Gene Discovery in *Catharanthus roseus* using Virus Induced Gene Silencing

Richard Michael Eric Payne

This thesis is submitted in fulfilment of the requirements of the degree of Doctor of Philosophy at the University of East Anglia

Department of Biological Chemistry

John Innes Centre

Norwich

September 2015

Abstract

This thesis presents the use of Virus Induced Gene Silencing (VIGS) for the discovery of enzymes and transporters involved in monoterpene indole alkaloid (MIA) metabolism in the medicinal plant *Catharanthus roseus.*

C. roseus is the source of a number of MIAs that are used as chemotherapeutic agents in the treatment of a variety of cancers, however the complete biosynthetic pathway for these metabolites remains to be elucidated. Additionally, this metabolic pathway is subcellulary compartmented with the key branch point enzyme, strictosidine synthase, localised to the plant vacuole. There is therefore a need for the import of the substrates for strictosidine biosynthesis; secologanin and tryptamine, across the vacuolar membrane, and export of the product, strictosidine, for synthesis of the downstream alkaloids.

This thesis presents the identification of two proteins that act as trans-tonoplastic transporters in MIA metabolism. The multidrug and toxic compound extrusion (MATE) protein, CrMATE1952, was localised to the vacuolar membrane and silencing its expression *in planta* resulted in the accumulation of a secologanin derivative. This implicates CrMATE1952 in the transport of secologanin into the vacuole and highlights the importance of the spatial organisation of the pathway in preventing secologanin derivatisation. Secondly silencing the expression of a tonoplast localised nitrate/peptide (NPF) transporter, CrNPF2.9, resulted in the 20-fold accumulation of strictosidine, suggesting this transporter is the exporter of strictosidine from the vacuole. Furthermore, VIGS also allowed the identification of a reticuline oxidase like protein, CrRO, which resulted in the accumulation of two new MIAs in leaf tissue upon silencing.

This thesis highlights a reverse genetics strategy for gene identification in metabolic pathways and is the first time the MATE and NPF transporters, and the reticuline oxidase like enzymes, have been shown to be involved in MIA metabolism in *C. roseus.*

Acknowledgements

First and foremost, I would like to thank Prof. Sarah O'Connor for taking me into your lab after completing the rotation programme at the John Innes Centre. You have given me the freedom to explore interesting science and have independence, yet always offered guidance, support and discussion, that has kept me going when things weren't working, and inspired me when things were.

I would like to thank Nick Brewin, Mike Merrick and my rotation supervisors Paul O'Maille, Matthew Moscou and Brande Wulff. The first year of the PhD helped to shape how I think about science and made me realise the type of questions that drive me.

To Fernando Geu-Flores, Nat Sherden, Wes Glenn (the tom to my jerry) and John Cheng; you were new to the UK when I started at JIC, and took me under your collective wings. The initial rotation and supervision with you four shaped me as a scientist, and I can't thank you enough for gently persuading me to re-join the O'Connor group, and your continued friendship. To Franziska and Anna; it's been a pleasure being a PhD student with you two, and I look forward to riding the Norwich SOS bus together once we have all graduated. It seems crazy that the last three years have gone by so quickly. To Vangelis Tatsis, Hajo Kries, Lorenzo Caputi, Dorota Jakubczyk, Thu-Thuy Dang, Mohammed Kamileen, Stephanie Brown, Marc Jones and all O'Connor group members past and present, thank you for all your help in the lab, for fruitful discussion and criticism, and making it a slightly crazy but very enjoyable working environment.

To collaborators at other institutes and universities; I would like to thank Vincent Courdavault, Alex Van Moerkercke, Alain Goosens, Hussam Hassan Nour-Eldin, Joanne Nettleship and Ray Owens for continuing support with our research.

At the JIC we are blessed to have a great metabolomics facility, and I would like to especially thank Lionel Hill for all his guidance and support with mass spectrometry, as well as Gerhard Saalbach and Lorenzo Caputi with support with mass spectrometry imaging. Additionally I would like to thank the support of both Grant Calder and Alison Pendle for guidance with microscopy, Andrew Davies in the photography department, as well as the support of the media kitchen staff, horticultural services and stores for keeping the labs running so efficiently.

The Biological Chemistry department at JIC is a great place to work, even if we occasionally lose in the football tournament, and I would like to thank everyone past and present that has made it that way.

I was lucky enough to join the JIC when a great bunch of people were also starting their PhD's, and I would like to thank you all for your friendship and for all the good times that we have had over the four years; Robert Ietswaart, Jo Harrison, Rachel Goddard, Chris Judge, Artemis Giannakopoulou, Olu Shorinola, Ania Kowalski, Rowena Fung and Scott Berry. Annis Richardson, Lucy Gannon and Claire Drurey; you three are awesome. You have kept me well fed for four years, gave me a home for 3 months, and improved my bakery skills, I can't ask for more than that. John Steele; I look forward to keeping on playing guitar with you and enjoying the culinary delights that you bring back from your travels. Mike Rugen; I've had to put up with sharing a lab with you for four years, and to be honest I have enjoyed it, especially when it has involved BBQ, beer, punk music and rugby. Thanks for being my dinner date for all those late nights thesis writing in the library.

To all my friends at the JIC, it has been a great experience; Alex Holden (thanks for tea and grapes and making thesis writing bearable), Hadrien Peyret (one day you might win at poker and thanks

for being a great housemate), Farzana Miah, Philippa Borrill, Tilly Eldridge, Ben Wagstaff, John Cocker, Fred Collin, Lukasz Langowski, Michael Piperakis and Daniel Garcia Seco.

To my friends from university and home, thank you for keeping me grounded. It's always nice to visit and get away from the PhD bubble; Rachel Cruickshank, Jenny Ellis, Marianne Alton, Georg Hochberg, Paul Chappell, Alec Symes, Alan Gibbs, Josh Dack, Sam Binding, Aaron Carter-Whittley, Chris Scutt, Steve Flower, Daniel Ellis and Lloyd Stickels.

To my teachers at secondary school and university professors; you have made education enjoyable and inspired me to pursue a career in science. I would like to thank Mr Bowman, Mr Tebbutt, Dr Wilson, Mr Burney, Prof Andre Furger and Prof Nick Kruger. Thank you for instilling in me the value of constantly learning.

Mostly I would like to thank my family. To my Mum and Dad; for your continuing love and support, believing that I can achieve what I want to and always being there when I have needed you. To my little brother, Alex, I wish you the best of luck as you finish school and start your university career; I hope I haven't put you off science too much! To my Grandpa, A. Sue, and Nan and late Grandad, thank you for always being there, it's always nice to see you and speak to you, and have your encouragement.

Finally to Nana; you were one of the few people to have read my master's thesis cover to cover, and you have gotten away lightly not having to read this one. This is dedicated to you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

This thesis uses standard abbreviations for nucleic acids (one letter code) and amino acids (one and three letter codes). Standard SI units are also employed unless otherwise stated.

16-OMT	16-hydroxytabersonine O-methyltransferase
7-DLGT	7-deoxyloganetic acid glucosyltransferase
7-DLH	7-deoxyloganic acid hydroxylase
8-HGO	8-hydroxygeraniol reductoisomerase
ACN	Acetonitrile
BBE	berberine bridge enzyme
BCA	bicinchoninic acid
BLAST	basic local alignment search tool
bp	base pair
BSA	bovine serum albumin
CBDA	cannabidiolic acid
CBGA	cannabigerolic acid
CCCP	carbonyl cyanide *m*-chlorophenyl hydrazone
cDNA	complementary DNA
CFP	cyan fluorescent protein
CHOX	carbohydrate oxidase
CoA	Coenzyme A
cRNA	complementary ribonucleic acid
D4H	desacetoxyvindoline 4-hydroxylase
DAPI	4',6-diamidino-2-phenylindole
DAT	deacetylvindoline acetyltransferase
DBOX	dihydrobenzophenanthridine oxidase
DESI	desorption electrospray ionisation
DMAPP	dimethylallylpyrophosphate
DMN	dihydrometanicotine
DMSO	dimethyl sulfoxide
DMT	drug/metabolite transporter
dNTPs	deoxynucleotide triphosphates
dsRNA	double stranded ribonucleic acid
DTT	dithiothreitol
DXR	1-deoxy-D-xyulose 5-phosphate reductoisomerase
DXS	1-deoxy-D-xyulose 5-phosphate synthase
EDTA	ethylenediaminetetraacetic acid
EIC	extracted ion chromatogram
ESI	electrospray ionisation
EtOH	ethanol
EV	empty vector
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
FPKM	fragments per kilobase of transcript per million reads mapped
G8H	geraniol-8-hydroxylase
GALDH	L-galactono-1,4-lactone dehydrogenase

GFP	green fluorescent protein
Glc	glucose
GOOX	glucooligosaccharide oxidase
GPP	geranyl pyrophosphate
GSH	glutathione
GST	glutathione *S*-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
IAA	indole acetic acid
IO	iridoid oxidase
IPAP	internal phloem associated parenchyma
IPP	isopentyl pyrophosphate
IPTG	isopropyl β-D-galactopyranoside
IS	iridoid synthase
IT-ToF	Ion trap time of flight
kDa	kilodalton
LAMT	loganic acid methyltransferase
LB	lysogeny broth
LC-MS	liquid chromatography mass spectrometry
MATE	multidrug and toxic compound extrusion
MBS	Modified Barths Solution
MECS	2C-methyl-D-erythritol 2,4 cyclodiphosphate synthase
MeJA	methyl jasmonate
MeOH	methanol
MEP	2-C-methyl-D-erythritol 4-phosphate
MES	2-(*N*-morpholino)ethanesulfonic acid
MeV	multiexperiment viewer
MFS	major facilitator superfamily
MIA	monoterpene indole alkaloid
MOPS	3-(N-morpholino)propanesulfonic acid
MPGR	medicinal plant genomics resource
MSI	mass spectrometry imaging
MVA	mevalonate
NAD +	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide (reduced)
NADP +	nicotinamide adenine dinucleotide phosphate
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NMR	Nuclear Magentic Resonance
NMT	16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase
NPF	nitrate/peptide transporter family
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
pI	protein isoelectric point
PTFE	polytetrafluroethylene
qPCR	quantitative PCR

RNAi	ribonucleic acid interference
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sf9	*Spodoptera frugiperda* clonal cell line
SGD	strictosidine β-glucosidase
siRNA	small interfering ribonucleic acid
SLS	secologanin synthase
SOM	self-organising map
ssRNA	single stranded ribonucleic acid
STOX	(*S*)-tetrahydroprotoberberine oxidase
STR	strictosidine synthase
T16H	tabersonine 16-hydroxylase
T3O	tabersonine 3-oxygenase
T3R	tabersonine 3-reductase
TDC	tryptophan decarboxylase
TE	Tris-EDTA buffer
THCA	tetrahydrocannabinolic acid
TIC	total ion chromatogram
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
TRV	tobacco rattle virus
USER	uracil specific excision reagent
VAO	vanillyl alcohol oxidase
VIGS	virus induced gene silencing
WT	wild type
YFP	yellow fluorescent protein

1 Introduction

1.1 Plant Alkaloids

Plant-derived alkaloids are a diverse class of compounds that have had profound societal impact throughout human history (Dewick, 2002). These compounds contain a basic nitrogen, and are typically derived from an amino acid precursor, such as tryptophan, tyrosine, lysine, arginine or histidine (Dewick, 2002). From a medical perspective these compounds have been richly exploited for millennia due to their diverse range of pharmacological activities. These compounds have been used in the treatment of pain- morphine (Lipp, 1991), used as local anaesthetics- cocaine (Ruetsch et al., 2001), as anti-asthmatic agents- ephedrine (Maitai, 1975), anti-arrythmic agents- quinidine (Markel et al., 1993), anti-hyptertension drugs-reserpine (Doyle et al., 1955) and anti-cancer agents- vinblastine (Dyke and Nelson, 1977) (Figure 1-1).

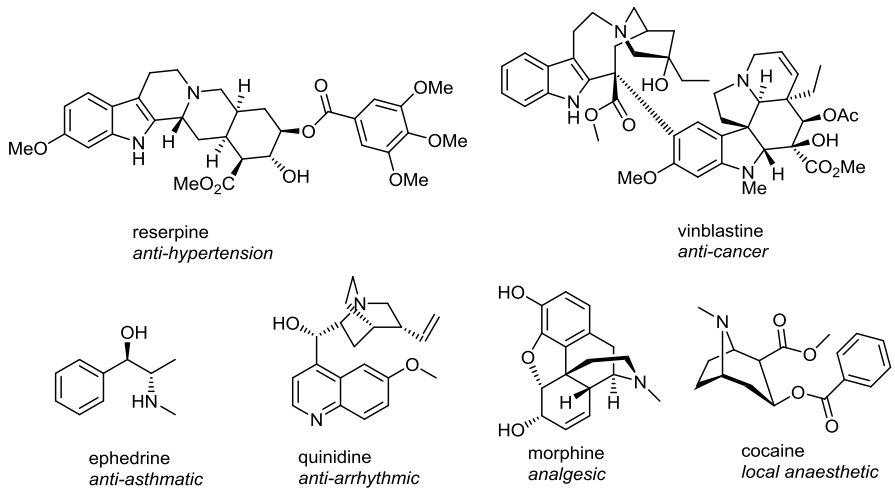


Figure 1-1 Representative plant alkaloids of medical importance
As highlighted, alkaloids are chemicals that contain a basic nitrogen, and have a broad range of structural scaffolds

As well as the diverse range of pharmacological functions, from a chemical perspective, these compounds are structurally diverse, as highlighted in Figure 1-1, and to date there has been concerted effort in the elucidation of the biosynthetic pathways of a number of plant alkaloids.

The chemical complexity of natural products, due to the diversity in ring structures, oxygenation patterns, and stereocentres, together with the fact that these natural product scaffolds can also be further modified by glycosylation and other tailoring reactions, makes these compounds

challenging for chemical synthesis on an industrial scale (Leonard et al., 2009). Additionally, the plant alkaloids are often isolated in low yields from the plant, and therefore the extraction of these compounds is labour intensive and requires the cultivation of large quantities of the plants, which in turn increases the cost of these small molecules (Leonard et al., 2009; Ro et al., 2006)

The elucidation of the biosynthetic pathway and subsequent reconstitution in a heterologous host offers the potential for a platform that would allow for the high production of expensive compounds that can be easily extracted and grown in a fermentable system on cheap inputs, such as glucose.

This has most recently been seen with the reconstitution of morphine biosynthesis in *Saccharomyces cerevisiae* (DeLoache et al., 2015; Fossati et al., 2015; Winzer et al., 2015), the reconstitution of artemisinin, a terpene natural product from *Artemisia annua* used in the treatment of malaria (Ro et al., 2006) in *S. cerevisiae*, as well as in the reconstitution of strictosidine biosynthesis in *S. cerevisiae,* which as will be highlighted subsequently in this chapter, is a key metabolic branchpoint for monoterpene indole alkaloid (MIA) biosynthesis, (Brown et al., 2015).

Additionally, such biosynthetic platforms allow diversification of the structural backbones of these natural products, either in the generation of 'unnatural' natural products, as highlighted through halogenation (Glenn et al., 2011) or through the combinatorial use of enzymes from diverse plant families to generate new products that would not exist in nature.

1.2 The pathway of monoterpene indole alkaloid biosynthesis

The MIA pathway is commonly encountered in the Apocynaceae, Loganiaceae and Rubiaceae plant families (De Luca et al., 2012) and the focus of this thesis will be centred on the biosynthesis of the MIAs present in the Madagascar periwinkle, *Catharanthus roseus* (Apocynaceae). *C. roseus* produces in excess of 130 MIAs, with a diverse range of structural scaffolds (O'Connor and Maresh, 2006). The two alkaloids of highest economic importance are the anti-cancer agents, vincristine and vinblastine, which are used in the clinical treatment of a variety of cancers. These compounds act by inhibiting cell division, by interacting with α and β tubulin, thereby preventing the polymerisation of microtubules during cell division (Bowman et al., 1988).

1.2.1 Biosynthesis of secologanin, tryptamine and strictosidine

All MIAs are derived from a convergent metabolic pathway that utilises an amino acid product from the shikimate pathway, tryptophan, as well as the seco-iridoid glycosylated terpene, secologanin (O'Connor and Maresh, 2006).

The entire pathway for the formation of secologanin has been elucidated. Secologanin is derived from the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway for terpene synthesis (Contin et al., 1998). Isopentyl diphosphate (IPP), the precursor for all terpenoids, can be produced by either the mevalonate or MEP pathways, however [1-^{13}C]-glucose feeding studies of *C. roseus* cell suspension cultures suggest that the dominant route of secologanin synthesis is via the MEP pathway (Contin et al., 1998). The MEP pathway produces geranyl-pyrophosphate (GPP) which is used for the biosynthesis of secologanin.

Geraniol, which is derived from GPP via geraniol synthase (Simkin et al., 2013), is hydroxylated by geraniol 8-hydroxylase (G8H), CYP76B6, to 8-hydroxygeraniol. G8H can further oxidise 8-hydroxygeranial, forming 8-oxogeraniol, together with the enzyme 8-hydroxygeraniol oxidoreductase, to the linear dialdehyde, 8-oxogeranial (Miettinen et al., 2014).

This molecule, 8-oxogeranial, is cyclised by iridoid synthase (Geu-Flores et al., 2012), resulting in the equilibrium mixture of cis-trans-iridodial and cis-trans-nepetelactol. A cytochrome p450, iridoid oxidase (IO), CYP76A26, oxidises cis-trans-neptelactol to 7-deoxyloganetic acid (Miettinen et al., 2014; Salim et al., 2014) which is subsequently glycosylated by 7-deoxyloganetic acid glucosyltransferase (Asada et al., 2013; Miettinen et al., 2014), hydroxylated by a cytochrome p450, 7-deoxyloganic acid hydroxylase (7DLH) (Salim et al., 2013), CYP72A224, and methylated by loganic acid methyltransferase (LAMT) to form loganin (Murata et al., 2008), before being oxidatively cleaved by the cytochrome p450 secologanin synthase (SLS), CYP72A1, to yield the iridoid terpene secologanin, (de Bernonville et al., 2015; Irmler et al., 2000; Kellner et al., 2015b) (Figure 1-2).

In parallel, the amino acid tryptophan is decarboxylated by the pyridoxal phosphate dependent enzyme, tryptophan decarboxylase (Pennings et al., 1989) to form tryptamine, (Figure 1-2).

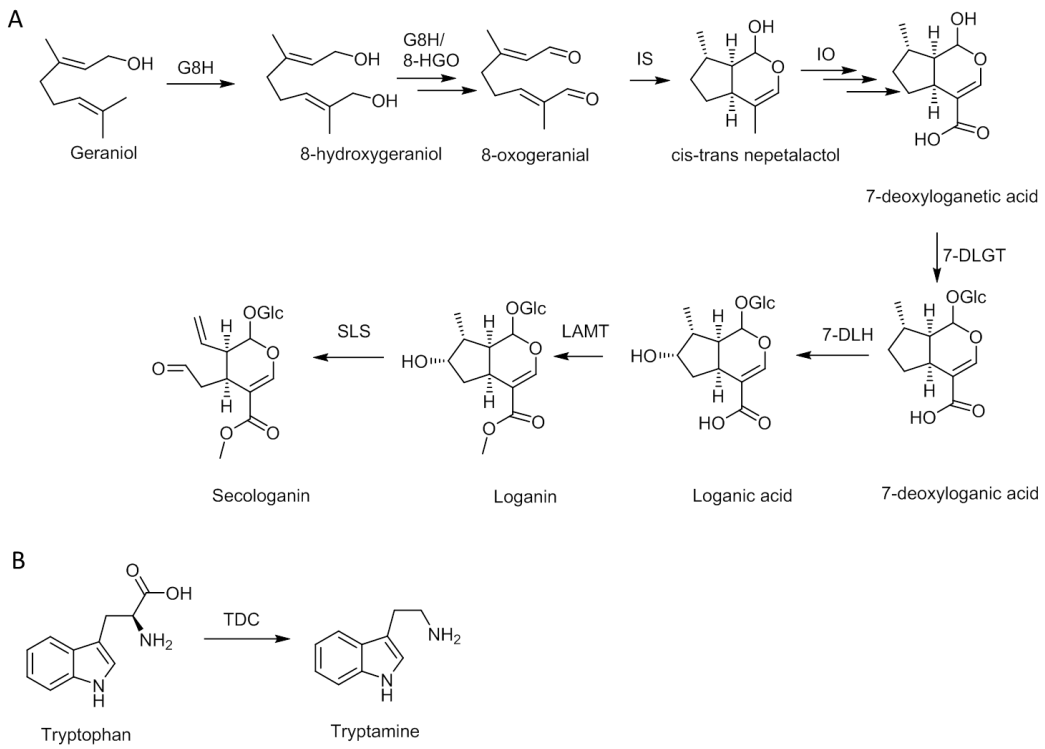


Figure 1-2 Overview of secologanin and tryptamine biosynthesis
A Secologanin is a glycosylated iridoid terpene derived from the MEP pathway. It is synthesised by a nine-step biosynthetic pathway comprising the enzymes G8H-geraniol 8-hydroxylase, 8-HGO- 8-hydroxygeraniol oxidoreductase, IS-iridoid synthase, IO-iridoid oxidase, 7-DLGT- 7deoxyloganetic acid glucosyltransferase, 7-DLH - 7-deoxyloganic acid hydroxylase, LAMT-loganic acid methyltransferase and SLS- secologanin synthase. B Tryptamine is synthesised by the decarboxylation of tryptophan by the enzyme tryptophan decarboxylase (TDC).

The first committed step in the formation of MIAs is the Pictet-Spengler condensation of secologanin and tryptamine to form the central metabolic intermediate, strictosidine, which represents a key branchpoint in this pathway (Maresh et al., 2008; Treimer and Zenk, 1979). The deglycosylation of this central intermediate results in a reactive di-aldehyde species, that subsequently rearranges to 4,21-dehydrogeissoschizine, an intermediate from which nearly all structural scaffolds of MIAs are derived (O'Connor and Maresh, 2006) (Figure 1-3). The reconstitution of strictosidine biosynthesis has been achieved in both *S. cerevisiae* and *N. benthamiana* (Brown et al., 2015; Miettinen et al., 2014).

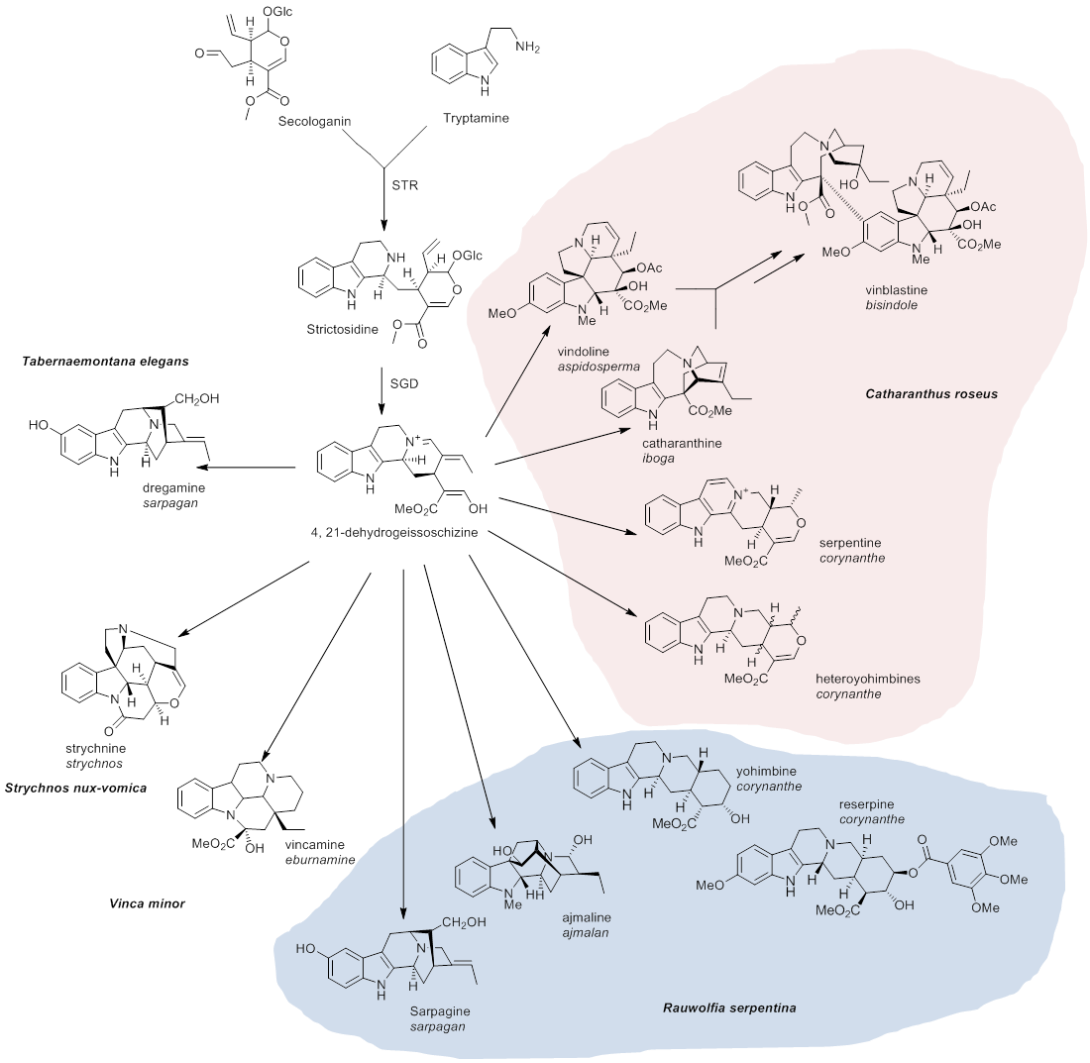


Figure 1-3 Strictosidine is the central branchpoint for MIA production in a broad range of plant species Secologanin and tryptamine condense by the action of strictosidine synthase (STR) to form strictosidine, which is deglycosylated and rearranged to 4,21-dehydrogeissoschizine. This forms the structural basis for a diverse range of alkaloids in multiple plant species. Highlighted in red are alkaloids present in *C. roseus*, highlighted in blue are alkaloids from *Rauwolfia serpentina*, and other structural scaffolds from *Vinca minor*, *Strychnos nux-vomica* and *Tabernaemontana elegans* are also represented. The common name of the alkaloid is shown, together with the structural class underneath in italic. *C. roseus* can also produce bisindole alkaloids which form as a result of the condensation of two MIAs, as shown with vinblastine, which is derived from catharanthine, an iboga alkaloid, and vindoline, an aspidosperma alkaloid.

1.2.2 Biosynthesis of the aspidosperma, iboga and corynanthe alkaloids

C. roseus synthesises vincristine and vinblastine by dimerising the aspidosperma alkaloid, vindoline, and the iboga alkaloid, catharanthine (Figure 1-3).

The enzymatic steps from deglycosylated strictosidine to catharanthine and the precursor to vindoline, tabersonine, are still unknown. In comparison, the conversion from tabersonine to vindoline has been fully characterised (Figure 1-4) and comprises seven enzymatic steps.

Tabersonine is hydroxylated by tabersonine 16-hydroxylase (T16H) (St-Pierre and De Luca, 1995). Interestingly this is one of the first enzymes in *C. roseus* to have two isoforms that show tissue specific localisation to the leaves and flowers (Besseau et al., 2013). This hydroxyl group is methylated by 16-hydroxytabersonine *O*-methyltransferase (Levac et al., 2008). 16-methoxytabersonine is hydroxylated and reduced by the concerted action of two enzymes, tabersonine 3-oxygenase (T3O), and tabersonine 3-reductase (T3R) (Kellner et al., 2015a; Qu et al., 2015). The nitrogen derived from the indole ring of tryptophan is methylated (Liscombe et al., 2010), and this substrate forms the basis for the hydroxylation of desacetoxyvindoline by desacetoxyvindoline 4-hydroxylase (D4H) (De Luca and Cutler, 1987), and its subsequent acetylation by deacetylvindoline acetyltransferase (DAT) (St-Pierre et al., 1998), to form the final product vindoline.

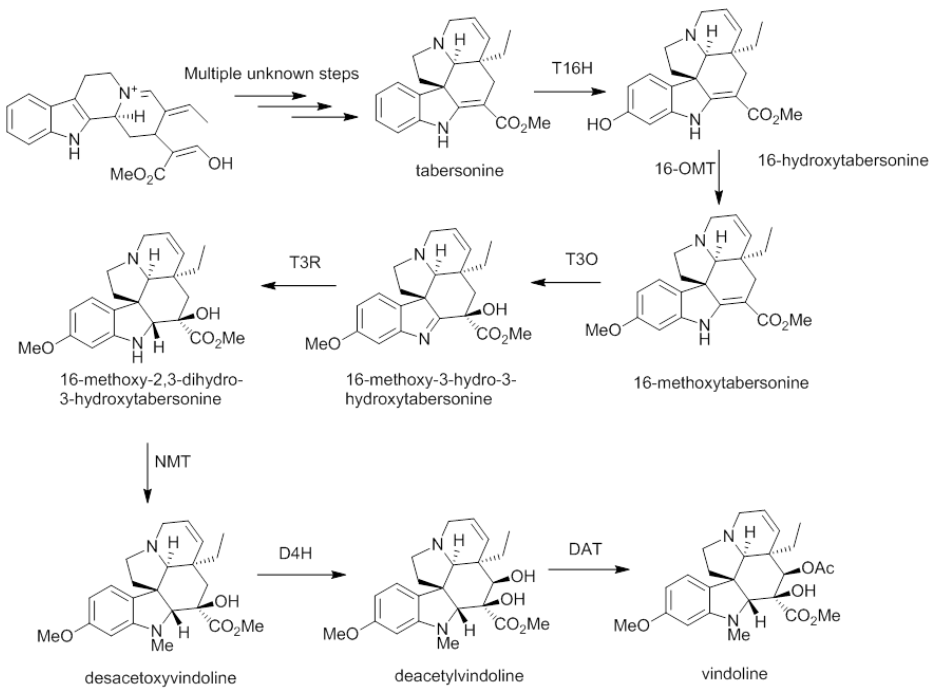


Figure 1-4 The seven step biosynthesis of vindoline from tabersonine.
T16H- tabersonine 16-hydroxylase, 16OMT- 16-hydroxytabersonine 16-O-methyltransferase, T3O- tabersonine 3-oxygenase, T3R- tabersonine 3-oxygenase, NMT- 16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase, D4H- desacetoxyvindoline 4-hydroxylase, DAT- deacetylvindoline acetyltransferase.

One of the major limitations in our understanding of the MIA pathway is in the discovery of enzymes that act on the major branchpoint of the pathway, namely the deglycosylated product of strictosidine. This part of the pathway is fundamental to the shuttling of intermediates into all MIA structural scaffolds in *C. roseus,* however to date the only enzyme that has been shown

to work on the deglycosylated product of strictosidine, is an alcohol dehydrogenase that is involved in the formation of the corynanthe class of alkaloids, and catalyses the conversion of strictosidine aglycone to tetrahydroalstonine (Stavrinides et al., 2015) (Figure 1-5).

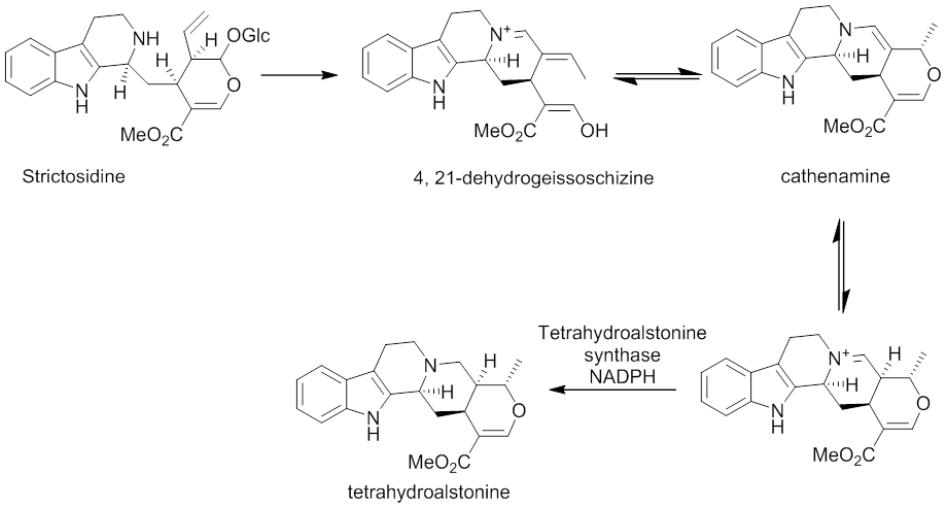


Figure 1-5 Reaction of tetrahydroalstonine synthase.
Tetrahydroalstonine synthase catalyses the formation of tetrahydoalstonine from the deglycosylated product of strictosidine in an NADPH dependent reaction.

There are therefore a number of key missing enzymatic steps in vinblastine and vincristine biosynthesis, and this opens up a plethora of interesting questions with regards this metabolic pathway. Firstly, these missing steps are predicted to catalyse novel chemistry and secondly it is interesting to understand how the plant diverts the common substrate, strictosidine aglycone, into a minimum of at least three separate metabolic pathways for the synthesis of the different structural scaffolds of MIAs present in *C. roseus* (corynanthe, iboga and aspidosperma), especially as this dialdehyde has shown to be very reactive with the potential to crosslink proteins, nucleic acids and free amines (Guirimand et al., 2010). An overview of the known steps in the pathway to vindoline biosynthesis is highlighted in Figure 1-6.

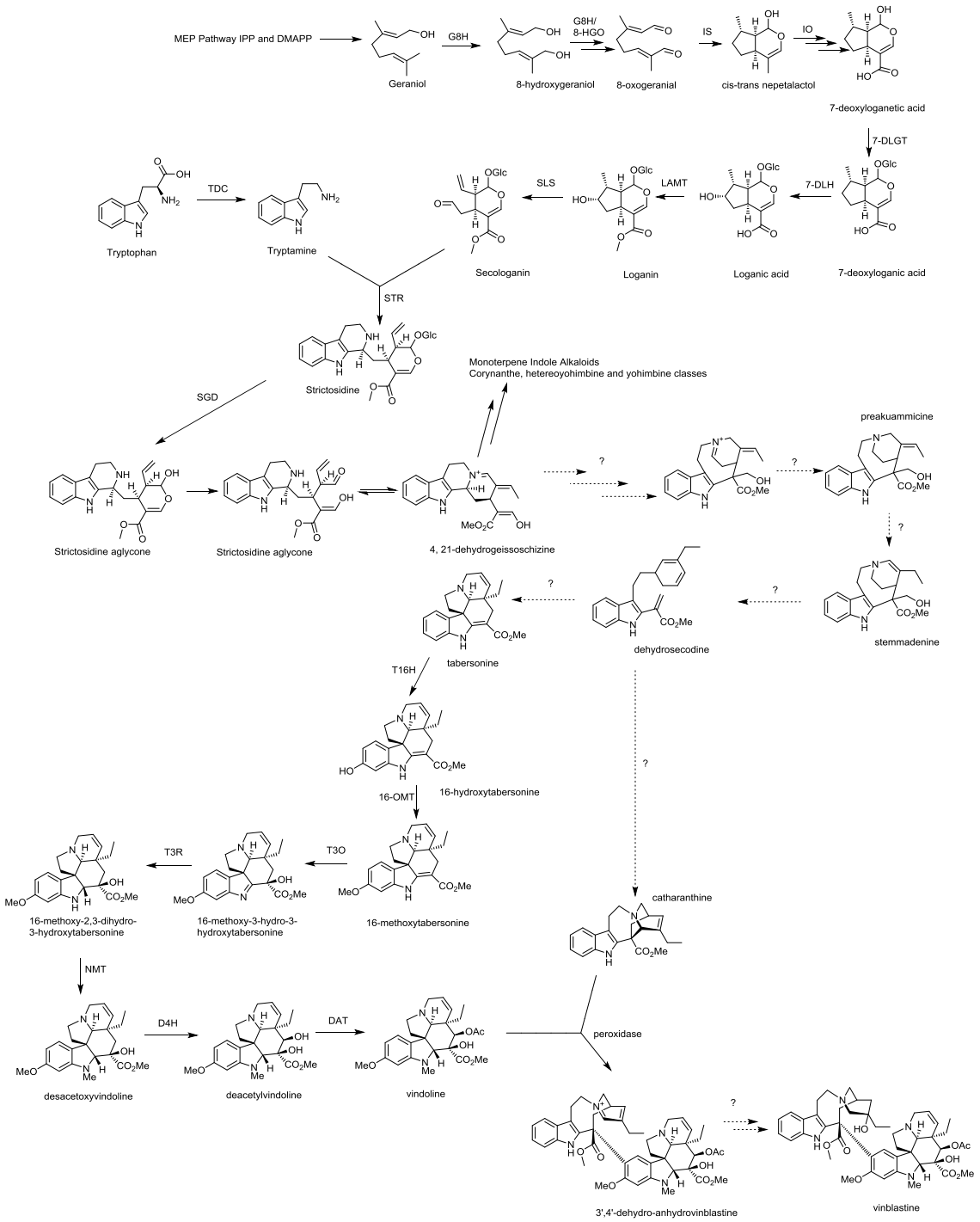


Figure 1-6 Overview of the pathway to vinblastine biosynthesis in *C. roseus*

Vinblastine biosynthesis encompasses a pathway that utilises in excess of 30 enzymes, through the convergence of seco-iridoid and amino acid metabolism, and the common branchpoint intermediate strictosidine. Known enzymes are highlighted above the reaction arrows, whilst unknown reactions are highlighted by a ?.

1.3 The inter- and intra-cellular distribution of MIA metabolism in *C. roseus*

The pathway for MIA biosynthesis is further complicated in *C. roseus* due to the fact that it is compartmentalised at both the cellular and intra-cellular level. This indicates that there is a requirement for transport of biosynthetic intermediates among different cell types, and between organelles (Courdavault et al., 2014).

1.3.1 Intercellular distribution of MIA metabolism

At least three cell types are necessary for MIA biosynthesis. The sequencing of the leaf "epidermome" to generate a cDNA library enriched in transcripts localised to leaf epidermis (Murata et al., 2008) suggests that the enzymes involved in secologanin biosynthesis, 1-deoxy-D-xyulose 5-phosphate synthase (DXS), 1-deoxy-D-xyulose 5-phosphate reductoisomerase (DXR) and 2C-methyl-D-erythritol 2,4-cyclodiphsophate synthase (MECS) of the MEP pathway and G8H for the generation of 8-hydroxygeraniol are not localised to the leaf epidermis. This observation is supported by *in situ* hybridisation studies (Burlat et al., 2004) which indicate these enzymes are localised to the internal phloem parenchyma (IPAP) cells of young organs. Additionally 8-hydroxygeraniol oxidoreductase, (Miettinen et al., 2014), iridoid synthase (Geu-Flores et al., 2012), 7-deoxyloganetic acid glucosyltransferase and 7-deoxy loganic acid hydroxylase (Miettinen et al., 2014) have also been shown to be preferentially expressed in the IPAP tissue by *in situ* hybridisation.

