was restricted to the title and abstract, and only three for the title alone. The general terms for tangibles, such as alkaloid and terpenoid, were also too broad and retrieved too many patents in full text searches. However, in many cases restriction to title/abstract/claims reduced the number of hits to zero or near zero (Table 2).

Table 2. Number of hits achieved by restricting searches to title/abstract/claims in the principal patent databases

Term	EPO			USPTO		CAMBIA	
	EP	**World**	**WIPO**	**Abstract**	**Title**	**Abstract**	**Title**
General concept							
Metabolic pathway regulation	1	11	18	0	0	153	5
Metabolic regulation	18	354	215	6	0	1478	252
Secondary metabolism	2	179	28	0	3	201	23
Secondary product	603	15916	779	65	9	8616	580
Cell factory	12	246	11	33	2	123	7
Metabolic engineering	17	262	52	17	15	259	176
Metabolic pathway	44	826	194	116	16	1200	177
Metabolic pathway engineering	1	51	14	2	1	92	9
Metabolic pathway extension	0	1	0	0	0	14	0
Tangibles (general)							
P450	167	1599	363	199	105	968	637
Sesquiterpenoid	4	37	6	5	1	29	21
Shikimate pathway	0	19	8	5	1	25	6
Terpenoid	27	543	59	88	21	290	117
Transporters	181	2275	404	313	121	4	1
Alkaloid	131	2948	220	241	71	1314	702
Binary vector	79	1359	86	25	3	722	54
Diterpenid	7	138	8	15	15	62	55
Indole alkaloid	2	178	6	10	9	90	33
MEP pathway	1	8	5	4	1	18	6
Mevalonate pathway	10	115	27	21	2	123	47
Monoterpenoid	3	58	6	15	1	41	5
Tangibles (genes/enzymes)							
CPR	14	111	27	9	3	76	33
AS(ASA2, ASA)	7	37	5	10	6	33	22
TDC	4	17	3	240	9	14	6
DXPS	0	5	3	2	0	7	9
DXPR	0	1	0	0	0	8	12
Tangibles (product/intermediates)							
Strictosidine	1	4	1	0	0	2	2
Tryptamine	22	373	30	35	16	109	74
Tryptophan	173	3240	296	464	109	1473	388
Vinblastine	26	419	36	64	22	146	41
Vincristine	18	387	48	49	13	173	24
Vindoline	4	44	4	13	0	25	0
10-Hydroxygeraniol	0	0	0	0	0	0	0

Table 2. Number of hits achieved by restricting searches to title/abstract/claims in the principal patent databases **(cont.)**

Term	EPO			USPTO		CAMBIA	
	EP	**World**	**WIPO**	**Abstract**	**Title**	**Abstract**	**Title**
Ajmalicine	0	15	0	2	1	3	1
Camtothecin	4	9	1	1	0	1	2
Catharanthin	0	1	0	0	0	8	3
Secologanin	1	5	0	0	0	1	0
Serpentine	413	7540	523	2160	245	4819	491
Process terms							
Plant transformation	134	2306	344	52	35	5926	483
Gene transfer to plant	11	48	37	0	0	248	29
Genetic transformation (+plants)	11	539	59	1	0	969	40
Genetic engineering (+plants)	4	353	37	9	5	438	46
Transgenic plants	352	7760	1536	1317	291	7946	1816
Transformation system (+plants)	6	169	49	1	0	477	13
Transformation strategy (+plants)	4	3	0	0	0	5	0
Agrobacterium	123	1837	225	240	54	780	271
Particle bombardment	30	377	24	58	6	537	27
Biolistic	4	38	8	13	5	40	22
Microprojectile	8	71	14	28	5	97	24
Gene gun	3	112	12	3	2	26	9
Direct DNA transfer (+plants)	2	12	2	1	0	40	0
Selection + regeneration + plants	2	84	29	0	0	395	1
Cotrasnformation	31	13	10	5	0	48	0
Multigene transformation	11	0	2	0	0	8	0

7.5.2 Development of an ideal search string

To address these challenges and develop a non-redundant screening strategy that captured all relevant patents with few false positives, we set out to develop an 'ideal' search string, including all four general terms, the restriction to model and medicinal plants and individual tangibles such as genes/enzymes, intermediates and products. We tested these strings using broader tangibles such as alkaloid and terpenoid to assess the impact of different strings on the recovery of relevant and irrelevant patents. Four terms in addition to the tangible were found to reduce the number of false positives without eliminating relevant patents, and the addition of more search terms did not improve the results any further. We found that 'plants' improved the relevance of the hits and that adding various species names increased the number of hits marginally, suggesting that patent authors occasionally use plant species names without stating they are plants. The progressive focusing of the search results using this approach is summarized in Table 2.

In Boolean terms, the ideal search string was:

"X AND metabolic engineering AND (natural product* OR secondary) AND (plant* OR Y)"

Where X is the tangible primary term, Y is a list of plant species separated by OR operations, and * is a wildcard character that can replace any number of actual characters. The Boolean phrase (*natural product* OR secondary*) therefore encompasses 'natural product(s)', 'secondary product(s)', 'secondary metabolite(s)' and 'secondary metabolism'.

The search was implemented differently in each database because the format for entering Boolean search terms is database-specific. For example, USPTO requires phrases to be placed inside quotes and does not accept wildcards in this context, so plural phrases were listed separately from the singular phrase. This could be

accommodated because up to 20 search terms can be used in one string, so the text of the corresponding search string was:

X$ AND "metabolic engineering" AND ("natural products" OR "natural AND product" OR secondary) AND (plant$ OR tobacco OR Catharanthus OR Arabidopsis OR Y)

Where X is the tangible primary term, Y is an optional list of additional plant species and $ is the USPTO wildcard symbol.

EPO only allows 11 search terms but accepts wildcards within quotes, hence the corresponding search string was:

X* AND 'metabolic engineering' AND (secondary OR 'natural product*') AND (plant* OR tobacco OR Catharanthus OR Arabidopsis OR Y)

CAMBIA has very similar rules to EPO, and JPO only allows three terms entered in separate boxes for their title/abstract searches.

7.5.3 Quality control

Initial unrestricted searches yielded many false positives in all the databases but restriction to the title and abstract removed many relevant patents. The aim of the refinements discussed above was to reduce the false positive rate without losing relevant patents. We therefore carried out an exhaustive one-by-one screen of the broad patent search term "terpenoid" to determine the number of relevant patents and then devised a quality control system to score the refined searches (Table 3).

Table 3. Quality control analysis of each search strategy using the tangible term 'terpenoid'.

Strategy	USPTO database				EPO database			
	H	**FP**	**FN**	**QS**	**H**	**FP**	**FN**	**QS**
Full text USPTO	2256	2197	0	**2.5%**	2055	2014	0	**2%**
Restrict to title/abstract	121	79	17	**7%**	92	51	12	**16%**
Full text restricted by refinement term "metabolic engineering"	114	55	0	**52%**	72	31	0	**56%**
Full text restricted by refinement terms "metabolic engineering" and "plants"	114	55	0	**52%**	70	29	0	**59%**
'Perfect text' string using "plant(s)"	73	14	0	**80%**	50	10	1	**79%**
'Perfect text' string using "plant(s)" and list of individual plant species	73	14	0	**80%**	51	10	0	**80%**

Number of relevant patents in USPTO = 59
Number of relevant patents in EPO = 41
H – Hits; **FP**- False Positive; **FN**- False Negative; **QS**- Quality Score.