In comparison tryptophan decarboxylase, loganic acid methyltransferase, secologanin synthase and strictosidine synthase, as well as the enzymes downstream in the biosynthetic pathway, taersonine 16-hydroxylase and 16-*O*-methyltransferase involved in late stage vindoline biosynthesis, are preferentially expressed in the epidermis (Irmler et al., 2000; Murata and De Luca, 2005; Murata et al., 2008; St-Pierre et al., 1999). There is therefore a requirement for intercellular transport of a terpene intermediate from IPAP cells to epidermal cells in the leaf. To date, based on the localisation of the biosynthetic enzymes in the pathway, the most likely intermediate for inter-cellular transport between the IPAP cells and the epidermis is loganic acid (Miettinen et al., 2014).

The biosynthesis of vindoline from tabersonine is also likely to be inter-cellularly compartmented. Metabolite analysis of leaf epidermis enriched extracts, generated by carborundum abrasion, demonstrated an accumulation of tabersonine and 16-methoxytabersonine in comparison to vindoline (Murata and De Luca, 2005). Expression studies have shown that the vindoline biosynthetic genes T3O, T3R and NMT are found in higher levels

in the epidermis than in mesophyll cells (Qu et al., 2015), whilst the enzymes catalysing the last two steps in vindoline biosynthesis, D4H and DAT, are exclusively localised to the mesophyll (St-Pierre et al., 1999). This implies the requirement for the transport of the intermediate desacetoxyvindoline between the epidermis and mesophyll cells (Qu et al., 2015).

Biosynthesis of the iboga alkaloid catharanthine is thought to be localised to the leaf epidermis, as catharanthine is known to accumulate in leaf wax exudates. However there is also the requirement for transport between the epidermis and mesophyll for this compound, as the bisindole dimeric alkaloids, such as vinblastine, which form as a result of the condensation between vindoline and catharanthine, accumulate in the mesophyll, laticifers and idioblast cells.

1.3.2 Intracellular distribution of MIA metabolism

As well as the spatial distribution of the metabolic pathway between different cell types, the MIA pathway is localised to different subcellular organelles.

The enzymes of the MEP pathway are localised to the plastid for synthesis of the terpene precursor IPP (Hunter, 2007). Geranyl pyrophosphate is synthesised from the condensation of IPP and DMAPP by the enzyme GPPS. Geraniol synthase catalyses the formation of geraniol from geranyl pyrophosphate, which is then exported to the cytosol where it is hydroxylated by G8H, a cytochrome P450 associated with the endoplasmic reticulum membrane (Guirimand et al., 2009). Although (Contin et al., 1999a) suggested that the action of secologanin synthase (SLS) was localised to the vacuole, proposing a pathway for the vacuolar conversion of loganin to secologanin, it has since been established that SLS is anchored by an N-terminal helix to the cytosolic side of the endoplasmic reticulum, leading to the cytosolic production of secologanin (Guirimand et al., 2011a).

Tryptophan is synthesised from chorismate via an anthranilate intermediate in the plastid and is exported to the cytosol. Tryptophan decarboxylase (TDC) which catalyses the conversion of tryptophan to tryptamine is a cytosolic enzyme (Stevens et al., 1993) and it has been demonstrated by a combination of green fluorescent protein imaging, bimolecular fluorescence complementation assays and yeast two-hybrid analysis that both LAMT and TDC form homodimers in the cytosol (Guirimand et al., 2011a).

A key step for which substrate transport is required is in the biosynthesis of strictosidine. Strictosidine synthase is localised to the vacuole (Guirimand et al., 2010) and therefore its precursors, secologanin and tryptamine, must be transported from the cytosol across the tonoplast membrane. Similarly the product of the strictosidine synthase reaction, strictosidine,

subsequently needs to be exported from the vacuole as strictosidine β-glucosidase is localised to the nucleus (Guirimand et al., 2010).

The subcellular localisation for many of the late-stage vindoline biosynthetic enzymes is known. The enzyme tabersonine-16 hydroxylase (T16H), which catalyses the formation of 16-hydroxytabersonine in the cytosol, is a cytochrome P450 localised to the endoplasmic reticulum membrane (St-Pierre and De Luca, 1995) and *O*-methyltransferase (16-OMT) (Guirimand et al., 2011b), which directs synthesis of 16-methoxytabersonine, is cytosolic. The N-methyltransferase (NMT) which catalyses the formation of deascetoxyvindoline is associated with the peroxisomal membrane (Liscombe; personal communication), whilst the last two steps in the pathway desacetoxyvindoline-4-hydroxylase (D4H) and deacetylvindoline-4-*O*-acetyltransferase (DAT), are localised to the cytosol (De Luca and Cutler, 1987).

The bisindole alkaloids vincristine and vinblastine are generated from the coupling of catharanthine and vindoline, perhaps by the peroxidase CPRX1 generating the intermediate 3, 4-anhydrovinblastine. The peroxidase has been demonstrated to be localised to the vacuole (Sottomayor et al., 1998) suggesting the need for transport of both vindoline and catharanthine across the tonoplast membrane. However, the physiological role of this peroxidase has not been tested by silencing or knock-down experiments.

An overview of the current model of MIA inter- and intra- cellular transport is presented in Figure 1-7.

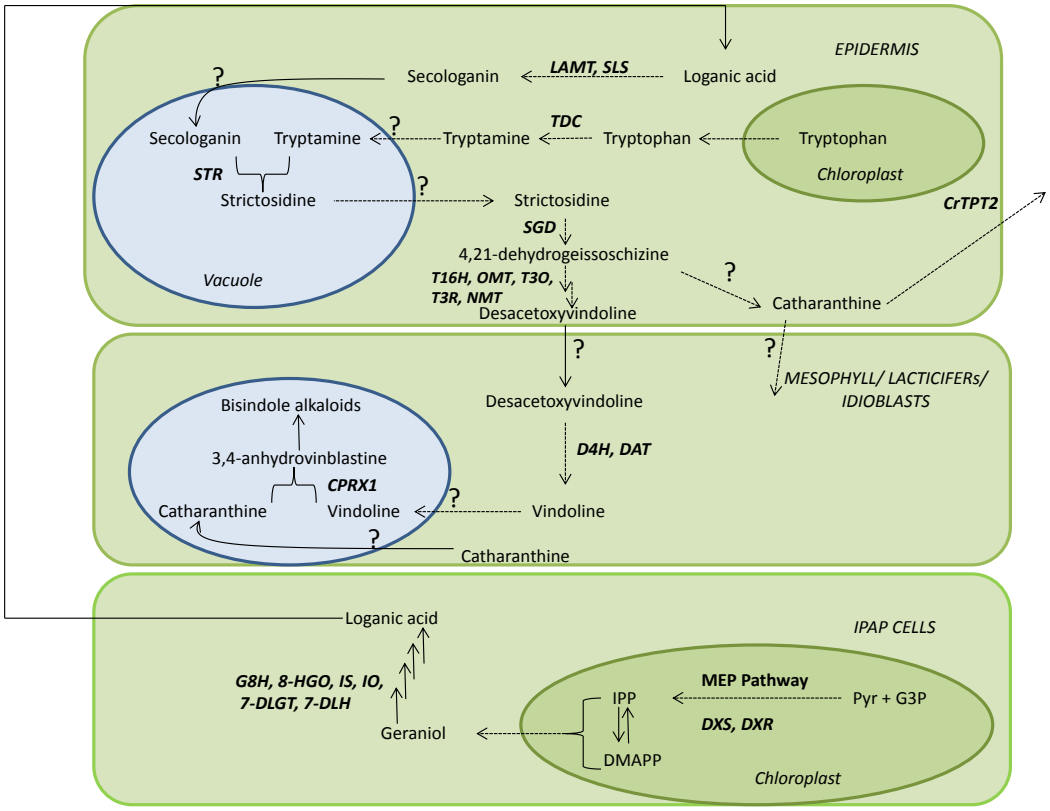


Figure 1-7 The current model of inter- and intra- cellular compartmentation in MIA biosynthesis.
The metabolic pathway to MIA biosynthesis in *C. roseus* spans three cell types, the internal phloem associated parenchyma (IPAP) cells, epidermis and the mesophyll cells. Unknown transport steps that require transmembrane movement across different organelles are highlighted by a ?.

1.4 Characterised *C. roseus* transporters

Based on the localisation of the known biosynthetic genes in the MIA pathway, there are therefore a number of hypothesised transporters that are required for the transmembrane transport of biosynthetic intermediates in MIA metabolism in *C. roseus*. This encompasses both the early part of the pathway involved in seco-iridoid biosynthesis, as well as in the formation of the downstream alkaloids.

Firstly, a terpene intermediate, which is likely to be loganic acid, is exported from the IPAP cells and imported into the epidermis (Miettinen et al., 2014).

Secondly, the intermediates secologanin and tryptamine must be imported into the vacuole, and the product of their condensation, strictosidine, must be exported into the cytosol. These intracellular transport activities occur in the epidermal cells (Courdavault et al., 2014).

Thirdly, the iboga alkaloid, catharanthine, needs to be exported to the waxy exudate on the leaf from the epidermis (Courdavault et al., 2014; Yu and De Luca, 2013).

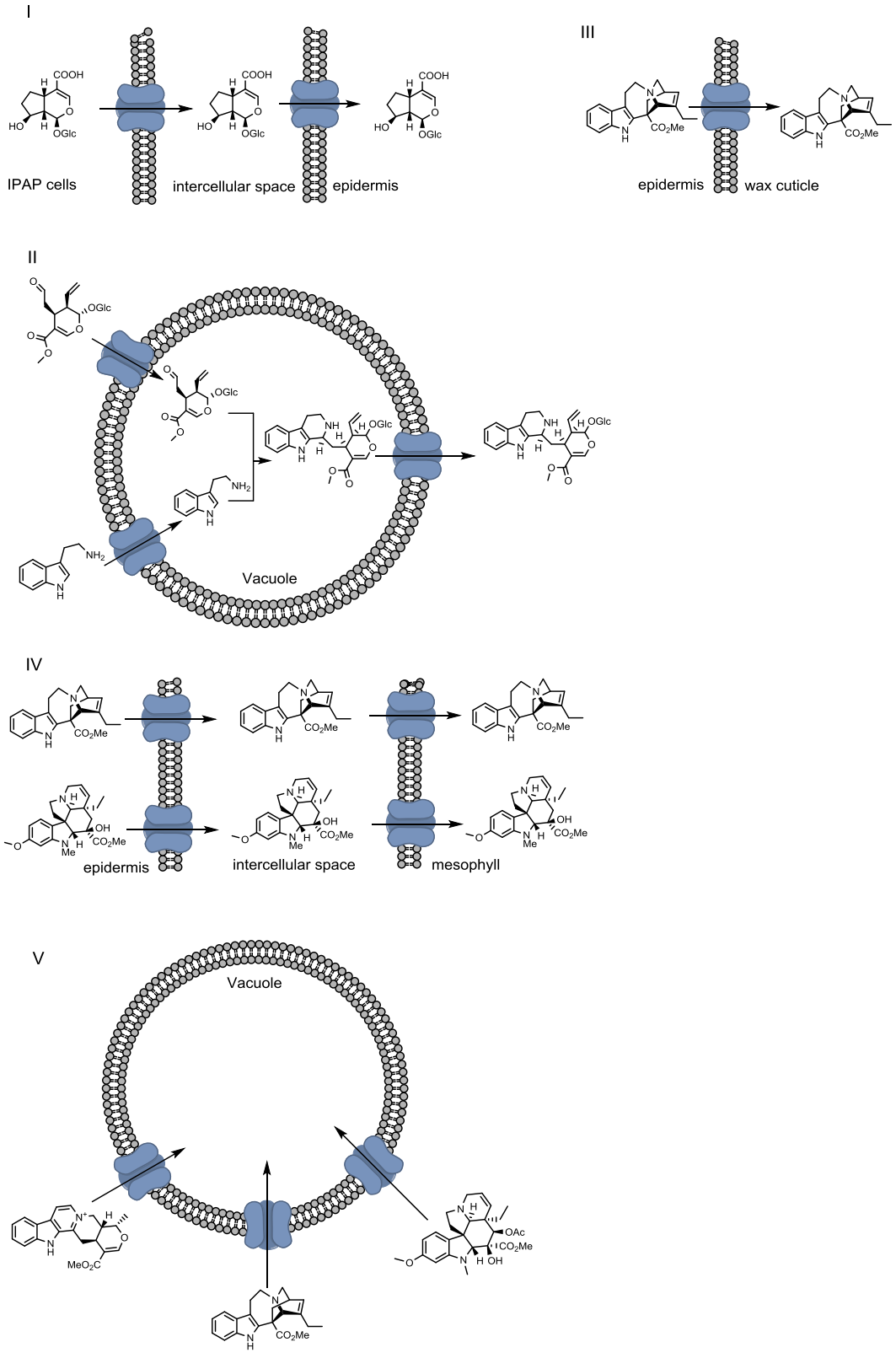
Fourth, both catharanthine, and the intermediate in the tabersonine-vindoline pathway, desacetoxyvindoline, need to be transported from the epidermis and into the mesophyll (Courdavault et al., 2014; Qu et al., 2015).

Lastly, the end products of the biosynthesis of MIAs, vindoline and catharanthine, as well as the other MIAs such as serpentine, are stored in the vacuole of the mesophyll (Courdavault et al., 2014).

To date there has been one report of a functionally characterised transporter from *C. roseus.* The exporter of catharanthine from the epidermis, to the waxy exudate, has been shown to be an ABC transporter, CrTPT2 (Yu and De Luca, 2013). This transporter was shown to be enriched in the leaf epidermome and could be induced by methyl jasmonate induction, which is a hormone that is known to lead to the upregulation of MIAs in *C. roseus*. This transporter localised to the plasma membrane and transports catharanthine when heterologously expressed in *S. cerevisiae*. Furthermore, transient silencing of this transporter by Virus Induced Gene Silencing (VIGS), resulted in a decrease in leaf catharanthine levels, indicating its physiological relevance to this pathway (Yu and De Luca, 2013).

Figure 1-8 Overview of the necessary transmembrane transport steps in the biosynthesis of MIAs in *C. roseus*

Overleaf: There are five main regions of the pathway that require transport I.) Movement of loganic acid between the IPAP cells and the epidermis II.) Movement of secologanin and tryptamine into the vacuole, and strictosidine out of the vacuole in the epidermis III.) Movement of catharanthine from the epidermis to the wax cuticle IV.) Movement of catharanthine and desacetoxyvindoline from the epidermis to the mesophyll and V.) Movement of downstream MIAs such as serpentine, catharanthine and vindoline into the mesophyll vacuole. The alkaloids presented here are likely to be protonated.


I
IPAP cells
intercellular space
epidermis
III
epidermis
wax cuticle
II
Vacuole
IV
epidermis
intercellular space
mesophyll
V
Vacuole

Additionally it has been shown utilising intact protoplasts and vacuolar membranes isolated from *C. roseus* leaf tissue that the transporters for vindoline and catharanthine uptake into the mesophyll vacuole are likely to be secondary active transporters that utilise the proton gradient (Carqueijeiro et al., 2013). Uptake of vindoline and catharanthine into the vacuole was shown to be ATP dependent, with the transport inhibited by NH_4Cl and carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP), which dissipate the proton gradient across the tonoplast membrane. This transport activity was not abolished upon treatment with the ABC transporter inhibitor, vanadate. This subsequently suggests that these vindoline and catharanthine importers are likely to be secondary active transporters. However, to date, no transporter has been identified that can perform this role (Carqueijeiro et al., 2013; Roytrakul and Verpoorte, 2007).

The remaining transport steps have yet to be elucidated, and the functional role this spatial distribution of MIA metabolism plays in *C. roseus* is still an enigma.

1.5 The role and importance in the spatial distribution of metabolism

The spatial distribution of metabolic networks is integral to their function, as metabolite transport between subcellular compartments fundamentally affects the structure of the metabolic network.

A genome wide compartmentalized model of the metabolic network in *Arabidopsis* using the known network structure and localisation of reactions in the cytosol, plastid, mitochondrion, endoplasmic reticulum, peroxisome, vacuole and Golgi apparatus, predicts that for a metabolic network comprising 1078 metabolites, there is a need for 772 intracellular transporters (Mintz-Oron et al., 2012). This therefore indicates that in excess of 70% of the metabolites in *Arabidopsis* are engaged in some form of transport. Clearly the role of transporters cannot be overlooked when considering engineering of plant metabolism, as the kinetics of transport is likely to directly affect metabolic fluxes.

There are two main hypotheses to explain why subcellular localisation of metabolic pathways occurs in eukaryotic cells. Firstly, subcellular localisation of metabolic pathways is likely to be related to the evolutionary history of the metabolic compartment. Organelles derived from endosymbiosis, such as the mitochondrion and the chloroplast, would result in the duplication of certain enzymes. If there is no selective pressure to keep both enzymes, it is feasible that the loss of enzyme functionality resulted in the mixed compartmentation of metabolic pathways,

and the need for metabolite transport between different organellar compartments. Examples of such pathways are in vitamin biosynthesis, such as with folic acid, as well as with cysteine biosynthesis (Sweetlove and Fernie, 2013; Takahashi et al., 2011).

However the spatial distribution of metabolism can also have direct physiological relevance to the metabolic pathway. The importance of localisation occurs at 1.) the organellar level and 2.) between organelles.

1.) The archetypal example of the importance of organellar compartmentation is with the specialised bioenergetic organelles such as the mitochondria and the chloroplasts. The requirement for strict control of redox potential necessitates the localisation of oxidative phosphorylation and the tricarboxylic acid cycle in the mitochondria. Coupling the carbon cycle to the electron transport chain is essential to minimise reactive oxygen species (ROS), and to control the redox chemistry of the electron transport chain (Lane and Martin, 2010).

2.) The spatial separation of duplicated metabolic pathways in different subcellular compartments also allows for the local control and regulation of the metabolic pathways. Separating parallel pathways in different organelles prevents the potential for futile cycling of intermediates, and secondly, it also has the potential to divert the end product substrate into different metabolic pathways. This is seen for the duplication in the biosynthesis of isopentenyl pyrophosphate (IPP), a key intermediate in terpene synthesis for secondary metabolism in plants. IPP is derived from either the canonical mevalonate (MVA) pathway, which utilises acetyl-CoA as substrate for biosynthesis of IPP and is localised to the cytosol, or the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway, which utilises D-glyceraldehyde 3-phosphate and pyruvate and is localised to the chloroplast. Compartmentalisation allows for differential flux through each parallel pathway and the maintenance of separate pools of IPP (Heinig et al., 2013).

In *C. roseus*, the functional role that spatial organisation plays in controlling flux through the MIA metabolic pathway is unclear (Courdavault et al., 2014).

The localisation and biosynthesis of strictosidine in the vacuole may control its contact with the next enzyme in the pathway, strictosidine β-glucosidase (Guirimand et al., 2010). Although the conversion of strictosidine to strictosidine aglycone is necessary in the cell for the formation of the downstream alkaloids such as vindoline and catharanthine, it is possible that the localisation of strictosidine in the vacuole provides a buffering mechanism to prevent accumulation of strictosidine in the cytoplasm and nucleus (Guirimand et al., 2010).

This in turn restricts the potential uncontrolled deglycosylation of strictosidine to the reactive dialdehyde product, strictosidine aglycone. The spatial separation of strictosidine, and the strictosidine β-glucosidase therefore acts as a pre-emptive buffer to prevent uncontrolled formation of the dialdehyde, that has the potential to crosslink free amines in the cell (Guirimand et al., 2010).

This metabolic scenario is analagous to the 'mustard oil bomb' or 'glucosinolate-myrosinase' complex in *Brassicaceae* (Bones and Rossiter, 1996). The glucosinolates are a class of glycosylated compounds that contain nitrogen and sulphur. Upon deglycosylation by a myrosinase enzyme these compounds can re-arrange to form a thiocyanate, an isothiocyanate or a nitrile, which are compounds active in plant defense against herbivory (Winde and Wittstock, 2011). The spatial separation at the cellular level, of the myrosinase and glucosinolates prevents the uncontrolled formation of these reactive compounds, however upon herbivory, this spatial distribution is destroyed resulting in their formation (Figure 1-9).

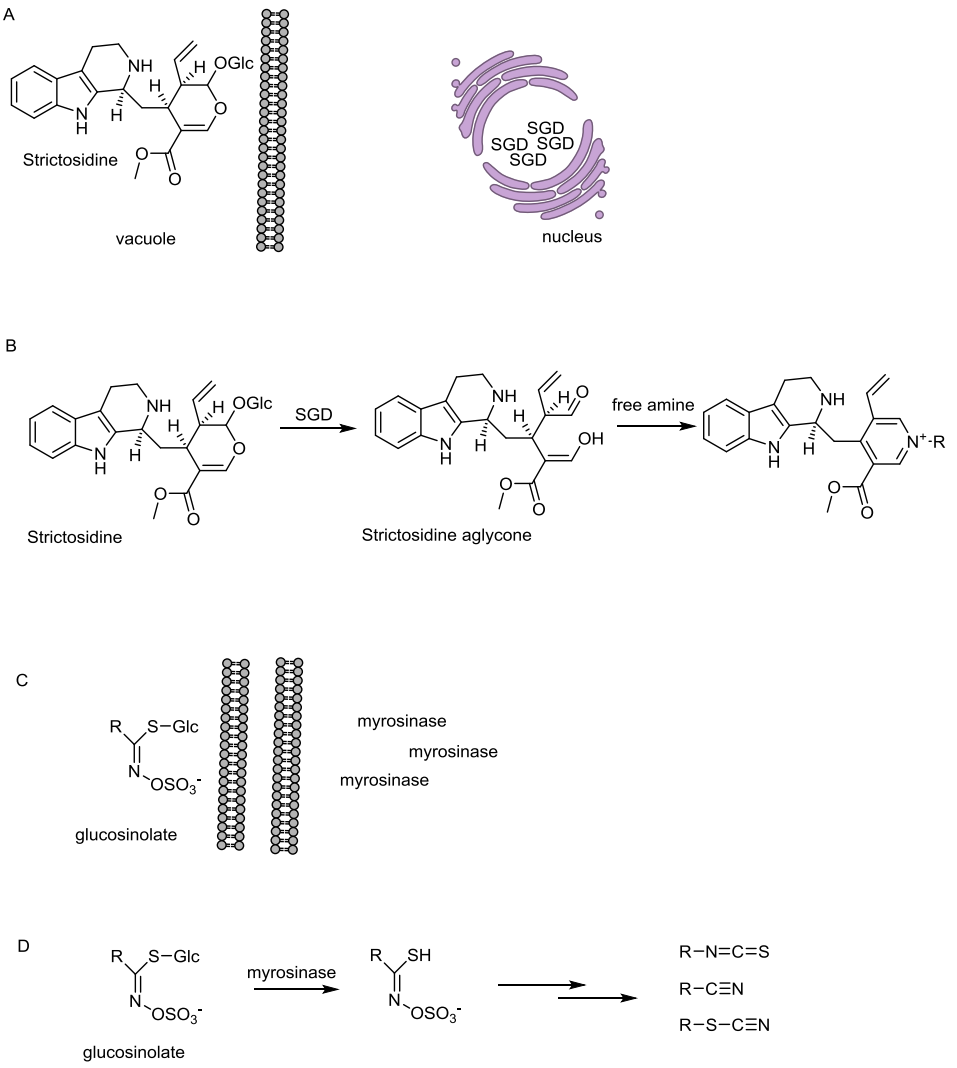


Figure 1-9 Compartmentation of strictosidine may prevent its uncontrolled deglycosylation to a reactive intermediate in an analogous system to the 'glucosinolate-myrosinase' complex.
A. Spatial separation of strictosidine and strictosidine β-glucosidase (SGD) in separate organelles prevents B. The uncontrolled formation of deglycosylated strictosidine which can crosslink free amine. This is analogous to C. the cellular separation of glucosinolates and the enzyme myrosinase which can D. result in the formation of thiocyanates, isothiocyanates and nitriles.

1.6 Identification of transporters involved in secondary metabolic pathways

There are a four main transporter classes in plants that are known to transport some of the diverse array of plant metabolites that are produced in nature and have been functionally characterised: The ABC transporters (Kang et al., 2011), The Major Facilitator Superfamily (MFS) (Pao et al., 1998), the Drug/Metabolite transporter (DMT) family (Jelesko, 2012) and the Multidrug and Toxic Compound Extrusion Proteins (MATEs) (Omote et al., 2006). Briefly, the ABC transporters are primary transporters that directly utilise the hydrolysis of ATP as an energy source for the transport of a solute against its concentration gradient (Kang et al., 2011) whilst,

in comparison, the MFS (Pao et al., 1998), DMT (Jelesko, 2012) and MATE (Omote et al., 2006) transporters are secondary transporters that either utilise the energy stored as an electrochemical gradient across the lipid bilayer for active transport, or act as facilitative transporters of solutes down their electrochemical gradient. A more thorough review of these transporter classes and their role in plant metabolism is presented in Chapter II, III and IV.

There are a number of limitations to the high-throughput identification of plant natural product transporters.

Traditional approaches to the identification of primary metabolic substrate transporters have utilised the functional complementation in yeast knockouts, using plant cDNA libraries. However, secondary plant metabolites are often not amenable to the use of such approaches as they are not essential for yeast growth (Nour-Eldin and Halkier, 2013).

An alternative method is a functional genomics screen of plant cDNA libraries utilising an assay that is amenable to measure metabolite transport. This has been applied to the identification of *Arabidopsis* glucose and glucosinolate transporters in *Xenopus laevis* oocytes (Nour-Eldin et al., 2012; Nour-Eldin et al., 2006), by expression screening of 239 full length transporters in oocytes for uptake activity of the substrate of interest.

As will be highlighted in subsequent chapters, there are two major limitations of utilising a functional screening approach to the identification of plant metabolite transporters, and other enzymes.

I.) The successful heterologous expression of eukaryotic membrane proteins is challenging. A large scale functional screening approach, although incredibly useful in identifying new substrate specificities of transporter classes (Nour-Eldin et al., 2012; Nour-Eldin et al., 2006), may be limited by whether the membrane protein is functional in the heterologous host, and as such this screening approach may result in a false negative result. It also requires the optimisation of the assay for each substrate to be tested.

II.) Screening for *in vitro* activity, by analysing transport uptake, does not assess whether the transport activity detected is physiologically relevant to the plant, and whether the spatial distribution of the metabolic reaction is integral to the metabolic pathway. Additionally this screening approach may also lead to false positive results, in which a transporter may display *in vitro* transport capacity for a substrate, however not be the physiologically relevant transporter *in vivo.*

An alternative approach for the identification of metabolite transporters is through reverse genetics. A reverse genetics screen of metabolite transporters and assessment for a metabolic phenotype results in a clear indication of the physiological relevance of the transporter.

As well as the identification of the transporter involved for a particular substrate, a reverse genetics screen provides more information with regards to the importance to the role of the spatial organisation of the metabolic pathway, and whether substrate compartmentation is critical in the metabolic pathway.

Furthermore positive results from the reverse genetics screen can subsequently be confirmed by heterologous expression for the *in vitro* characterisation of transport activity.

1.7 Virus Induced Gene Silencing

One of the most common methods for reverse genetics in non-model plant species is Virus Induced Gene Silencing (VIGS), which has proven to be a versatile tool for gene discovery in plants (Burch-Smith et al., 2004). It has been successfully used in a number of medicinal plant species, including the benzylisoquinoline alkaloid producing plants: Californian poppy, *Eschscholzia californica* (Wege et al., 2007) and opium poppy, *Papaver somniferum* (Hileman et al., 2005) as well as the tropane alkaloid producing plant *Hyoscyamus niger* (Li et al., 2006).

VIGS is used to transiently silence an endogenous plant gene, utilising an RNA interference mechanism that results in sequence specific post transcriptional gene silencing (PTGS). This process effectively hijacks the innate plant anti-viral defence response (Baulcombe, 1999) (Burch-Smith et al., 2004).

Upon infection of a plant, plant viruses produce double stranded RNA (dsRNA). Typically, this double stranded RNA is formed through the action of host dependent RDR (RNA-dependent RNA polymerase)-dependent synthesis of dsRNA from single stranded RNA viral templates (Qu et al., 2008). This dsRNA is recognised and cleaved by ribonuclease Dicer-like (DCL) proteins into siRNAs. In *Arabidopsis*, there are four DCL proteins, with DCL4 and DCL2 primarily used for viral siRNA biogenesis, which cleave the dsRNA into 21 nucleotide and 22 nucleotide fragments respectively. The DCL4 cleavage of dsRNA is aided by the dsRNA binding protein DRB4 (Adenot et al., 2006). The viral siRNAs generated are recognised by the Argonaute proteins (AGO), of which there are ten homologues in *Arabidopsis*. AGO1 and AGO2 (Baumberger and Baulcombe,

2005; Bhattacharjee et al., 2009), are primarily involved in anti-viral silencing. Binding of the AGO proteins to the siRNA results in the formation of the RNA-induced silencing complex (RISC) (Baumberger and Baulcombe, 2005). This results in the complementary base pairing of the siRNA to its RNA target and leads to either cleavage of the mRNA or translational repression of the mRNA, and the gene encoded by this mRNA is not expressed (Figure 1-10).

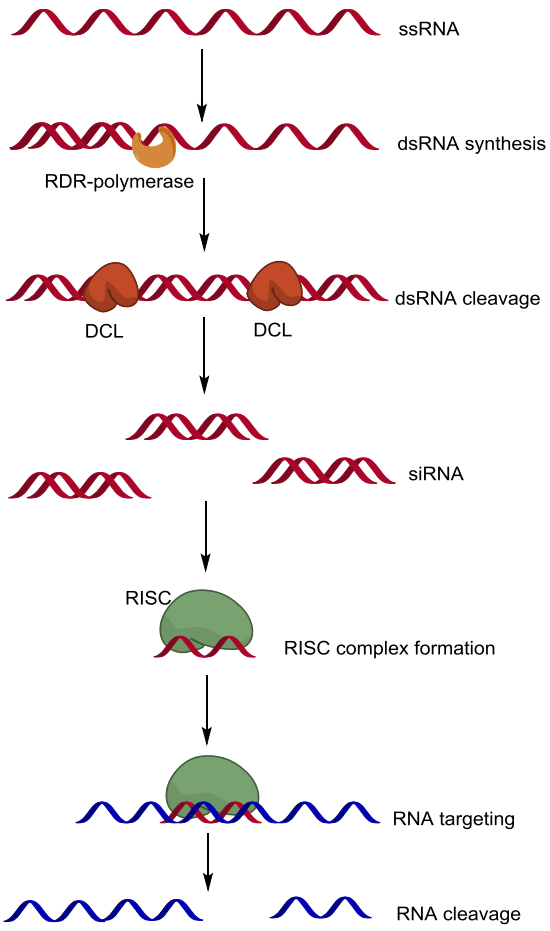


Figure 1-10 Overview of RNA interference in plants.
Viral ssRNA is converted to dsRNA by RNA Ddependent RNA-polymerase (RDR-polymerase), which is cleaved into siRNA by Dicer like enzymes (DCL). These siRNAs are loaded into the RISC complex which contains the argonaute proteins, and directs cleavage of endogenous RNA (blue) through recognition by complementary base pairing. Figure adapted from (Watson and Wang, 2012).

VIGS utilises modified viral vectors for the silencing of endogenous plant mRNAs and thereby provides a method for transiently silencing gene of interest in the plant (Ratcliff et al., 2001).

Although there are many plant viruses that have been used to develop VIGS vectors, the most common VIGS vector is derived from the tobacco rattle virus, pTRV (Ratcliff et al., 2001). Tobacco

rattle virus (TRV) has a broad host range, and the virus spreads quickly and uniformly in most plants, resulting in efficient silencing in a variety of plant species (Ratcliff et al., 2001).

The tobacco rattle virus is an RNA virus that is encoded by two RNAs, RNA1 and RNA2. RNA1 encodes a replicase, and proteins involved in viral movement. RNA2 encodes for the viral coat proteins and two non-structural proteins. A set of bipartite VIGS vectors was generated from this base structure (Liu et al., 2002). pTRV1 contains the cDNA clones of RNA1 cloned between the cauliflower mosaic virus (CaMV) 35S promoter and a nopaline synthase (NOS) terminator in a T-DNA cassette. pTRV2 contains the cDNA clones of RNA2 cloned between cauliflower mosaic virus (CaMV) 35S promoter and a nopaline synthase (NOS) terminator in a T-DNA cassette, with the genes encoding the two non-structural proteins replaced by a multiple cloning site (Liu et al., 2002).

The multiple cloning site in the pTRV2 plasmid allows insertion of a fragment of the gene of interest (200 bp-1500 bp) to be silenced in the plant. Cultures of *Agrobacterium tumefaciens* that have been transformed with the pTRV1 and pTRV2 constructs can then be mixed, and infiltrated into the plant of choice. This T-DNA cassette inserts into the plant genome, and results in the formation of a systemic virus that can infect the plant, and induce the anti-viral RNAi silencing pathway (Liu et al., 2002).

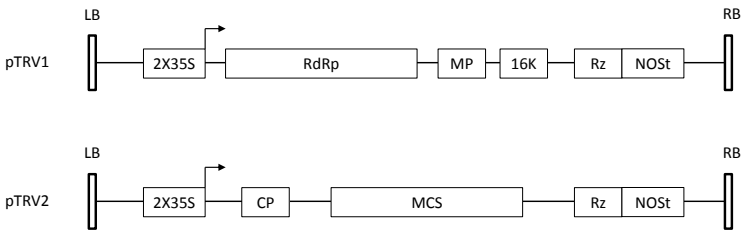


Figure 1-11 Structure of the TRV based VIGS vectors.
The RNA1 and RNA2 cDNA clones were cloned between two CaMV 35S promoters (2X35S) and NOS terminator (NOSt) in a T-DNA vector, Rz is a self-cleaving ribozyme that is common to both vectors and LB and RB highlight the left and right borders of the T-DNA. RNA1, which results in the formation of pTRV1 contains an RNA-dependent RNA polymerase (RdRp), a movement protein (MP, and a 16Kda cysteine-rich protein (16K). In turn RNA2, which results in the formation of pTRV2 encodes for the viral coat protein (CP) and contains a multiple cloning site (MCS) for cloning of the VIGS targeting region for the gene of interest (Liu et al., 2002). *Figure adapted from (Watson and Wang, 2012).*

Upon viral infection, the recombinant VIGS vector results in the formation of viral derived dsRNA which activates the anti-viral RNA silencing pathway. Because the VIGS vector contains a region from a target gene sequence, this results in the formation of siRNAs that target the endogenous

mRNA for degradation or translational repression, and thereby silences endogenous genes in the plant.

1.8 Virus Induced Gene Silencing in *C. roseus*

VIGS has been utilised using the pTRV2 system for the identification and characterisation of a number of genes that are involved in the MIA pathway since this technique was first adapted to *C. roseus* leaf tissue in 2011 (Liscombe and O'Connor, 2011).

This has included the characterisation of genes of a number of steps involved in secologanin biosynthesis; iridoid synthase (Geu-Flores et al., 2012), 7-deoxyloganetic acid glucosyltransferase (Asada et al., 2013), 7-deoxyloganic acid hydroxylase (Salim et al., 2013), 7-deoxyloganetic acid synthase, or iridoid oxidase (Salim et al., 2014), genes involved in vindoline biosynthesis; including desacetoxyvindoline 4-hydroxylase (D4H) (Liscombe and O'Connor, 2011), deacetylvindoline 4-O-acetyltransferase (DAT) (Liscombe and O'Connor, 2011), 16-methoxy-2,3-dihydrotabersonine N-methyltransferase (NMT) (Liscombe and O'Connor, 2011), tabersonine 16-hydroxylase (Besseau et al., 2013), tabersonine 3-oxygenase and tabersonine 3-reductase (Kellner et al., 2015a; Qu et al., 2015) as well as genes involved in heteroyohimbine biosynthesis such as tetrahydroalstonine synthase (Stavrinides et al., 2015).

The ABC transporter CrTPT2 that is involved in extrusion of catharanthine to the leaf wax exudate was subjected to VIGS to confirm its role as the catharanthine exporter. The use of a reverse genetics strategy, followed by screening for a metabolic phenotype, demonstrated that these genes are physiologically relevant to MIA metabolism *in planta* (Yu and De Luca, 2013).

Although a variety of infiltration systems have been used for VIGS in *C. roseus* including 'agrodrench', vacuum infiltration, and syringe infiltration, the most efficient method for silencing genes in *C. roseus* is through the use of the pinch wounding method (Liscombe and O'Connor, 2011). This method uses plants that have at least one pair of full expanded leaves, and the stem is wounded with a needle or pair of forceps, and a mixture of *Agrobacterium* cultures harbouring pTRV1 and pTRV2 plasmids are placed on the wound site.

The *Agrobacterium* is uptaken by the xylem, causing the systemic infection of the virus in the newly developing leaves in the plant. This in turn triggers the post-transcriptional RNAi silencing that downregulates the viral RNA, as well as the endogenous mRNA of the gene present on the pTRV2 viral vector.

An overview of this approach is presented in Figure 1-12. The phenotype that is observed in the positive control is caused by the downregulation of *Magnesium chelatase*, a gene integral to chlorophyll biosynthesis, where the bleaching of the leaf tissue is a result of the decreased amount of chlorophyll. Genes that affect MIA metabolism usually do not have a visibly observable phenotype, so the newly emerging leaves that occur 21-25 days post pinching must be analysed by liquid chromatography mass spectrometry (LC-MS) to assess whether silencing has caused a change in the metabolic phenotype. For these experiments a set of *Magnesium chelatase* plants is used as a positive control as a visual indication that the silencing has worked.

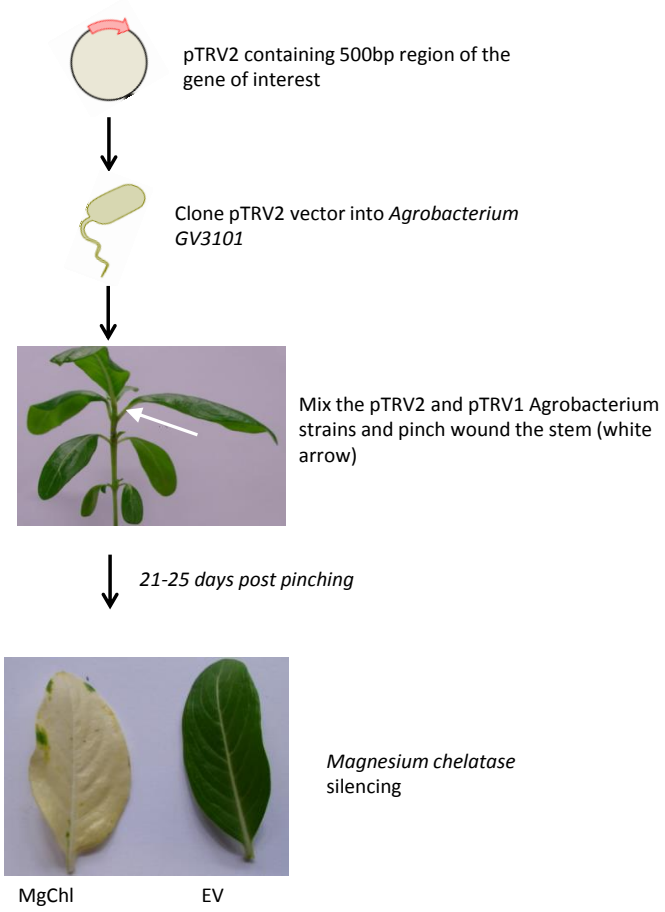


Figure 1-12 Overview of the VIGS method applied to *C. roseus*

It has also recently been shown that efficient VIGS silencing can be achieved independently of *Agrobacterium* mediated infiltration using a biolistic mediated approach (Carqueijeiro et al., 2015). This methodology resulted in comparable levels of gene silencing in comparison to the *Agrobacterium* mediated methods.