The quality score was calculated by setting a score of 100% where all relevant patents were found and there were no false positives. The score was reduced proportionately for false negatives, as a percentage of relevant patents, and false positives, as a percentage of hits. Assuming that the number of relevant patents for a search term X is exactly 100, a search string that recovered all these patents and these alone would achieve a quality score of 100%. A search that recovered 90 relevant patents and no false positives would score 90%. And a search that recovered all 100 relevant patents, but also 50 irrelevant ones would score 66% (33% of the total are irrelevant). The penalties for false positives and false negatives are additive, so a search retrieving 95 of the relevant patents and 35 irrelevant ones would score 72% (100% – 5% – 23%). This approach attributes a greater penalty to a false negative than a false positive, reflecting the greater importance of missing a relevant patent than including an irrelevant one that can be manually discarded from the dataset at a later point. It is also possible for poorly-designed searches to achieve negative scores.

7.5.4 Evaluation of the search results

The focused search strategy described above was applied with a small number of key words representing genes/enzymes, products/intermediates and processes (Table 1). This reduced the volume of patents to a manageable number suitable for manual screening. The relevance of each patent was then evaluated by reading the title, the abstract and finally the claims. The title provided preliminary guidance on relevance but this still resulted in some borderline cases that needed to be evaluated in more detail to consider their impact on the project. The claims were the most important part of the patent document because they determine the scope of protection, and these were regarded as the 'gold standard' by which to judge the relevance of the patent. This analysis resulted in a core set of approximately 300 relevant patent documents, resolving to 49 when redundant members of the same patent family were eliminated (Table 4). These were added to an IP database developed as part of the project website. The IP database can be searched by patent number, title, filing date, assignee, inventor and other categories, including potential impact on commercial products arising from the project (Figures 8 and 9).

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants".

Patent number	Patent title	Assignee	Filing date	Main claims
A. Genes/Proteins				
CN101250543	Japan snakeroot strictosidine synthase gene and its coding protein and application	Univ Shanghai	02/06/2010	*Not available in English*
EP 0156267 B1	Method for the preparation of strictosidine	Gyogynöveny KUTATO INTEZET	15/03/1985	Method for the preparation of strictosidine
US 2007/0212745 A1	Beta-glucosidase and a process for extraction thereof	Council Of Scientific & Industrial Research, India	12/09/2006	β glucosidase enzyme useful for the cleavage of β 1,4 linkage of p-nitrophenyl β-D-glucopyranoside (PNPG)
US 2009/0165167 A1	Means and methods to enhance the production of vinblastine and vincristine in *Catharanthus roseus*	Univ Gent, Valtion Teknillinen, Vib Vzw	14/09/2006	Isolated polynucleotide that enhances the production of a compound selected from the group consisting of tabersonine, catharanthine, vinblastine, vincristine, and any combination thereof in *Catharanthus roseus*
US 6303330	Geranyl diphosphate synthase large subunit, and methods of use	Washington State University Research Foundation	18/10/1999	Isolated nucleic acid molecule encoding a geranyl diphosphate synthase large subunit protein Replicable expression vector comprising a nucleic acid molecule of claim Host cell comprising expression vector comprising a nucleic acid molecule encoding a geranyl diphosphate synthase large subunit protein
US 6841717	Methyl-d-erythritol phosphate pathway genes	Monsanto Technology	06/08/2001	Substantially purified nucleic acid molecule comprising a DNA molecule that encodes a protein comprising the amino acid sequence consisting of SEQ ID Host cell or transgenic plant comprising the nucleic acid molecule of claim
US 7067647	Nucleic acid sequences to proteins involved in isoprenoid synthesis	Calgene LLC	13/11/2001	Isolated nucleic acid molecule comprising a nucleic acid sequence encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase DNA construct comprising a nucleic acid molecule of the claim Host cell or plant comprising the DNA construct of the claim
US 7390642	Salutaridinol 7-o-acetyltransferase and derivatives thereof	Donald Danforth Plant Science Center	11/06/2002	Isolated nucleic acid comprising consecutive nucleotides Method for producing a protein having salutaridinol 7-O-acetyltransferase activity

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
US 7405343	Methyl-d-erythritol phosphate pathway genes	Monsanto Technology	27/10/2004	Substantially purified nucleic acid molecule Transformed cell or transgenic plant comprising the nucleic acid molecule of the claim
US 7473538	Sequence-determined DNA fragments encoding cytochrome P450 proteins	Ceres Inc	24/02/2006	An isolated polynucleotide having a cDNA sequence that encodes a polypeptide having cytochrome P450 activity
WO 2002/064764 A3R4	Terpene synthase/cyclase and olefin synthase and uses thereof	Plant Research International	12/02/2002	Nucleic acid or functional fragment thereof according to proteinaceous molecule comprises a geranyl diphosphate (GPP), and/or sesquiterpenoid synthase/cyclase and/or terpene synthase/cyclase Vector comprising a nucleic acid according to anyone of the claims Host comprising a nucleic acid according to anyone of the claims Method for producing a flavor, fragrance and/or bio-control agent comprising a) transforming or transfecting a suitable host with at least one nucleic acid encoding a proteinaceous molecule according to anyone of the claims b) expressing said nucleic acid in the presence of a suitable substrate c) optionally isolating the formed product. Use and composition of the compounds
WO 2002/083720 A2	Intermediates and enzymes of the non-mevalonate isoprenoid pathway	Bacher A, Rohdich F	10/04/2002	Protein in a form that is functional for the enzymatic conversion of 2C-methyl-D-erythhtol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate notably in its (£)-form
WO 2003/097790 A2	Genes and uses thereof to modulate secondary metabolite biosynthesis	Vlaams Interuniversitair Instituut Voor Biotechnologie, Vtt Biotechnology	16/05/2003	An isolated polypeptide that modulates the production of at least one secondary metabolite in an organism or cell derived thereof

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
WO 2004/011667 A2	Geranyl diphosphate synthase molecules, and nucleic acid molecules encoding same	Washington State University Research Foundation	23/07/2003	Isolated nucleic acid molecule that encodes a geranyl diphosphate synthase protein Vector comprising a nucleic acid molecule that encodes a geranyl diphosphate synthase protein Host cell, cultured plant tissue or transgenic plant comprising a vector comprising a nucleic acid molecule that encodes a geranyl diphosphate synthase protein Isolated geranyl diphosphate synthase protein that is least 70% identical to a geranyl diphosphate synthase protein consisting of the amino acid sequence claimed here
WO 2007/031556 A2	Means and methods to enhance the production of vinblastine and vincristine in *Catharanthus roseus*	Vib Vzw, Universiteit Gent, Valtion Teknillinen Tutkimuskeskus	14/09/2006	Isolated polynucleotide that enhances the production of tabersonine and/or catharanthine and/or vinblastine and/or vincristine in *Catharanthus roseus* Recombinant DNA vector comprising at least one of the polynucleotide sequences according to the claims Transgenic plant or cell derived thereof that is transformed with a recombinant DNA vector according to the claim Use of at least one of the polynucleotides according to claims 1 or 2 to enhance the biosynthesis of tabersonine and/or catharanthine and/or vinblastine and/or vincristine in Catharanthus roseus
WO 2008/101238 A2	Natural products from vinca	Rice University	18/02/2008	Pharmaceutical composition comprising an isolated compound
B. Synthetic Tia Pathway				
US 7393946	Method of modulating metabolite biosynthesis in recombinant cells	Rijksuniversiteit Leiden	07/02/2000	Method of increasing in a *Catharanthus* plant cell the expression of one or more genes involved in the biosynthesis of tryptophane or tryptamine
US 7405057	Cytochrome P450s and uses thereof	Board Of Trustees Of The University Of Kentucky	08/03/2002	Method for producing an isoprenoid compound Host cell expressing a recombinant isoprenoid synthase and a recombinant CYP71 family cytochrome P450 polypeptide
US 7795503	Modulating plant alkaloids	Ceres Inc	22/02/2006	Plant cell comprising an exogenous nucleic acid Method of producing one or more alkaloids in a plant cell comprising growing the plant cell expressing an endogenous sequence of interest of the claim