There are a number of benefits of utilising VIGS in *C. roseus* as a reverse genetic approach for biosynthetic pathway discovery (Burch-Smith et al., 2004).

1.) VIGS is a relatively fast reverse genetics tool, with an experiment in *C. roseus* completed within 3 months, from the time of planting the *C. roseus* seed. It does not rely on the need for transgenic plant formation and selection, as with the formation of stable transgenics, and this considerably shortens the screening time.

2.) The leaf tissue of the *C. roseus* plant, which is an important site of alkaloid biosynthesis, is strongly affected by VIGS. Before the utilisation of VIGS, the most common methods gene silencing in *C. roseus* used transgenic cell suspension cultures (Papon et al., 2004), or hairy root cultures (Runguphan et al., 2009). Although these studies provided a great deal of information on MIA biosynthesis in *C. roseus*, these approaches were limited by the fact that many medicinally important alkaloids are not produced in hairy roots or cell suspension cultures. For example, the downstream alkaloid vindoline is not synthesised in root tissue or cell suspension culture, and therefore attempting to study its biosynthesis in these systems is not possible. Furthermore, an understanding of the role the spatial distribution of the metabolic network plays, between the different leaf tissues, is not possible in these systems.

3.) VIGS can be used to silence multiple genes at the same time by stacking regions of different genes into the pTRV2 vector. This allows for the combinatorial silencing of genes in order to probe the potential for functional redundancy in the plant, as well as the silencing of whole gene families in the plant (Burch-Smith et al., 2004).

4.) VIGS is a knockdown, and not a knock out of the gene of interest. Although this means that complete silencing of the gene to 100% is never observed, and there is therefore always some basal residual activity of the gene of interest targeted, this methodology does allow for the analysis of genes that might be embryonically lethal if not present in the plant (Burch-Smith et al., 2004).

1.9 VIGS of transport proteins

As highlighted in the previous section, VIGS has been successfully applied to the characterisation of a number of MIA pathway genes, and the use of this reverse genetics strategy is well precedented for the elucidation of physiologically relevant enzymes in *C. roseus.* However, apart from CrTPT2, there has been less precedent for VIGS of transport proteins in this medicinal plant.

There are a number of examples where VIGS has been used successfully in the characterisation of transport proteins. With the exception of the catharanthine transporter CrTPT2, most of these have primarily been used to study the role of transport proteins in their physiological response to pathogen resistance and plant development, as opposed to screening for novel metabolic phenotypes, and the identification of secondary metabolite transporters.

For example, an ABC subfamily C transporter (ABCC) from wheat was shown to be involved in the transport of the mycotoxin deoxynivalenol, which is a fungal virulence factor. VIGS in wheat utilising the barley stripe mosaic virus (BSMV)-derived VIGS vectors, showed that this transporter resulted in susceptibility to deoxynivalenol, however the precise metabolite it was transporting was not determined (Walter et al., 2015).

The TRV VIGS system was used to silence a hexose transporter, LeHT1, in tomato, which was shown to contribute to the resistance of tomato to a variety of viruses including tomato yellow leaf curl virus (TYCLV), bean dwarf mosaic virus (BDMV), cucumber mosaic virus (CMV) and tobacco mosaic virus (TMV) (Eybishtz et al., 2010).

S-adenosylmethionine is synthesised in the cytosol, however is utilised in the chloroplast by a number of plastid methyltransferases as a methyl donor in the synthesis of chlorophylls, plastoquinone, tocopherol and phylloquinone. The SAM therefore needs to be imported in to the chloroplast for biosynthesis of these small molecules (Bouvier et al., 2006). Silencing of the chloroplast localised *S*-adenosylmethonine transporter NbSAMT1 using VIGS, in *Nicotiana benthamiana*, resulted in a stunted growth phenotype, together with pale green leaves as a result of a decrease in chlorophyll and plastoquinone content. This is consistent with the equivalent transporter in *Arabidopsis* which was analysed by T-DNA insertional mutagenesis (Bouvier et al., 2006).

There is therefore precedent for the use of the VIGS system, in multiple plant species, to assess the importance of both enzymes and transport proteins involved in secondary metabolic pathways *in vivo.*

1.10 Aims and Scope of Thesis

The primary aim of this thesis is to utilise a reverse genetics approach, VIGS, for the identification of genes that are involved in MIA metabolism in *C. roseus*. As will be shown, although a diverse range of genes have been assessed for a metabolic phenotype in this thesis, there is a particular emphasis on the identification of transport proteins that could be involved in MIA metabolism in *C. roseus.*

Chapter II assesses the use of transcriptome mining for the generation of candidate genes that could be involved in the MIA pathway, as well as discussing the results from a variety of VIGS experiments, which highlights a number of the pitfalls and limitations of utilising this approach.

Chapter III and Chapter IV discuss the identification of two transporters that have been shown to be physiologically relevant to *C. roseus* MIA metabolism, together with attempts at their *in vitro* characterisation and highlights how the use of this approach can be used for transporter identification.

Chapter V discusses the identification of a reticuline oxidase like enzyme that is involved in MIA metabolism in *C. roseus*. This chapter also highlights the benefits of utilising a reverse genetics screen for the identification of novel effects on MIA metabolism that would not necessarily be predictable or unlikely to be identified by a functional genetics screen.

The overarching theme of this research is to utilise VIGS in *C. roseus* in order to identify genes involved in MIA metabolism that may not be amenable to a functional genomics screening approach, as well as assess the integral role the spatial organisation of the metabolic pathway has on MIA biosynthesis.

2 Transcriptome mining and characterisation of gene candidates in *C. roseus*

2.1 Transcriptomic resources in medicinal plants

The utilisation of reverse genetic methods, such as VIGS in medicinal plants, requires the prior knowledge of the gene target and its sequence. The field of gene discovery in medicinal plants has been greatly advantaged due to the availability of next generation sequencing technologies. For *C. roseus*, there are three full transcriptomes that are available in the public domain, the Medicinal Plant Genomic Resources (MPGR, http://medicinalplantgenomics.msu.edu; (Góngora-Castillo et al., 2012)), Phytometasyn (http://www.phytometasyn.ca; (Xiao et al., 2013)), as well as Cathacyc (http://www.cathacyc.org; (Van Moerkercke et al., 2013)).

Within this thesis the primary source of sequence and expression data has been from the Medicinal Plant Genomics Resource, however the other databases have also been used for cross-referencing gene sequences.

The MPGR *C. roseus* transcriptome contains 32,607 unique transcripts that likely represent unique genes expressed in *C. roseus* (unigenes). These transcripts have been annotated using similarity searches using the *Arabidopsis thaliana* proteome, the UniRef100 database and Pfam domain composition (Góngora-Castillo et al., 2012).

One of the benefits of using this transcriptomic resource is the amount of sequencing data derived from different tissues and different experimental conditions. In the MPGR database, for *C. roseus*, transcriptomic sequence data has been obtained from 23 samples taken from different tissues under various conditions. These are summarised in Table 1.

Table 1 Tissues and experimental conditions available in the *C. roseus* MPGR transcriptomic resource
Sequence data is available from the developmental tissues, as well as sterile seedlings, *C. roseus* suspension cultures and *C. roseus* hairy root lines under a number of different elicitor treatments. The suspension cultures were treated with 0.3 mg/ml yeast extract (YE), and the sterile seedlings and hairy roots were incubated with 100 µM methyl jasmonate (MeJA). The TDCi hairy roots are antisense lines for tryptophan decarboxylase, whilst the RebH hairy root is a line expressing a bacterial halogenase. For subsequent graphs of expression presented in this thesis the different lines will be represented by the numbers 1-23.

Number	Tissue/ experimental condition
1	Flowers
2	Suspension culture YE 6 h
3	Suspension culture YE 12 h
4	Suspension culture YE 24 h
5	Sterile seedlings MeJA 0 h
6	Sterile seedlings MeJA 6 h
7	Sterile seedlings MeJA 12 h
8	Sterile seedlings MeJA 24 h
9	Sterile seedlings 5 d
10	Sterile seedlings 12 d
11	Sterile seedlings (control)
12	Suspension culture (control)
13	Mature leaf
14	Immature leaf
15	Stem
16	Root
17	Wild type hairy roots
18	Wild type hairy roots MeJA 0 h
19	Wild type hairy roots MeJA 24 h
20	TDCi hairy root
21	TDCi hairy root MeJA 0 h
22	TDCi hairy root MeJA 24 h
23	RebH hairy root

The expression level of each of the 32,607 unigenes, represented as fragments per kilobase of transcript per million reads mapped (FPKM), has been calculated in each of the libraries, thereby providing expression information on each of the unigenes in different tissues and experimental conditions (Góngora-Castillo et al., 2012).

For gene discovery of MIA pathway genes in *C. roseus,* the diversity of sequence data across different tissues and experimental conditions allows for filtering the dataset based upon expression (Geu-Flores et al., 2012).

Due to the known localisation of the metabolic pathway to leaf tissue, any gene not expressed in leaf can be discarded. Additionally the seedlings that have been elicited with methyl jasmonate (MeJA) allows for identification of genes that are upregulated upon MeJA induction. The MIA pathway responds to MeJA elicitation (Vazquez-Flota and De Luca, 1998), and therefore genes that are preferentially expressed in response to MeJA elicitation are preferred candidates.

Additionally this large dataset provides the opportunity to analyse genes that are co-expressed with known MIA genes under all of the tissues and experimental conditions, and this has been the primary basis for candidate gene selection in this study.

2.2 Co-expression analysis

Gene discovery utilising co-expression analysis assumes that a metabolic pathway, or part of a metabolic pathway, is co-regulated at the transcriptional level, and therefore the genes in the pathway would have a similar expression pattern.

The data can then be arranged or clustered based on co-expression profiles, with the unigenes that have similar expression profiles to known genes in the metabolic pathway of interest, highlighted. Candidate genes that have a similar expression pattern to these known genes are then characterised by either a functional assay or reverse genetics method.

There is precedence for the co-regulation of a number of metabolic pathways in plants. This has been shown in primary metabolism in *Arabidopsi*s, with the genes in the pathways of glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway tightly co-expressed (Wei et al., 2006).

Arabidopsis also shows co-expression of a number of secondary metabolic pathways, including flavonoid biosynthesis (Wei et al., 2006; Yonekura-Sakakibara et al., 2008), as well as in the biosynthesis of monoterpenes (Ginglinger et al., 2013). For example the terpene synthases and

cytochrome p450s that are involved in the formation of the monoterpene alcohol linalool were identified by co-expression analysis (Ginglinger et al., 2013).

In *C. roseus*, this methodology has been primarily used for gene discovery for the identification of genes in the seco-iridoid branch of the metabolic pathway. Co-expression analysis was utilised for the successful identification of iridoid synthase (Geu-Flores et al., 2012) as well as multiple unknown genes in the seco-iridoid branch of the MIA pathway (Miettinen et al., 2014).

2.3 Results and Discussion

2.3.1 Candidate selection for VIGS

Two methods of selection by co-expression analysis were employed to determine genes of interest for this study; Hierarchical Clustering and Self-Organising Maps.

2.3.1.1 Hierarchical Clustering

Hierarchical clustering was implemented using the Multiple Experiment Viewer (MeV) v4.7 (Eisen et al., 1998; Saeed et al., 2003). The 32,607 unigenes present in the *C. roseus* dataset, were filtered based on expression level in leaf tissue. To decrease the number of unigenes under consideration; unigenes with a $\log_2$FPKM value of less than 2 in the leaf were discarded. Multiple hierarchical clustering dendograms were generated using different distance metrics to calculate the differences between unigene expression profiles; Pearson Correlation, Euclidean Distance and Manhattan Distance, as well as the average linking criterion as the linking method. Known genes involved in the MIA pathway were highlighted, and unigenes in the surrounding cluster were selected for further interrogation by the VIGS, with the candidates presented in Table 2.

As highlighted in both the work of Miettinen et al., 2014 and Geu-Flores et al., 2012, the genes involved in the biosynthesis of geraniol via the MEP pathway, the biosynthesis of the seco-iridoid secologanin, the biosynthesis of tryptamine and the biosynthesis and deglycosylation of strictosidine appear to be co-regulated, whilst genes involved in the downstream MIA pathway such as those involved in the conversion of tabersonine to vindoline do not cluster with this method. This is supported by the hierarchical clustering analysis presented in this thesis (Figure 2-1).

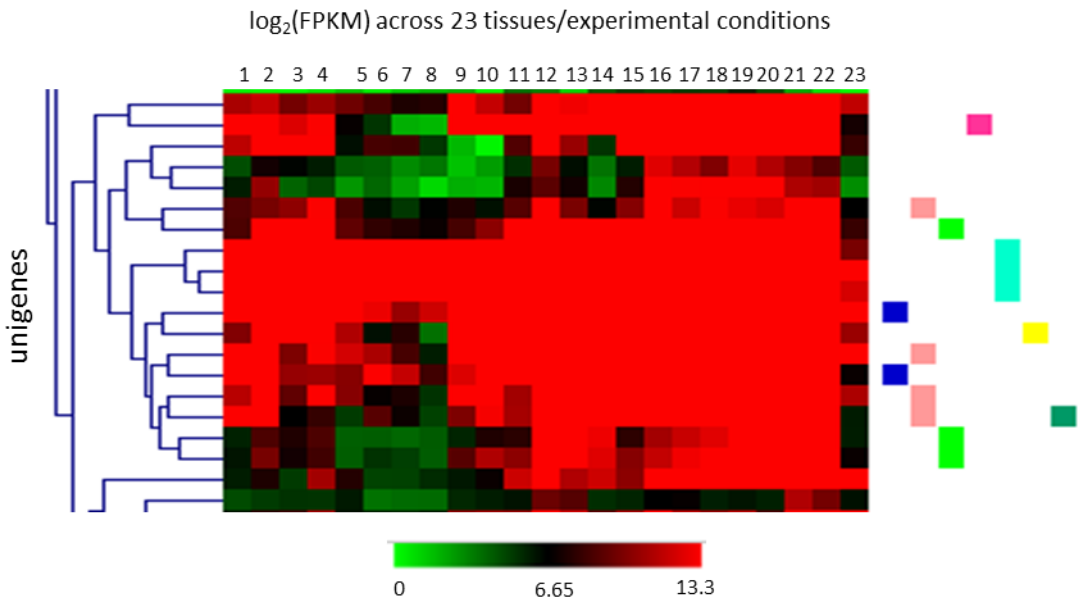


Figure 2-1 Representative hierarchical clustering dendogram.
Hierarchical clustering using the Pearson correlation distance metric with average linkage was implemented using the MeV v4.7 software. Clusters of unigenes with similar expression profile are represented on the y-axis. The heatmap represents the log_2(FPKM) across the 23 different tissues and experimental conditions highlighted in Table 1. This region of the dendogram highlights the co-expressed cluster involved in seco-iridoid and strictosidine biosynthesis. Dark pink- tryptophan decarboxylase, light green-Genes in the MEP pathway, Light blue- secologanin synthase isoforms, Dark blue- strictosidine synthase isoforms, Dark green- 7-deoxyloganetic acid glucosyltransferase, Yellow- Loganic acid methyltransferase, Light Pink-Candidate genes annotated as transport proteins.

2.3.1.2 Self-Organising Maps

To corroborate the genes selected by hierarchical clustering, an alternative method for clustering transcriptomic expression data was also utilised, self-organising maps (SOM) (Kohonen, 1990). This method for application to the *C. roseus* transcriptomic dataset was developed by Marc Jones, John Innes Centre.

For each node in the SOM there are approximately 40 unigenes presented, with the spatial proximity of each node on the map representative of the similarity in expression profile. In the SOM, strictosidine synthase, secologanin synthase, tryptophan decarboxylase, strictosidine glucosidase, geraniol 8-hydroxylase and loganic acid methyltransferase localised to 5 neighbouring nodes (Figure 2-2). This indicates that these six genes showed similar expression profiles and supports the hierarchical clustering method used in this study, as well as the clustering methods utilised in Geu-Flores et al., 2012 and Miettinen et al., 2014, which indicate that genes of seco-iridoid and strictosidine biosynthesis are co-regulated. Similarly genes

involved in the downstream MIA pathway such as those involved in the conversion of tabersonine to vindoline do not cluster in the SOM analysis.

A full list of the genes highlighted in these 5 nodes is presented in the Appendix, with the genes selected for assessment by VIGS in this thesis highlighted in Table 2.

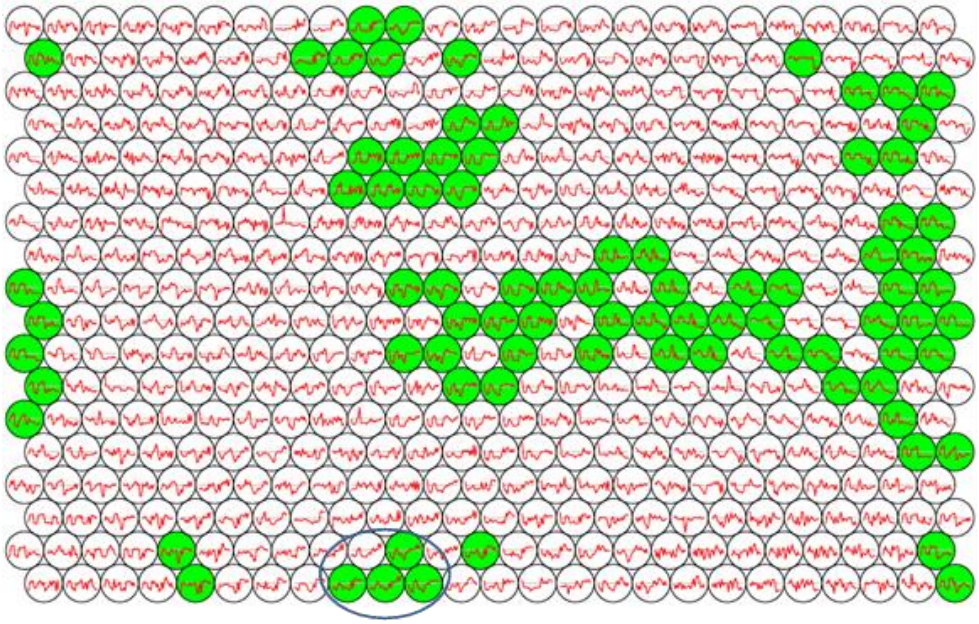


Figure 2-2 Self-organising map for *C. roseus* transcriptomic data.
Each node represents around 40 unigenes, with each node containing genes with the most similar expression profile. Furthermore neighbouring nodes are related to each other by the similarity of their expression profile. The average expression profile of genes in the node is plotted as a red line on each node. The genes strictosidine synthase, secologanin synthase, tryptophan decarboxylase, strictosidine β-glucosidase, geraniol 8-hydroxylase and loganic acid methyltransferase all localise to the nodes highlighted by the blue circle. Nodes coloured green represent the highest quality nodes as defined in the Materials and Methods.

The gene list generated by the SOM analysis was very similar to that generated as a result of hierarchical clustering, indicating the robustness of both approaches. The best candidates for enzymology and transport in the MIA pathway in both of these datasets were two MATE proteins, CrMATE1952 and CrMATE10740 which will be discussed in Chapter III, an NPF transporter, CrNPF2.9, which will be discussed in Chapter IV, and a reticuline oxidase like enzyme, CrRO, which will be discussed in Chapter V. In subsequent chapters the expression profile of these genes are plotted as a line graph to demonstrate co-expression with the known MIA genes.

2.3.1.3 Alternative candidate selection criteria

Although there is precedence for genes in a secondary metabolic pathway to be co-expressed, this does not always occur. The hierarchical clustering analysis and self-organising map analysis in this thesis, together with the work of Geu-Flores et al., 2012 and Miettinen et al., 2014, highlights this. The genes in the early part of the MIA biosynthetic pathway appear to be co-regulated in the MPGR transcriptomics dataset; namely the biosynthesis of geraniol via the MEP pathway, the biosynthesis of the seco-iridoid secologanin, the biosynthesis of tryptamine and the biosynthesis and deglycosylation of strictosidine.

In contrast the expression pattern of the known genes that act after the deglycosylation of strictosidine, that are involved in the biosynthesis of vindoline from tabersonine, as well as in the formation of the alkaloid tetrahydroalstonine, do not cluster with each other or these early genes.

The genes that have been selected by co-expression analysis for investigation in this study would therefore be predicted to act on the earlier steps of the MIA pathway. To obtain candidates that might act on the later stages of the pathway is difficult utilising the transcriptomic resources available. The genome of *C. roseus* has been published (Kellner et al., 2015b), and this work highlights alternative routes to gene discovery utilising the spatial proximity of genes on the genome, as there are growing reports of the physical clustering of genes in secondary metabolism in plants (Osbourn, 2010). This methodology was successful in the identification of tetrahydroalstonine synthase (Stavrinides et al., 2015).

Additionally, an alternative approach in the selection of gene candidates is to utilise homology to known genes in other species that have also been implicated in secondary metabolism, together with their expression level in leaf tissue, as this is the known site of MIA biosynthesis in *C. roseus.*

An overview of the gene candidates selected for VIGS, together with their locus number in the MGPR transcriptome, the reason for selection and the chapter in which they are discussed in this thesis is presented in Table 2.

Table 2 VIGS candidates and their reason for selection in thesis.
Locus number represents the gene number present in the final transcriptomic dataset in the MPGR *C. roseus* transcriptome

Gene class	locus number	Gene family	Name	Chapter	Reason for Selection
Transporter	locus_1763	ABC transporter		II	Expression level in leaf tissue
	locus_1443	Purine Permease		II	Expression level in leaf tissue and homology to purine permeases
	locus_24210	P-DME transporter		II	Co-regulation with MIA pathway
	locus_2720	MATE protein		III	Co-regulation with MIA pathway
	locus_1952	MATE protein	CrMATE_1952	III	Co-regulation with MIA pathway
	locus_10740	MATE protein	CrMATE_10740	III	Co-regulation with MIA pathway
	locus_3327	MATE protein		III	homology to CrMATE_1952
	locus_8527	MATE protein		III	homology to CrMATE_1952
	locus_9753	NPF transporter	CrNPF2.4	IV	Co-regulation with MIA pathway and known *in vitro* activity
	locus_8319	NPF transporter	CrNPF2.6	IV	Co-regulation with MIA pathway and known *in vitro* activity
	locus_9086	NPF transporter	CrNPF2.9	IV	Co-regulation with MIA pathway
Enzyme	locus_7197	Glutathione s-transferase		II	Co-regulation with MIA pathway
	locus_6502	Glutathione s-transferase		II	Homology to locus_7197
	locus_2878	Glutathione s-transferase		II	Homology to locus_7197
	locus_5608	Reticuline oxidase	CrRO	V	Co-regulation with MIA pathway
	locus_5372	Reticuline oxidase		V	Homology to locus_5608 and expression in leaf tissue
	locus_2174	Reticuline oxidase		V	Homology to locus_5608 and expression in leaf tissue
	locus_13163	Progesterone β-reductase	CrP5βR2	II	Homology to iridoid synthase
	locus_240	Progesterone β-reductase	CrP5βR4	II	Homology to iridoid synthase
	locus_2665	Progesterone β-reductase	CrP5βR5	II	Iridoid synthase
Transcription factor	locus_12822	bHLH transcription factor	BIS1	II	Co-regulation with MIA pathway *in vitro* activity

2.3.2 Reproducibility of the C. roseus alkaloid profile

To assess whether silencing by VIGS of a candidate gene causes a change in the metabolite profile, the VIGS-Empty vector (VIGS-EV) control tissue must have a reproducible metabolic profile, in which there is minimal variation in alkaloid content between individual plants in a given VIGS experiment. The stability of this alkaloid profile has been presented previously in (Asada et al., 2013; Besseau et al., 2013; Geu-Flores et al., 2012; Kellner et al., 2015a; Liscombe and O'Connor, 2011; Qu et al., 2015; Salim et al., 2014; Salim et al., 2013)

A representative liquid chromatography-mass spectrometry (LC-MS) total ion chromatogram for a methanol extract of the VIGS-EV control is shown in Figure 2-4. For all VIGS experiments presented in this thesis, the LC-MS peak areas of the known metabolites secologanin, strictosidine, vindoline, catharanthine, and *m/z* 349 (this contains a mixture of the diastereomers of serpentine and alstonine) (Figure 2-3) were calculated, and compared to the VIGS samples of the gene of interest to assess the effect gene silencing has on MIA metabolism.

For each individual VIGS experiment between 8-15 plants were used for each VIGS pTRV2 construct under consideration.

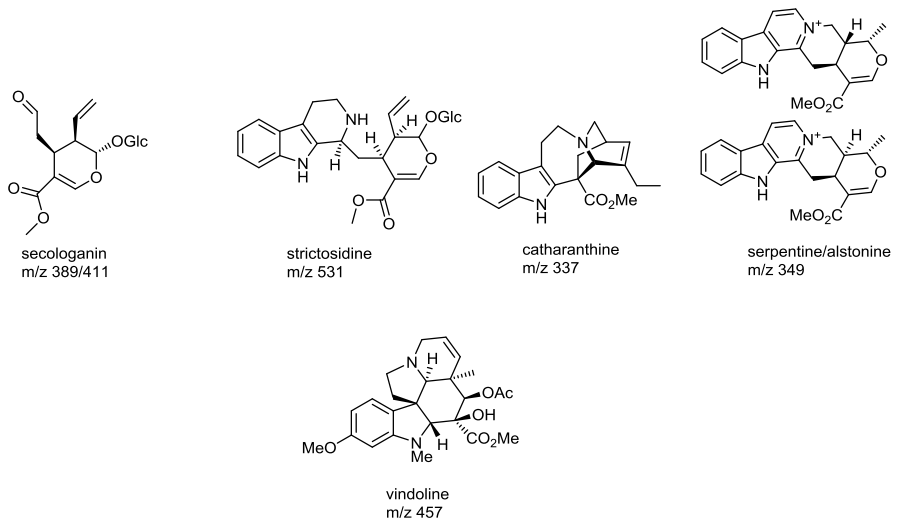


Figure 2-3 Chemical structures of the compounds analysed by LC-MS in methanolic extracts of *C. roseus* VIGS tissue

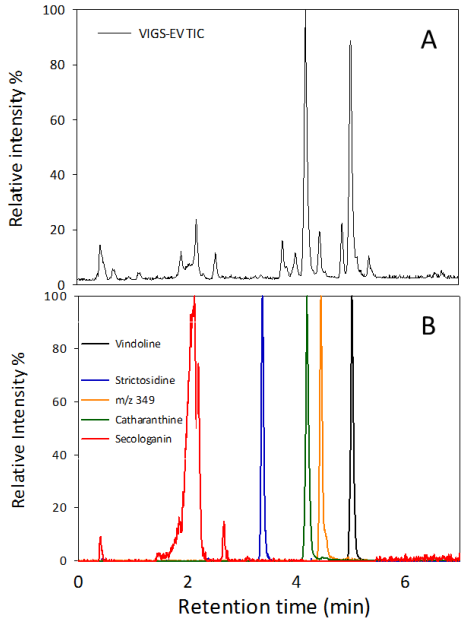


Figure 2-4 Representative VIGS-EV LC-MS chromatogram.
A.) Representative total ion chromatogram of VIGS-EV methanol extract B.) Extracted ion chromatogram for secologanin (Na^+ adduct); *m/z* 411, strictosidine; *m/z* 531, serpentine/alstonine; *m/z* 349, catharanthine; *m/z* 337, vindoline; *m/z* 457.

2.3.3 Criteria for assigning a metabolic phenotype resulting from VIGS

As well as changing the levels of the known alkaloids in the MIA pathway, a VIGS phenotype is often characterised by the accumulation of new metabolites that do not appear in VIGS-EV control tissue (Asada et al., 2013; Besseau et al., 2013; Geu-Flores et al., 2012; Kellner et al., 2015a; Liscombe and O'Connor, 2011; Qu et al., 2015; Salim et al., 2014; Salim et al., 2013). For the identification of new compounds that accumulate in VIGS tissue as a result of silencing, but are not present in VIGS-EV tissue, untargeted metabolomics using the peak finding algorithm XCMS (Smith et al., 2006) was utilised. This software calculates the peak area for all ions present in the total ion chromatogram for each VIGS sample.

For each individual VIGS sample, XCMS results in the extraction and identification of ~100 peaks. The average LC-MS peak area for each identified compound was calculated and compared between the VIGS-EV and VIGS-gene of interest samples, utilising a Student's two sample t-test of equal variance to assess whether an identified peak is significantly different between VIGS-

EV control tissues and VIGS as a result of gene silencing. Three levels of stringency were applied in defining whether a new peak has been identified that is a result of silencing the gene of interest as opposed to random variation in the *C. roseus* leaf samples.

1.) Firstly, the new peak must accumulate a minimum of 3-fold higher in the gene silenced samples relative to the VIGS-EV control samples.
2.) Secondly, the p-value defined by the t-test is corrected for multiple testing by the Benjamini-Hochberg method for determining the False Discovery Rate (FDR), and only peaks that satisfy the corrected level of significance are assessed to be significant (Benjamini and Hochberg, 1995). This is to correct for the rate of false positive results that are a statistical artefact of multiple testing of the peaks discovered by XCMS.
3.) Third, if VIGS of a gene results in the formation of new peaks that satisfy the two above criteria, the whole VIGS experiment is repeated a minimum of 3 separate times utilising a targeted method to identify these accumulating compounds. If the compounds reproducibly accumulate in these experiments relative to VIGS-EV controls then the metabolic phenotype is adjudged to be due to VIGS of the gene of interest. qPCR in these samples is subsequently performed to assess the extent of knockdown of the gene of interest relative to empty vector controls.

2.3.4 VIGS Screening of Candidate Genes

2.3.4.1 ABC transporters

There are five main families of transporters that function as multidrug resistance pumps, the Drug metabolite transporter (DMT) superfamily (Jelesko, 2012), Resistance/nodulation/division (RND) family (Remy and Duque, 2014), ATP-Binding Cassette (ABC) family (Kang et al., 2011), Major Facilitator Superfamily (MFS) (Pao et al., 1998), and the Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) flippase Superfamily (Omote et al., 2006) , in accordance with the nomenclature from the transporter classification database (TCB) (Saier et al., 2009). In higher plants the DMT superfamily, ABC family, MFS, and MOP flippase superfamily, have all been implicated in the movement of secondary metabolites. The gene families for these transporters are particularly expanded *in planta,* for example in *A. thaliana* there are approximately 130 ABC transporters, 120 MFS and 58 Multidrug and toxic compound extrusion protein (MATE) transporters, which is a subfamily of the MOP flippase superfamily (Remy and Duque, 2014).

ABC transporters are a large class of transport proteins that utilise the direct hydrolysis of ATP as an energy source to move solutes across a lipid bilayer (Kang et al., 2011; Rea, 2007). Most ABC transporters contain multiple transmembrane domains, which are hydrophobic and form the membrane pore and usually consist of 4-6 α-helices, as well as two cytosolic domains. The two cytosolic domains are the nucleotide binding domains (NBD), which contain the ATP-binding Walker A and B motifs (Kang et al., 2011; Rea, 2007). The NBD's bind and utilize MgATP, coupling the hydrolysis of ATP and release of ADP to the energization of transport. Most classes of ABC transporter contain two transmembrane domains and two NBDs, and have a pseudotwofold symmetry. These transporters can either be in the form of one contiguous peptide, or expressed as two half molecules that subsequently associate as either homo- or hetero- dimers (Kang et al., 2011; Rea, 2007). There are eight main families of ABC transporters in plants, named ABC A-I. The ABC transporter classes A-D have a forward orientation in which the transmembrane domain is at the N-terminus, whilst the ABCG class has a reverse orientation with the NBD at the N-terminus of the protein. To date, only the full length ABCB (Multidrug resistance/P-glycoprotein like), ABCC (Multidrug resistance associated protein like) and ABCG (Pleiotropic drug resistance protein) families have been implicated in the transport of plant secondary metabolites (Figure 2-5) (Kang et al., 2011; Rea, 2007).

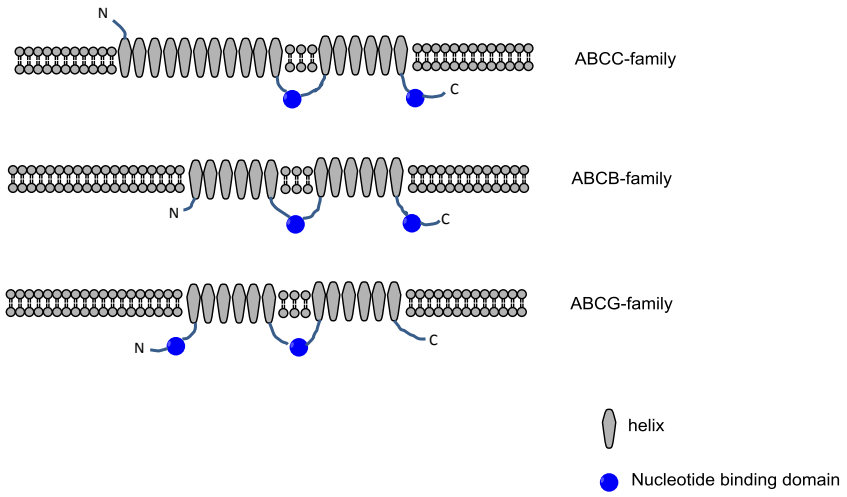


Figure 2-5 Structure of the ABC transporter families that are involved in transport of secondary metabolites *in planta*.
The ABCC family and ABCB family are forward orientated ABC transporters, with the ABCC family comprising an extra N-terminal domain comprising 5 α-helices. The ABCG family are a reverse orientated transporter family.

The ABCB (MDR/PGP) family of proteins consist of both full length, as well as half size ABC transporters, with a forward orientation (Kang et al., 2011). Most of the full length ABCB transporters localise to the plasma membrane, whilst some of the half transporters have been localised to other organelles such as the mitochondria. The proteins that have been characterised have a diverse substrate repertoire ranging from auxin (Cho and Cho, 2013), to their involvement in iron-sulphur cluster assembly (Schaedler et al., 2014). To date, only one plant MDR transporter has been implicated in the transport of secondary metabolites, the MDR1 from Coptis japonica (CjMDR1). *Coptis japonica* is a medicinal plant that produces the benzylisoquinoline alkaloid berberine. The transporter CjMDR1 is localised to the plasma membrane in *C. japonica* cells and when heterologously expressed in *Xenopus laevis* oocytes, results in the import of berberine (Shitan et al., 2003).

The ABCC (MRP) family of transporters in plants only consist of full length proteins, with no half transporters encoded in the family (Kang et al., 2011). They have a forward orientation, however unlike the ABCB family, many also contain an additional N-terminal transmembrane domain (TMD0) that has 3-5 helices. In terms of subcellular localisation these membrane transporters have been shown to reside on both the plasma membrane and the tonoplast, and they are the only full length ABC transporters that have been shown to localise to this vacuolar membrane (Kang et al., 2011). The localisation to the vacuolar membrane has profound implications for this transporter family's role in detoxification of compounds. Many xenobiotic compounds are conjugated to hydrophilic molecules such as glucose, glucuronide or glutathione, and this

conjugation has been thought to play an important role in recognition of the substrate for transport (Kang et al., 2011). In turn this family of transporters has been directly implicated in the transport of glutathione conjugated compounds (GS-X) such as N-ethylmaleimide-GS (NEM-GS), S-(2,4-dinitrophenyl)-GS, (DNP-GS) as well as the glutathionylated herbicide metolachor-GS (Li et al., 1995; Martinoia et al., 1993). Furthermore this family has the capacity to transport a number of chlorophyll catabolite products for detoxification (Lu et al., 1998), folate derivatives (Chen et al., 2002) and phytate (inositol hexakisphosphate, $InsP_6$) (Shi et al., 2007).

As well as in the detoxification of exogenously supplied xenobiotic compounds, the ABCC family has been implicated in the transport of endogenous secondary natural products such as flavonoids. Although there are multiple mechanisms hypothesised by which flavonoids can be sequestered in the vacuole, including through the use of MATE transporters and vesicular trafficking (Zhao, 2015), ABC transporters have also been implicated. A transgenic antisense construct for the ABCC transporter, ZmMRP3, in maize results in altered anthocyanin biosynthesis and reduced pigmentation (Goodman et al., 2004). Additionally the VvABCC1 transporter in grapevine, *Vitis vinifera*, has also been shown to be involved in anthocyanin vacuolar sequestration. VvABCC1 is localised to the tonoplast membrane, and was shown to transport malvidin 3-*O*-glucoside when overexpressed in yeast vesicles (Francisco et al., 2013). Interestingly this is only transported in the presence of glutathione (GSH), and is co-transported by VvABCC1. This is an indication that although GSH is needed for transport of the anthocyanin, conjugation of the substrate to glutathione is not necessary for co-transport. (Francisco et al., 2013).

The final ABC transporter family implicated in secondary metabolite transport in planta are the ABCG (PDR) type family, which again are either full or half transporters with a reverse orientation (Kang et al., 2011). In *Arabidopsis* one of the major functions of the ABCG family is in the export of lipid precursors for the formation of the cuticular wax (McFarlane et al., 2010), however there is growing evidence that these transporters are also involved in the movement of secondary metabolites. In *Nicotiana tabacum* and *Nicotiana plumbaginifolia* this transporter class has been shown to be involved in the export of diterpenes, such as sclaerol (Crouzet et al., 2013), and, as has been discussed previously, this transporter class also contains the *C. roseus* transporter CrTPT2 which is involved in the export of catharanthine from the epidermis to the cuticle wax (Yu and De Luca, 2013).

Interestingly the human homologues of the ABCB and ABCC family members, MDR1 and MRP1/MRP2 have the capacity to transport the *C. roseus* bisindole alkaloids vincristine and

vinblastine (Loe et al., 1998; Souza et al., 2011). These family members therefore are of potential interest in the transport of the MIAs in *C. roseus*.

2.3.4.2 ABC transporters in C. roseus

There are in excess of 100 ABC transporters present in *C. roseus*, however most do not appear to co-regulate with the MIA pathway. A member of the ABCC subfamily, locus_1763, was selected for silencing based on its expression level in leaf tissue. A ~500 bp fragment of this gene was cloned into the pTRV2u vectors for silencing in *C. roseus* as described in (Geu-Flores et al., 2012; Liscombe and O'Connor, 2011).