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
WO 2010/004584 A2	Process for production of anti-diabetic compound in root culture of *Catharanthus roseus*	National Institute of Plant Genome Research, New Delhi	09/06/2009	Process of production of serpentine in hairy root culture of *Catharanthus roseus*, said process comprises Composition for treatment of diabetes and insulin resistance, said composition comprising a therapeutically effective amount of serpentine and a pharmaceutically or nutritionally acceptable carrier, wherein the serpentine is obtained by transforming a cell, tissue or any part of *Catharanthus roseus* with *Agrobacterium rhizogenes* to obtain hairy roots
C. Genetically Engineered Organisms				
US 2002/0025517 A1	Methods and compositions for cellular and metabolic engineering	Maxygen Inc.	20/03/1997	Method of evolving a biocatalytic activity of a cell by recombining at least a first and second DNA segment from at least one gene conferring ability to catalyze a reaction of interest Method of evolving a gene to confer ability to catalyze a reaction of interest Modified form of a cell, wherein the modification comprises a metabolic pathway evolved by recursive sequence recombination Method of evolving a biosensor for a compound of interest
US 2009/0205083 A1	Engineered zinc finger proteins targeting 5-enolpyruvyl shikimate-3-phosphate synthase genes	Dow Agrosciences, Llc, Sangamo Bio Sciences, Inc.	25/09/2008	Non-naturally occurring zinc finger protein (ZFP) that binds to an EPSPS target genomic region of interest, said ZFP comprising one or more engineered zinc finger binding domains Plant host cell comprising one or more polynucleotides according to the claim Method for cleaving one or more EPSPS genes in a plant cell Donor vector comprising first and second DNA sequences Method for introducing and expressing an exogenous nucleic acid sequence into the genome of a plant cell Transgenic plant cell or transgenic plant obtained according to the method of the claim Method for stimulating intramolecular homologous recombination in the genome of a plant cell
US 7129392	Materials and methods for increasing isoprenoid production in cells	Kizer, Jeffery, Lance, The Regents Of The University Of California	28/04/2004	Method for providing transformed cells having increased isoprenoid production

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
WO 1999/006581 A1	Transgenic plants using the *tdc* gene for crop improvement	Sanford Scientific Inc.	31/07/1998	A transgenic plant tissue having improved resistance to fungi, bacteria, nematodes or combinations thereof, comprising a plant tissue whose cells are transformed with and express a *tdc* and/or *as* transgene Nucleic acid construct according to the claims Method of enhancing resistance to phytopathogenic bacteria, fungi, and/or nematodes in a plant tissue
EP 1707641 (A2)	Evolution of whole cells and organisms by recursive sequence recombination	Maxygen, INC	15/07/1999	Method of evolving a cell to acquire a desired property, comprising: (i) forming protoplasts of a population of different cells; (ii) fusing the protoplasts to form hybrid protoplasts, in which genomes from the protoplasts recombine to form hybrid genomes; (iii) incubating the hybrid protoplasts under conditions promoting regeneration of cells, thereby producing regenerated cells; (iv) repeatedly forming protoplasts from the regenerated cells, fusing the protoplasts to form hybrid protoplasts, in which genomes from the protoplasts recombine to form additional hybrid genomes; incubating the additional hybrid protoplasts under conditions promoting regeneration of cells, thereby producing additional regenerated cells; and (v) selecting or screening to isolate additionally regenerated cells that have evolved toward acquisition of the desired property
WO 2009/139626 A1	Increased production of health-promoting compounds in plants	Expressive Research B.V	13/05/2009	Method to enhance the production of plant- specific compounds (with health-promoting properties) by increasing expression of a regulatory gene product
D. Targeted Methods				
WO 2003/075129 A2	Methods, systems, and software for identifying functional bio-molecules	Maxygen, Inc	03/03/2003	Method for identifying amino acid residues for variation in a protein variant library in order to affect a desired activity Computer program product comprising a computer readable medium on which is provided program instructions for identifying amino acid residues for variation in a protein variant library in order to affect a desired activity Method of identifying members of a population of biopolymer sequence variants most suitable for artificial evolution Computer system or program for predicting sequences that comprise desired properties Method of predicting at least one property of at least one target polypeptide sequence System for predicting sequence activities Method of producing libraries of desired sizes

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
WO 2004/079006 A1	Method for identifying the function of a gene	Max-Planck-Gesellschaft Zur Förderung Der Wissenschaften	04/03/2004	Method for determining the function of a gene Method for identifying a gene which is capable of modifying the amount of a metabolite in an organism Method for identifying a metabolite which is capable of modifying the amount of a transcript in an organism Use of a gene the function of which has been determined by the method of any one of the claims Use of a metabolite identified by the method of the claims for modifying the amount of a transcript in an organism
WO 2005/059556 A1	Method for analysing metabolites	Max Planck Geselllschaft Zur Förderung Der Wissenschaften	17/12/2004	Method for analyzing the metabolites of a biological sample Set of isotopically labeled metabolites obtainable from a sample Use of the set of isotopically labeled metabolites as a quantitative standard for determining the amount of one or more metabolites in a biological sample Kit comprising an isotopically labeled metabolizable compound and a manual for use in carrying in out the method of any one of the claims Use of an isotopically labeled compound that can be metabolized by a cell for labeling the metabolites or the quantitative determination of metabolites in a biological sample
WO 2005/064002 A1	Method for mass production of secondary metabolites in plant cell culture by treatment of an alkanoic acid or salt thereof	Samyang Genex Corporation	29/12/2004	Method for production of secondary metabolites in plant cell culture Medium for producing secondary metabolites in plant cell culture
AU 2000/016960 A1	Strictosidine glucosidase from *Catharanthus roseus* and its use in alkaloid production	Universiteit Leiden	02/12/1999	Method for producing terpenoid-indole alkaloids using an in vitro culture system, wherein said culture system comprises plant materials as a source for precursors and partly as a source for glucose, for said terpenoid-indole alkaloids Culture system for producing terpenoid-indole alkaloids or precursors thereof in vitro Use of a culture system according to the claim in the production of pharmaceutical terpenoid-indole alkaloids Isolated and/or recombinant nucleic acid encoding a strictosidine glucosidase or a functional fragment and/or derivative thereof, which strictosidine glucosidase in *Catharanthus roseus* has a coding sequence Vector comprising a nucleic acid according to the claim

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
EP 0354778 B1	Method for production of alkaloid	Mitsui Petrochemical Industries	09/08/1989	Method for producing a dimeric alkaloid, which comprises reacting catharanthine with 20 vindoline in the presence of Fe3+, adding oxygen and a dicarboxylic acid or derivative thereof to the reaction system and reacting the reaction product with a hydride source
US 6703539	Method and compositions for modifying levels of secondary metabolic compounds in plants	National Research Council Of Canada	22/01/1999	Method for improving a nutritional profile of a plant Genetically altered plant or a descendant thereof, comprising a recombinant nucleic acid molecule stably incorporated into the genome of said plant Animal feed derived at least in part from the genetically modified plant or descendant thereof according to the claim, or from a cell, seed or component thereof
US 6855866	Polynucleotides useful for modulating transcription	University Of California	28/11/2000	Isolated promoter polynucleotide which specifically initiates transcription in a plant suspensor cell and/or basal region of a plant embryo Method of introducing an isolated polynucleotide into a host cell comprising Expression cassette comprising the promoter polynucleotide of the claim operably linked to a heterologous polynucleotide Host cell or plant comprising the expression cassette of the claim
US 7161061	Metabolite transporters	Monsanto Technology	09/05/2002	Substantially purified nucleic acid molecule that encodes an adenylate transporter Nucleic acid molecule of the claim further defined as comprising a 3′ non-translated sequence that functions in a plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3′ end of the mRNA molecule Transformed plant transformed with the substantially purified nucleic acid molecule of the claims Method of producing a plant having seeds with an increased tocopherol level Feedstock comprising the transformed plant of the claim or part thereof Meal comprising plant material manufactured from the transformed plant of the claim
US 7402667	Promoter, promoter control elements, and combinations, and uses thereof	Ceres, Inc	30/09/2004	Plant comprising a vector construct
US 7563863	Plastid transit peptides	Pioneer Hi-Bred International, Inc	16/01/2008	Isolated nucleic acid molecule from the group encodes to the peptide of SEQ ID, a polypeptide comprising the nucleic acid claimed Vector comprising the nucleic acid of the claim Plant cell or transgenic plant comprising the nucleic acid of the claim Method for targeting a polypeptide to a plastid

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
WO 2000/042200 A1	Strictosidine glucosidase from *Catharanthus roseus* and its use in alkaloid production	Universiteit Leiden	02/12/1999	Method for producing terpenoid-indole alkaloids using an in vitro culture system, wherein said culture system comprises plant materials as a source for precursors and partly as a source for glucose, for said terpenoid-indole alkaloids Culture system for producing terpenoid-indole alkaloids or precursors thereof in vitro Use of a culture system according to the claim in the production of pharmaceutical terpenoid-indole alkaloids Isolated and/or recombinant nucleic acid encoding a strictosidine glucosidase or a functional fragment and/or derivative thereof, which strictosidine glucosidase in Catharanthus roseus has a coding sequence Vector comprising a nucleic acid according to the claim Cell comprising a nucleic acid, a vector or a gene encoding a strictosidine synthase or a functional fragment and/or derivative thereof
WO 2000/067558 A1	Regulation of gene expression in tobacco for manipulation of plant growth and secondary metabolism	Timko M	05/05/2000	Isolated DNA molecule comprising nucleotide sequence Vector comprising the isolated DNA molecule of the claim Cultured transgenic tobacco cell or transgenic tobacco plant stably transformed with the vector of the claims Method for regulating gene expression in a plant comprising functionally linking an alkaloid gene promoter to a nucleic acid encoding a protein
WO 2002/083888 A2	The use of genes encoding membrane transporter pumps to stimulate the production of secondary metabolites in biological cells	Vlaams Interuniversitair Instituut Voor Biotechnologie Vzw	18/04/2002	Use of an expression cassette comprising a gene encoding an ABC-transporter to induce or to enhance the production or the secretion of at least one secondary metabolite, alkaloids by plant cells Transgenic plant cell culture, with an enhanced production or secretion of an at least one secondary metabolite, transformed with an expression vector comprising an expression cassette according to the claims Isolated polypeptide selected from the groups
WO 2002/083888 A3R4	The use of genes encoding abc transporters to stimulate the production of secondary metabolites in biological cells	Vlaams Interuniversitair Instituut Voor Biotechnologie Vzw	18/04/2002	Use of expression cassette comprising a gene encoding an ABC-transporter to induce r to enhance the production or the secretion of at least one secondary metabolite by plant cells Transgenic plant or transgenic plant cell cultures Isolated polupeptide

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
WO 2002/090506 A3R4	Metabolite transporters	Monsanto Technology Llc	09/05/2002	Substantially purified nucleic acid molecule Transformed plant having an exogenous nucleic acid molecule that encodes an adenylate transporter Method of producing plants of the claims Oil derived from a seed of a transformed plant
WO 2003/025193 A1	Plant enzymes for bioconversion	Plant Research International B.V.	17/09/2002	Method for converting a substrate and generating a conversion product comprising subjecting said substrate to an enzyme derived from a species or organism containing sesquiterpenes Isolated or purified enzyme derived from a species or organism containing sesquiterpenes Isolated and/or recombinant nucleic acid encoding an enzyme according to anyone of the claims Host cell comprising a nucleic acid according to the claim
WO 2006/111512 A1	Improved methods controlling gene expression	Basf Plant Science Gmbh	13/04/2006	Method for transgenic expression with enhanced specificity in an eukaryotic organism such as human, animal or plant Chimeric ribonucleotide sequence comprising at least one sequence capable to confer a preferred phenotype or beneficial effect to a eukaryotic organism, and at least one sequence substantially complementary to a microRNA sequence naturally occurring in a eukaryotic organism Expression construct of the claim Expression vector comprising an expression construct of the claims Transformed cell or non-human organism comprising a chimeric ribonucleotide sequence of the claims, an expression construct of any of the claims or an expression vector of any of the claims Pharmaceutically preparation of at least one expression construct of the claims, a chimeric ribonucleotide sequence of the claims, or a vector according to any of the claims

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
E. Industrial Production				
WO 2008/073348 A2	Microgravity bioreactor systems for production of bioactive compounds and biological macromolecules	Marshall University Research Corporation, The United States Of America As Represented By The Administrator Of The National Aeronautics And Space Administration	10/12/2007	Method for continuous culture of plant cells from the group consisting of: *Catharanthus roseus* G. Don, *Scopolia solanaceae, Duboisia solanaceae, Atropa solnanaceae, Datura solanaceae, Taxus baccata, Nothapodytes foetida, Artemisia annua, Lithospermum erγthrorhizon, Aloe barbadensis, Atropa belladonna, Cinchona ledgeriana, Datura metel, Digitalis lanata, Dioscorea deltoida, Dioscorea composita, Glycyrrhiza glabra, Hyoscyamus Niger, Hyoscyamus albus, Panax ginseng, Papaver somniferum, Rheum officinale, Rouwolfia serpentina, Eucalyptus globulus, Eugenia caryophyllata, Jasminum, Lavandula angustffolia, Mentha pzerita, Pelargonium, Thaumatocoeus danielli, Vetiver, Santalum album* L. and *Solanum Xanthocarpum*; comprising growing the cells in a hydro focusing bioreactor (HFB) under conditions sufficient for growth Method for producing one or more bioactive compounds, comprising continuously culturing plant cells in a hydrofocusing bioreactor under conditions sufficient for production of one or more bioactive compounds by said plant cells, and isolating said bioactive compounds Method for assaying the presence of one or more bioactive plant compounds, comprising continuously culturing plant cells in a hydrofocusing bioreactor, whereby said plant cells produce the bioactive compounds Process for obtaining a tissue-like, three-dimensional plant cell construct in a hydrofocusing bioreactor, comprising filling the culture chamber of said hydrofocusing bioreactor with a medium and plant cells of one or more distinct types to establish a culturing environment within the culture chamber and continuously culturing the plant cells from at least about 3 days to about 35 days Tissue-like three-dimensional plant cell construct grown in a hydro focusing bioreactor, wherein the tissue-like three-dimensional plant cell construct has a reorganized and degraded cytoskeleton and swollen chloroplasts Method for continuous culture of fungal cells comprising growing the cells in a hydrofocusing bioreactor (HFB) under conditions sufficient for growth Method for continuous culture of bacterial cells comprising growing the cells in a hydrofocusing bioreactor (HFB) under conditions sufficient for growth