VIGS of this ABC transporter did not lead to an observable metabolic phenotype, with no statistically significant difference in the alkaloid peaks of strictosidine, catharanthine, serpentine/alstonine or vindoline. Additionally there was no change in the seco-iridoid glycoside, secologanin (Figure 2-6). Therefore, to date, this transporter has not been implicated in MIA biosynthesis in *C. roseus*

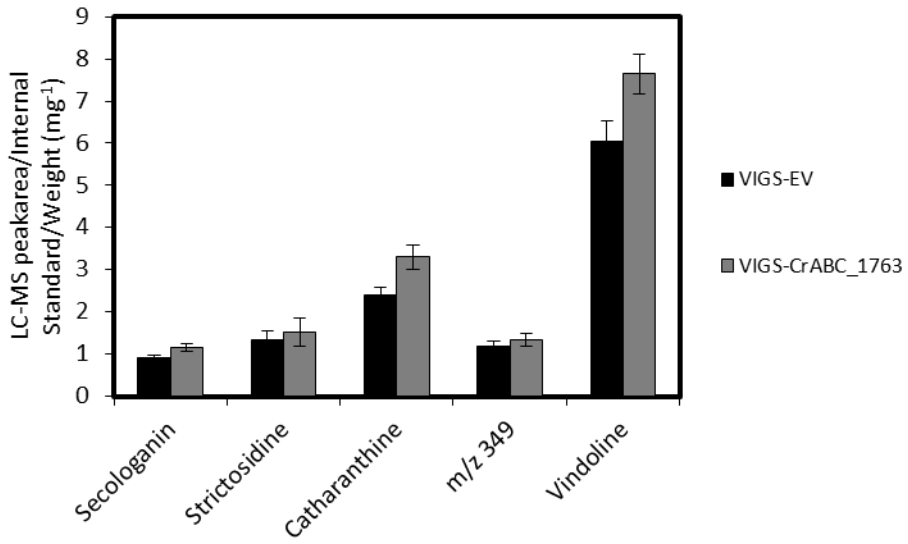


Figure 2-6 Metabolite analysis upon silencing of the ABC transporter locus_1763 relative to EV control
Alkaloid profile for VIGS of CrABC_locus_1763 relative to Empty Vector control tissue in *C. roseus* cv. Sunstorm Apricot, VIGS-CrABC_locus_1763 (n=9), VIGS-EV (n=8) ** $p<0.01$. All data presented is mean $\pm$ SEM.

2.3.4.3 Glutathione S-transferases

Glutathione transferases (GST) are a large group of enzymes that are found all forms of life, from bacteria to humans. These enzymes have primarily been characterised as having the catalytic ability to detoxify electrophilic xenobiotic compounds through the conjugation with the tripeptide glutathione (GSH; gamma-glu-cys-gly) (Figure 2-7) (Dixon et al., 2010).

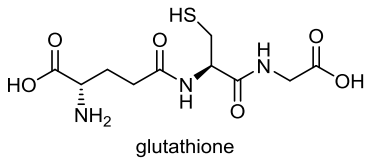


Figure 2-7 Structure of glutathione

In plants the soluble GST family are a group of dimeric enzymes that have the capacity to form either homo- or heterodimers. Each protein in the dimer has a glutathione (GSH) interacting domain, which is generally at the N-terminus, as well as a hydrophobic binding pocket in the C-terminus that is used to bind the substrate (Axarli et al., 2009).

In plants there are seven main families, dehydroascorbate reductase (DHAR), tetrachlorohydroquinone dehalogenase like (TCHQD), GST phi (F), GST tau (U), GST lambda (L), GST zeta (Z) and GST theta (T). The GSTF and GSTU classes are large and plant specific (Dixon et al., 2010).

Plant GSTs have been primarily demonstrated to detoxify exogenously applied electrophilic herbicides through conjugation with GSH however a definitive physiological role for this large protein family has remained elusive (Dixon et al., 2010).

There are three main roles this plant family has been hypothesised to play *in planta:*

Firstly, these enzymes may catalyse the glutathionylation of endogenous metabolites. The evidence for this is fairly limited, as not many *S*-glutathionylated metabolites have been shown to accumulate in plant extracts. There is evidence that a number of flavonoid derivatives, such as the chalcone isoliquirtigenein and medicarpin are able to be glutathionylated *in vitro* (Cummins et al., 2003). Additionally oxylipin compounds such as oxophytodienoic acid (OPDA) are readily *S*-glutathionylated (Figure 2-8). This was demonstrated by metabolomic analysis of plant extracts upon feeding OPDA to plants (Davoine et al., 2006), as well as shown *in vitro* in which the enzymes AtGSTF8 and AtGSTU18 were able to conjugate glutathione to OPDA (Dixon

and Edwards, 2009). Additionally GSH conjugates of an unstable protoporphyrinogen as well as GSH conjugated fatty acid thioesters have also been detected *in planta* (Dixon et al., 2008). A key difference between the xenobiotic conjugation of GSH and the conjugation of endogenous ligands, is that, due to the reaction mechanism of conjugation, the conjugation to the endogenous ligand, is likely to be unstable and reversible. One of the limitations of detecting endogenous glutathionylated conjugates is therefore the potential for the reversibility of conjugation, as well as the fact these compounds can be rapidly degraded and detoxified (Dixon et al., 2010) (Figure 2-8).

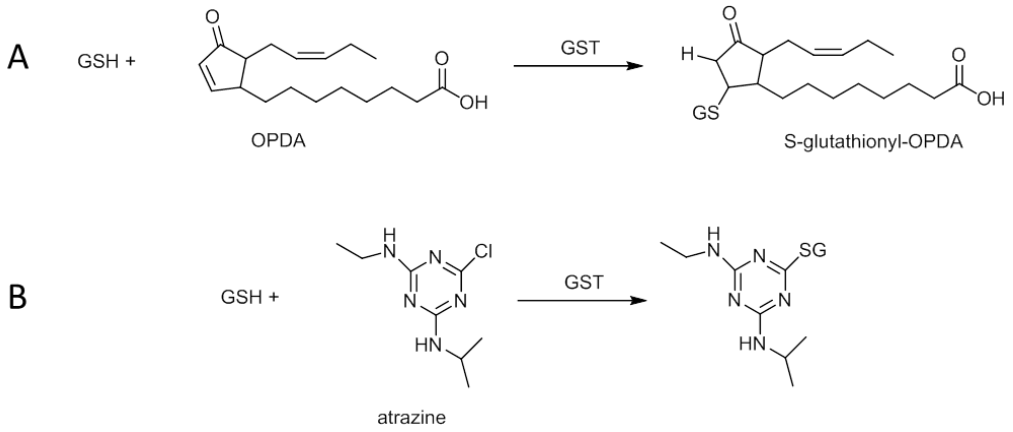


Figure 2-8 Glutathionylation of endogenous compounds and xenobiotics.
A.) Glutathionylation of the oxylipin OPDA has been demonstrated *in planta* and is a nucleophilic conjugate addition B.) Glutathionylation of atrazine is a substitution reaction releasing HCl. Figure adapted from Dixon et al., 2010.

Secondly, as well as the catalytic ability to conjugate reactive species with glutathione some families of glutathione *S*-transferases have evolved novel catalytic function. For example the dehydroascorbate reductase (DHAR) class is able to catalyse the conversion of dehydroascorbate to ascorbic acid (Dixon et al., 2002) (Figure 2-9). GSTs of the F, U and T classes have been shown to reduce hydroperoxide substrates to their respective alcohols with concomitant release of GSSG (Dixon and Edwards, 2009). Additionally there are a number of examples of this family catalysing cis-trans isomerisation reactions (Fernández-Cañón and Peñalva, 1998) (Figure 2-9).

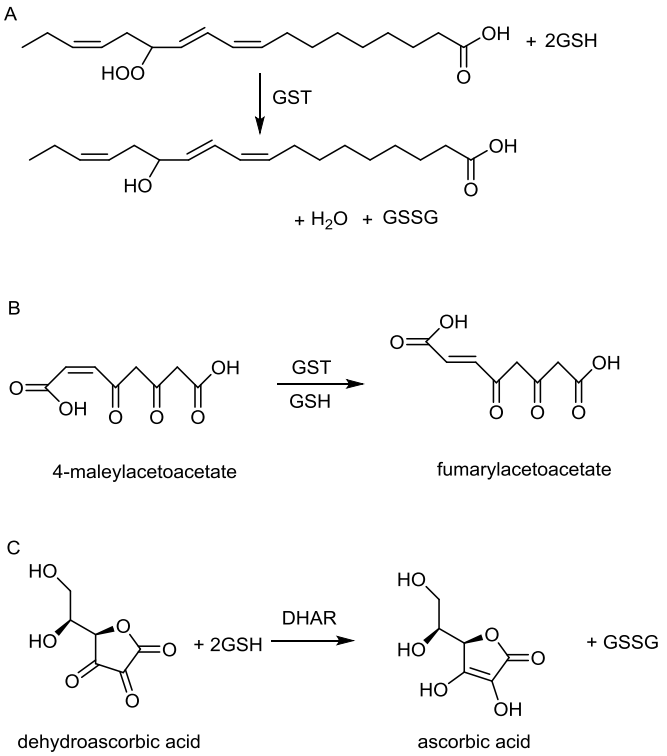


Figure 2-9 Non-canonical enzymatic reactions catalysed by GSTs.
Glutathione *S*-transferases can catalyse non-canonical reactions such as A.) Reduction of hydroperoxides B.) cis-trans isomerisation and C.) conversion of dehydroascorbate to ascorbic acid (Dixon et al., 2010).

Third, it is feasible that the glutathione *S*-transferases do not play a catalytic role at all in the conjugation of either glutathione, or the alternative catalytic capabilities highlighted above. Instead their role could be transient, whereby they bind small molecule intermediates, in the presence of GSH, and in doing so, act as ligandins. This acts to either protect reactive small molecule pathway intermediates by binding, or direct transport of the small molecule to the correct subcellular location.

There have been a number of examples in which *Arabidopsis* glutathione transferases of the L, F and U class have been able to bind to biologically relevant ligands. A ligand fishing approach, in which wheat GSTLs were recombinantly expressed in *E. coli*, affinity- immobilised and incubated with plant extracts from *Arabidopsis*, *N. benthamiana* and wheat, demonstrated that TaGSTL1 selectively bound a number of flavonoids derived from kaempferol and quercetin, such as rutin, kaempferol-3-*O*-rutinoside, and taxifolin (Dixon and Edwards, 2010). Additionally the *Arabidopsis* GSTF family member AtGSTF2, utilising a similar ligand fishing approach, was shown to specifically bind a number of heterocyclic compounds, including norharmane, harmane,

indole-3-aldehyde, 1-acetyl-β-carboline and the indole derived phytoalexin, camalexin (Figure 2-10) (David et al., 2011). To date the physiological relevance of such interactions is not known.

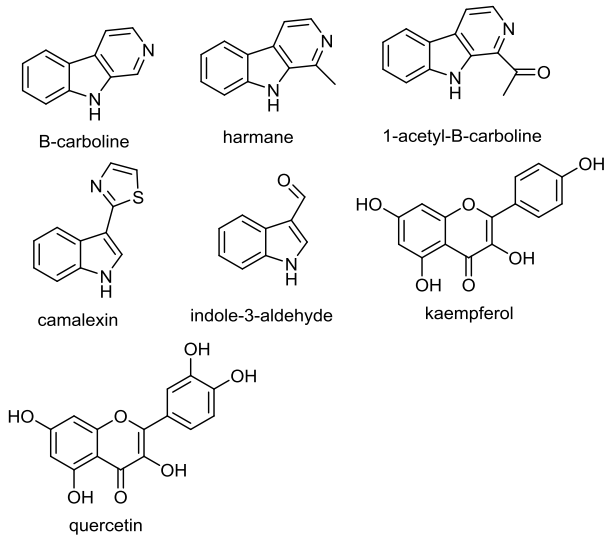


Figure 2-10 Compounds binding GSTs ascertained by ligand fishing

In a similar ligandin role, a number of glutathione *S*-transferases are directly involved in flavonoid transport and their sequestration in the vacuole in a variety of plant species. As discussed in the previous section, ABCC type transporters have been implicated in flavonoid transport, where they are thought to transport either flavonoid-glutathione conjugates, or flavonoids together with glutathione as a co-substrate.

GSTs may bind both the flavonoid and glutathione and act in the delivery of the substrate to the transporter. In petunia flowers a mutant allele in anthocyanin biosynthesis that had a decrease level of cyanidin 3-*O*-glucoside but no effect on flavonol accumulation, was identified by transposon mutagenesis, *An9*, and shown to have homology to the glutathione *S*-transferase gene family (Alfenito et al., 1998). Although this protein was shown to have glutathione conjugating activity towards the substrate 1-chloro-2,4-dinitrobenzene (CDNB), it was not able to efficiently conjugate cyanidin 3-*O*-glucoside, which would be expected to be its *in vivo* substrate (Alfenito et al., 1998). It is possible therefore that this protein is involved in binding and delivery of cyanidin 3-*O*-glucoside for its import into the vacuole. Similarly in maize, the *Bz2* mutant results in the same metabolic phenotype as *An9*, with the accumulation of cyanidin 3-*O*-glucoside. Bz2 is also a glutathione *S*-transferase, however in a different structural class to An9 (Marrs et al., 1995). The An9 and Bz2 proteins are able to reciprocally complement the

phenotypes observed, with overexpression of *An9* in *bz2* deficient maize, and overexpression of *Bz2* in *an9* deficient petunia, being able to complement the phenotype (Alfenito et al., 1998). This demonstrates that these structurally diverse GSTs are likely to be performing the same function in anthocyanin transport. Similarly an *Arabidopsis* mutant, *transparent testa 19*, *tt19*, also displayed a decrease in anthocyanin pigmentation colouration, and was also shown to be involved in proanthocyanidin sequestration to the vacuole (Kitamura et al., 2004). This phenotype could be partially complemented by the petunia *An9*. Additionally in grapevine a GST was identified that could complement the maize *bz2* mutant (Ageorges et al., 2006) and when silenced by RNAi in hairy root resulted in differential compartmentation of anthocyanins.

To date the precise role by which GSTs function to sequester anthocyanins in the vacuole is still unclear, despite the fact that this mechanism is conserved across multiple plant species. There is a need for experimental evidence for how its action as a flavonoid carrier protein facilitates transport. The two primary hypotheses are that it is involved in either vesicular trafficking or providing vacuolar transporters with their co-substrates (Zhao, 2015).

As has been highlighted, GSTs *in planta* have been shown to have a range of physiological functions, from the detoxification of exogenously supplied herbicides by conjugation with the tripeptide glutathione, to flavonoid transport. Often it has been hard to determine the mechanistic basis behind the role the glutathione transferase plays *in planta.* Similarly due to the differential roles these enzymes have been shown to display *in planta*, it has been particularly hard to predict the physiological function of orphan GSTs.

One mechanism by which it is possible to predict function is through transcriptomic profiling, and analysis of the coordinated expression of GSTs with known enzymatic activities. A key example of this is with *TT19* in *Arabidopsis*, which is transcriptionally co-regulated with genes in the anthocyanin biosynthetic pathway (Kitamura et al., 2004).

This methodology was also elegantly displayed in Sappl et al., 2009 which surveyed the induction of GSTs under differential stress conditions, and was able to highlight four GSTs that were regulated in accordance with application of salicylic acid. A reverse genetics approach silencing all four GSTs resulted in no clear metabolic phenotype, indicating that there might be functional redundancy in these GSTs role *in vivo* (Sappl et al., 2009). This contrasts with the role of the GSTs in flavonoid transport, where silencing of one locus results in a clear metabolic phenotype, therefore indicating that such functional redundancy may be dependent on the physiological role of the GST.

2.3.4.4 Glutathione S-transferases in C. roseus.

As highlighted, the diverse nature in which GSTs are involved in plant metabolism, the fact they have been shown to interact with indole based substrates *in vivo*, as well as the fact that they have been implicated in the subcellular transport of plant secondary metabolites, makes this an interesting category of enzymes to assess their involvement in MIA metabolism in *C. roseus*.

Based on the precedence for the GST *TT19* to be co-regulated with the anthocyanin biosynthetic pathway, GSTs were analysed for co-regulation with genes known in the MIA pathway.

There are 50 glutathione *S*-transferases annotated in the *C. roseus* transcriptome. The most abundantly expressed GST in leaf tissue, locus_7197, is also co-ordinately regulated with the known genes in the *C. roseus* transcriptome, as highlighted in Figure 2-11.

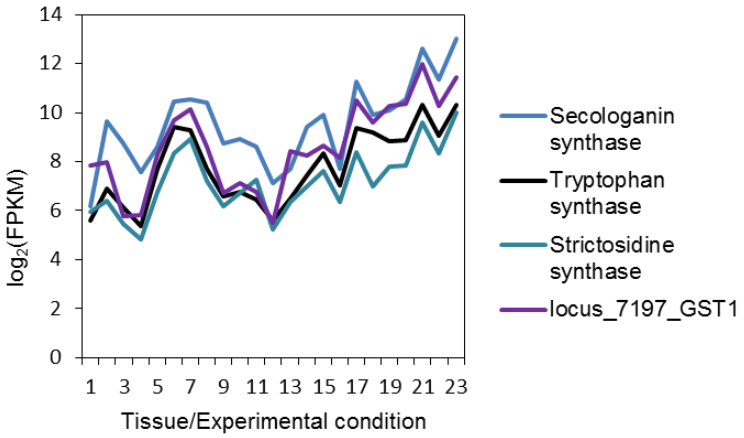


Figure 2-11 Co-regulation of locus_7197 GST1 with known genes in the biosynthetic pathway
Expression profile of locus_7197_GST1 measured as the log_2(FPKM) in the 23 different experimental conditions highlighted in Table 1. The expression profile of the known co-regulated genes that are involved in MIA biosynthesis; secologanin synthase, tryptophan synthase and strictosidine synthase are also shown for comparison.

2.3.4.5 Virus Induced Gene Silencing of Glutathione S-transferases in C. roseus

The locus_7197 together with two other glutathione *S*-transferases, locus_6502_GST2 and locus_2878_GST3, that, although were not co-ordinately regulated with the MIA pathway, were abundantly expressed in leaf tissue, were targeted by VIGS, utilising pTRV2u constructs containing ~500 bp of each GST of interest. The silencing experiment was repeated twice in two different cultivars of *C. roseus*, cv. Little Bright Eyes and cv. Sunstorm Apricot (Figure 2-12), to assess whether the lack of a phenotype was due to the cultivar used.

As can be seen in both cultivars, there was no statistically significant difference in the alkaloid profile of strictosidine, catharanthine, vindoline or serpentine/alstonine, as well as no significant difference in the iridoid terpene precursor secologanin, in either cultivar. Additionally, there was no reproducible accumulation of any new peaks when analysed by untargeted metabolomics. Therefore there is no definitive evidence that the GSTs analysed are involved in the MIA pathway.

It is possible that the GSTs may have functional redundancy, and therefore to assess whether there is a metabolic phenotype there would need to be combinatorial silencing of multiple genes. However, this contrasts with the specificity of the role that the GSTs play in flavonoid transport.

These GSTs could be heterologously expressed, and assessed whether they have binding capacity to any of the known MIAs, as has been described for a number of *Arabidopsis* GSTs, however the functional relevance of this interaction would still be physiologically unclear.

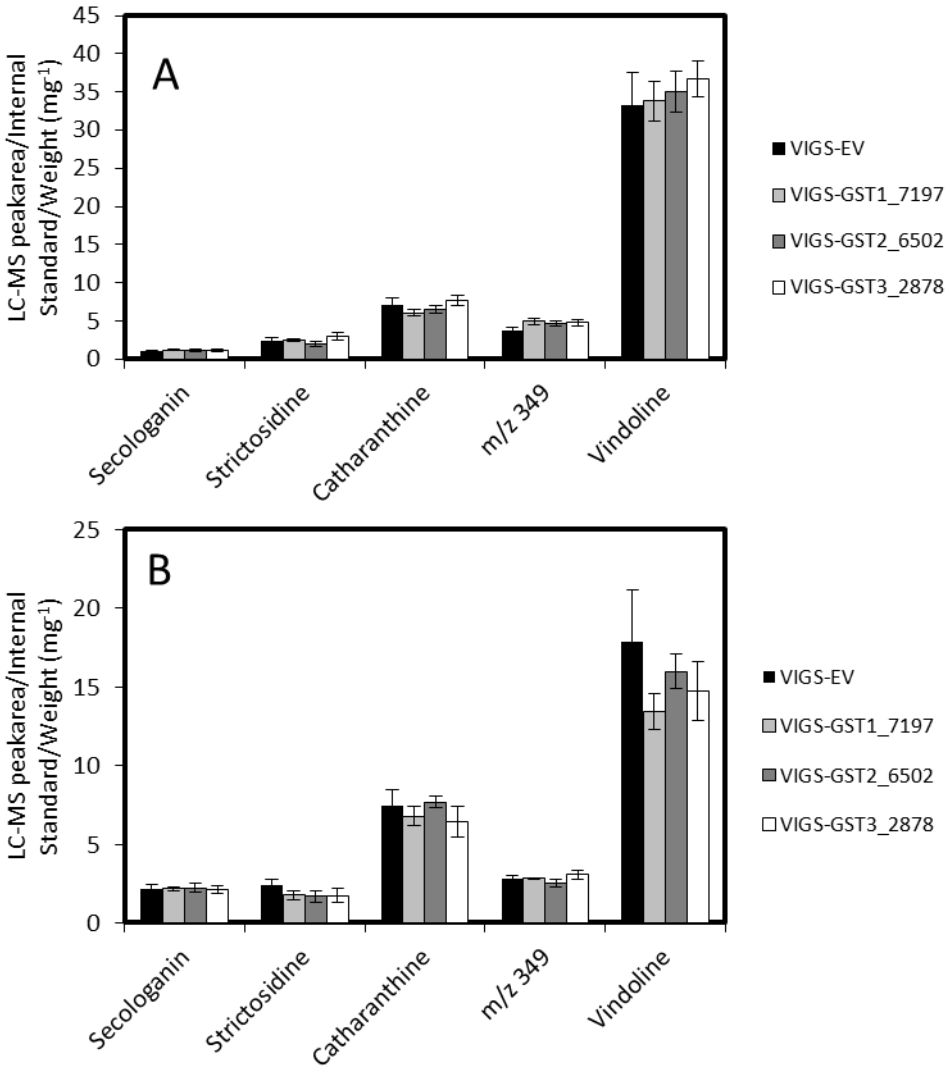


Figure 2-12 Metabolite analysis upon silencing Glutathione *S*-transferases relative to EV control
A.) Alkaloid profile for VIGS of GST1_7197, GST2_6502, GST3_2878 relative to Empty Vector control tissue in cv. Little Bright Eyes VIGS-GST1_7197(n=11), VIGS-GST2_6502 (n=10), VIGS-GST3_2878 (n=10), VIGS-EV (n=10) ** $p<0.01$ B.) Alkaloid profile for VIGS of GST1_7197, GST2_6502, GST3_2878 relative to Empty Vector control tissue in cv. Sunstorm Apricot VIGS-GST1_7197(n=8), VIGS-GST2_6502 (n=8), VIGS-GST3_2878 (n=8), VIGS-EV (n=8) ** $p<0.01$. All data presented is mean $\pm$ SEM.

2.3.4.6 The Drug/Metabolite Transporter Superfamily

As mentioned there are five main families of proteins that have been shown to be involved in multidrug resistance, and this includes a drug/metabolite transporter (DMT) superfamily (Jack et al., 2001). The DMT superfamily is a large ancient family that encompasses 14 subfamilies, including the small multidrug resistance (SMR) transporter family, which has previously been shown to result in multidrug resistance in bacteria. In plants these subfamilies include the plant drug/metabolite exporter (P-DME)- family, the triose phosphate transporter (TPT)-family, the nucleotide sugar transporters, which are selective for different nucleotide sugars, and the plant purine permease family (PUP) (Jack et al., 2001).

2.3.4.7 Purine permease family

The purine permease (PUP) family of transporters is a distinct group of transport proteins in the DMT superfamily. These transporters typically contain between 9-10 transmembrane helices, and are secondary transporters using the electrochemical gradient for energy.

These plant specific purine transporters have been shown to be involved in the transport of a number of alkaloid compounds, including the purine adenine, the purine based phytohromones such as the cytokinins (Bürkle et al., 2003; Gillissen et al., 2000) as well as the pyridine alkaloid nicotine (Hildreth et al., 2011).

There are 21 PUP like genes present in the model plant species *Arabidopsis*, however, to date, only 3 PUP-like transporters have been characterised from this species, AtPUP1-3 (Jelesko, 2012). AtPUP1 was identified as an adenine transporter through functional complementation of an *Arabidopsis* cDNA library into a yeast strain *fcy2* that was deficient in adenine and cytosine transport. AtPUP1 and AtPUP2 have subsequently been shown to be able to transport the cytokinins, trans-zeatin and kinetin, whilst the substrate for AtPUP3 is still unknown (Bürkle et al., 2003; Gillissen et al., 2000). Furthermore AtPUP1 has also been shown to be able to transport pyridoxine (vitamin B6) (Szydlowski et al., 2013).

The purine permeases have also been implicated in the transport of secondary metabolites in tobacco. The pyridine alkaloid nicotine is a key chemical defense compound of tobacco. The biosynthesis of nicotine is upregulated in response to wounding and herbivory, as well as in response to the exogenous elicitation of MeJA, and the regulation of this biosynthetic pathway is under the coordinate control of two regulatory loci called *NIC1* and *NIC2* (Hibi et al., 1994).

Nicotine is generated exclusively in the root tips, however is localised to other plant tissues, including the stem, and leaves. This implicates the need for organ specific transport. Nicotine is loaded in the xylem for transport to foliar tissues and, in this leaf tissue nicotine accumulates in the vacuoles of mesophyll cells (Morita et al., 2009). The transport of nicotine into the vacuole is due to the action of a number of MATE proteins which will be reviewed in Chapter III.

The purine permease NUP1 has been implicated in nicotine transport in tobacco (Hildreth et al., 2011). NUP1 shares 56% identity to the *Arabidopsis thaliana* protein, AtPUP1, and was shown to be co-ordinately regulated by the *NIC1* and *NIC2* loci in *N. tabacum*, and induced upon elicitation with MeJA (Hildreth et al., 2011). Transport of radiolabelled nicotine into *Schizosaccharamyces pombe* was shown for cells heterologously expressing the purine permease (NUP1). This appeared to be substrate specific as transport was not efficiently competed for with the tropane alkaloids, atropine and scopolamine, the pyridine alkaloids, anabasine and anatabine (Hildreth et al., 2011).

Overexpression of NUP1 as a GFP-fusion in tobacco BY2-cells indicated that this protein localised to the plasma membrane, and therefore implicates NUP1 as a nicotine uptake permease (Hildreth et al., 2011). Nicotine levels in hairy root cultures of *N. tabacum* with an NUP-RNAi silencing construct resulted in a lower nicotine content in the hairy roots and a comparative increase in the nicotine levels in the root culture media. The role of NUP1, as a plasma membrane nicotine uptake transporter in roots, that does not influence root to shoot transport, suggests it may be involved controlling the capacity of the roots to retain nicotine. This therefore implicates this family in controlling the reabsorption of nicotine from the rhizosphere or involved in controlling the apoplastic nicotine concentration, which may be involved in maintaining whole plant nicotine levels (Hildreth et al., 2011).

Recently NUP1, has also been shown to have the capacity to transport vitamin B6, as well as having substrate preference for the naturally occurring (*S*)-isomer of nicotine (Kato et al., 2015).

The role of the purine permeases in the transport of purine and pyridine alkaloids make them an intriguing family to decipher their role in other metabolic pathways.

2.3.4.8 Purine permeases in C. roseus

In *C. roseus* there are 30 members of the purine permease superfamily. The putative purine permease, CrPUP_1443 does not coregulate with the gene expression profile for known genes in the pathway, however is the most highly expressed purine permease present in leaf tissue.

VIGS of the putative purine permease CrPUP_1443 did not lead to a metabolic change in comparison to empty vector controls (Figure 2-13). As can be seen in both replicates, there was no statistically significant difference in the alkaloid profile of strictosidine, catharanthine, vindoline or *m/z* 349, as well as no significant difference in the iridoid terpene precursor secologanin, in either cultivar. Together with this, there was no reproducible accumulation of any new peaks when analysed by untargeted metabolomics.

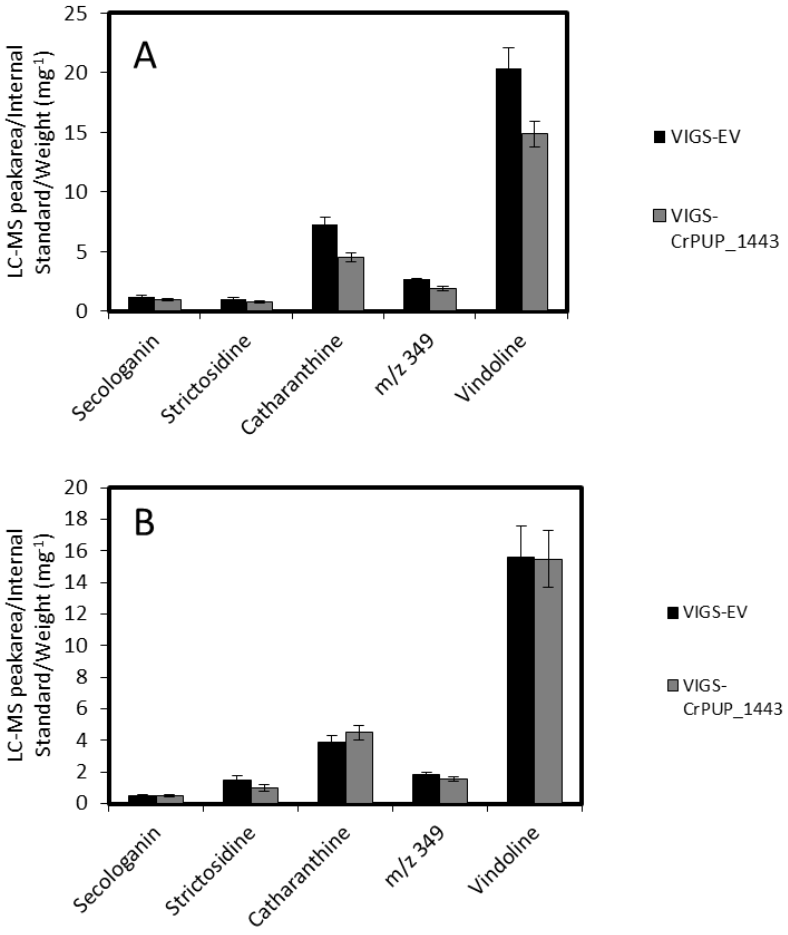


Figure 2-13 Metabolite Analysis upon silencing CrPUP locus_1443 relative to EV control
A.) Alkaloid profile for VIGS of CrPUP_1443 relative to Empty Vector control tissue in Little Bright Eyes VIGS CrPUP_1443 (n=9) VIGS-EV (n=8) ** $p<0.01$ **B.)** Alkaloid profile for VIGS of CrPUP_1443 relative to Empty Vector control tissue in Little Bright Eyes VIGS CrPUP_1443 (n=8) VIGS-EV (n=8) ** $p<0.01$. All data presented is mean $\pm$ SEM.

As will be shown in Chapter III this putative purine permease, when expressed as a C-terminal GFP fusion in *S. cerevisiae* localises to the yeast plasma membrane. Therefore even though this VIGS experiment does not give a characteristic metabolic phenotype, it would be useful to test and screen this yeast strain for uptake of known MIAs to assess whether this protein family has substrate specificity for other alkaloids other than the purine based cytokinins, or pyridine based alkaloids such as nicotine. To date this reverse genetics approach does not indicate the involvement of this transporter in MIA biosynthesis.

2.3.4.9 The plant drug/metabolite exporter (P-DME) Family

The plant drug/metabolite exporter (P-DME) Family is a subfamily of the DMT superfamily that is found exclusively in plants, and is most related to the drug/metabolite exporter (DME) family that are found in bacteria (Jack et al., 2001). The majority of proteins in this family contain 10 transmembrane helices. To date, this protein family has been underrepresented in their functional characterisation. In *Medicago truncatula* a P-DME transporter is thought to be involved in nodulation and was designated MtNodulin21 (Gamas et al., 1996), whilst in *Arabidopsis* only one transporter has been thoroughly investigated for its function, WAT1 (Ranocha et al., 2010).

WAT1 is a transport protein that is likely to be involved in the transport of auxin. Cellular localisation of WAT1 expression demonstrates that this transporter is expressed ubiquitously throughout the plant (Ranocha et al., 2010). Overexpression of WAT1 as a C-terminal GFP fusion in *Arabidopsis* and *Nicotiana benthamiana* demonstrated that WAT1 is localised to the tonoplastic membrane (Ranocha et al., 2010).

Mutant lines carrying a T-DNA insertion in *wat1* resulted in a number of physiological phenotypes. The mutants had a smaller leaf size, with increased chlorophyll content, shorter stems, and irregular organisation of epidermal cells. The *wat1* mutant had less lignin, and in particular was deficient in S unit lignin present in xylem fibres, together with a decrease in the non-cellulosic polysaccharide, xylan, suggesting its involvement in secondary cell wall and fibre development (Ranocha et al., 2010).

Tryptophan metabolism was significantly modified in *wat1* mutants, with the free amino acid being reduced relative to wild type, together with the accumulation of other indolic compounds (Ranocha et al., 2010). This was accompanied by a decrease in the free indole acetic acid (IAA) concentration in the basal portion of the stem. This is the first time a tonoplastic membrane

protein has been shown to be involved in secondary cell wall formation as well as involved in tryptophan and auxin metabolism (Ranocha et al., 2010).

Isolated vacuoles from wild type and *wat1* mutant plants have been fed with ^{14}C-IAA. The *wat1* mutants retained a higher level of ^{14}C-IAA indicating that WAT1 is likely to act as an exporter of IAA from the vacuole. Moreover heterologous expression in yeast and *Xenopus laevis* oocytes demonstrated transport of IAA and implicates this transporter as an exporter of auxin from the vacuole (Ranocha et al., 2010).

2.3.4.10 C. roseus WAT1 homologue

In *C. roseus* a homologue of WAT1 co-regulates with known genes in the metabolic pathway of MIA biosynthesis (Figure 2-14). Given that an *Arabidopsis* WAT1 homologue transports auxin, which has an indole moiety much like the MIAs, this candidate appeared to be a good prospect for involvement in MIA metabolism.

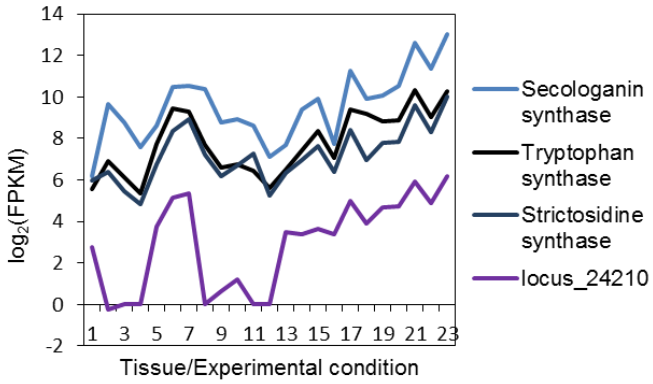


Figure 2-14 Co-regulation of locus_24210 with known genes in the biosynthetic pathway.
Expression profile of locus_24210 measured as the $\log_2$(FPKM) in the 23 different experimental conditions highlighted in Table 1. The expression profile of the known co-regulated genes that are involved in MIA biosynthesis; secologanin synthase, tryptophan synthase and strictosidine synthase are also shown for comparison

VIGS of the WAT1 homologue, locus_24210, which was annotated as an auxin responsive factor 5NG4, did not result in a metabolic phenotype when targeted by VIGS, in multiple biological replicates, suggesting that this transporter is not involved in MIA metabolism in *C. roseus* (Figure 2-15).

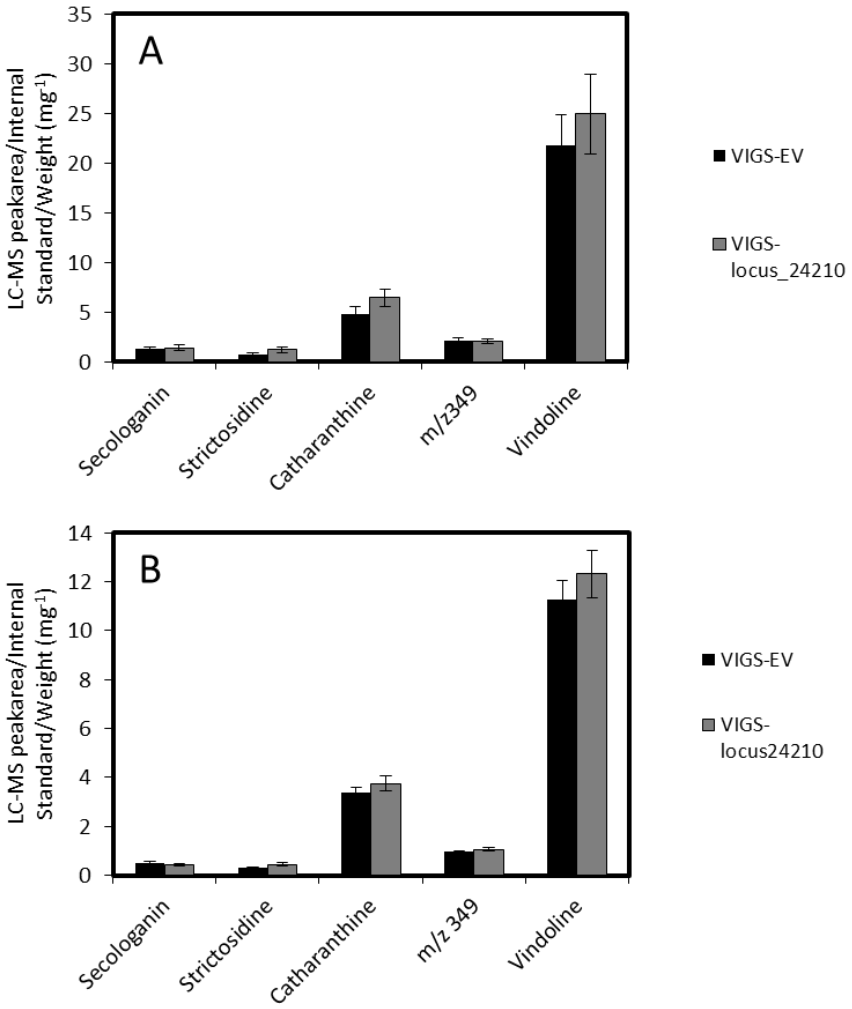


Figure 2-15 Metabolite analysis VIGS locus_24210 relative to EV Control.
A.) Alkaloid profile for VIGS of WAT1 locus_24210 relative to Empty Vector control tissue in cv. Little Bright Eyes VIGS-24210 (n=12) VIGS-EV (n=12) ** $p<0.01$ B.) Alkaloid profile for VIGS of WAT1 locus_24210 relative to Empty Vector control tissue in cv. Little Bright Eyes VIGS-24210 (n=16) VIGS-EV (n=14) ** $p<0.01$. All data presented is mean $\pm$ SEM.