Table 4. List of patents relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Assignee	Filing date	Main claims
WO 2004/006657 A1	Transgenic plants used as a bioreactor system	Bureau Of Sugar Experiment Stations, The University Of Queensland	11/07/2003	Method for modulating the levels of a metabolic or biosynthetic product in a plant, including introducing a product into a plant Vector comprising one or more of the genetic sequences recited in any one of the claims Genetically modified cell comprising one or more of the genetic sequences recited in any one of the claims Product produced in transgenically modified plant by the method of any one of the claims Plant based bioreactor system used for the production of a metabolic or biosynthetic product, said bioreactor comprising one or more sugarcane (*Saccharum* sp.) plant(s) produced according to the method of any one of the claims
JP19910254784 19910904	Production of plant metabolite		04/09/1991	Not available in english
WO 2009/038922 A1	Forming embolic particles	Boston Scientific Limited	20/08/2008	Method comprising: freeze-drying a mixture comprising a polymer to form a particle Method comprising: forming a particle having a maximum dimension of 5,000 microns by irradiating a polymer to crosslink the polymer

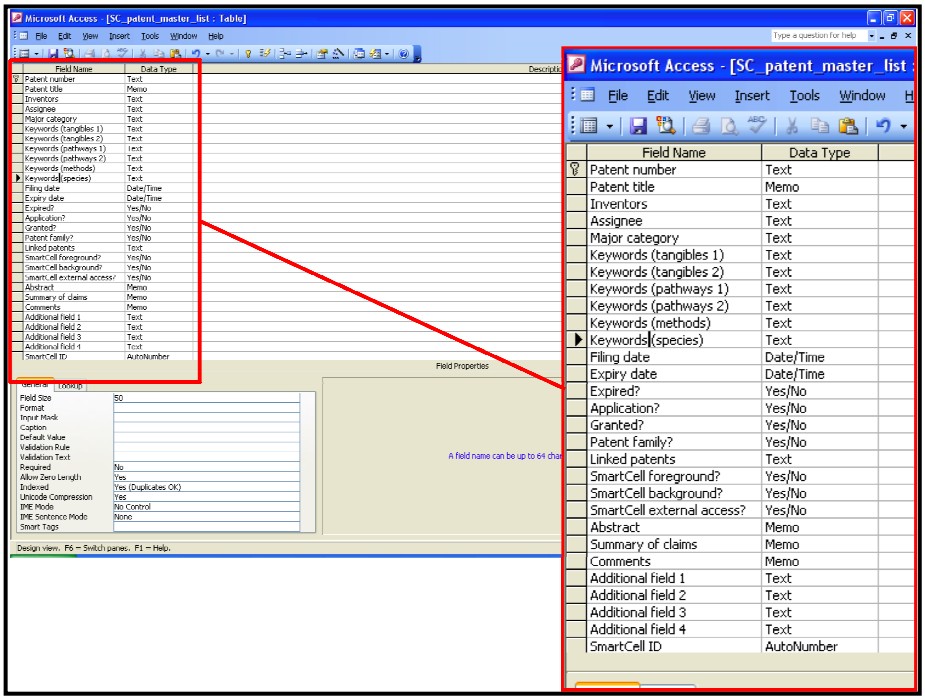

Figure 8. Categories in the SmartCell IP database.

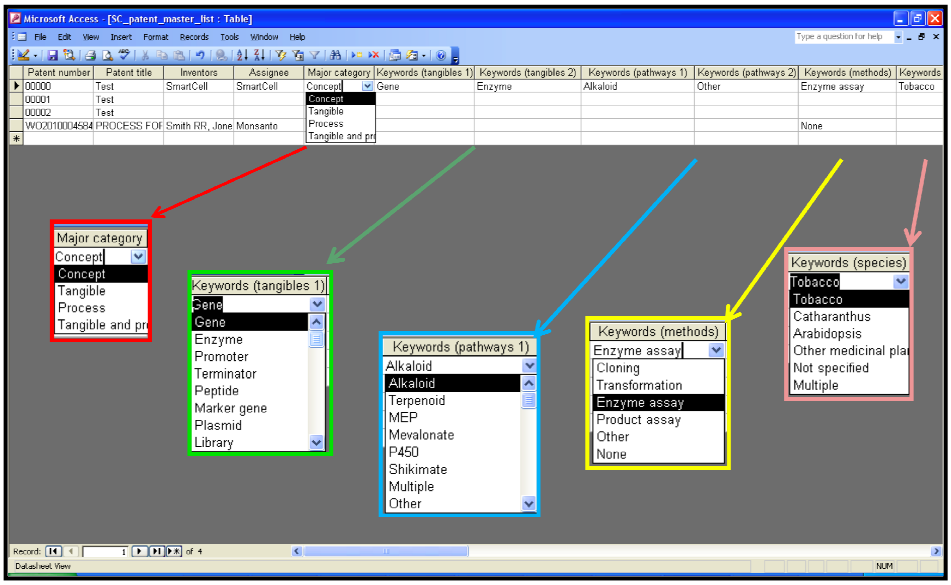

Figure 9. The keywords used to classify relevant patents in the SmartCell IP database.

To evaluate the impact of the claims in each of these 49 patents, we organized them into five categories, as shown in Table 4:

- **Genes/Proteins** – this included claims covering specific isolated or recombinant DNA sequences and corresponding products and their uses. Here we identified 16 relevant patents covering 11 individual genes/proteins and 5 individual products or intermediates from the TIA pathway.
- **Synthetic TIA pathway** – this included claims related to synthetic versions of the TIA pathway used to generate particular metabolites. Here we identified five relevant patents.
- **Genetically engineered organisms** – this included claims related to plant cells, tissues and whole plants containing or expressing parts of the TIA pathway. Here we identified five relevant patents.
- **Specific methods** – this included claims related to cloning and expression methods that specifically target the TIA pathway, as opposed to generic cloning methods, e.g. specific vectors, constructs, libraries, probes and also methods for gene transfer and tissue culture, where appropriate. Here we identified 19 relevant patents.
- **Industrial production** – this included claims for scaling up production and processing. Here we identified four relevant patents.

7.6. Discussion

Searchable patent databases provide a valuable resource allowing researchers to investigate the prior art without incurring the costs of patent specialists or wasting funds on fruitless applications. However, with more than 40 million patents and applications now published, even searching free databases can be labor intensive without a focused search strategy (WIPO, 2006). As a case study, we therefore analyzed patents related to the engineering of terpenoid indole alkaloid (TIA) metabolism in plants. We developed a focused search strategy to remove redundancy and reduce the workload without missing important and relevant patents. This resulted in the identification of 49 key patents associated with TIA metabolic engineering in plants which were divided into five major categories.

The Genes/Proteins category contained 16 of the 49 relevant patents and could be divided into four subcategories that are described in more detail in Table 5: (a) isolated sequence; (b) isolated sequence, protein production method and use; (c) enzyme and its reaction; and (d) method for the preparation of a compound. All patents were classified except CN101250543 because the claims were not available in English. The claims in subcategory (a) are polynucleotides that enhance the production of at least one secondary metabolite, such as tabersonine, catharanthine, vinblastine, vincristine, and any combination thereof, as well as a sequence that encodes a polypeptide with cytochrome P450 activity. The claims in subcategory (b) protect sequences such as geranyl diphosphate synthase, 1-deoxy-D-xylulose 5-phosphate reductoisomerase, sesquiterpenoid synthase/cyclase and terpene synthase/cyclase. They also protect the vectors containing these sequences, as well as host cells or transgenic plants transformed with such vectors. The claims in subcategory (c) protect enzymes that cleave the β(1,4)linkage of p-nitrophenyl β-D-glucopyranoside and that convert 2C-methyl-D-erythhtol 2,4-cyclodiphosphate [sic] into 1-hydroxy-2-methyl-2-butenyl 4-diphosphated. The claims in subcategory (d) relate to methods for the preparation of strictosidine and natural products from *Vinca* species.

The synthetic TIA pathway category comprised four patents protecting different methods for the production of molecules related to the TIA pathway such as tryptamine, isoprenoids and serpentine, using transgenic plants or cells transformed with genes, but unlike the Genes/Proteins category, the claims hereunder did not cover the genes or proteins themselves.

The genetically engineered organisms category comprised five generic patents claiming methods to produce compounds by metabolic engineering, including methods to transform cells, protoplasts, plant tissues or whole plants to acquire a specific metabolic profile.

Table 5. Subdivision of relevant patents in the Gene/Protein category relevant to "engineering the TIA metabolic pathway in plants".

Patent number	Patent title	Main claims
A. Isolated sequence		
US 2009/0165167 A1	Means and Methods to Enhance the Production of Vinblastine and Vincristine in *Catharanthus roseus*	Isolated polynucleotide that enhances the production of a compound selected from the group consisting of tabersonine, catharanthine, vinblastine, vincristine, and any combination thereof in *Catharanthus roseus*
US 7473538	Sequence-determined DNA fragments encoding cytochrome P450 proteins	An isolated polynucleotide having a cDNA sequence that encodes a polypeptide having cytochrome P450 activity
WO 2003/097790 A2	Genes and uses thereof to modulate secondary metabolite biosynthesis	An isolated polypeptide that modulates the production of at least one secondary metabolite in an organism or cell derived thereof
B. Isolated sequence, method of production and uses		
WO 2004/011667 A2	Geranyl diphosphate synthase molecules, and nucleic acid molecules encoding same	Isolated nucleic acid molecule that encodes a geranyl diphosphate synthase protein Vector comprising a nucleic acid molecule that encodes a geranyl diphosphate synthase protein Host cell, cultured plant tissue or transgenic plant comprising a vector comprising a nucleic acid molecule that encodes a geranyl diphosphate synthase protein Isolated geranyl diphosphate synthase protein that is least 70% identical to a geranyl diphosphate synthase protein consisting of the amino acid sequence claimed here
US 6303330	Geranyl diphosphate synthase large subunit, and methods of use	Isolated nucleic acid molecule encoding a geranyl diphosphate synthase large subunit protein Replicable expression vector comprising a nucleic acid molecule of claim Host cell comprising expression vector comprising a nucleic acid molecule encoding a geranyl diphosphate synthase large subunit protein
US 6841717	Methyl-D-erythritol phosphate pathway genes	Substantially purified nucleic acid molecule comprising a DNA molecule that encodes a protein comprising the amino acid sequence consisting of SEQ ID Host cell or transgenic plant comprising the nucleic acid molecule of claim
US 7067647	Nucleic acid sequences to proteins involved in isoprenoid synthesis	Isolated nucleic acid molecule comprising a nucleic acid sequence encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase DNA construct comprising a nucleic acid molecule of the claim Host cell or plant comprising the DNA construct of the claim
US 7390642	Salutaridinol 7-O-acetyltransferase and derivatives thereof	Isolated nucleic acid comprising consecutive nucleotides Method for producing a protein having salutaridinol 7-O-acetyltransferase activity
US 7405343	Methyl-D-erythritol phosphate pathway genes	Substantially purified nucleic acid molecule Transformed cell or transgenic plant comprising the nucleic acid molecule of the claim

Table 5. Subdivision of relevant patents in the Gene/Protein category relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Main claims
WO2002064764	Terpene synthase/cyclase and olefin synthase and uses thereof	Nucleic acid or functional fragment thereof according to proteinaceous molecule comprises a geranyl diphosphate (GPP), and/or sesquiterpenoid synthase/cyclase and/or terpene synthase/cyclase Vector comprising a nucleic acid according to anyone of the claims Host comprising a nucleic acid according to anyone of the claims Method for producing a flavor, fragrance and/or bio-control agent comprising a) transforming or transfecting a suitable host with at least one nucleic acid encoding a proteinaceous molecule according to anyone of the claims b) expressing said nucleic acid in the presence of a suitable substrate c) optionally isolating the formed product Use and composition of the compounds
WO 2007/031556 A2	Means and methods to enhance the production of vinblastine and vincristine in *Catharanthus roseus*	Isolated polynucleotide that enhances the production of tabersonine and/or catharanthine and/or vinblastine and/or vincristine in *Catharanthus roseus* Recombinant DNA vector comprising at least one of the polynucleotide sequences according to the claims Transgenic plant or cell derived thereof that is transformed with a recombinant DNA vector according to the claim Use of at least one of the polynucleotides according to claims 1 or 2 to enhance the biosynthesis of tabersonine and/or catharanthine and/or vinblastine and/or vincristine in Catharanthus roseus
C. Enzyme and its reaction		
US 2007/0212745 A1	Beta-glucosidase and a process for extraction thereof	β glucosidase enzyme useful for the cleavage of β 1,4 linkage of p-nitrophenyl β-D-glucopyranoside (PNPG).
WO 2002/083720 A2	Intermediates and enzymes of the non-mevalonate isoprenoid pathway	Protein in a form that is functional for the enzymatic conversion of 2C-methyl-D- erythhtol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate notably in its (£)-form
E. Method to prepare a compound		
EP 0156267 B1	Method for the preparation of strictosidine	Method for the preparation of strictosidine
WO 2008/101238 A2	Natural products from vinca	Pharmaceutical composition comprising an isolated compound

The specific methods category contained 18 of the 49 relevant patents, claiming specific methods to (a) identify genes/proteins in the TIA pathway, (b) analyze the corresponding metabolites, (c) produce those metabolites and (d) modulate production by manipulating gene expression (Table 6). The claims in subcategory (a) protect methods that identify genes, gene products and metabolites or determine the function of a gene (including computer systems and software used to predict enzyme activities). Subcategory (b) comprises a single patent that protects a method for analyzing the metabolites in a biological sample, including the preparation of isotopically-labeled metabolites using a kit. The claims in subcategory (c) relate to six patents that protect methods for the production of secondary metabolites *in vitro* using plant cell cultures including the bioreactors, medium and culture systems, and methods for converting a substrate into a relevant product using enzymes. The claims in subcategory (d) include protection for promoters, transporters, vectors and transit peptides and the vectors containing them.

Finally, the industrial production category included four relevant patents, three of which protect the use of engineered cells or plants as scaled-up bioreactors for the production of important metabolites. The fourth patent was only available in Japanese and will need to be evaluated by a Japanese-speaking colleague to determine its relevance.

Table 6. Subdivision of relevant patents in the Methods category relevant to "engineering the TIA metabolic pathway in plants".