2.3.4.11 The role of the transcription factor BIS1 in MIA metabolism

This work was performed in collaboration with Alain Goossens, University of Ghent. VIGS constructs were cloned by Alex Van Moerkercke and the VIGS experiment and metabolite analysis was performed by Richard Payne. This work has been published in (Van Moerkercke et al., 2015)

As well as the enzymes and transporters that are integrally involved in MIA biosynthesis, a reverse genetic approach can be utilised to identify genes that are involved in the regulation of the MIA metabolic pathway. VIGS of transcription factors involved in plant secondary metabolism has been successfully applied in a number of plant species, including *Nicotiana attenuata* (Woldemariam et al., 2013), *Nicotiana benthamiana* (Todd et al., 2010) and *Capsicum eximium* (Aguilar-Barragán and Ochoa-Alejo, 2014).

MIA biosynthesis in *C. roseus* is induced by the phytohormone jasmonate (JA). This induction is controlled by the action of transcription factors, many of which are also inducible themselves by JA (De Geyter et al., 2012). The best characterised transcription factor that is known to regulate MIA metabolism in *C. roseus* is ORCA3, which controls expression of the indole branch of the MIA pathway (van der Fits and Memelink, 2000). However the iridoid genes upstream of LAMT, (Figure 2-16), despite being JA inducible, are not under regulatory control by ORCA3, suggesting other transcription factors may be involved in this regulatory network (van der Fits and Memelink, 2000).

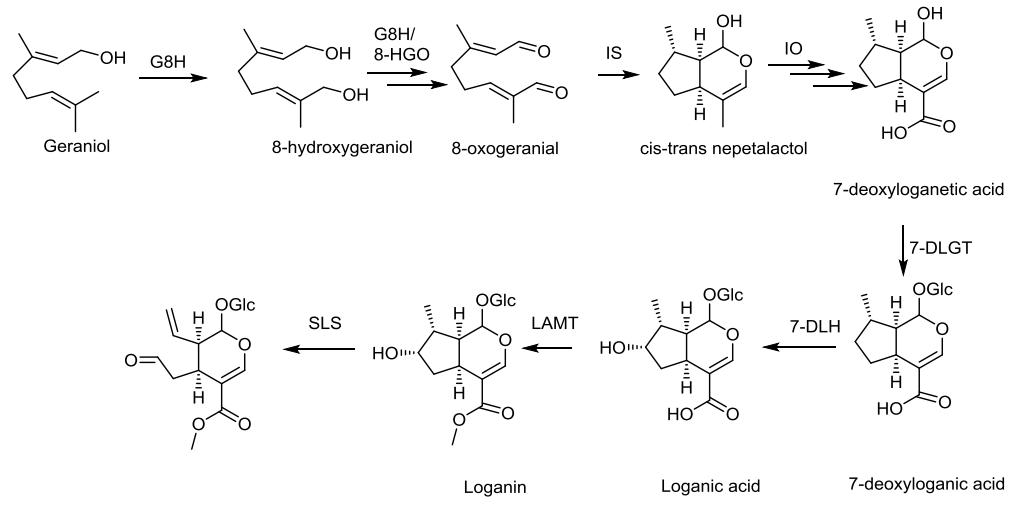


Figure 2-16 Overview of secologanin biosynthesis.
Secologanin is a glycosylated iridoid terpene derived from the MEP pathway. It is synthesised by a nine-step biosynthetic pathway comprising the enzymes G8H-geraniol 8 hydroxylase, 8-HGO- 8-hydroxygeraniol oxidoreductase, IS-iridoid synthase, IO-iridoid oxidase, 7-DLGT- 7deoxyloganetic acid glucosyltransferase, 7-DLH - 7-deoxyloganic acid hydroxylase, LAMT-loganic acid methyltransferase and SLS- secologanin synthase. The genes upstream of LAMT are not under transcriptional control of ORCA3.

A basic helix-loop-helix (bHLH) transcription factor from *C. roseus*, BIS1, was identified that co-regulates with these early iridoid genes and is induced by JA (Van Moerkercke et al., 2015). The bHLH transcription factor family has been implicated in terpene biosynthesis in other plant species such as *Arabidopsis thaliana* (Hong et al., 2012), *Artemisia annua* (Ji et al., 2014), and *Solanum lycopersicum* (tomato) (Spyropoulou et al., 2014).

The bHLH transcription factor BIS1, was shown to specifically transactivate the promoters of the genes upstream of LAMT in iridoid biosynthesis from *C. roseus*, namely geraniol synthase, 8-hydroxygeraniol oxidase, iridoid synthase, 7-deoxyloganetic acid glucosyltransferase, and 7-deoxyloganic acid hydroxylase, in a *Nicotiana tabacum* protoplast assay system (Van Moerkercke et al., 2015).

Overexpression of the transcription factor was shown to boost MIA production in *C. roseus* cell suspension cultures, and affect iridoid gene expression when overexpressed in *C. roseus* hairy root lines, however with a lack of effect on MIA production (Van Moerkercke et al., 2015).

As with the other secoiridoid genes, this transcription factor is preferentially expressed in the internal phloem associated parenchyma (IPAP), indicating that this transcription factor localises to the same cell type as the early part of the secoiridoid pathway (Van Moerkercke et al., 2015). To assess the role of this transcription factor in the whole plant, the bHLH transcription factor, BIS1, was silenced by VIGS.

As highlighted in the Figure 2-17 there was no clear phenotype on the MIAs upon silencing, or on the level of secologanin in the leaf tissue. Furthermore there was no accumulation of any new peaks when analysed by an untargeted metabolomics approach.

There are a number of reasons why no major effect on MIA levels was observed by VIGS. It is possible that there may be functional redundancy with other transcription factors, and therefore despite transactivation of the metabolic pathway *in vitro*, silencing of this transcription factor can be compensated by other genes in the MIA regulatory network. This study therefore highlights the need to utilise a reverse genetics approach to determine the physiological relevance of a gene of interest in metabolic pathways.

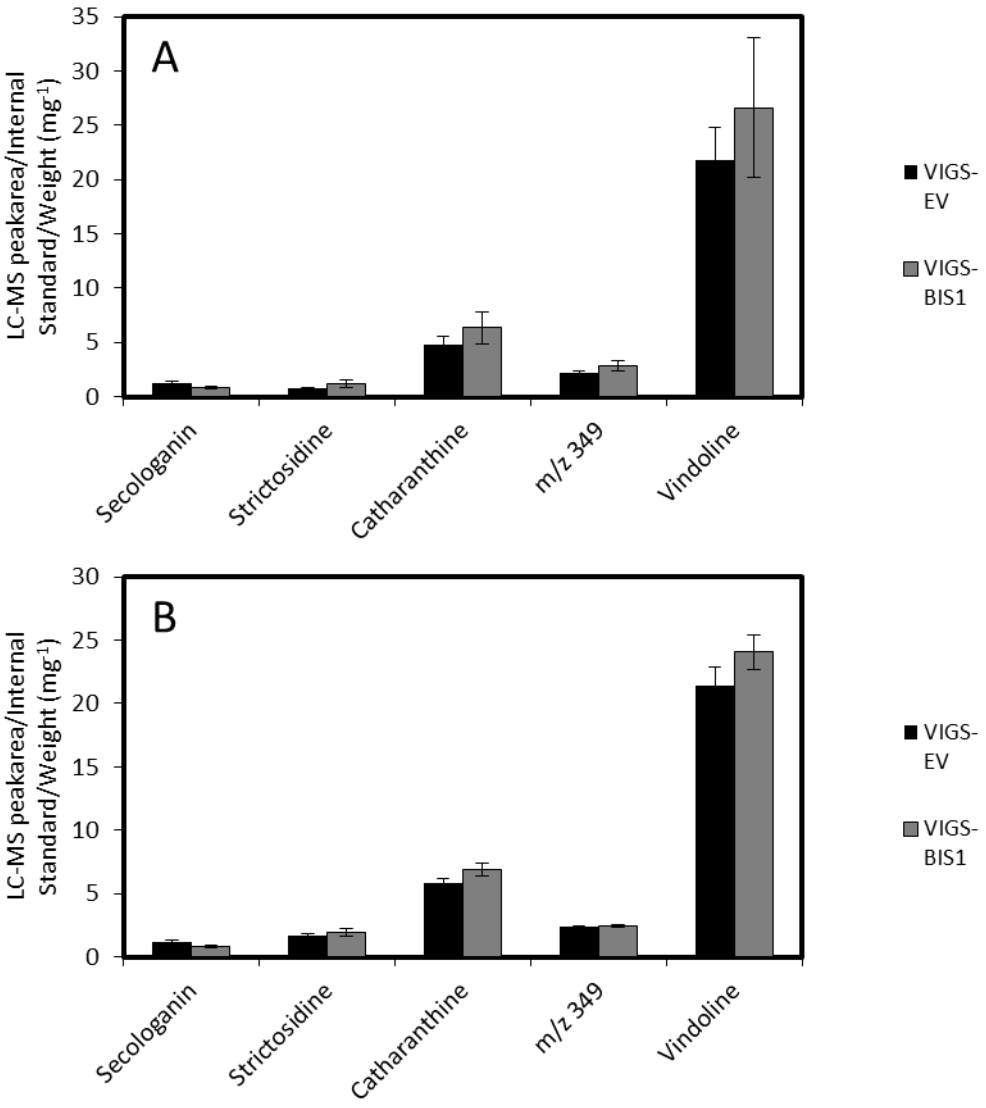


Figure 2-17 Metabolite analysis VIGS-BIS1 relative to EV Control
A.) Alkaloid profile for VIGS of BIS1 relative to Empty Vector control tissue in cv. Little Bright Eyes VIGS-BIS1 (n=13) VIGS-EV (n=12) ** $p<0.01$ B.) Alkaloid profile for VIGS of BIS1 relative to Empty Vector control tissue in cv. Little Bright Eyes VIGS-BIS1 (n=12) VIGS-EV (n=12) ** $p<0.01$. All data presented is mean $\pm$ SEM.

2.3.4.12 The progesterone 5β-reductase family in C. roseus.

This work was performed in collaboration with Alain Goossens, University of Ghent. VIGS constructs were cloned by Alex Van Moerkercke and the VIGS experiment and metabolite analysis was performed by Richard Payne. This work has been published in (Munkert et al., 2015)

A key early step in the *C. roseus* MIA pathway is the generation of the seco-iridoid terpene moiety secologanin, and, as has been discussed in Chapter I the complete seco-iridoid pathway has been recently elucidated (Miettinen et al 2014).

A key enzyme in the pathway is iridoid synthase, which catalyses the non-canonical cyclisation of 8-oxogeranial to cis-trans nepetelactol (Geu-Flores et al., 2012). This enzyme is a member of the progesterone 5β-reductases, which are known to catalyse the reduction of the double bond of the α,β-unsaturated carbonyl of progesterone (Figure 2-18).

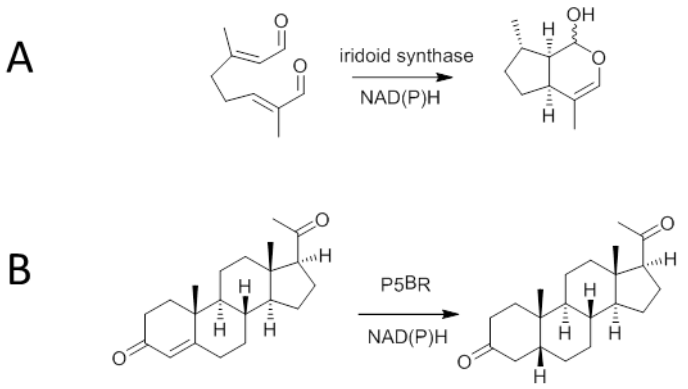


Figure 2-18 Cyclisation of 8-oxogeranial to cis-trans nepetelactol catalysed by iridoid synthase
A.) Reaction catalysed by iridoid synthase B.) Reaction catalysed by progesterone 5β-reductase

As well as iridoid synthase there are five other homologues of the progesterone 5β-reductase family in the *C. roseus* transcriptome, CrP5βR1-CrP5βR6, with CrP5βR5 being the iridoid synthase identified by (Geu-Flores et al., 2012). These five homologues, together with iridoid synthase are all localised to the cytosol when overexpressed as N-terminal GFP fusions in tobacco BY2 cells (Munkert et al., 2015).

It has been shown that CrP5βR5 is co-expressed with a number of known genes in the biosynthetic pathway, including geraniol synthase, geraniol 8-oxidase, and 8-hydroxygeraniol oxidoreductase (Munkert et al., 2015). Similarly CrP5β5R4 also co-expresses with the genes involved in MIA biosynthesis, including LAMT, secologanin synthase, strictosidine synthase, and strictosidine β-glucosidase. The homologues CrP5βR1, CrP5βR3, CrP5βR6 and CrP5βR2 however

are constitutively expressed and do not follow the expression pattern of MIA known genes, and are relatively lowly expressed in leaf tissue. Cellular localisation of CrP5βR4 and CrP5βR5/iridoid synthase demonstrate that both localise to the IPAP cells (Munkert et al., 2015).

Recombinant CrP5βR4, CrP5βR5 and CrP5βR2 could turnover 8-oxogeranial *in vitro*, to form cis-trans nepetelactol, and the steady state kinetic constants were similar for all three enzymes (Munkert et al., 2015). This demonstrated that other homologues of the progesterone 5β-reductase family in *C. roseus* are capable of catalysing the cyclisation of 8-oxogeranial (Munkert et al., 2015).

It has previously been shown in Geu-Flores et al., 2012 that silencing of CrP5βR5 by VIGS results in the net decrease in vindoline and catharanthine, together with the accumulation of three new peaks when monitored by LC-MS with *m/z* values of 526, 528 and 484, which were shown to be derivatized products of 8-oxogeranial (Geu-Flores et al., 2012).

To assess whether CrP5βR4 and CrP5βR2 also contribute to the MIA metabolic pathway, these genes were silenced, together with CrP5βR5 as a positive control. VIGS of the two genes CrP5βR4 and CrP5βR2 did not result in a decrease in vindoline, catharanthine, strictosidine or secologanin, and the characteristic iridoid peaks of *m/z* 526, 528 and 484 did not accumulate Figure 2-19. In comparison, the positive control CrP5βR5, although not resulting in a decrease in the MIAs as was observed in Geu-Flores et al., 2012, did result in the accumulation of the three diagnostic iridoid derived peaks (Figure 2-19).

This work demonstrates that while CrP5βR4 is able to catalyse the same reaction as CrP5βR5 *in vitro* with similar kinetic efficiency, is co-expressed in the metabolic pathway and localised to the same cell type as CrP5βR5, it appears CrP5βR5 is the dominant iridoid synthase enzyme in *C. roseus* MIA biosynthesis for which a metabolic phenotype is observed upon silencing. In the presence of CrP5βR5, CrP5βBR4 contributes little to the MIA metabolic pathway. To further assess this, combinatorial silencing of these genes by VIGS would be necessary to deduce the impact on the metabolic phenotype. This is consistent with the fact that progesterone 5β-reductases can be promiscuous and therefore *in vitro* activity and physiological function are not necessarily linked (Bauer et al., 2010).

This result in particular highlights the importance of a reverse genetics approach in determining the functional relevance of enzymatic activity *in planta.*

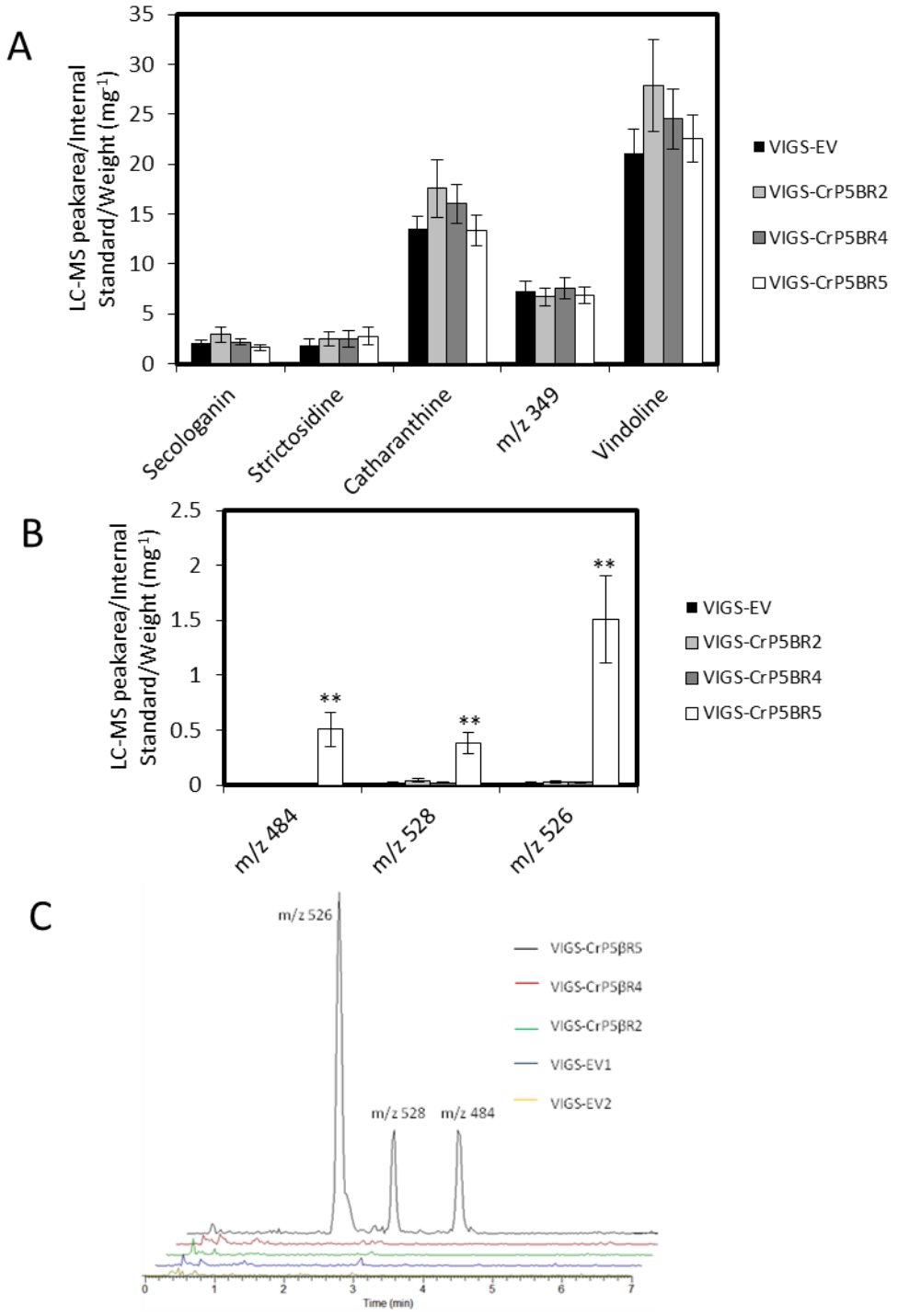


Figure 2-19 Metabolite Analysis upon silencing Progesterone 5β-reductases relative to EV control
A.) Alkaloid profile for VIGS of the progesterone 5β-reductases family relative to Empty Vector control tissue in cv. Little Bright Eyes VIGS-CrP5βR5 (n=10), VIGS-CrP5β4 (n=10), VIGS-CrP5βR2 (n=12),VIGS-EV (n=9) ** p<0.01 B.) Iridoid profile for VIGS of the progesterone 5β-reductases family relative to Empty Vector control tissue in cv. Little Bright Eyes, measuring *m/z* 484, *m/z* 526, *m/z* 528; VIGS-CrP5βR5 (n=10), VIGS-CrP5β4 (n=10), VIGS-CrP5βR2 (n=12), VIGS-EV (n=9) ** p<0.01. All data are mean ± SEM C.) Representative extracted ion chromatogram for *m/z* 484, *m/z* 526 and *m/z* 528 in VIGS-CrP5βR5, VIGS-CrP5β4, VIGS-CrP5βR2 and VIGS-EV tissue.

2.4 Conclusion

2.4.1 VIGS as a system for gene discovery

Overall this chapter highlights the use of transcriptome mining for the identification of potential candidates for screening in *C. roseus* by VIGS. The principle we have used is the assumption that genes within a secondary metabolic pathway are likely to be co-expressed. Utilising multiple independent strategies to look at co-regulation we identified a number of candidates with expression profiles that cluster with known genes in the MIA pathway. Genes that were ultimately disregarded due to a negative metabolic phenotype upon silencing are reported in this chapter. Successful candidates will be discussed in Chapters III, IV and V.

Co-regulation is clearly observed for the early stages in the biosynthesis of MIAs. Co-expression analysis has been employed successfully in the identification and characterisation of a number of genes in the seco-iridoid pathway (Geu-Flores et al., 2012; Miettinen et al., 2014). After the branchpoint intermediate strictosidine, the known genes which are involved in the biosynthesis of vindoline do not have a discernible co-expression profile.

This suggests that although the co-expression transcript mining approach is a good first attempt to identify gene candidates that might be involved in the MIA pathway, there are likely to be genes that cannot be identified using this approach and alternative strategies are needed.

This is likely to be true in the identification of enzymes involved in the branches of MIA biosynthesis that lead to downstream alkaloids. Alternative approaches to candidate identification include genome mining based strategies (Kellner et al., 2015b) whereby genes in secondary metabolic pathways may be spatially close on the genome, and this approach has been successful in the identification of an alcohol dehydrogenase involved in tetrahydroalstonine biosynthesis (Stavrinides et al., 2015).

Nevertheless this transcriptome mining approach has resulted in the identification of a number of genes that will form the basis for this thesis. Additionally the fact we are identifying genes that are co-regulated transcriptionally, indicates that the genes we have identified are likely to be involved in the early stages of MIA biosynthesis.

2.4.2 VIGS screening in C. roseus

In this chapter, we assessed the metabolic phenotypes after targeting a number of candidate genes, which have either been shown to co-regulate with known genes in the MIA pathway, are highly expressed in *C. roseus* leaf tissue (the site of alkaloid biosynthesis), have homology to genes involved in the MIA pathway, or have been implicated in MIA metabolism due to *in vitro* assays.

Silencing of the ABC transporter CrABC_1763, three glutathione *S*-transferases, two DMT family proteins, the transcription factor BIS1 and the progesterone 5β-reductases, CrP5βR4 and CrP5βR2, did not result in the identification of a gene that results in a discernible metabolic phenotype in *C. roseus*

There are a number of reasons as to why we do not see a metabolic phenotype.

Firstly the gene may not be involved in the MIA metabolic pathway, and therefore silencing of this gene does not lead to a metabolic phenotype.

Secondly it is possible that there is inefficient silencing using the vectors we have designed. As has been described in Chapter I, a region of the gene of interest is cloned into the pTRV2 vector, and this acts as a template for the formation of dsRNA and subsequently siRNA that can target the endogenous mRNA of the gene for degradation and therefore silencing. It is possible that the region selected for of the gene of interest does not lead to the formation of efficient siRNAs.

To compensate for this it would be necessary to redesign the region used for silencing and repeat the silencing experiments with a new vector to assess whether an alternative metabolic phenotype is generated. In this study, for each gene, a ~500 bp fragment has been used to generate the vector for silencing. This is a fairly extensive region for each gene of interest and increases the chance for the generation of potential silencing dsRNAs in the VIGS tissue as the 500 bp region is cleaved into 21-24 bp fragments. Additionally this 500 bp region in our study was from the centre of the coding sequence and did not contain any homopolymeric sequences. This conforms to the observation that for silencing of genes using the pTRV vectors in *N. benthamiana* efficient silencing has been observed with insert lengths in the region of ~200 bp to ~1300 bp, with most efficient silencing using regions from the centre of the coding sequence and without any homopolymeric sequences (Liu and Page, 2008).

A third reason for the negative VIGS phenotype may be due to functional redundancy of the gene of interest. If multiple genes can perform the same reaction in the plant, then silencing one of the genes might be compensated by the action of others.

For the transcription factor candidate BIS1, which was shown to transactivate the iridoid gene promoters *in vitro*, and induce MIA production when overexpressed in *C. roseus* cell suspension cultures, silencing of the gene by VIGS does not lead to a metabolic phenotype. Despite this *in vitro* demonstration that this transcription factor is likely to be involved in MIA biosynthesis, the regulatory network of MIA metabolism in *C. roseus* may be more complex than initially assumed, and the physiological role of BIS1 could be compensated for by another transcription factor.

Furthermore despite the fact that the progesterone 5β-reductases CrP5βR4 and CrP5βR2 have the catalytic capacity to cyclise 8-oxogeranial, and have similar kinetic constants to the characterised iridoid synthase, CrP5βR5, the VIGS screen demonstrates that only the silencing of CrP5βR5 results in a physiologically relevant metabolic phenotype. This demonstrates that it is necessary to utilise reverse genetics approaches in order to determine the physiological relevance of enzymatic activity *in planta*.

This work therefore highlights that it is necessary to assess the relationship of the gene of interest to other genes in the plant, and determine whether there are any homologues that could compensate for enzymatic activity.

It has been shown that the VIGS silencing approach can be used for the simultaneous silencing of multiple genes *in planta* by cloning multiple gene regions into the pTRV2 vector. This has been utilised for the redundant metabolic pathways involved in starch biosynthesis in *N. benthamiana* (George et al., 2012). Therefore for genes where there is the potential for genetic redundancy in *C. roseus*, the combinatorial silencing of the genes may be necessary to induce a metabolic phenotype. This approach will be highlighted in Chapters III, IV and V.

To definitively assess whether the lack of a metabolic phenotype is due to functional redundancy or inefficient silencing, qPCR of the gene of interest is necessary between VIGS-EV and VIGS-gene silenced plants, to validate that the silencing has failed. In this study we have utilised VIGS to efficiently screen for metabolic phenotypes, with genes that result in negative metabolic phenotype not being optimised further. If further evidence implicates these genes involvement in MIA metabolism, then qPCR of the VIGS tissue, and the generation of alternative VIGS vectors targeting a different region of the cDNA could be generated.

This work documents a number of the pitfalls of using VIGS and other reverse genetic engineering strategies for gene discovery. If a positive phenotype is observed then this is indicative that the gene is likely to be involved in the pathway. However a negative phenotype does not necessarily indicate that the gene is not involved in the pathway, as it is possible that there could be extentuating circumstances such as genetic compensation that could preclude an observable phenotype.

Ideally VIGS should be utilised as a methodology in tandem with combinatorial gene silencing, as well as functional characterisation of the genes of interest, to definitively assess involvement in the metabolic pathway. To date the work presented in this chapter indicates that the genes silenced do not have a role in MIA metabolism in *C. roseus*. If other data supports involvement of these genes, such as through a functional screen, then these candidates could be revisited using a VIGS approach that utilises a different region of the gene for silencing, or a method for combinatorial silencing.

Despite the negative phenotypes addressed in this chapter, VIGS screening has allowed for the relatively fast assessment of whether there is evidence for these genes to be physiologically involved in MIA metabolism by reverse genetics. This would not be feasible in the same timeframe as other approaches. As will be highlighted in Chapters III, IV and V, this transcriptomic mining and VIGS screening approach, has resulted in a number of positive phenotypes that indicate involvement in MIA metabolism in *C. roseus*.

3 Characterisation of *C. roseus* MATE proteins

3.1 Introduction to MATEs

The MATE family of transporters is the most recently characterised family of multidrug resistance transporters, and is a subfamily of the MOP flippase superfamily in accordance with the transporter classification database (Brown et al., 1999; Jack et al., 2001). The first MATE was identified in *Vibrio parahaemolyticus* NorM, along with its homologue YdhE in *E. coli* (Morita et al., 1998). These proteins possess 12 transmembrane helices, and utilise the electrochemical gradient of H^+ or Na^+ across the cellular membrane for transport, properties that are also found in the Major Facilitator Superfamily (MFS), though MATES do not possess sequence identity with the MFS transporter family (Brown et al., 1999). These proteins are therefore secondary transporters utilising the energy of the electrochemical gradient across the lipid bilayer, as opposed to direct hydrolysis of ATP as with the ABC transporter family. In bacteria and yeast MATE transporters have been shown to confer multidrug resistance to a variety of cationic drugs including ethidium bromide, tetraphenylphosphonium, berberine, acriflavine, norfloxacin and ethionine (Omote et al., 2006).

Two MATE proteins, SLC47A1 and SLC47A2, have been identified in the human genome. SLC47A1 is highly expressed in the kidney and liver, as well as in the first trimester placenta, whilst SLC47A2 is preferentially localised to the kidney. In the excretory organs, liver and kidney, the MATEs act in concert with organic cation transporters, for the transcellular movement of a diverse variety of substrates, including drugs, toxins, and a number of physiological compounds such as creatinine and thiamine (Motohashi and Inui, 2013; Staud et al., 2013). SLC47A1 knockout mice are viable and fertile without major genetic abnormalities. To date the major role of these proteins in humans is thought to be in excretion and detoxification, where they act in concert with other transporter families (Motohashi and Inui, 2013; Staud et al., 2013).

3.1.1 MATE transporters in plants: roles in growth and development

In plants the MATE family of transporters is large, with approximately 58 family members identified in *Arabidopsis* (Remy and Duque, 2014). This class of transporter has been shown to have direct involvement in a number of transport processes *in planta*, with primary roles so far elucidated in heavy metal detoxification (Green and Rogers, 2004), disease resistance (Nawrath

et al., 2002), responses to environmental stress (Zhang et al., 2014), growth and development (Li et al., 2014), as well as secondary metabolite transport (Morita et al., 2009).

Salicylic acid is a hormone synthesised by plants after exposure to pathogens, or stress, such as UV-light and leads to the expression of a number of pathogenesis related proteins. In *Arabidopsis*, *eds5* is a mutant that is hypersusceptible to pathogens and accumulates very little salicylic acid upon inoculation with virulent or avirulent pathogens, and after abiotic stress. *eds5* exhibits low level expression of pathogenesis related proteins and is therefore unable to establish a defense response. *EDS5* was cloned and could complement *eds5* mutants when expressed in *Arabidopsis*, and was shown to be a MATE transporter based on sequence homology (Nawrath et al., 2002). The transcript of *EDS5* was inducible by environmental stress, application of an avirulent strain of *P. syringae*, as well as upon exogenous treatment of salicylic acid (Nawrath et al., 2002). A C-terminal GFP fusion of this protein expressed in *Arabidopsis*, showed that the protein was localised to the chloroplast envelope, and this was confirmed by immunoblotting of the chloroplast envelope membranes with anti-GFP antibody (Serrano et al., 2013; Yamasaki et al., 2013). Transport assays of radiolabelled ^{14}C-salicylic acid, using chloroplasts isolated from both *eds5* mutant and *EDS5* overexpressing *Arabidopsis* lines, together with transport assays in yeast suggest this transporter can bidirectionally transport salicylic acid. This is further supported by a genetic study in which an ethanol inducible, chloroplastic salicylic acid synthase is expressed in chloroplasts of *eds5* mutants and EDS5 overexpressing plants. Upon ethanol induction *eds5* mutant chloroplasts contained a higher level of salicylic acid, indicating EDS5 is required for salicylic acid export from the chloroplast (Serrano et al., 2013). A homologue of EDS5, EDS5H, has been shown to also localise to the chloroplast envelope, but does not appear to function in salicylic acid biosynthesis (Parinthawong et al., 2015) and its physiological function remains cryptic.

In contrast to EDS5, ADS1, another *Arabidopsis* MATE protein, that has low homology to EDS5, has been shown to be a negative regulator of plant disease. Over expression of this MATE in *Arabidopsis* results in susceptibility to *P. syringae*, whilst expression of an RNAi construct leads to the same susceptibility as wild type plants. ADS1 is also likely to be involved in salicylic acid biosynthesis, as *ADS1* overexpressing plants have a lower SA content, however the specific transporter substrate has not been determined (Sun et al., 2011).

Overexpression of two rice MATEs, *OsMATE1* and *OsMATE2* in *Arabidopsis* have been implicated in pathogen susceptibility, together with altered floral morphology and sensitivity to arsenic, demonstrating that MATEs can have pleiotropic effects on plant physiology (Tiwari

et al., 2014). Again, the specific substrates for the transporters have not been identified, highlighting that deconvolution of the molecular function of transporters identfied by *in planta* assays is not always straightforward.

The role of MATE transporters in hormonal transport is not limited to salicylic acid. Abscisic acid (ABA) has been hypothesised to be transported in *A. thaliana* by a MATE transporter annotated as AtDTX50. Abscisic acid plays a variety of roles in plant growth and development, however is fundamental in response to environmental stress conditions. Under drought conditions an increase in the ABA level results in stomatal closure and an upregulation of stress genes (Hetherington and Woodward, 2003; Tuteja, 2007). A T-DNA insertion mutant in the gene of *AtDTX50* resulted in plants significantly smaller and yellow compared to wild type. In this insertion mutant no *AtDTX50* mRNA was detectable, and the phenotype was complemented by overexpression of *AtDTX50* (Zhang et al., 2014). The *dtx50* mutant lines displayed growth inhibition when grown on ABA containing medium, together with bleached leaves and root defects, and germinated more slowly than wild type seeds in the presence and absence of ABA. *dtx50* mutants grown in soil contained higher levels of ABA, and this was accompanied by up regulation of ABA marker genes in the mutant when analysed by qPCR. AtDTX50 localised to the plasma membrane, as determined by overexpression of a AtDTX50-GFP fusion protein in *Arabidopsis* mesophyll protoplasts, as well as in onion epidermal cells (Zhang et al., 2014). The fact that the *dtx50* mutant plants accumulated ABA and that the protein localised to the plasma membrane is indicative that AtDTX50 is likely to be an ABA efflux transporter from the cell. This was confirmed by transport assays in three different systems including *E. coli*, *Xenopus laevis* oocytes and *Arabidopsis* (Zhang et al., 2014).

Additionally, a MATE protein ADP1 has been implicated in regulating local auxin biosynthesis in *Arabidopsis.* A dominant *Arabidopsis* mutant *adp1-D* resulted in an increase in the number of axillary branches, flowers and lateral roots, and is the result of a gain of function mutation that overexpresses ADP1 (Li et al., 2014). Transgenic over expression of *ADP1* in *Arabidopsis* resulted in these same pleiotropic phenotypes. This protein is specifically expressed in the shoot apical meristem, and associated with endo-membrane organelles. This mutant is associated with a decrease in auxin levels in the plant (Li et al., 2014).

Furthermore MATE transporters have also been implicated in plant development in *Arabidopsis*. In a screen for abnormal shoot mutants (*abs*), two mutants resulted in loss of apical dominance, and demonstrated early flowering, termed *abs3-1 D* and *abs4-1D*, and these mutants were both shown to have increased expression of two MATE transporters, *ABS3* and *ABS4*. Overexpression

of *ABS3* and *ABS4* together with two other homologues *ABS3L1* and *ABS3L2*, recapitulated this phenotype (Wang et al., 2015). Overexpression of *ABS3* and *ABS4* also resulted in a decrease in the length of the hypocotyls by affecting cell elongation. In comparison a quadruple loss of function mutant of *ABS3*, *ABS4*, *ABS3L1* and *ABS3L2* resulted in longer hypocotyls relative to wild type controls (Wang et al., 2015).

Similarly the *Arabidopsis* ZRZ gene is a MATE protein that localises to the mitochondria, whose overexpression results in growth of axillary buds, the production of leaves much faster than wild type plants, and is involved in specifying lateral organ initiation rate (Burko et al., 2011).

As well as the MATEs involvement in hormone transport, that subsequently impacts growth, and plant responses to environmental stress, the MATE protein family has also been implicated in multidrug transport and heavy metal detoxification.

An *Arabidopsis* cDNA library was screened for multidrug resistance in an *E. coli* mutant strain KAM3 (Li et al., 2002). This mutant strain lacks endogenous multidrug efflux carriers, and will not grow on the common antibiotic norfloxacin. *Arabidopsis* cDNA clones that allow KAM3 to grow in the presence of norfloxacin were selected and sequenced, and this screening methodology identified a MATE transporter annotated as AtDTX1. AtDTX1 is able to efflux norfloxacin out of KAM3 cell relative to wild type controls, and norfloxacin accumulates at lower levels in AtDTX1 expressing KAM3 cells. Consistent with the MATE mechanism of action, this efflux was dependent on the plasma membrane proton gradient across the *E. coli* plasma membrane, as addition of CCCP, a proton gradient decoupler, abolishes norfloxacin efflux (Li et al., 2002). Moreover the efflux activity was dependent on the external pH of the medium. Other ions, such as sodium, do not affect transport, suggesting this protein is dependent on the proton gradient and not other cations. AtDTX1 has a wide substrate range, including ethidium bromide as well as the plant alkaloids berberine and palmitine. This strain also displays enhanced tolerance to heavy metal ions such as cadmium. Expression of AtDTX1 as a GFP fusion *in planta*, localised the transporter to the plasma membrane, implicating a role in cellular export as a means for multidrug efflux in plant cells (Li et al., 2002).

The MATE protein ALF5 has also been implicated in the multidrug resistance of plants to environmental toxins (Diener et al., 2001). Loss of function mutants in *alf5* in *Arabidopsis* resulted in aberrant lateral root formation and were hypersensitive to treatment with polyvinylpyrrolidone and pyrrolidinone, whilst expression of *ALF5* in yeast resulted in resistance

to the toxic cation tetramethylammonium. ALF5 localised to the root epidermis, and is therefore likely to act as an efflux pump involved in detoxification (Li et al., 2002).

A key mechanism that has been identified for heavy metal detoxification in soils is through the secretion of organic acids such as malate, oxalate and citrate. These organic acids are able to complex heavy metals to form non-toxic complexes, thereby detoxifying acidic soils where heavy metals become solubilised. As well as this, the release of organic acids into the vascular tissues allows for chelation of heavy metals such as iron which is necessary for their correct distribution in the plant (Green and Rogers, 2004).

In *Arabidopsis*, *ferric reductase defective 3* (*frd3*) mutants display signs of chlorosis, accumulate iron and other heavy metals in their roots, and have a different distribution of iron in plant tissues relative to wild type controls. The FRD3 protein is a MATE transporter that localises to the pericycle of root cells (Green and Rogers, 2004), and exudes citrate into the plant vasculature. Citrate chelates 99.5% of iron in the xylem, and this chelation is integral for correct localisation of iron and its translocation in the plant. *frd3* mutant plants contain less citrate in the xylem, than wild type plants, and the chlorotic and growth phenotype of *frd3* mutants can be compensated by exogenous supply of citrate (Durrett et al., 2007). Transport of citrate *in vitro* by FRD3 was demonstrated utilising *Xenopus laevis* oocytes. Due to the orientation of the transporter in the *Xenopus* plasma membrane and the fact FRD3 is an exporter from the cytosol, ^{14}C-citrate was injected into the oocyte, and *FRD3* expressing oocytes were shown to have a greater rate of citrate efflux than non-injected controls (Durrett et al., 2007). The closest homologue of *AtFRD3* in the *Arabidopsis* genome is *AtFRDL1*, and is involved in citrate efflux from the root, into the rhizosphere, which is required for aluminium tolerance (Liu et al., 2009). An *Arabidopsis* MATE transporter, BCD1 that is localised to the Golgi apparatus has also been implicated in iron homeostasis during stress (Pil et al., 2012).

In barley, a MATE transporter (HvAACT1), with homology to AtFRD3, was positionally cloned and associated with a QTL that conferred aluminium tolerance to barley (Furukawa et al., 2007). This transporter was able to transport citrate *in vitro* in *Xenopus laevis* oocytes, and preferentially expressed in the root epidermal cells, at the plasma membrane. Unlike AtFRD3, which secretes citrate into the xylem, HvAACT1, like AtFRDL1, secretes citrate into the rhizosphere which increases tolerance to Al^{3+} in acidic soils (Furukawa et al., 2007). Overexpression of HvAACT1 increased citrate efflux and aluminium tolerace in both wheat and barley (Zhou et al., 2013). Similarly in *Sorghum bicolor* a major aluminium tolerance locus, Alt_{SB}, was cloned and shown to be a MATE protein, *SbMATE,* that localised to the plasma membrane and can transport citrate

in *Xenopus laevis* oocytes. This MATE is therefore thought to confer aluminium tolerance by a similar mechanism to HvAACT1(Magalhaes et al., 2007). Furthermore, overexpression of *SbMATE*, as well as *FRD3*, conferred aluminium tolerance in barley (Zhou et al., 2014). The homologues of *AtFRD3*, *SbMATE* and *HvAACT1*, have been cloned from maize (Maron et al., 2010), rice (Yokosho et al., 2010; Yokosho et al., 2009), rice bean (Yang et al., 2011) and Eucalyptus (Sawaki et al., 2013) and shown to either confer tolerance to aluminium, or play a role in the translocation of heavy metals in the xylem.

Interestingly, a citrate transporter that is a homologue of AtFRD3, is utilised specifically in symbiotic nitrogen fixation within legume root nodules (Takanashi et al., 2013). *LjMATE*, which is only expressed in nodules, is capable of transporting citrate in *Xenopus laevis* oocytes, and RNAi lines of this transporter resulted in defective legume nodules, with an altered iron localisation. Inductively coupled plasma mass spectrometry (ICP-MS) demonstrated that the iron content of the *ljmate* mutant was significantly lower than wild type root nodules. This suggests that LjMATE is integral for correct localisation of iron during the process of nodulation (Takanashi et al., 2013).

3.1.2 Role of MATEs in transport of secondary metabolites

Secondary metabolism is spatially separated in many plant species, and there is an increasing appreciation of the role transporters play in biosynthesis of these compounds. MATE proteins in particular have been implicated in the biosynthesis of flavonoids and alkaloids.

3.1.2.1 Flavonoid transport

Flavonoids are a diverse family of aromatic secondary metabolites that include the anthocyanins, flavanols and proanthocyanidins (Ferreyra et al., 2012; Winkel-Shirley, 2001). Proanthocyanidins accumulate exclusively in the endothelium of the seed coat, and although they are colourless, upon oxidation they turn brown, which gives the testa its characteristic colour upon maturity. Mutants of *Arabidopsis* that display an altered colouration of the testa, termed transparent testa mutants, have been isolated and are altered in enzymes involved in flavonoid biosynthesis or in the regulation of flavonoid metabolism (Appelhagen et al., 2014). A mutant, *transparent testa 12* (*tt12*), was identified in a gene that encodes for a member of the MATE family of secondary metabolite transporters (Debeaujon et al., 2001). The *tt12* seeds displayed reduced seed dormancy, and the phenolic compounds normally present in the endothelium layer were not detected. A time course of proanthocyanidin accumulation *in planta*

demonstrated a decrease in pigmentation and a diffuse pattern of localisation in the *tt12* mutants with respect to control plants. The gene conferring the *tt12* phenotype was cloned by T-DNA tagging, and the resulting full length gene cloned was able to complement the *tt12* mutation, and shown to encode a protein with sequence identity to known MATE transporters. *In situ* hybridization, as well as fusing the promoter of *TT12* to a GUS reporter, demonstrated that this gene is expressed in the endothelium layer of the seed coat (Debeaujon et al., 2001) and that *TT12* is expressed specifically in proanthocyanidin producing cells (Marinova et al., 2007). TT12 is likely, therefore, to be a vacuolar transporter for proanthocyanidin precursors such as catechin and leucocyanidins (Debeaujon et al., 2001).

Targeted LC-MS of *tt12* mutant and wild type seeds demonstrated that neither the precursor epicatechin, nor proanthocyanidin polymers, were detected in *tt12* seeds. Additionally, glycosylated quercetin and kaempferol derivatives were reduced in *tt12* seeds (Marinova et al., 2007) further suggesting that *TT12* is implicated in flavonoid metabolism. Overexpression of both an N-terminal and C-terminal GFP fusion of TT12, together with a tonoplastic marker protein TPK1 fused to DsRed2, in onion epidermal cells, resulted in co-localisation, implicating TT12 as a tonoplastic protein, and therefore likely to be involved in the movement of flavonoids into the vacuole (Marinova et al., 2007).

Overexpression of the MATE TT12 in *S. cerevisiae* as a GFP-fusion resulted in its localisation to both the vacuolar and other endomembranes. Transport across the yeast membrane vesicles was performed with the flavonoid glycoside cyanidin-3-*O*-glucoside, which was uptaken in a time dependent manner relative to empty vector controls. No uptake of the aglycone cyanidin was observed, suggesting that the transport by TT12 may be dependent upon glycosylation of the core flavonoid backbone (Marinova et al., 2007). It was subsequently shown that this *Arabidopsis* TT12 protein is also able to transport epicatechin 3-*O*-glycoside, more efficiently than cyanidin 3-*O*-glycoside (Zhao and Dixon, 2009). Transport of the flavonoids epicatechin, catechin and catechin 3-*O*-glycoside was not observed, highlighting the substrate specificity of this transporter (Zhao and Dixon, 2009).

The transport of cyanidin 3-*O*-glucoside by TT12 into yeast vesicles was dependent on the proton gradient, as uptake was inhibited by $NH_4^+Cl^-$, which is used to disrupt the pH gradient across the membrane, as well as inhibited by bafilomycin A1, an inhibitor of the vacuolar H^+-ATPase (Marinova et al., 2007).

A functional orthologue of *Arabidopsis* TT12 has also been identified in *Medicago truncatula* (Zhao and Dixon, 2009). In *M. truncatula*, a glycosyltransferase has been shown to specifically glycosylate epicatechin, forming epicatechin 3-*O*-glycoside, which is likely to be a good candidate for vacuolar transport in proanthocyanidin biosynthesis in the seed coat. Expression of a transcription factor from *Arabidopsis TT2*, lead to upregulation of epicatechin 3-*O*-glycoside production in *Medicago* hairy roots when ectopically expressed and upregulation of a transporter in vacuoles that promotes epicatechin 3-*O*-glycoside transport (Zhao and Dixon, 2009).

This *Medicago* MATE transporter, homologous to *Arabidopsis TT12*, was named MATE1, and was shown to be up regulated by the transcription factor TT2. Notably, its localised expression in flowers, young pods and the seed coat paralleled other *Medicago* proanthocyanidin biosynthetic genes, suggesting a role in anthocyanin metabolism. MATE1-GFP fusions demonstrated localisation to the tonoplastic membrane *in planta*, as well as potential pre-vacuolar vesicle structures in tobacco epidermal cells, which is consistent with the localisation properties of *Arabidopsis* TT12 protein (Zhao and Dixon, 2009). Overexpression of the MATE1-GFP in yeast demonstrated its localisation to internal vesicles, and uptake assays using the microsomal fraction of the yeast cells demonstrated the proton gradient dependent transport of both epicatechin 3-*O*-glycoside and cyanidin 3-*O*-glycoside (Zhao and Dixon, 2009). Much like the *tt12* mutant in *Arabidopsis*, mutation of *MATE1* in *Medicago* resulted in a decrease in proanthocyanidin accumulation in the endothelial layer (Zhao and Dixon, 2009). The *MATE1* from *Medicago* was also able to complement the *tt12* mutation in *Arabidopsis*, and resulted in restored seed coat colouration in the endothelial layer (Zhao and Dixon, 2009). Conserved TT12 proteins in other species, such as *Brassica napus*, *Brassica oleracea*, and *Brassica rapa* have also been shown to contribute to quantitative traits in yellow seed coats (Chai et al., 2009).

As well as in the transport of flavonoids involved in proanthocyanidin deposition in the seed coat, MATE transporters have also been shown to be involved in the transport of other flavonoid moieties.

In grapevine (*Vitis vinifera*) two MATE transporters annotated as anthoMATE1 and 3 (AM1 and AM3) were highlighted as potential anthocyanin transporters (Gomez et al., 2009). Ectopic expression of C-terminal GFP fusions of AM1 and AM3 in *Vitis vinifera* hairy roots, and onion epidermal cells, demonstrated that these transporters localise to the tonoplast membrane, and co-localise with a known tonoplast protein from *Arabidopsis*, TPK1 (Gomez et al., 2009). Expression of *AM3* correlated with acylated anthocyanin biosynthesis and was localised to the

berry skin. Expression of *AM1* and *AM3* in *S. cerevisiae* and transport assays across yeast microsomal membranes demonstrated that these transporters were not able to transport cyanidin 3-*O*-glycoside or malvidin 3-*O*-glycoside, however were able to transport the acylated variants, such as malvidin 3-*p*-coumaroyl-glucoside, indicating acylation is integral for substrate recognition (Gomez et al., 2009). Again, like TT12 and MATE1 from *Arabidopsis* and *Medicago* respectively, AM1 and AM3 were shown to utilise proton driven transport. The physiological role of AM1 and AM3 through knock down of the gene is yet to be demonstrated.

Similarly a second MATE from *Medicago* named *MATE2*, which shows closer homology to *AM1* and *AM3* than to *TT12*, was shown to be involved in the transport of malonylated and acylated anthocyanins (Zhao et al., 2011). When expressed in *S. cerevisiae,* MATE2 demonstrated substrate preference for a wider variety of flavonoid glucosides than MATE1, and in particular showed substrate preference for malonylated versus non-manolyated flavonoid glucosides, and transported these substrates in a proton gradient dependent fashion (Zhao et al., 2011). *In situ* hybridization and qPCR demonstrated that *MATE2* expression localised to floral and leaf tissues, and in particular, in the mesophyll of young leaves. Again, this transporter localised to the tonoplast when expressed as a C-terminal GFP fusion in *Arabidopsis*. Tnt1 retrotransposon insertion mutants in *Medicago* for *MATE2*, which were used to abolish *MATE2* expression resulted in a decrease in anthocyanin and flavonoid glycoside content in both leaves and flowers relative to wild type controls (Zhao et al., 2011). Proanthocyanidin biosynthesis was also affected in the seed coat of *mate2* mutants, with a higher epicatechin 3-*O*-glycoside content relative to wild type controls. This implicates that blocking the transport of anthocyanins into the vacuole may lead to accumulation of epicatechin 3-*O*-glycoside due to feedback in the phenylpropanoid pathway (Zhao et al., 2011).

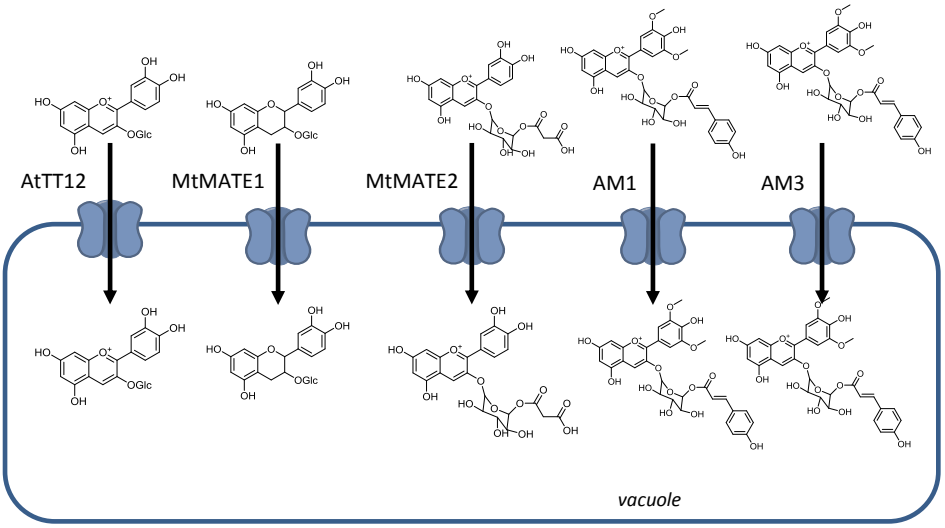


Figure 3-1 Substrate preferences for known MATE proteins involved in flavonoid transport
AtTT12 from *Arabidopsis* and MtMATE1 from *Medicago* preferentially transport the glycosylated intermediates epicatechin 3-*O*-glycoside and cyanidin 3-*O*-glycoside. In comparison the MtMATE2 from *Medicago* and AM1 and AM3 from Grapevine primarily transport malonylated and acylated derivatives such as malonylated cyanidin 3-*O*-glycoside and malvidin 3-*O*-p-coumaroyl-glucoside.

There have been numerous reports of MATEs being potentially involved in flavonoid transport, in other species such as in blueberry and tomato fruit, due to co-regulation of MATEs with known flavonoid pathway genes, a correlation between MATE expression and anthocyanin accumulation, or upregulation of MATEs by transcription factors known to induce flavonoid biosynthesis. However *in vitro* and *in vivo* evidence to rigorously establish the function of these MATE proteins is still required (Chen et al., 2015; Di Matteo et al., 2013; Mathews et al., 2003).

3.1.2.2 Alkaloid and nicotine transport

As well as the plasma membrane localised transporter NUP1 that is involved in nicotine import at the plasma membrane (Jelesko et al 2012), there are four tonoplastically localised MATE proteins, NtMATE1, NtMATE2, NtJAT1 and NtJAT2 that have been implicated in nicotine transport (Morita et al., 2009; Shitan et al., 2014; Shoji et al., 2009). The MATE protein NtJAT1 (*Nicotiana tabacum* jasmonate inducible alkaloid transporter) is co-regulated with a number of genes in the nicotine biosynthetic pathway, including arginine decarboxylase, ornithine decarboxylase, aspartate oxidase, and quinolinate phosphoribosyltransferase, and is induced by MeJA (Morita et al., 2009). This transporter is also expressed in a number of different plant tissues including leaves, stems and roots and has been shown to localize to the tonoplastic membrane in leaf tissue (Morita et al., 2009). In contrast the MATE protein NtJAT2, unlike NtJAT1, is expressed at low levels in leaves, stems and roots, but shows a specific and strong

response to MeJA elicitation, after which it is preferentially expressed in leaves. Furthermore, NtJAT2 also localised to the tonoplastic membrane, when overexpressed in tobacco BY2 cells (Shitan et al., 2014).

Although NtJAT1 and NtJAT2 localise to the tonoplastic membrane in plant cells, overexpression of NtJAT1 and NtJAT2 as C-terminal GFP fusions in yeast resulted in its localisation to the plasma membrane for NtJAT1-GFP, and a dual localisation of the plasma membrane and vacuolar lumen for NtJAT2, in the yeast strain AD12345678 (Morita et al., 2009; Shitan et al., 2014). When incubated with nicotine, NtJAT1, NtJAT1-GFP, NtJAT2 and NtJAT2-GFP yeast cultures accumulated less nicotine than control cells, indicating that the yeast cells expressing NtJAT1 and NtJAT2 had a higher rate of efflux of nicotine from the yeast, and that NtJAT1 and NtJAT2 are likely acting as nicotine efflux transporters from the plasma membrane (Morita et al., 2009; Shitan et al., 2014). The NtJAT1 and NtJAT2 had a similar level of accumulation of nicotine in these cells (Shitan et al., 2014). *In planta*, due to the localisation to the tonoplast membrane, these transporters would therefore be importers of nicotine into the vacuole. The NtJAT2 expressing yeast strain was tested with the other alkaloids, anabasine, anatabine and berberine, demonstrating transport for these compounds, however this yeast strain was not able to transport cyanidin 3-*O*-glucoside or rutin indicating that there may be substrate specificity of transport, perhaps conferred by the basic, positively charged nitrogen present in all of these compounds (Shitan et al., 2014).

NtJAT1 was expressed in Sf9 insect cell culture and reconstituted into proteoliposomes containing a bacterial F-ATPase to generate a proton gradient across the membrane. ^{14}C-tetraethylammonium (TEA) was utilised as a model substrate and could be transported by NtJAT1. Treatment, of the proteoliposomes with azide, ammonium sulphate, nigericin and valinomycin, indicated that NtJAT1 resulted in electro-neutral H^+/TEA exchange. Additionally nicotine inhibited the uptake of TEA in these proteoliposomes indicating that nicotine may competitively inhibiting TEA transport, whilst in comparison flavonoids, such as kaempferol and quercetin, did not affect TEA transport, suggesting that NtJAT1, like NtJAT2, is specific for nitrogen containing plant alkaloids (Morita et al., 2009). Despite the fact that NtJAT1 and NtJAT2 both transport nicotine and are both co-ordinately controlled by MeJA elicitation, these two proteins only have 32% amino acid identity (Shitan et al., 2014).

NtMATE1 and *NtMATE2* are also co-regulated with the nicotine biosynthetic pathway, and were identified as genes that were downregulated in *nic1nic2* nicotine biosynthesis regulatory mutants, relative to wild type tobacco roots (Shoji et al., 2009). The two transporters MATE1

and MATE2 share 96.4% sequence identity however are in a different clade than either NtJAT1 or NtJAT2 (Shitan et al., 2014).

The *NtMATE1/2* genes are preferentially expressed in the root, with very low levels present in leaves, stems and flowers. GUS-staining using the *NtMATE1* 5' flanking region showed expression only at the root tip in the outer cortex cells. This localised expression was upregulated upon MeJA induction, though did not affect the spatial localisation of expression. This expression pattern is very similar to that of the known nicotine biosynthetic genes (Shoji et al., 2009).

Overexpression of an NtMATE1-GFP fusion in tobacco BY-2 cells, as well as immunogold electron microscopy of tobacco BY2 cells with anti-NtMATE1 antibodies and immunolabelling of organelle fractions from tobacco roots with anti-NtMATE1 antibodies, show this protein is localised to the tonoplast membrane (Shoji et al., 2009).

RNAi lines of *NtMATE1/2* did not lead to significant changes in nicotine and nornicotine alkaloid profiles in leaves and roots compared to a vector control line, however there was a subtle root growth phenotype of the RNAi lines upon exogenous treatment of nicotine, suggesting NtMATE1/2 contribute to nicotine transport into the vacuole. In tobacco BY2 cells, cell lines constitutively overexpressing NtMATE1 led to acidification of the cytoplasm upon elicitation with MeJA, as well as upon exogenous supply of nicotine, indicating a potential role in transport (Shoji et al., 2009). As with NtJAT1 and NtJAT2, when NtMATE1 was expressed in yeast, transport assays showed slower nicotine uptake relative to wild type controls, suggesting it is a nicotine exporter when expressed at the yeast plasma membrane (Shoji et al., 2009).

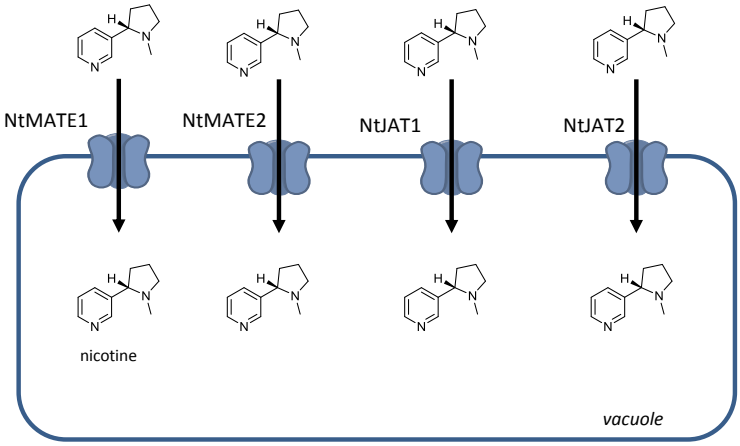


Figure 3-2 Role of MATE Transporters in nicotine transport
NtJAT1, NtJAT2, NtMATE1 and NtMATE2 have been shown to localise to the vacuole and transport nicotine.

The localisation of NtMATE1/2 to the root tip, together with their ability to transport nicotine *in vitro*, suggests that these transporters are involved in vacuolar sequestration of nicotine in the root, potentially preventing the retardation of root growth as a result of newly synthesized nicotine in the root tip.

To date, no metabolic phenotypes in *N. tabacum* have been reported that demonstrate the physiological relevance of NtJAT1 or NtJAT2, and the phenotype for the NtMATE1/2 is relatively subtle. Silencing all four tonoplastic MATE transporters could more definitively demonstrate the effect these transporters have on nicotine biosynthesis, and would rule out potential functional redundancies. If a subtle metabolic phenotype still persists in this quadruple mutant, it may implicate other transporters or transporter families in being functionally relevant to tonoplastic nicotine import which have yet to be identified. The differential subcellular localisation of NtJAT1, NtJAT2, and NtMATE1/2 suggests that different tissues may utilise different proteins for the vacuolar sequestration of nicotine (Morita et al., 2009; Shitan et al., 2014; Shoji et al., 2009).

3.1.2.3 Summary to Introduction

The ability of MATE proteins to transport a wide array of compounds and the diverse role for these proteins in the transport of secondary natural products such as flavonoids and alkaloids, makes this protein family a good candidate for prospective transport of secondary natural products in *C. roseus*. It is noteworthy that while there is a large gene family in *Arabidopsis* and other plant species, only a few of these transporters have been characterised, suggesting that there may be more functional diversity in this transporter class that has not yet been identified.

3.2 Aims of this Chapter

1.) To assess whether there are any MATE proteins that are involved in MIA metabolism in *C. roseus* using VIGS

2.) Functionally characterise the MATE transporters that display a metabolic phenotype after transient silencing *in planta*.

3.3 Results and Discussion

3.3.1 MATE Proteins in Catharanthus roseus

In the transcriptome of *C. roseus* there are 57 genes that are annotated as MATE proteins or have a synonymous annotation such as MATE efflux family, transparent testa 12 homologues, or Multidrug resistance pumps. The 14 highest expressed in *C. roseus* leaf tissue, together with the known characterised MATE transporters (AM1, AM3, NtMATE1, NtMATE2, NtJAT1, NtJAT2, MtMATE1, MtMATE2, TT12, FRD3, EDS5, ALF5, DTX1) are shown in the phylogenetic tree (Figure 3-3).

The sequence and phylogeny analysis does not necessarily give us a good indication of substrate preference for the MATE transporters. As has been demonstrated for the nicotine MATE transporters, proteins with diverse sequences are able to transport the same substrate. However this analysis does demonstrate that *C. roseus* contains homologues that are expressed in leaf tissue, the major site of MIA biosynthesis, for all major clades of MATE proteins that have been characterised to date.

There are two MATE proteins that are highly expressed in leaf tissue, and also cluster in expression profile with known biosynthetic genes in the MIA pathway (Figure 3-4), as discussed in Chapter II. These are annotated as CrMATE10740 and CrMATE1952, and were seen to be good candidates for further study.

Moreover analysis of the *C. roseus* genome demonstrates that the CrMATE1952 protein is on the same genomic contig as two biosynthetic genes in the MIA pathway, tryptophan decarboxylase, the enzyme that catalyses the formation of tryptamine, as well as strictosidine synthase (Kellner et al., 2015b).

There is precedent for secondary metabolic pathways to be clustered on the genome in a number of plant species, (Nützmann and Osbourn, 2014), however the presence of a transporter in these metabolic clusters is rare. It is debatable as to whether the putative cluster of tryptophan decarboxylase, strictosidine synthase and the CrMATE1952 is a true functional secondary metabolic cluster, as it only consists of three genes in a small part of the known alkaloid metabolic pathway, however it is known at the transcriptomic level that these genes are co-regulated. It is possible that the spatial proximity on the genome is involved in controlling transcriptional regulation.

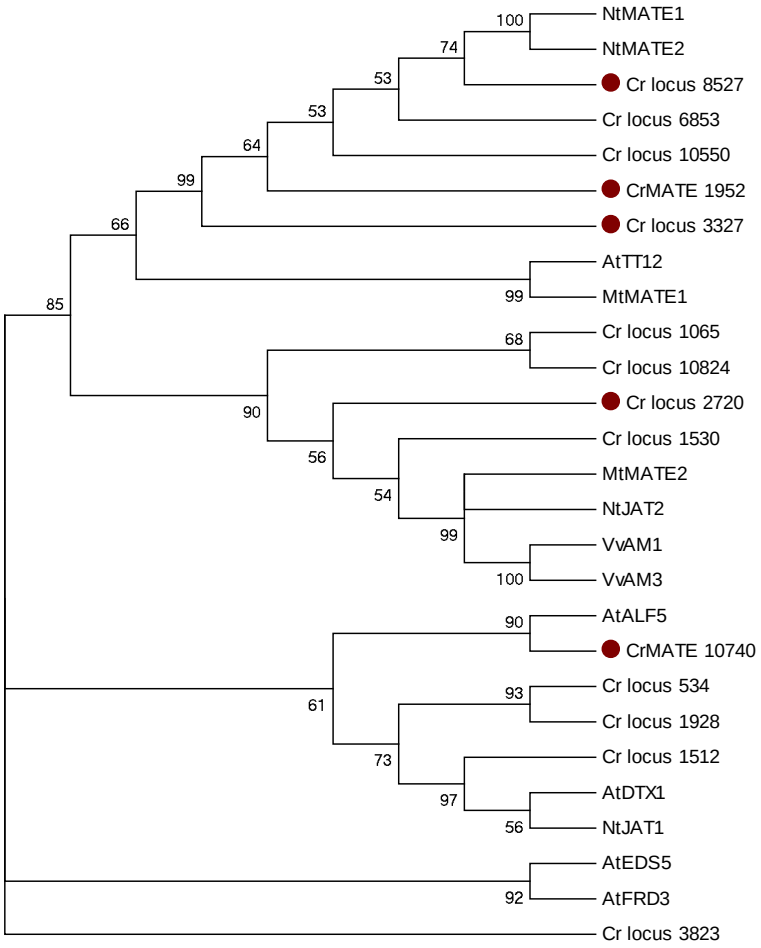


Figure 3-3 Phylogenetic tree of MATE proteins.
The 14 most abundantly expressed MATE proteins in *C. roseus* are highlighted as Cr_locus_contig no. The two *Catharanthus* MATE proteins that co-regulate with the early steps in the MIA biosynthetic pathway are highlighted as CrMATE10740 and CrMATE1952. The other sequences used in the tree are the functionally characterised MATE proteins from *Arabidopsis*; AtFRD3, AtDTX1, AtALF5 and AtTT12, *Medicago*: MtMATE1 and MtMATE2, Grapevine: VvAM1 and VvAM3, and *N. tabaccum*: NtMATE1, NtMATE2, NtJAT1 and NtJAT2. Highlighted sequences by red dots are *Catharanthus* loci which have been used for VIGS in this study. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model, and the bootstrap consensus tree was inferred from 500 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed.

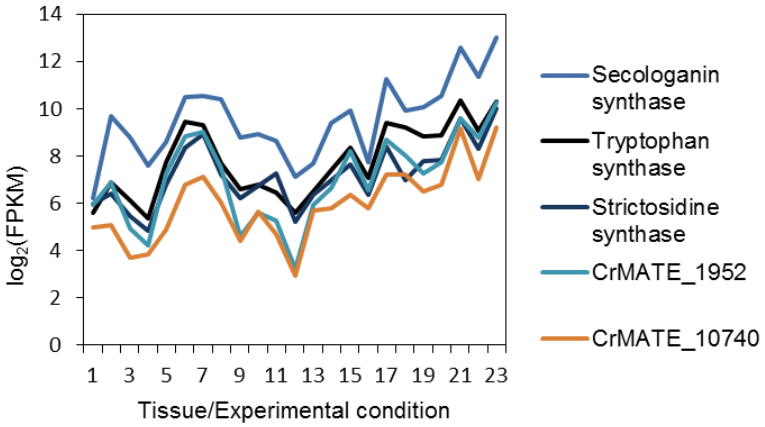


Figure 3-4 Transcriptomic profile of CrMATE1952 and CrMATE10740.
Analysis of the expression profile of CrMATE1952 and CrMATE10740 across 23 different tissues and experimental conditions, which have previously been presented in Table 1, demonstrates that they have a similar expression pattern as known genes in the MIA biosynthetic pathway.

3.3.2 MATE protein localisation and overexpression in N. benthamiana

Transient overexpression under the control of the CaMV35S promoter, of the CrMATE1952-YFP and CrMATE10740-YFP fusions in *C. roseus* cell suspension cultures resulted in their localisation to the tonoplastic membrane as well as other smaller vacuole like compartments (Figure 3-5). Both of these proteins co-localised with a known tonoplastic marker protein, the potassium channel TPK1 from *Arabidopsis* (Bertl et al., 2008; Gomez et al., 2009; Marinova et al., 2007).

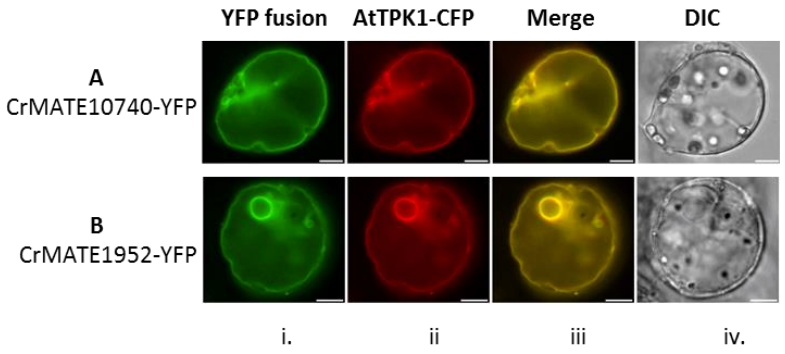


Figure 3-5 Overexpression of CrMATE10740-YFP and CrMATE1952-YFP in *C. roseus* cell suspension cultures
A.) Co-localisation of CrMATE10740 with the *Arabidopsi*s potassium channel TPK1 i.) CrMATE10740-YFP ii.) *Arabidopsis* TPK-CFP iii.) Merge iv.) Differential interference contrast B.) Co-localisation of CrMATE1952 with the *Arabidopsi*s potassium channel TPK1 i.) CrMATE1952-YFP ii.) *Arabidopsis* TPK-CFP iii.) Merge iv.) Differential interference contrast. Scale bar = 10 µm. *Cloning of the C. roseus CrMATE genes and Arabidopsis TPK1 gene into pEAQ-HT DEST1 was performed by R. Payne, subcloning of the genes, transformation by particle bombardment into C. roseus cell suspension cultures and visualisation was performed by Vincent Courdavault, University of Tours.*

For CrMATE1952-GFP, this localisation is also observed with overexpression in *N. benthamiana* epidermal cells (Figure 3-6). A CrMATE1952-GFP C-terminal GFP fusion of the protein was cloned into the pEAQ-HT DEST1 vector for transient overexpression in *N. benthamiana* using *Agrobacterium tumefaciens* (Sainsbury et al., 2009). This resulted in localisation of CrMATE1952-GFP to the tonoplast membrane, as well as to other small vacuole compartments when visualised by confocal microscopy (Figure 3-6).

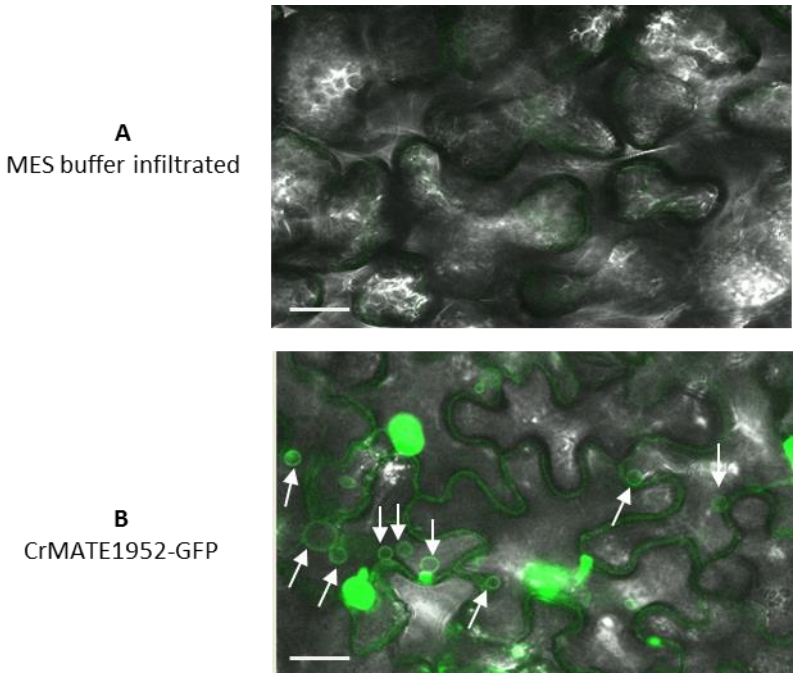


Figure 3-6 Expression of CrMATE1952-GFP in *Nicotiana benthamiana* epidermal cells.
Overexpression of CrMATE1952-GFP using the pEAQ-HTDEST1 expression vector coupled with *Agrobacterium tumefaciens* infiltration of *N. benthamiana* and visualisation by confocal microscopy. A.) MES buffer infiltrated *N. benthamiana* leaf tissue B.) Overexpression pf CrMATE1952-GFP using the pEAQ-HTDEST1 expression vector. White arrows indicate the localisation of CrMATE1952-GFP to smaller vacuoles in *N. benthamania* epidermal cells. Scale bar = 20 µm.

These smaller vacuoles have also been observed when transiently overexpressing other tonoplast localised transport proteins in *N. benthamiana*, such as the ABCC1, an ATP binding cassette protein that is involved in anthocyanidin 3-*O*-glycoside transport (Francisco et al., 2013). The localisation in *N. benthamiana* epidermal cells is also reminiscent of the tonoplastic MATE1 and MATE2 from *Medicago* (Zhao and Dixon, 2009; Zhao et al., 2011).

The localisation of CrMATE10740-GFP in *Nicotiana benthamiana* epidermal cells was not feasible due to the toxicity of overexpressing this protein and this will be discussed in more detail in Section 3.3.11.

3.3.2.1 Hypothetical roles of MATE proteins at the tonoplast in C. roseus

To date all MATE proteins that have been characterised direct the transport of the substrate away from the cytoplasm. Therefore if they are expressed at the plasma membrane they export the substrate from the cell, whilst expression at the tonoplast membrane results in vacuolar import from the cytosol. Therefore the localisation of these transporters, CrMATE1952 and CrMATE10740, to the tonoplast membrane limits the potential substrates that these transporters might be involved in transporting.

As discussed in Chapter I, the MIA pathway in *C. roseus* is subcellularly compartmented, with the pathway occurring in multiple organelles, and there are multiple points in the pathway in which the vacuole is known to be involved.

The vacuole is thought to be the site of storage of many of the downstream alkaloids in MIA metabolism in *C. roseus* such as vindoline, serpentine, and catharanthine in mesophyll and idioblasts cells (Figure 3-7). These compounds are transported into the vacuole using a proton dependent secondary transport system in *C. roseus* cells (Carqueijeiro et al., 2013). A member of the MATE protein family could potentially function as an importer of the downstream alkaloids into the vacuole. This family has been hypothesised to function in this storage role (Carqueijeiro et al., 2013), however to date no gene has been identified via either *in vitro* activity or through reverse genetics that substantiates this view (Carqueijeiro et al., 2013; Roytrakul and Verpoorte, 2007).

Alternatively, the vacuole is also known to be the site of biosynthesis of strictosidine in the epidermis, which is an earlier step in the pathway, and a key branchpoint step in the biosynthesis of MIAs. Strictosidine synthase, which is localised to the vacuole, catalyses the Pictet Spengler condensation of tryptamine and secologanin to form strictosidine. This therefore requires the transport of the two substrates for strictosidine synthesis, tryptamine and secologanin into the vacuole (Figure 3-7). Although there is debate in the literature with regards to whether tryptamine is able to be transported across the tonoplast membrane through an ion-trap mechanism, and whether a transport step is needed to facilitate transport (Deus-Neumann and Zenk, 1984; Deus-Neumann and Zenk, 1986), secologanin is a glycosylated iridoid, and is therefore likely to be impermeable to the lipid bilayer and requires a transporter to be moved into the vacuole (Courdavault et al., 2014).

As has been discussed in Chapter II, the transcriptional co-regulation of the genes involved in MIA biosynthesis has centred on the early steps in the pathway, namely the formation of tryptamine, secologanin and strictosidine. Enzymes downstream that are involved in different branches of MIA biosynthesis, such as the steps between tabersonine and vindoline, have not been shown to be strongly co-regulated.

The fact that the two MATE transporters co-regulate with the genes involved in secologanin, tryptamine and strictosidine biosynthesis, as well as the fact that CrMATE1952 is clustered on the genome with tryptophan decarboxylase and strictosidine synthase is strongly indicative that these transporters are likely to be involved in the early steps in the pathway.

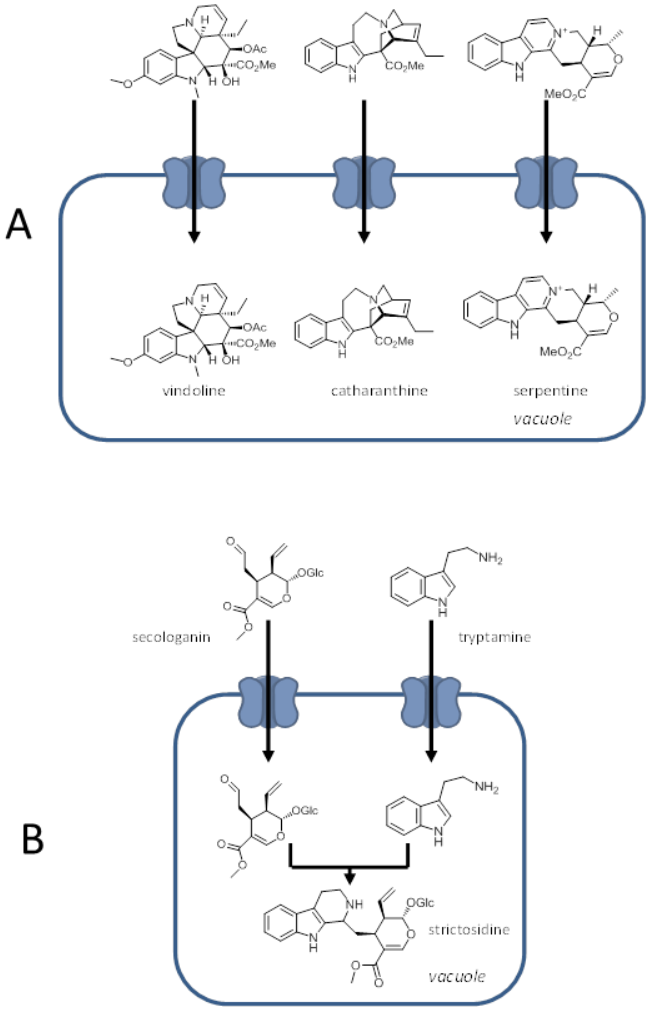


Figure 3-7 Potential role of MATES at the tonoplast membrane.
MATES are secondary transporters that move substrates away from the cytoplasm and therefore at the tonoplast membrane are vacuolar importers. There are two main hypothesised roles for the MATES in MIA transport. A.) These transporters could be transporting the final products of MIA biosynthesis, such as serpentine, vindoline and catharanthine into the vacuole for storage or B.) These transporters are involved in the transport of secologanin or tryptamine into the vacuole in order for the biosynthesis of strictosidine

Furthermore, CrMATE1952 gene expression has been shown to be preferentially enriched in epidermal tissue by qPCR (Vincent Courdavault; personal communication). This epidermal enrichment is consistent with the MATEs involvement in the early steps of the pathway as opposed to transporting the downstream MIAs into the vacuole in mesophyll and laticifer cells.