Patent number	Patent title	Main claims
A. Methods to identified Genes/Proteins		
WO 2003/075129 A2	Methods, systems, and software for identifying functional bio-molecules	Method for identifying amino acid residues for variation in a protein variant library in order to affect a desired activity Computer program product comprising a computer readable medium on which is provided program instructions for identifying amino acid residues for variation in a protein variant library in order to affect a desired activity Method of identifying members of a population of biopolymer sequence variants most suitable for artificial evolution Computer system or program for predicting sequences that comprise desired properties Method of predicting at least one property of at least one target polypeptide sequence System for predicting sequence activities Method of producing libraries of desired sizes
WO 2004/079006 A1	Method for identifying the function of a gene	Method for determining the function of a gene Method for identifying a gene which is capable of modifying the amount of a metabolite in an organism Method for identifying a metabolite which is capable of modifying the amount of a transcript in an organism Use of a gene the function of which has been determined by the method of any one of the claims Use of a metabolite identified by the method of the claims for modifying the amount of a transcript in an organism
B. Methods to metabolite analysis		
WO 2005/059556 A1	Method for analysing metabolites	Method for analyzing the metabolites of a biological sample Set of isotopically labeled metabolites obtainable from a sample Use of the set of isotopically labeled metabolites as a quantitative standard for determining the amount of one or more metabolites in a biological sample Kit comprising an isotopically labeled metabolizable compound and a manual for use in carrying in out the method of any one of the claims Use of an isotopically labeled compound that can be metabolized by a cell for labeling the metabolites or the quantitative determination of metabolites in a biological sample

Table 6. Subdivision of relevant patents in the Methods category relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Main claims
C. Methods to produce secondary metabolites		
WO 2005/064002 A1	Method for mass production of secondary metabolites in plant cell culture by treatment of an alkanoic acid or salt thereof	Method for production of secondary metabolites in plant cell culture Medium for producing secondary metabolites in plant cell culture
AU 2000/016960 A1	Strictosidine glucosidase from catharanthus roseus and its use in alkaloid production	Method for producing terpenoid-indole alkaloids using an in vitro culture system, wherein said culture system comprises plant materials as a source for precursors and partly as a source for glucose, for said terpenoid-indole alkaloids Culture system for producing terpenoid-indole alkaloids or precursors thereof in vitro Use of a culture system according to the claim in the production of pharmaceutical terpenoid-indole alkaloids Isolated and/or recombinant nucleic acid encoding a strictosidine glucosidase or a functional fragment and/or derivative thereof, which strictosidine glucosidase in Catharanthus roseus has a coding sequence Vector comprising a nucleic acid according to the claim Cell comprising a nucleic acid, a vector or a gene encoding a strictosidine synthase or a functional fragment and/or derivative thereof
WO 2000/042200 A1	Strictosidine glucosidase from catharanthus roseus and its use in alkaloid production	Method for producing terpenoid-indole alkaloids using an in vitro culture system, wherein said culture system comprises plant materials as a source for precursors and partly as a source for glucose, for said terpenoid-indole alkaloids Culture system for producing terpenoid-indole alkaloids or precursors thereof in vitro Use of a culture system according to the claim in the production of pharmaceutical terpenoid-indole alkaloids Isolated and/or recombinant nucleic acid encoding a strictosidine glucosidase or a functional fragment and/or derivative thereof, which strictosidine glucosidase in Catharanthus roseus has a coding sequence Vector comprising a nucleic acid according to the claim Cell comprising a nucleic acid, a vector or a gene encoding a strictosidine synthase or a functional fragment and/or derivative thereof

Table 6. Subdivision of relevant patents in the Methods category relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Main claims
US 6703539	Method and compositions for modifying levels of secondary metabolic compounds in plants	Method for improving a nutritional profile of a plant Genetically altered plant or a descendant thereof, comprising a recombinant nucleic acid molecule stably incorporated into the genome of said plant Animal feed derived at least in part from the genetically modified plant or descendant thereof according to the claim, or from a cell, seed or component thereof
WO 2003/025193 A1	Plant enzymes for bioconversion	Method for converting a substrate and generating a conversion product comprising subjecting said substrate to an enzyme derived from a species or organism containing sesquiterpenes Isolated or purified enzyme derived from a species or organism containing sesquiterpenes Isolated and/or recombinant nucleic acid encoding an enzyme according to anyone of the claims Host cell comprising a nucleic acid according to the claim
EP 0354778 B1	Method for production of alkaloid	Method for producing a dimeric alkaloid, which comprises reacting catharanthine with 20 vindoline in the presence of Fe3+, adding oxygen and a dicarboxylic acid or derivative thereof to the reaction system and reacting the reaction product with a hydride source
D. Methods to modulate gene expression		
US 6855866	Polynucleotides useful for modulating transcription	Isolated promoter polynucleotide which specifically initiates transcription in a plant suspensor cell and/or basal region of a plant embryo Method of introducing an isolated polynucleotide into a host cell comprising Expression cassette comprising the promoter polynucleotide of the claim operably linked to a heterologous polynucleotide Host cell or plant comprising the expression cassette of the claim
US 7161061	Metabolite transporters	Substantially purified nucleic acid molecule that encodes an adenylate transporter Nucleic acid molecule of the claim further defined as comprising a 3′ non-translated sequence that functions in a plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3′ end of the mRNA molecule Transformed plant transformed with the substantially purified nucleic acid molecule of the claims Method of producing a plant having seeds with an increased tocopherol level Feedstock comprising the transformed plant of the claim or part thereof Meal comprising plant material manufactured from the transformed plant of the claim

Table 6. Subdivision of relevant patents in the Methods category relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Main claims
US 7402667	Promoter, promoter control elements, and combinations, and uses thereof	Plant comprising a vector construct
US 7563863	Plastid transit peptides	Isolated nucleic acid molecule from the group encodes to the peptide of SEQ ID, a polypeptide comprising the nucleic acid claimed Vector comprising the nucleic acid of the claim Plant cell or transgenic plant comprising the nucleic acid of the claim Method for targeting a polypeptide to a plastid
WO 2000/067558 A1	Regulation of gene expression in tobacco for manipulation of plant growth and secondary metabolism	Isolated DNA molecule comprising nucleotide sequence Vector comprising the isolated DNA molecule of the claim Cultured transgenic tobacco cell or transgenic tobacco plant stably transformed with the vector of the claims Method for regulating gene expression in a plant comprising functionally linking an alkaloid gene promoter to a nucleic acid encoding a protein
WO 2002/083888 A2	The use of genes encoding membrane transporter pumps to stimulate the production of secondary metabolites in biological cells	Use of an expression cassette comprising a gene encoding an ABC-transporter to induce or to enhance the production or the secretion of at least one secondary metabolite, alkaloids by plant cells Transgenic plant cell culture, with an enhanced production or secretion of an at least one secondary metabolite, transformed with an expression vector comprising an expression cassette according to the claims Isolated polypeptide selected from the groups
WO 2002/083888 A3R4	The use of genes encoding abc transporters to stimulate the production of secondary metabolites in biological cells	Use of expression cassette comprising a gene encoding an ABC-transporter to induce r to enhance the production or the secretion of at least one secondary metabolite by plant cells Transgenic plant or transgenic plant cell cultures Isolated polupeptide
WO 2002/090506 A3R4	Metabolite transporters	Substantially purified nucleic acid molecule Transformed plant having an exogenous nucleic acid molecule that encodes an adenylate transporter Method of producing plants of the claims Oil derived from a seed of a transformed plant

Table 6. Subdivision of relevant patents in the Methods category relevant to "engineering the TIA metabolic pathway in plants" **(cont.)**

Patent number	Patent title	Main claims
WO 2006/111512 A1	Improved methods controlling gene expression	Method for transgenic expression with enhanced specificity in an eukaryotic organism such as human, animal or plant Chimeric ribonucleotide sequence comprising at least one sequence capable to confer a preferred phenotype or beneficial effect to a eukaryotic organism, and at least one sequence substantially complementary to a microRNA sequence naturally occurring in a eukaryotic organism Expression construct of the claim Expression vector comprising an expression construct of the claims Transformed cell or non-human organism comprising a chimeric ribonucleotide sequence of the claims, an expression construct of any of the claims or an expression vector of any of the claims Pharmaceutically preparation of at least one expression construct of the claims, a chimeric ribonucleotide sequence of the claims, or a vector according to any of the claims

7.7. Conclusions

Patent applications must include a discussion of the prior art to show how the invention is novel, non-obvious and applicable in an industry setting. The prior art must therefore be analyzed for pre-existing inventions or overlapping claims that may affect the chance of commercial success. Large organizations retain patent experts to conduct due diligence but this is often beyond the capabilities of publically-funded research. Therefore, many researchers make use of free patent databases maintained by organizations such as EPO, USPTO, JPO and CAMBIA, which are believed to cover more than 80% of the patent documents worldwide (Eurostat, 2013).