3.3.3 Virus Induced Gene Silencing

VIGS is a reverse genetic approach that can be used in *C. roseus* to assess whether a gene has a functional role in MIA metabolism by screening for a metabolic phenotype. For individual silencing of the genes CrMATE1952 and CrMATE10740 as well as another MATE gene, locus_2720, in *C. roseus*, 500 bp regions of the genes were cloned into the pTRV2u vector for transformation into *A. tumefaciens* strain GV3101 and transient gene silencing.

3.3.3.1 VIGS CrMATE10740 and locus_2720

VIGS for CrMATE10740 and locus_2720 did not result in a statistically significant difference in any of the major alkaloid peaks vindoline, catharanthine, serpentine/alstonine, strictosidine or the iridoid glycoside secologanin and therefore to date, there is no evidence that these transporters are functionally relevant in MIA biosynthesis in *C. roseus* (Figure 3-8).

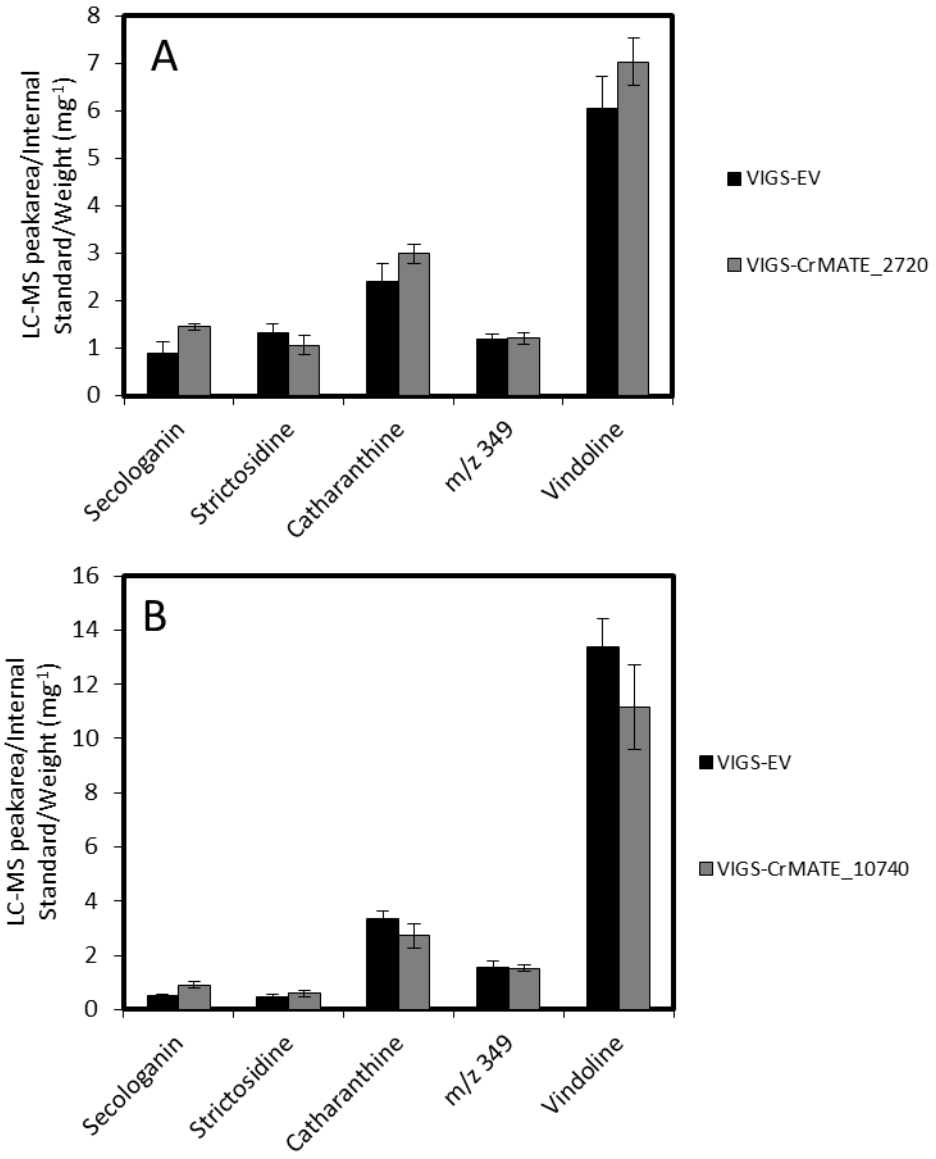


Figure 3-8 Metabolite profile upon VIGS of CrMATE10740 and locus_2720
A.) Alkaloid profile for VIGS of locus_2720 relative to empty vector control tissue in Sunstorm Apricot VIGS-CrMATE1952 (n=12) VIGS-EV (n=12) ** $p<0.01$ B.) Alkaloid profile for VIGS of CrMATE10740 relative to empty vector control tissue in Little Bright Eyes VIGS-CrMATE10740 (n=12) VIGS-EV (n=12) ** $p<0.01$. All data shown are mean $\pm$ SEM.

3.3.3.2 VIGS CrMATE1952

VIGS for the CrMATE1952 resulted in the reproducible accumulation of a new compound with *m/z* 391 (H^+ adduct) and *m/z* 413 (Na^+ adduct) (Figure 3-9, Figure 3-10). This peak is not observed in VIGS-EV controls, and is also not present in any VIGS experiments attempting to transiently silence other genes from *C. roseus*. The accumulation of this peak is reproducible across multiple VIGS experiments for silencing CrMATE1952 (Figure 3-11 and Appendix). Through qPCR we also observed that there is a decrease in the expression of CrMATE1952 in the VIGS-CrMATE1952 silenced lines relative to the VIGS-EV controls, confirming the knockdown of the gene by VIGS (Figure 3-11).

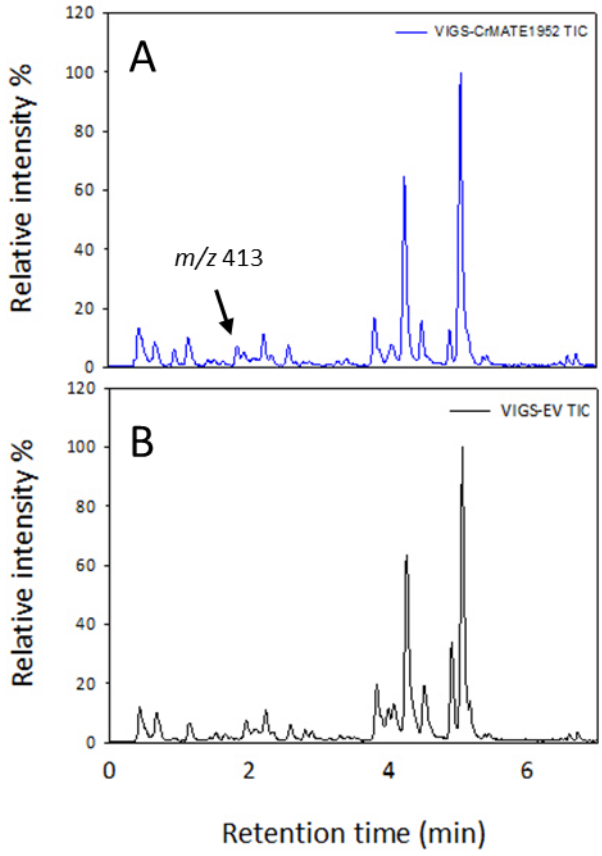


Figure 3-9 Representative Total Ion Chromatogram for VIGS-EV and VIGS-CrMATE1952 leaf tissue. Methanol extracts of leaf tissue from A.) VIGS-CrMATE1952 and B.) VIGS-EV were analysed by LC-MS. Presented are the total ion chromatograms. Highlighted is the new peak *m/z* 413 that reproducibly accumulates in VIGS-CrMATE1952 silenced tissue. It is necessary to note that secologanin elutes as *m/z* 411 (Na^+ adduct) on the Shimadzu-IT-TOF mass spectrometer and *m/z* 389 (H^+ adduct) on the Thermo-Finnigan instrument with a Deca XP ion trap detector. Similarly the phenotype observed in VIGS-CrMATE1952 silenced tissue runs as *m/z* 413 (Na^+ adduct) on the Shimadzu-IT-TOF mass spectrometer and *m/z* 391 (H^+ adduct) on the Thermo-Finnigan instrument with a Deca XP ion trap detector.

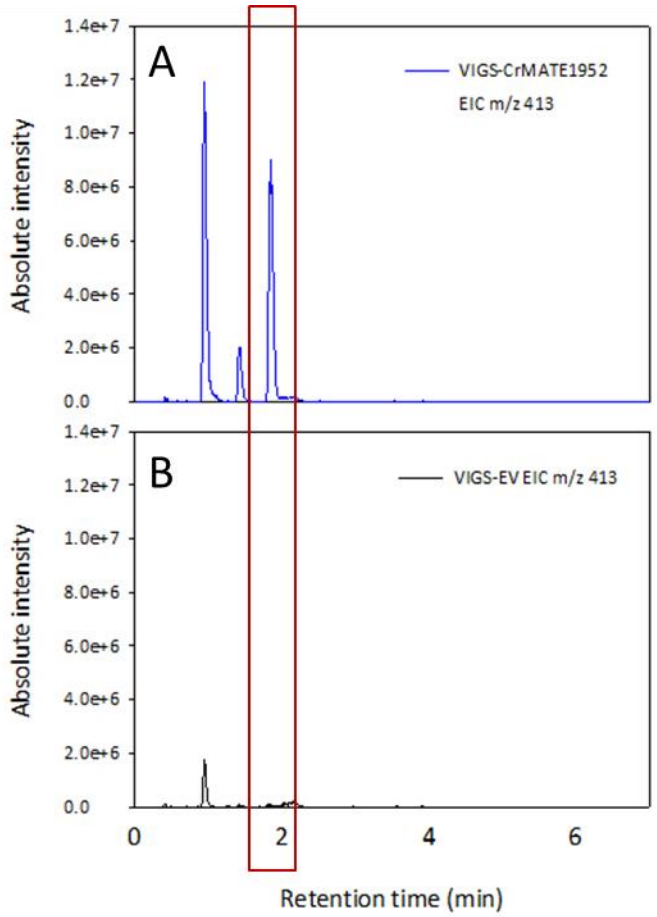


Figure 3-10 Representative Extracted Ion Chromatogram of *m/z* 413 for VIGS-EV and VIGS-CrMATE1952 leaf tissue.

Methanol extracts of leaf tissue from A.) VIGS-CrMATE1952 and B.) VIGS-EV leaf tissue were analysed by LC-MS. Presented are the extracted ion chromatograms for the *m/z* 413 that accumulates in the VIGS-CrMATE1952 silenced tissue. Although there are multiple *m/z* 413 peaks it is only the peak highlighted in red that reproducibly accumulates relative to VIGS-EV control.

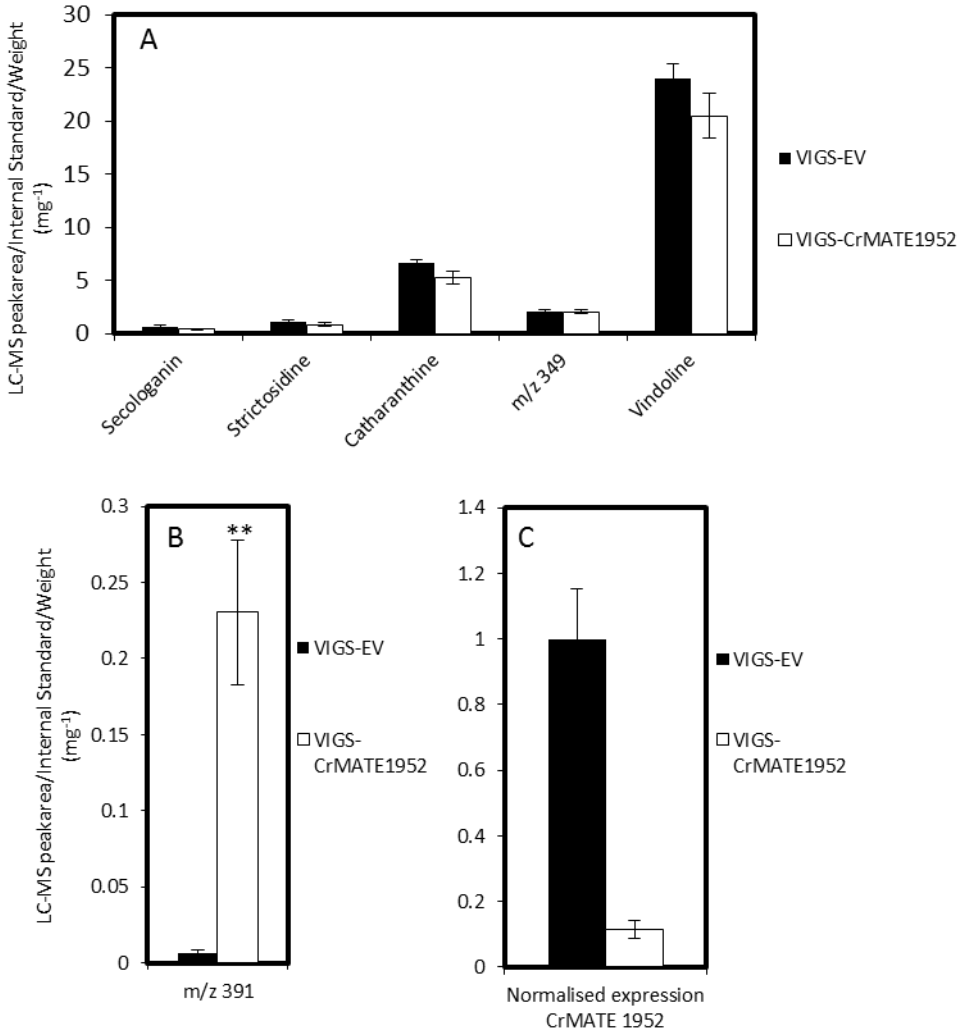


Figure 3-11 Metabolite and qPCR profile upon VIGS of CrMATE1952

A.) Alkaloid profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes VIGS-CrMATE1952 (n=12) VIGS-EV (n=8) ** p<0.01 B.) *m/z* 391 profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes, VIGS- CrMATE1952 (n=12) VIGS-EV (n=8) ** p<0.01 C.) qPCR normalised expression of the CrMATE1952 gene in VIGS-EV and VIGS-CrMATE1952 silenced plants VIGS-CrMATE1952 (n=8) VIGS-EV (n=8). All data shown are mean ± SEM.

3.3.4 Secologanin feeding to wild type C. roseus leaf tissue

The new peak *m/z* 391 (H^+ adduct) and *m/z* 413 (Na^+ adduct) that is accumulating in CrMATE1952 silenced tissue is likely to be an iridoid glycoside derived from secologanin.

Feeding wild type *C. roseus* leaf tissue with secologanin can recapitulate the metabolic phenotype observed upon CrMATE1952 silencing, namely the formation of a peak of *m/z* 391 (H^+ adduct) and *m/z* 413 (Na^+ adduct). This peak is at the same retention time as the peak that accumulates in VIGS-CrMATE1952 silenced tissue. Therefore, it is reasonable to assume that these compounds are derived from secologanin. Since secologanin has an *m/z* 389 (H^+ adduct) and *m/z* 411 (Na^+ adduct) the new compound that accumulates is likely to be the reduced form of secologanin in which the aldehyde is reduced to the alcohol, forming secologanol (Figure 3-12).

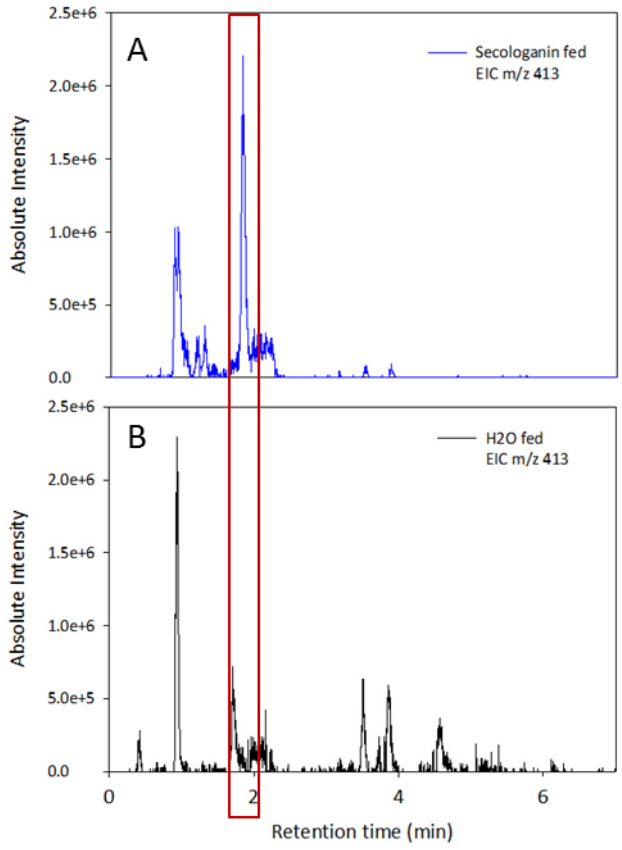


Figure 3-12 Representative Extracted Ion Chromatogram of *m/z* 413 for Wild Type *C. roseus* leaf tissue fed with secologanin or water.
Methanol extracts of leaf tissue from A.) Secologanin fed and B.) H_2O fed wild type *C. roseus* leaf tissue were analysed by LC-MS. Presented are the extracted ion chromatograms for the *m/z* 413 that accumulates upon secologanin feeding, and co-elutes with the peak that accumulates in CrMATE1952 silenced leaf tissue.

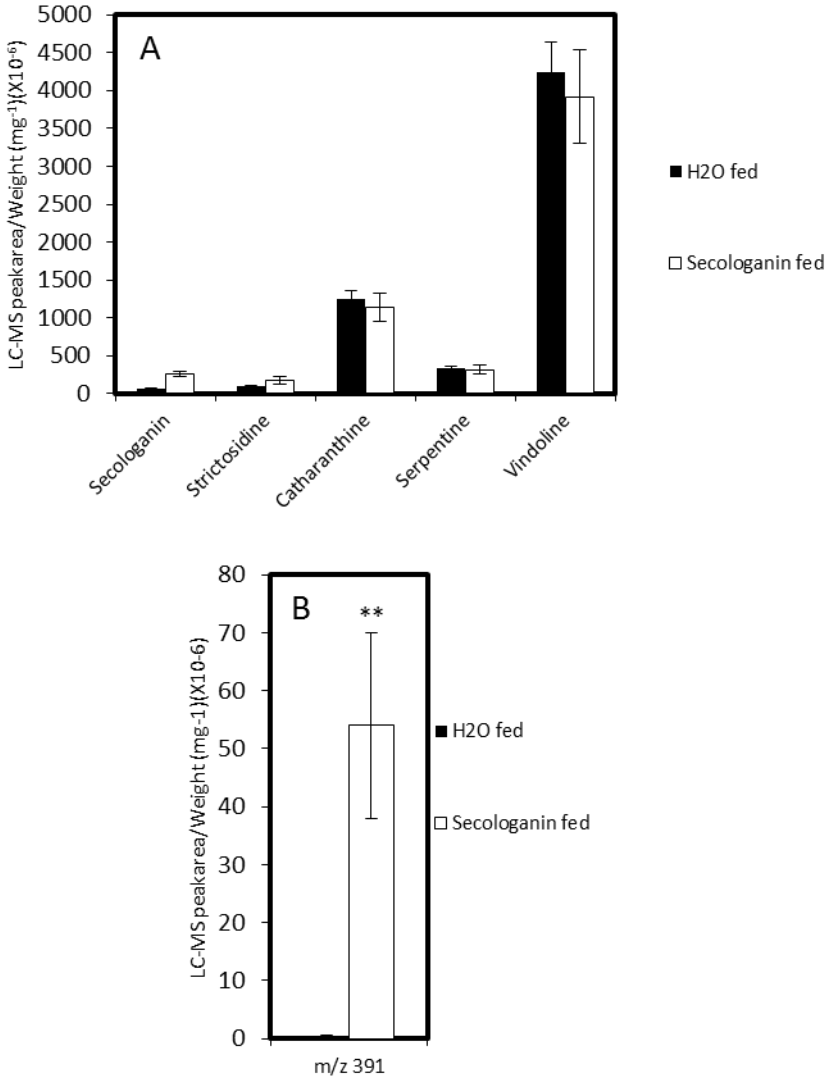


Figure 3-13 Metabolite profile upon feeding wild type *C. roseus* leaf tissue secologanin
A.) Alkaloid profile of *C. roseus* Little Bright Eyes leaf tissue fed either H_2O (n=4) or secologanin (n=4) ** $p<0.01$ B.) *m/z* 391 profile *C. roseus* Little Bright Eyes leaf tissue fed either H_2O (n=4) or secologanin (n=4) ** $p<0.01$. All data shown are mean $\pm$ SEM.

3.3.5 Sodium borohydride reduction of secologanin

To assess whether the peak of *m/z* 391 (H^+ adduct) and *m/z* 413 (Na^+ adduct) is the reduced product of secologanin, this iridoid glycoside can be chemically reduced to the alcohol, and the product analysed by LC-MS. Chemical reduction of secologanin to secologanol is feasible using sodium borohydride ($NaBH_4$) in aqueous methanol. Treatment of a standard of 1 mM secologanin with $NaBH_4$ resulted in the chemical reduction to the alcohol (Figure 3-14, Figure 3-15). This compound has the same retention time and *m/z* 413 as the peak that is accumulating in CrMATE1952 silenced plant tissue, and that is accumulating in wild type *C. roseus* leaf tissue upon exogenous secologanin feeding (Figure 3-16).

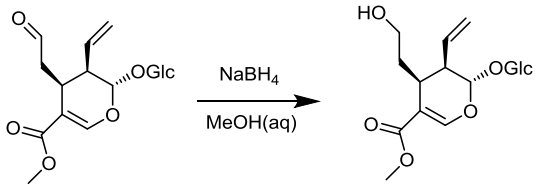


Figure 3-14 Conversion of secologanin to secologanol
Secologanin can be converted to its alcohol derivative secologanol by chemical reduction using sodium borohydride ($NaBH_4$). This reaction may also proceed enzymatically through an as of yet undiscovered reductase enzyme as is observed upon feeding secologanin to wild type leaf tissue

Furthermore treatment of the methanolic extract of an EV control tissue with sodium borohydride, supports this hypothesis. Sodium borohydride treated methanolic extract of VIGS-EV resulted in the accumulation of a peak of *m/z* 413 that has the same retention time as the peak accumulating in VIGS-CrMATE1952 tissue (Figure 3-15, Figure 3-16).

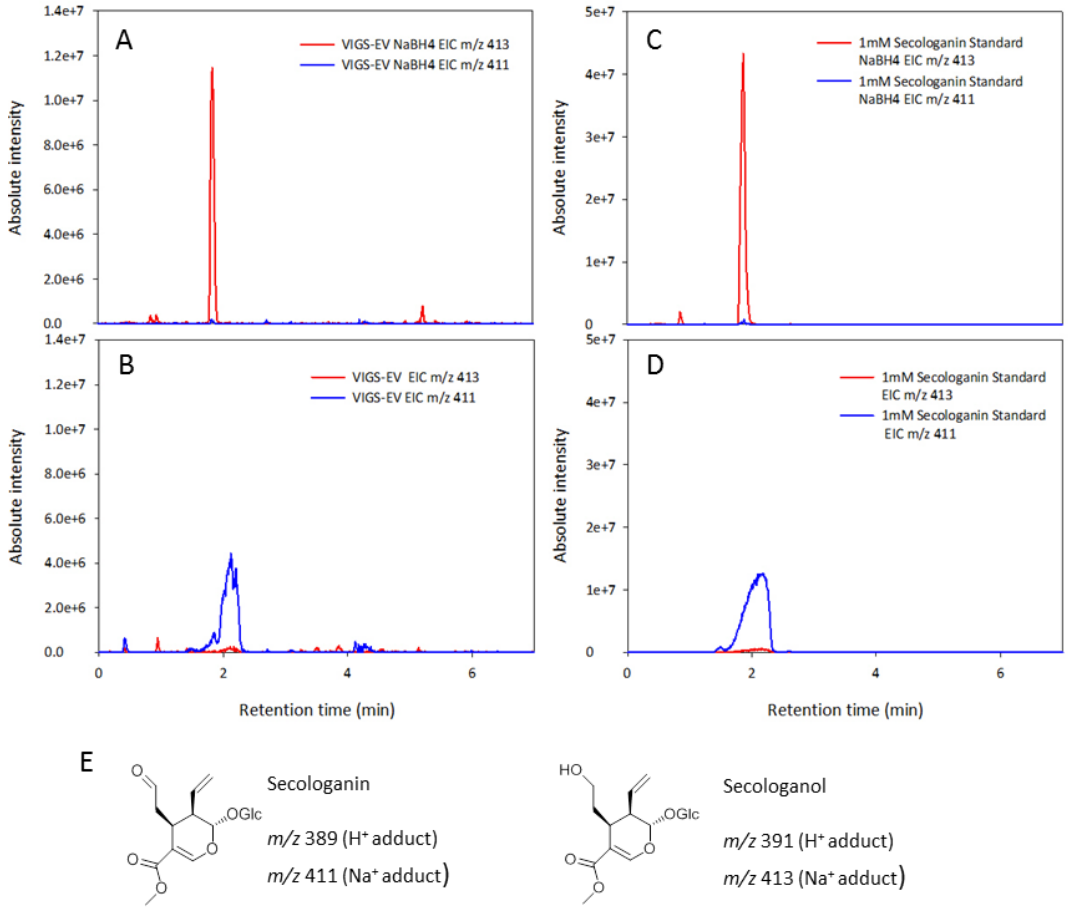


Figure 3-15 Representative Extracted Ion Chromatogram of *m/z* 413 and *m/z* 411 for VIGS-EV reduced with $NaBH_4$, and a secologanin standard reduced with $NaBH_4$.
Aqueous Methanol extracts of VIGS-EV leaf tissue treated A.) with $NaBH_4$ and B.) without $NaBH_4$. Aqueous Methanol extracts of a 1 mM secologanin standard treated C.) with $NaBH_4$ and B.) without $NaBH_4$. E.) Chemical structures of secologanin and secologanol.

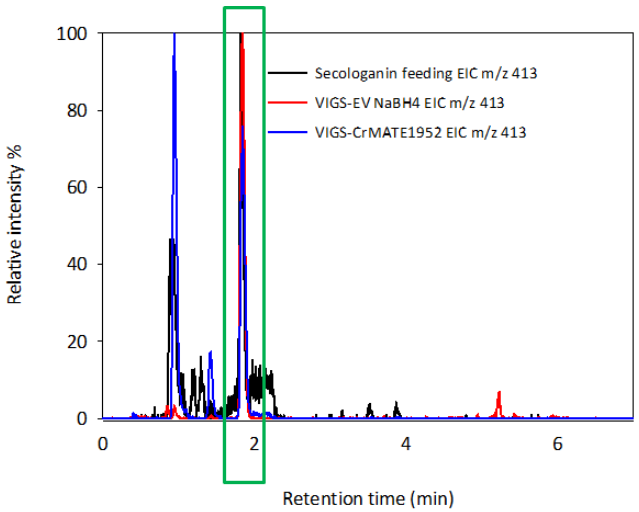


Figure 3-16 Extracted Ion Chromatograms *m/z* 413
Overlay of the extracted ion chromatograms of *m/z* 413 from the secologanin feeding experiments to wild-type leaf tissue (black), reduction of a VIGS-EV leaf tissue with $NaBH_4$ (red), and the VIGS-CrMATE1952 leaf tissue (blue). The green box indicates these compounds elute at the same retention time.

3.3.6 Proposed role of CrMATE1952 in MIA metabolism

Due to the localisation of the CrMATE1952 protein to the tonoplast membrane and the fact that a derivative of secologanin is accumulating in VIGS-CrMATE1952 silenced tissue, our current hypothesis is that CrMATE1952 could be acting as the secologanin importer into the vacuole.

The reason a derivative of secologanin is observed upon silencing is likely to be due to the effect silencing the CrMATE1952 transporter has on the spatial localisation of secologanin (Figure 3-17). Silencing of the transporter does not necessarily affect the secologanin concentration, but instead it would directly affect the relative ratio of secologanin in the cytoplasm and the vacuole, with the CrMATE1952 silenced leaves having a higher concentration of secologanin in the cytoplasm due to inhibition of vacuolar import.

Plants, and other living organisms, do not accumulate aldehydes as these are reactive species and potentially toxic (Kunjapur et al., 2014; LoPachin and Gavin, 2014). As mentioned previously, secologanin has an aldehyde moiety, and it is possible that the response of the plant to the accumulation of secologanin in the cytoplasm is its detoxification through reduction to the alcohol. Similarly it is also conceivable that this is not a direct plant response, but instead, if there is a promiscuous reductase in the cytoplasm, mislocalisation of secologanin to the cytosol, increases the relative chance that secologanin can act as the substrate for reduction. The phenotype observed is therefore likely to be due to the relative mislocalisation of secologanin, due to inhibition of vacuolar import, and its subsequent derivitisation to an alcohol.

This result is supported by the secologanin feeding experiments. Upon feeding secologanin to wild-type leaf tissue an accumulation of *m/z* 391 and *m/z* 413 is observed, and this peak that accumulates co-elutes with the compound observed upon CrMATE1952 silencing, as well as the *m/z* 391 and *m/z* 413 peak that is seen upon treatment of secologanin with $NaBH_4$. The reason for the observation of this peak can be explained in the same way as CrMATE1952 silencing. Feeding a large quantity of secologanin, that exceeds the concentration present in leaf tissue, potentially saturates the potential for secologanin import into the vacuole, and therefore again affects the relative distribution of secologanin in the vacuole and cytoplasm. Accumulation of secologanin in the cytoplasm upon secologanin feeding thereby recapitulates the metabolic phenotype observed upon CrMATE1952 silencing, and leads to secologanin reduction to secologanol (Figure 3-17).

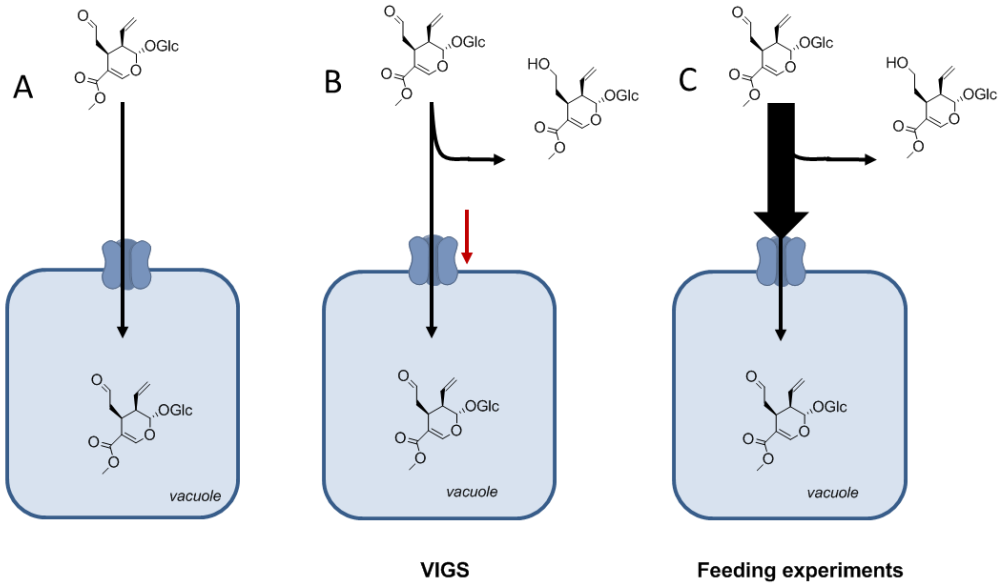


Figure 3-17 Proposed scheme for secologanol formation in *C. roseus.*
A.) Secologanin is an aldehyde that is usually stored in the vacuole B.) Silencing of the transporter (red arrow) results in differential localisation of secologanin and its subsequent reduction C.) Feeding secologanin also results in an increased cytoplasmic concentration of secologanin and its conversion to secologanol due to saturating the pathway.

This result highlights the importance of subcellular compartmentation in metabolic pathways when the substrate is potentially toxic or has the potential to be derivatised by enzymes from other metabolic pathways. Secologanin is an aldehyde that accumulates in leaf tissue, and its localisation to the vacuole may be necessary to prevent its derivatisation. The data we present here provides evidence for the involvement of a tonoplastic transporter that directly affects secologanin metabolism.

There is a disparate effect on the downstream alkaloids in the replicates of the VIGS experiments upon CrMATE1952 silencing. Although the accumulation of *m/z* 391 and *m/z* 413 is always observed upon CrMATE1952 silencing, there is variability between VIGS experiments with regards to the affect silencing this transporter has on the downstream alkaloids strictosidine, catharanthine, serpentine and vindoline (Figure 3-11 and Appendix). The fact that a decrease in these alkaloids is not always observed is indicative that there is enough flux through the pathway to maintain MIA biosynthesis, and therefore enough secologanin is able to enter the vacuole for strictosidine synthesis to maintain MIA biosynthesis in the CrMATE1952 silenced leaf tissue.

There are two main hypotheses that can account for the maintenance of MIA biosynthesis despite inhibiting the secologanin importer, CrMATE1952. Firstly it is possible that there is functional redundancy in secologanin transport activity and it is possible that multiple

transporters could also be transporting secologanin into the vacuole. The phenotype we observe upon CrMATE1952 silencing is therefore due to the fact that CrMATE1952 contributes enough towards secologanin import into the vacuole to result in the relative mislocalisation of secologanin, however there is enough transport activity of secologanin into the vacuole to maintain MIA biosynthesis due to the activity of other transporters.

Alternatively it is possible that CrMATE1952 is the only transporter involved, and despite 80-90% silencing at the transcript level by qPCR, there is still enough basal activity of secologanin import by CrMATE1952 to maintain flux to downstream alkaloids.

The pathway to MIA biosynthesis is through a convergent metabolic pathway between primary amino acid metabolism derived from tryptophan, and the monoterpene iridoid glycoside pathway. If there is a relative discrepancy in the relative flux between these two branches, and the flux through the terpene metabolism to secologanin is greater than the flux through to tryptamine, it is possible that the basal activity of CrMATE1952 that is present despite silencing has enough capacity to maintain flux to downstream alkaloids.

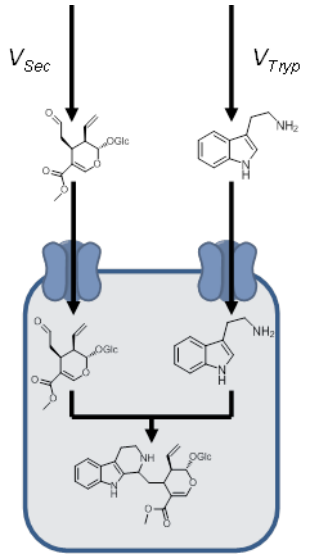


Figure 3-18 Monoterpene indole alkaloid biosynthesis is a convergent metabolic pathway.
MIA biosynthesis is a convergent metabolic pathway centred on the formation of the key branchpoint strictosidine in the vacuole. If rate of formation of secologanin (V_{sec}) greatly exceeds the formation of tryptamine (V_{Tryp}) it is possible that despite silencing transport of secologanin into the vacuole by VIGS there is enough residual basal activity of the transporter CrMATE1952 to maintain flux to the downstream alkaloids such as catharanthine or vindoline.

3.3.7 Isotopic labelling of the strictosidine pool

Given that strictosidine biosynthesis requires both secologanin and tryptamine substrates, we hypothesised that one substrate could be rate limiting in the leaf tissue. Isotopic labelling studies were used to explore whether tryptamine supply to the MIA pathway is potentially limiting.

Wild type leaf tissue was cut at the petiole and fed with 5 mM d_4-tryptamine over a period of four days. Feeding studies with isotopically labelled tryptamine demonstrated that feeding the substrate led to higher strictosidine accumulation relative to wild type controls. This resulted in a ~ 10 fold increase in the amount of strictosidine in the leaf tissue (Figure 3-19).

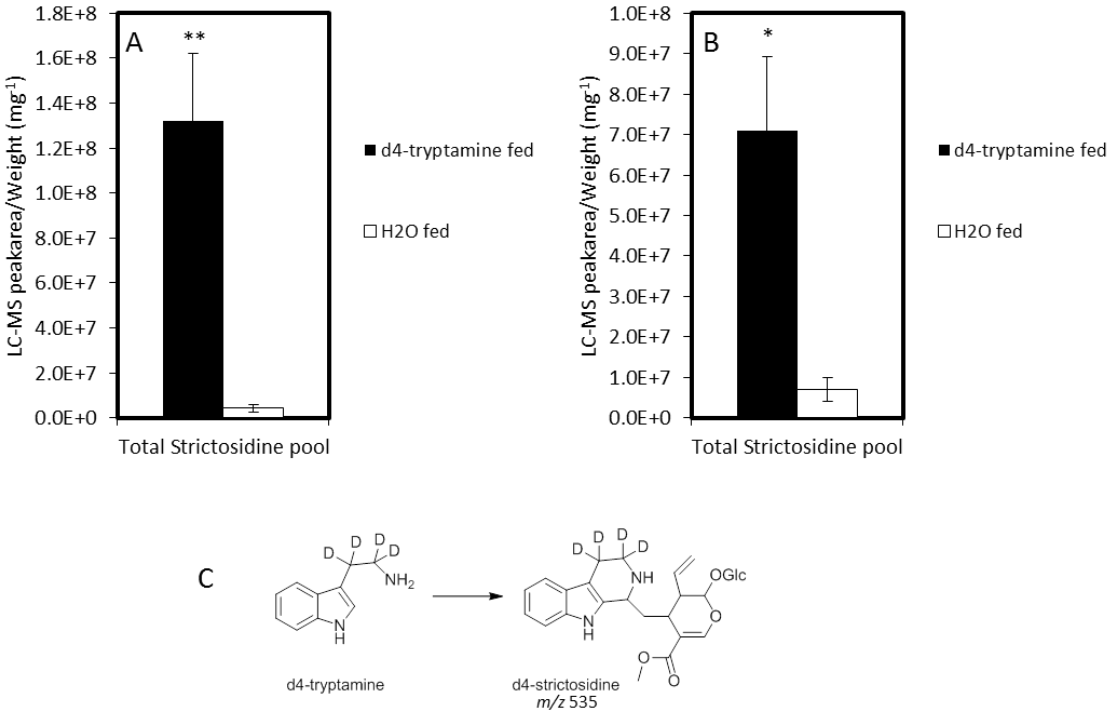


Figure 3-19 Feeding d_4-tryptamine resulted in an increase in the total strictosidine pool.
Wild type *C. roseus* leaves were cut at the petiole and fed 100 µl 5 mM d_4-tryptamine for four days. The total strictosidine pool is the combined peak area of the endogenous strictosidine *m/z* 531 and the labelled strictosidine *m/z* 535. A and B represent separate replicates of the feeding experiment A.) d_4-tryptamine fed leaves (n=5), H_2O fed leaves (n=5) B.) d_4-tryptamine fed leaves (n=3) H_2O fed leaves (n=3). $p < 0.05$ * $p < 0.01$ **. C.) Structure of d_4-strictosidine. All data presented is the mean ± SEM.