The availability of more than 40 million patent documents means that due diligence can be a labor-intensive and time-consuming process if patent searches are not focused. However, focusing on relevant patents must not carry a high risk of accidental exclusion. We have developed a tiered, high-quality patent search approach which allows the recovery of most relevant patents while excluding most false positives, as a preliminary approach for determining the patentability of an invention without recourse to expensive professional help. This high-quality search strategy allowed us to identify 49 patents relevant for engineering of the TIA pathway in plants, revealing the major genes/proteins, products, intermediates and methods which need to be negotiated, and the previously described synthetic pathways and genetically modified organisms relevant to the approach. The data thus gathered will allow to 'invent around' these restrictions and avoid patent infringement by ensuring no protected sequences, vectors, bioreactors and processing methods are used. Importantly, the strategy outlined in this chapter can easily be applied in other research projects by modifying the search string accordingly.

7.8. References

Adcock M. (2007) Intellectual property, genetic modified crops and bioethics. *Biotechnology Journal*, 2: 1088-1092.

CAMBIA (2010) CAMBIA, accessed online December 2009 - December 2010, http://www.cambia.org/daisy/cambia/home.html.

CAMBIA (2013a) CAMBIA, accessed online 07 January, http://www.cambia.org/daisy/cambia/home.html.

CAMBIA (2013b) Funders and partners, accessed online 07 January 2013, http://www.cambia.org/daisy/cambia/about/595.html.

CAMBIA (2013c) Location and facilities, accessed online 07 January 2013, http://www.cambia.org/daisy/cambia/about/594.html.

EPO (2013a) European Patent Organisation, accessed online 03 January 2013, http://www.epo.org/about-us/organisation.html.

EPO (2013b) Espacenet, accessed online 03 January 2013: 03-01-2013, http://www.epo.org/searching/free/espacenet.html.

Espacenet patent search (2010) Advanced search, accessed online December 2009-December 2010, http://worldwide.espacenet.com/advancedSearch?locale=en_EP.

Espacenet patent search (2013a) Smart search accessed online 03 January 2013, http://worldwide.espacenet.com/?locale=en_EP.

Espacenet patent search (2013b) Advanced search, accessed online 03 January 2013, http://worldwide.espacenet.com/advancedSearch?locale=en_EP.

Eurostat (2013) Patent statistic, accessed online 04 January 2013, http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Patent_statistics.

Fleck B., Baldock C. (2003) Intellectual property protection for plant-related inventions in Europe. *Nature*, 4: 834- 838.

IPDL (2010) IPDL- Industrial Property Digital Library, accessed online December 2009-December 2010, http://www.ipdl.inpit.go.jp/homepg_e.ipdl.

IPDL (2013) IPDL- Industrial Property Digital Library, accessed online 03 January 2013, http://www.ipdl.inpit.go.jp/homepg_e.ipdl.

JPO (2013) Japan Patent Office, accessed online 03 January 2013, http://www.jpo.go.jp/.

Kannan N. (2010) Importance of intellectual property rights. *International Journal of Intellectual Property Rights*, 1: 1-5.

Lei Z., Juneja R., Wright B.D. (2009) Patents versus patenting: implications of intellectual property protection for biological research. *Nature Biotechnology*, 27: 36-40.

Patentlens (2013) Fulltext, accessed online 07 January 2013, http://www.patentlens.net/patentlens/structured.html?noScroll.

USPTO (2010) Search for patents, accessed online December 2009- December 2010, http://www.uspto.gov/patents/process/search/#heading-6.

USPTO (2013a) USPTO offices, accessed online 03 January 2013, http://www.uspto.gov/about/offices/index.jsp.

USPTO (2013b) Search for patents, accessed online 03 January 2013, http://www.uspto.gov/patents/process/search/#heading-6.

WIPO (2006) Inventing the future, an introduction to patents for small and medium-sized enterprises. *Intellectual Property for Business Series*, 3: 1-47.

WIPO (2008) WIPO intellectual property handbook. *WIPO publications,* 1-488.

WIPO (2012) WIPO IP facts and figures. *WIPO Economics & Statistics Series,* 1-42.

WIPO (2013a) Patent scope, accessed online 03 January 2013, http://patentscope.wipo.int/search/en/search.jsf.

WIPO (2013b) What is WIPO?, accessed online 04 January 2013, http://www.wipo.int/about-wipo/en/.

GENERAL CONCLUSIONS

General Conclusions

- The metabolomic analysis of transgenic tobacco plants expressing different combinations of monoteprene secoiridoid pathway genes has provided greater insight into the development of effective metabolic engineering strategies for secondary metabolism.
- Monoterpenes such as geraniol, geraniol diphosphate and 10-hydroxygeraniol are toxic to plant cells. This could explain why they are sequestered into subcellular compartments (to avoid systemic toxicity) although this increases the complexity of engineering strategies for the terpenoid pathway.
- A population of transgenic plants expressing *AtGpps* and *VoGes* or *VoGes* and *Cyp76b6* confirmed that our transformation method was suitable for engineering the monoterpenoid pathway.
- The metabolomic analysis of transgenic tobacco plants expressing *AtGpps* and *VoGes* indicated that terpenoid biosynthesis was completely suppressed in the transgenic plants, highlighting the need to understand the multilevel regulation of endogenous metabolic pathways for the effective metabolic engineering of secondary products.
- The unexpected impact of *AtGpps* and *VoGes* expression on non-target pathways suggested that a stress response is triggered in the transgenic tobacco plants, resulting in the accumulation of protective metabolites such as nicotine and betaine, and the depletion of others such as choline and inositol.
- The quantitative analysis of the flower proteome in two transgenic tobacco plants expressing high levels of *VoGes* and *Cyp76b6* (and wild type plants for comparison) by iTRAQ/MALDI-MS/MS indicated there were no significant changes. The GES and CYP76B6 proteins were not detected in the transgenic plants despite the presence of the corresponding transcripts, suggesting that these proteins accumulate at levels below the detection threshold of our proteomic methods.
- The transient expression of *VoGes* showed that the gene is functional and can produce trace amounts of geraniol and its derivatives, as expected. However,

stable expression did not result in the production of geraniol presumably because the longer-term accumulation of this compound is toxic.

- The transient expression of downstream genes in the secoiridoid pathway did not induce earlier-acting enzymes because no downstream metabolites were detected (10-hydroxygeraniol, 10-hydroxygeranial, 10-oxogeranial and iridodial). However, LC-MS/TOF analysis revealed 60 new compounds in transgenic agroinfiltrated lines expressing *pineGpps*, *VoGes*, *cyp76b6*, *10-hgo* and *mpc*. Although these compounds could not be identified, preliminary data suggested that the expression of downstream genes led to the production of diverse unknown metabolites.
- The inability to regenerate transgenic plants expressing either the upstream or downstream combinations of secoiridoid biosynthesis transgenes prompted us to investigate the potential metabolomic impact of the introduced heterologous pathway.
- Further experiments showed that the *CrGes* gene product was toxic, preventing the regeneration of transgenic plants. The downstream combination of genes also had a negative impact, inducing a dwarf phenotype.
- A tiered, high-quality patent search strategy was developed, allowing the identification of patents relevant to the engineering and production of TIAs in plants.
- This strategy allowed us to identify approximately 50 relevant patents covering the major genes/proteins, products, intermediates and methods that need to be negotiated, as well as the previously-described pathways and genetically-modified organisms relevant to TIA metabolic engineering.