Additionally, the strictosidine pool in these d_4-tryptamine feeding experiments was reproducibly enriched in label to >90% demonstrating that the pool of strictosidine is effectively turned over in this time period (Table 3). This demonstrated that feeding the substrate for strictosidine biosynthesis, tryptamine, is able to alter the steady state of the metabolic pathway and result in the accumulation of strictosidine.

Table 3 Enrichment of d_4-strictosidine

The average enrichment of the d_4- strictosidine relative to unlabelled strictosidine is presented. Enrichment values were calculated as a percentage of the combined peak area of *m/z* 531 and *m/z* 535; (peak area *m/z* 535)/(∑(peak area *m/z* 535 + peak area *m/z* 531)) A.) d_4-tryptamine fed leaves (n=5), H_2O fed leaves (n=5) presented is the mean enrichment value ± SEM B.) d_4-tryptamine fed leaves (n=3) H_2O fed leaves (n=3) mean enrichment value ± SEM. The presence of d4-strictosidine in the water fed controls can be explained by baseline noise and natural isotope abundance.

Replicate		strictosidine	d_4-strictosidine
A	WT-H_2O fed leaf	0.983 ± 0.004	0.017 ± 0.004
	WT-d_4-tryptamine fed leaf	0.013 ± 0.004	0.987 ± 0.004
B	WT-H_2O fed leaf	0.989 ± 0.004	0.01 ± 0.004
	WT-d_4-tryptamine fed leaf	0.012 ± 0.006	0.988 ± 0.006

The accumulation of strictosidine suggests that its utilisation, either due to inefficiency in export from the vacuole or inefficiency in its utilisation by the enzymes downstream in the pathway, is a potential bottleneck in the MIA pathway under conditions where substrate is supplied exogenously.

The fact that the steady state of strictosidine is altered upon supply of tryptamine suggests that tryptophan decarboxylation to tryptamine may be limiting in the strictosidine biosynthesis, and by bypassing this enzyme by supplying tryptamine, it is possible to increase flux through the MIA pathway. The alteration in the steady state levels of strictosidine indicates the utilisation of strictosidine subsequently becomes a metabolic bottleneck.

For a linear pathway it could be unsurprising that exogenous supply of millimolar concentrations of substrate result in product increase, however the MIA metabolic pathway is a convergent metabolic network between seco-iridoid biosynthesis and tryptamine biosynthesis which merge at the branchpoint of strictosidine.

The fact that strictosidine is able to accumulate upon exogenous supply of tryptamine, is indicative that the seco-iridoid branch of the metabolic pathway is not limiting strictosidine biosynthesis. If the seco-iridoid branch of the pathway is not the limiting reagent in strictosidine biosynthesis, this further supports the argument that secologanin needs to be partitioned into the vacuole to protect it from derivatisation to its alcohol product in the cytoplasm.

To further assess the labelling of the strictosidine pool and the potential for tryptophan decarboxylase to be limiting, a series of labelling experiments were performed in which wild type *C. roseus* leaf tissue was fed d_4-tryptamine, d_5-tryptophan and H_2O. In all cases the size of the strictosidine pool, as well as the enrichment of the pool *m/*z 535 was monitored. *m/z* 535 is monitored for the d_5-tryptophan experiments as well, as the labelling of strictosidine with d_5-tryptophan results in the production of d_4-strictosidine (Figure 3-20).

Notably while feeding isoptopically labelled d_4-tryptamine resulted in an increase in the total pool size of strictosidine, feeding d_5-tryptophan did not lead to a statistically significant increase in the strictosidine pool relative to H_2O fed controls (Figure 3-20).

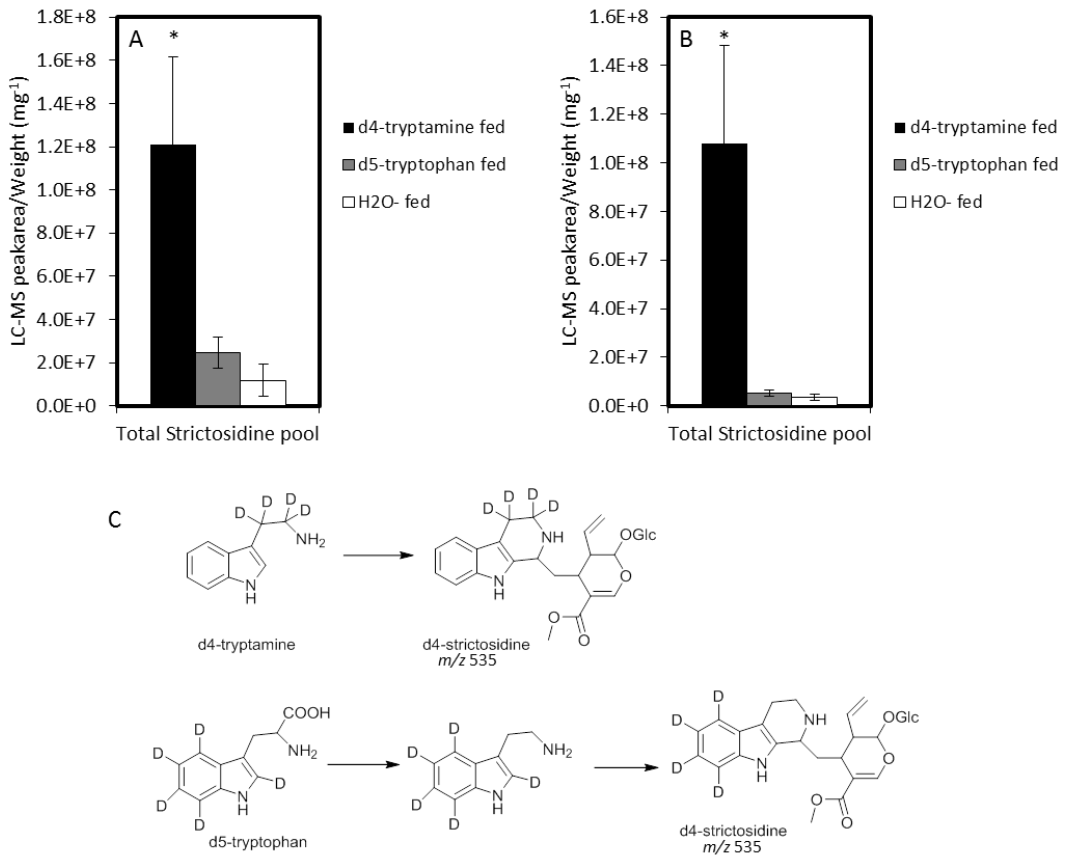


Figure 3-20 Isotopic labelling of strictosidine.
Wild type *C. roseus* leaves were cut at the petiole and fed either 100 µl 5 mM d_4-tryptamine, or 5 mM d_5-tryptophan daily for four days. The total strictosidine pool is the combined peak area of the endogenous strictosidine *m/z* 531 and the labelled strictosidine *m/z* 535. A and B represent separate replicates of the feeding experiment. A.) d_4-tryptamine fed leaves (n=5), d_5-tryptophan fed leaves (n=5), H_2O fed leaves (n=5) B.) d_4-tryptamine fed leaves (n=5), d_5-tryptophan fed leaves (n=5), H_2O fed leaves (n=5). $p < 0.05$*. All data represent the mean $\pm$ SEM. C.) Feeding both d_4-tryptamine and d_5-tryptopahn leads to the formation of d_4-strictosidine.

Feeding both isoptopically labelled d_4-tryptamine and d_5-tryptophan resulted in the labelling of the strictosidine pool with, as previously demonstrated, in excess of >90% labelling for d_4-tryptamine fed leaf tissue, and >65% labelling of the strictosidine pool for d_5-tryptophan fed leaf tissue in both experimental replicates (Table 4). Representative MS spectra for the strictosidine pool for d_4-tryptamine, d_5-tryptophan and H_2O-fed leaf tissue is presented in Figure 3-21.

Table 4 Enrichment of d_4-strictosidine

The average enrichment of the d_4-strictosidine relative to unlabelled strictosidine is presented. Enrichment values were calculated as a percentage of the combined peak area of *m/z* 531 and *m/z* 535. (peak area *m/z* 535)/(∑(peak area *m/z* 535 + peak area *m/z* 531)). A.) d_4-tryptamine fed leaves (n=5), d_5-tryptophan fed leaves (n=5), H_2O fed leaves (n=5) presented is the mean enrichment value ± SEM B.) d_4-tryptamine fed leaves (n=5), d_5-tryptophan fed leaves (n=5), H_2O fed leaves (n=5) mean enrichment value ± SEM. The presence of d4-strictosidine in the water fed controls can be explained by baseline noise and natural isotope abundance.

		strictosidine	d_4-strictosidine
A	WT-H_2O fed leaf	0.997 ± 0.001	0.001 ± 0.001
	WT-d_4-tryptamine fed leaf	0.057 ± 0.03	0.943 ± 0.003
	WT-d_5-tryptophan fed leaf	0.216 ± 0.029	0.784 ± 0.029
B	WT-H_2O fed leaf	0.981 ± 0.005	0.019 ± 0.005
	WT-d_4-tryptamine fed leaf	0.049 ± 0.02	0.951 ± 0.02
	WT-d_5-tryptophan fed leaf	0.324 ± 0.03	0.676 ± 0.03

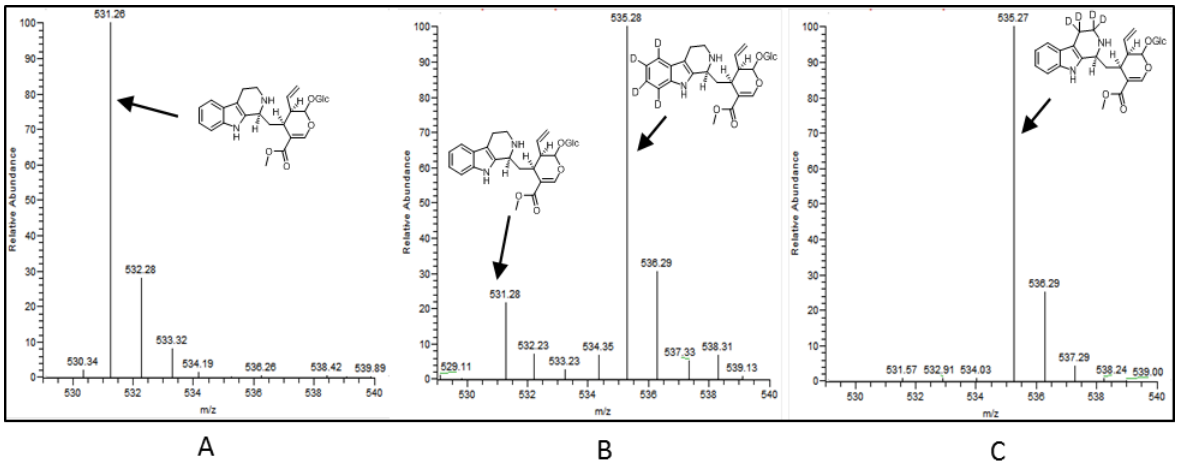


Figure 3-21 Representative mass spectra for strictosidine for H_2O fed, d_5-tryptophan fed and d_4-tryptamine fed *C. roseus* leaf tissue

Mass spectra for strictosidine A.) H_2O fed *C. roseus* leaf tissue, B.) d_5-tryptophan fed *C. roseus* leaf tissue and C.) d_4-tryptamine fed *C. roseus* leaf tissue.

This set of feeding experiments is further supportive of the hypothesis that tryptophan decarboxylase may be limiting MIA biosynthesis in *C. roseus*, and bypassing this enzyme by exogenous supply of tryptamine can result in the accumulation of strictosidine. Furthermore, the fact that exogenous tryptamine has such as substantial impact on the levels of strictosidine suggests that the seco-iridoid substrate is not the limiting one.

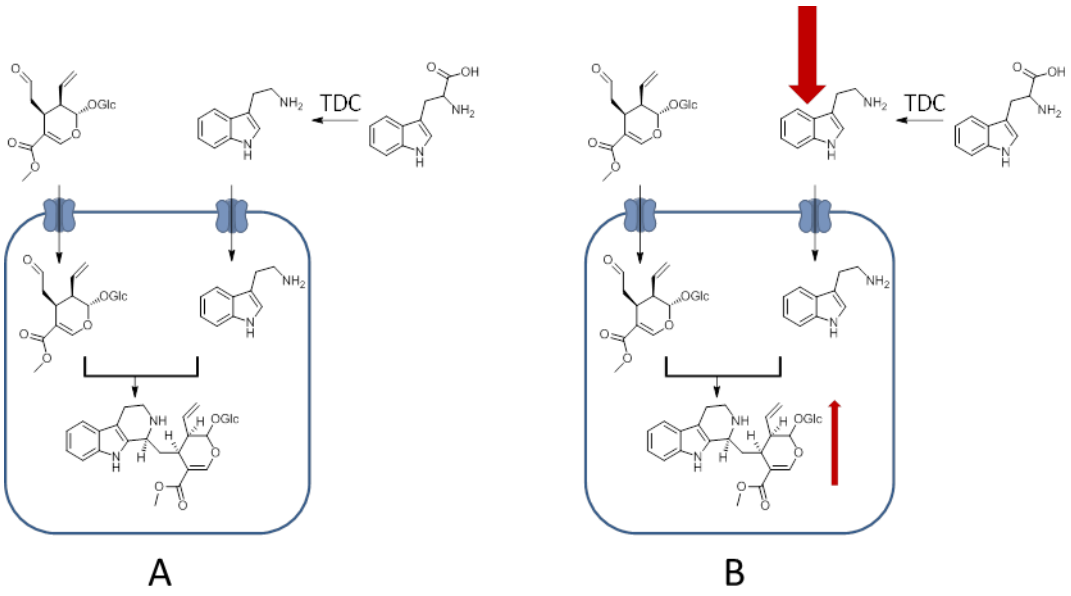


Figure 3-22 Exogenous supply of tryptamine results in an increase in strictosidine production
A.) Strictosidine biosynthesis is a convergent metabolic pathway between secoiridoid biosynthesis and tryptamine B.) Exogenous supply of tryptamine and bypassing tryptophan decarboxylase (TDC) can increase strictosidine, indicating secologanin biosynthesis is not a limiting factor in strictosidine synthesis as highlighted by red arrows

There is debate in the literature about whether tryptamine or the secoiridoid branch of MIA metabolism is limiting, with conflicting results about the increases in alkaloid accumulation upon exogenous feeding of different substrates (Canel et al., 1998; Contin et al., 1999b; Morgan and Shanks, 2000; Peebles et al., 2006; Whitmer et al., 1998). These previous studies have all been performed in cell suspension cultures which have a different alkaloid profile to leaf tissue, and the work presented in this thesis is the first demonstration that in the leaf tissue the tryptamine branch appears to be limiting. It is therefore possible that the limiting nature of either branch of the convergent metabolic branchpoint is dependent on the particular tissue or system under investigation.

3.3.8 Virus Induced Gene Silencing Fusions

Two sets of VIGS fusion pTRV2u vectors were generated to combinatorially silence multiple MATE genes, and assess whether there are functionally redundant transporters involved in secologanin import into the vacuole in *C. roseus*.

3.3.8.1 Combinatorial silencing of CrMATE1952 and CrMATE10740

A double fusion silencing vector between CrMATE1952 and CrMATE10740, pTRV2u-CrMATE1952_CrMATE10740, was generated by uracil specific excision reagent (USER) fusion cloning (Geu-Flores et al., 2007; Nour-Eldin et al., 2010), to transiently silence both MATE proteins CrMATE1952 and CrMATE10740.

As demonstrated previously, CrMATE10740 is tightly co-regulated with the MIA pathway, however individual silencing did not lead to a metabolic phenotype. Although CrMATE1952 and CrMATE10740 are not homologous, and occur on different clades in the phylogenetic tree (Figure 3-3), as has been demonstrated by the nicotine MATE transporters, MATEs with diverse sequences are able to transport the same metabolite, and therefore it is possible that both CrMATE1952 and CrMATE10740 might both be involved in secologanin import into the vacuole.

Silencing using the double vector for VIGS-CrMATE1952_CrMATE10740 resulted in the same result as individually silencing the CrMATE1952, Figure 3-23. Both sets of plants accumulated a metabolite with an *m/z* 391, and in this particular VIGS experiment both individually silencing CrMATE1952 and using the double vector for silencing both CrMATE1952 and CrMATE10740, also resulted in a decrease in the downstream alkaloids vindoline and catharanthine. qPCR confirmed silencing of the gene CrMATE1952 in both the CrMATE1952 individually silenced plants and CrMATE1952-CrMATE10740 double silenced plants (Figure 3-23), as well as silencing of the CrMATE10740 gene in the double fusion CrMATE1952-CrMATE10740 silenced plants (Figure 3-23).

The fact that there is not a statistically significant difference in the peak of *m/z* 391 between the individually silenced CrMATE1952 and the double silenced CrMATE1952_CrMATE10740 plants is indicative that the CrMATE10740 is unlikely to be involved in secologanin import, as there is no cumulative effect upon silencing both transporters.

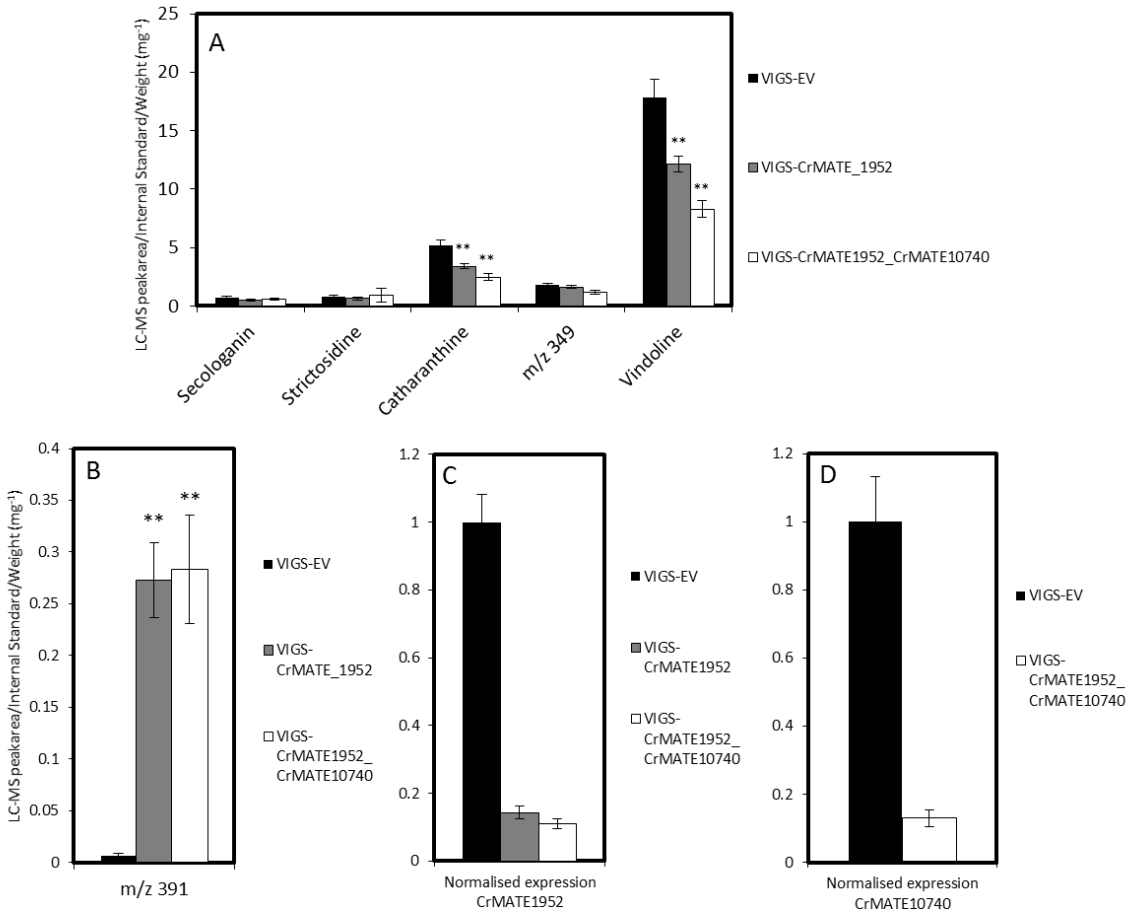


Figure 3-23 Metabolite and qPCR profile upon VIGS of CrMATE1952 and the double fusion CrMATE1952_CrMATE10740.

A.) Alkaloid profile for VIGS of CrMATE1952, and the double construct CrMATE1952_CrMATE10740, relative to empty vector control tissue in Little Bright Eyes VIGS-CrMATE1952 (n=12), VIGS-CrMATE1952_CrMATE10740 (n=12), VIGS-EV (n=12) ** p<0.01 B.) *m/z* 391 profile for VIGS of CrMATE1952 and the double construct CrMATE1952_CrMATE10740, relative to empty vector control tissue in Little Bright Eyes VIGS- CrMATE1952 (n=12), VIGS-CrMATE1952_CrMATE10740 (n=12), VIGS-EV (n=12) ** p<0.01 C.) qPCR normalised expression of the CrMATE1952 gene in empty vector, and CrMATE1952 and CrMATE1952_10740 silenced plants; VIGS-CrMATE1952 (n=8), VIGS-CrMATE1952_10740 (n=8), VIGS-EV (n=8). D.) qPCR normalised expression of the CrMATE10740 gene in empty vector and CrMATE1952_10740 silenced plants; VIGS-CrMATE1952_10740 (n=8), VIGS-EV (n=8). All data shown are mean ± SEM.

3.3.8.2 Combinatorial silencing of CrMATE1952 homologues.

As shown by the phylogenetic tree (Figure 3-3), CrMATE1952 has a number of homologues present in the transcriptome of *C. roseus*, namely locus_10550, locus_3327, locus_8527. Although the locus_10550 is closest in sequence alignment to CrMATE1952, it is not highly expressed in leaf tissue and was therefore excluded from further consideration. A triple silencing vector for the transient silencing of CrMATE1952, locus_3327 and locus_8527 in *C. roseus* by VIGS was generated. The genes at locus_3327 and locus_8527 do not co-regulate with the genes in the MIA pathway, however they are expressed in leaf tissue. It is therefore feasible that they may contribute basal activity towards secologanin import into the vacuole that can compensate for CrMATE1952 silencing.

Silencing utilising this triple silencing vector resulted in the same metabolic phenotype as that found by singly silencing CrMATE1952 relative to EV controls with the emergence of a peak of *m/z* 413 (Na^+ adduct).

There was no significant difference between the triple silencing vector and singly silenced CrMATE1952 with regards to effects on downstream alkaloids or secologanol accumulation, suggesting there was no major effect on phenotype as a result of cumulatively silencing the genes as opposed to singly silencing CrMATE1952 (Figure 3-24).

qPCR confirmed that in both the CrMATE1952 and the CrMATE_locus_3327_locus_8527 triple silencing plants the gene CrMATE1952 was silenced relative to VIGS-EV controls (Figure 3-24).

Further silencing experiments are therefore potentially needed to exacerbate this metabolic phenotype if more transporters are involved in secologanin import.

This work demonstrates that combinatorial VIGS vectors can be generated by stacking ~500 bp regions of genes onto the pTRV2u vector and cloned by a USER fusion cloning strategy, as highlighted by the CrMATE-CrMATE10740 double silenced plants. Further combinatorial silencing experiments could therefore utilise this approach to stack genes from disparate transporter classes and assess their effect on secologanol formation together with the effect on downstream MIAs.

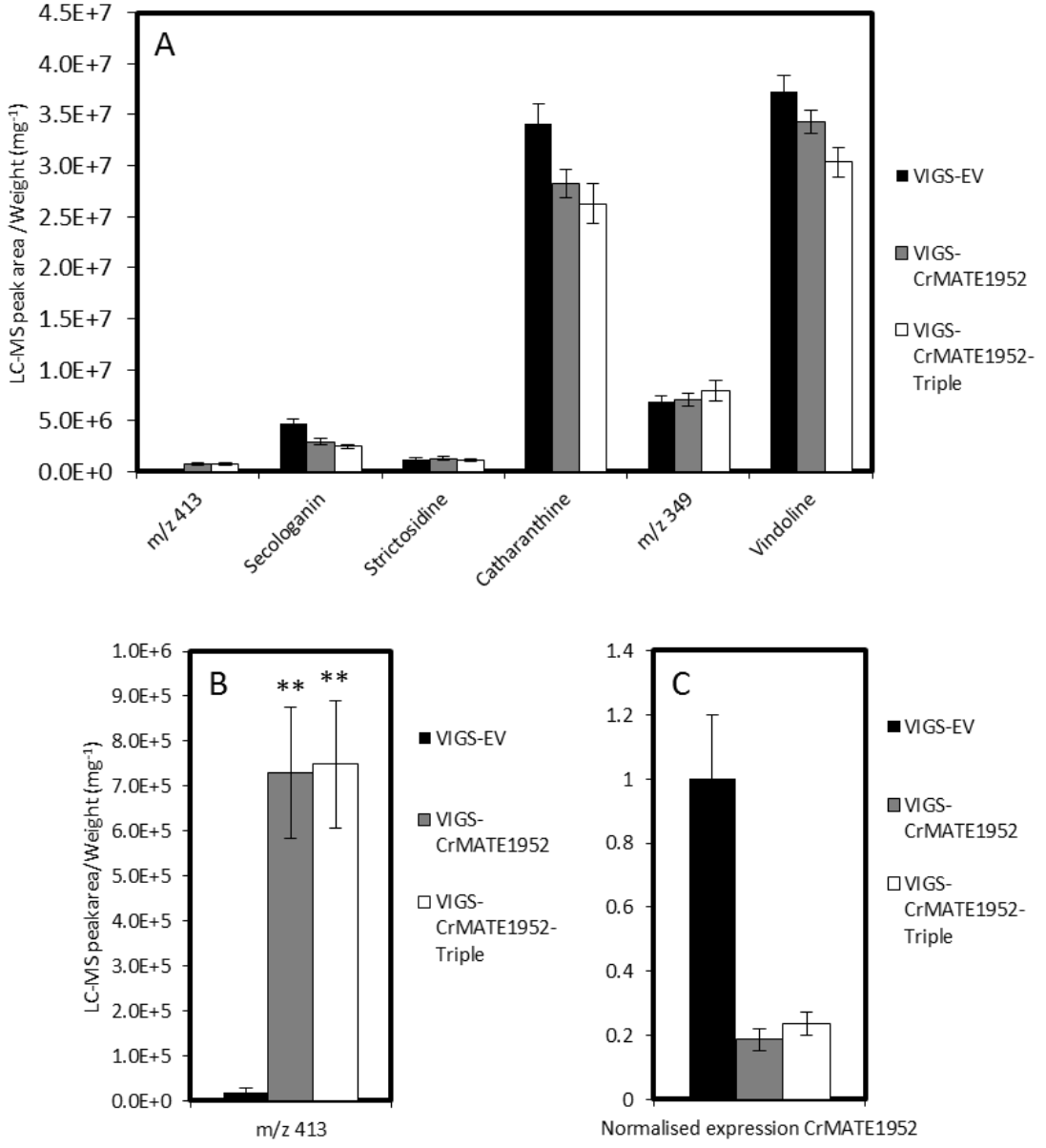


Figure 3-24 Metabolite and qPCR profile upon VIGS of CrMATE1952 and the triple fusion CrMATE1952_locus3327_locus8527

A.) Alkaloid profile for VIGS of CrMATE1952, and the triple construct CrMATE1952-Triple, relative to empty vector control tissue in Little Bright Eyes VIGS-CrMATE1952 (n=9), VIGS-CrMATE1952-Triple (n=14), VIGS-EV (n=9) ** p<0.01 B.) *m/z* 413 profile for VIGS of CrMATE1952 and the triple construct CrMATE1952-Triple, relative to empty vector control tissue in Little Bright Eyes VIGS- CrMATE1952 (n=9), VIGS-CrMATE1952-Triple (n=14), VIGS-EV (n=9) ** p<0.01 C.) qPCR normalised expression of the CrMATE1952 gene in empty vector, CrMATE1952 and CrMATE1952-Triple silenced plants VIGS-CrMATE1952 (n=8), VIGS-CrMATE1952-Triple (n=8), VIGS-EV (n=8). All data shown are mean ± SEM.

3.3.9 Heterologous expression of MATE proteins in S. cerevisiae

The function of a transporter also needs to be assessed by heterologous expression and *in vitro* biochemical assays. *S. cerevisiae* is a good host for the expression of eukaryotic membrane proteins and for the characterisation of substrate transport (Newstead et al., 2007). This system was used for the expression of CrMATE1952 and CrMATE10740 to assess whether they have the capacity to transport secologanin.

C-terminal tagging with green fluorescent protein, GFP, facilitates in the identification of whether a membrane protein is correctly folded. For an integral membrane protein fused to GFP, the GFP folds and becomes fluorescent only if the N-terminal membrane protein is integrated into the membrane (Newstead et al., 2007). Screening of C-terminally tagged GFP fusions of membrane proteins, together with analysing the localisation of expression in the yeast strain can therefore be used to assess whether *S. cerevisiae* is a suitable host for heterologous expression and transport assays (Drew et al., 2008; Newstead et al., 2007).

As discussed previously there are a number of plant MATEs that have been characterised for transport of secondary metabolites in *S. cerevisiae*; the four tonoplast localised nicotine transporters NtMATE1/2, NtJAT1, NtJAT2 (Morita et al., 2009; Shitan et al., 2014; Shoji et al., 2009) as well as the flavonoid transporters AtTT12, MtMATE1, MtMATE2, AM1 and AM3, (Gomez et al., 2009; Marinova et al., 2007; Zhao and Dixon, 2009; Zhao et al., 2011). The localisation of the membrane protein in the yeast dictates the type of transport assay that is amenable.

For the MATEs, transport universally occurs in the direction away from the cytoplasm. Therefore, if the MATE transporter is expressed at the plasma membrane, a whole yeast assay approach is necessary. Although the nicotine transporters are localised to the tonoplast in plants, overexpression of these proteins as C-terminal GFP fusions in *S. cerevisiae*, under control of a constitutive promoter, resulted in localisation to the plasma membrane (Morita et al., 2009; Shitan et al., 2014; Shoji et al., 2009).

In this whole cell assay, the yeast culture expressing the MATE is incubated with the substrate, relying on endogenous yeast transporters to pump the potential substrate from the media into the cell. The yeast expressing the MATE, which selectively pumps the substrate out, would be expected to have a lower intracellular concentration of the substrate than a negative control not expressing the transporter. This type of assay has been used in the assessment of transport

for the NtJAT1, NtJAT2, NtMATE1 and NtMATE2 (Morita et al., 2009; Shitan et al., 2014; Shoji et al., 2009).

In comparison, over-expression of the tonoplast flavonoid transporters, AtTT12 (Marinova et al., 2007), MtMATE1 and MtMATE2 (Zhao and Dixon, 2009; Zhao et al., 2011) in *S. cerevisiae* as C-terminal GFP fusions, resulted in the localisation of each protein to a defined internal membrane in the yeast. The assay for flavonoid uptake was therefore designed to measure uptake of flavonoids from purified yeast vesicles expressing each MATE protein. AtTT12 has been shown to be expressed in an active form from both a constitutive and inducible promoter system (Marinova et al., 2007; Zhao and Dixon, 2009; Zhao et al., 2011), whilst MtMATE1 and MtMATE2 have been characterised utilising a galactose inducible promoter.

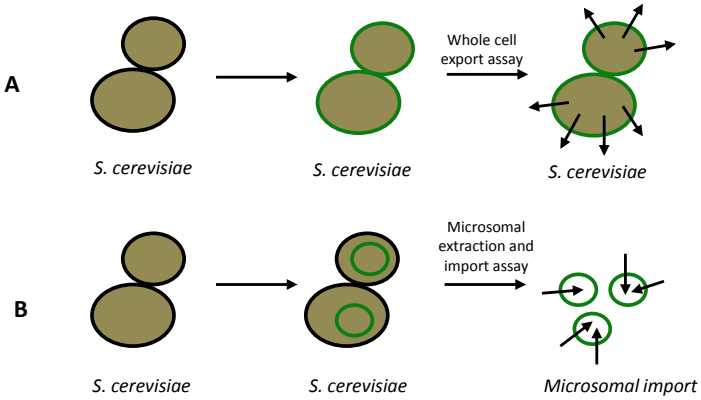


Figure 3-25 Transport assays in yeast
For the MATEs, if the proteins are expressed at the plasma membrane a whole yeast assay approach is needed. A.) The yeast culture is incubated in substrate and the yeast expressing the MATE would be expected to have a lower intracellular concentration due to export from the cell. B.) If the proteins are localised to internal membranes, the yeast microsomes can be extracted for an import assay.

3.3.9.1 Overexpression of CrMATE1952 and CrMATE10740 in yeast

CrMATE1952 and CrMATE10740 were cloned as C-terminal GFP fusions into the gateway compatible yeast expression vector pYES-DEST52. This plasmid is used for inducible protein expression under control of the galactose inducible promoter *GAL1* and has been used in the characterisation of MtMATE2, AtTT12 and MtMATE1 (Zhao and Dixon, 2009; Zhao et al., 2011). The use of an inducible *GAL1* promoter was chosen as it has been shown to increase the amount of protein expressed for eukaryotic membrane proteins, relative to using constitutive promoters (Newstead et al., 2007). The CrMATE1952-GFP and CrMATE10740-GFP fusions were expressed in three different yeast strains, InvSCI, WATII, and ybtI.

Unfortunately overexpression of CrMATE1952-GFP and CrMATE10740-GFP proteins as C-terminal GFP fusions in *S. cerevisiae*, upon induction by galactose, did not result in the defined membrane localisation of these proteins to either the plasma membrane, or an internal membrane, as has been observed with the other functionally characterised MATE proteins; NtMATE1/2, NtJAT1, NtJAT2, AtTT12, MtMATE1, MtMATE2, AM1 and AM3 (Gomez et al., 2009; Marinova et al., 2007; Morita et al., 2009; Shitan et al., 2014; Shoji et al., 2009; Zhao and Dixon, 2009; Zhao et al., 2011).

Expression of CrMATE10740-GFP or CrMATE1952-GFP resulted in localisation primarily to the lytic vacuole lumen in *S. cerevisiae* when visualised using confocal microscopy in the three different yeast strains, InvSCI, WAT11 and ybt1 (Figure 3-26).

This mislocalisation has also been observed for the nicotine transporting MATE protein NtJAT2 (Shitan et al., 2014). The NtJAT2-GFP fusion has a dual targeting to the plasma membrane, as well as the vacuolar lumen. The localisation of the GFP signal to the vacuolar lumen was hypothesised to be due to membrane protein mistargeting and the subsequent digestion of the protein by yeast proteases. As GFP is stable to proteolytic degradation it therefore localised to the vacuolar lumen. This was confirmed by a western blot with a GFP antibody which demonstrated the presence of two bands for the size of the NtJAT2-GFP fusion and free soluble GFP (Shitan et al., 2014).

A western blot of crude membranes from CrMATE10740-GFP in InvSC1 and WAT11 only resulted in a band with the same size as GFP when targeted with an anti-GFP antibody. For the CrMATE1952-GFP expressing yeast strains InvSCI or WATII no band was detectable with this anti-GFP antibody (Figure 3-27).

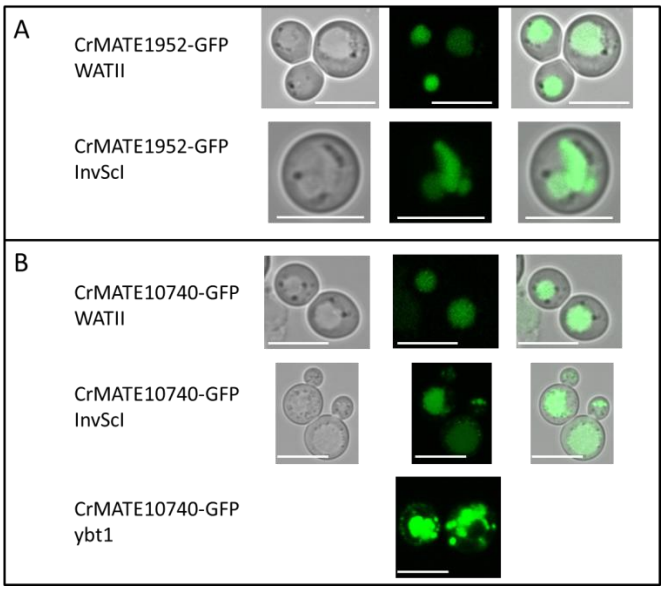


Figure 3-26 Heterologous expression of CrMATE1952-GFP and CrMATE10740-GFP in *S. cerevisiae*.
A.) Expression of CrMATE1952-GFP in two different strains of *S. cerevisiae* WATII and InvSCI and visualisation by confocal microscopy. B.) Expression of CrMATE10740-GFP in three different strains of *S. cerevisiae* WATII, InvSCI and ybtI and visualisation by confocal microscopy. (l-r) Brightfield, GFP-fluorescence, Merge. Scale bar = 5 µm.

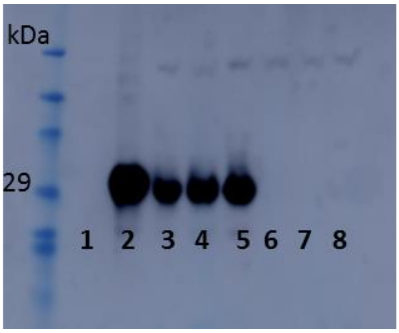


Figure 3-27 Western blot of microsomal membranes from GFP, CrMATE1952-GFP and CrMATE10740-GFP expressing yeast strains.
Western blot of yeast microsomal membranes extracted from 1. Untransformed yeast 2. GFP expressing WATII 3. CrMATE10740-GFP InvSCI colony1 4. CrMATE10740-GFP InvSCI colony2 5. CrMATE10740-GFP WATII colony1 6. CrMATE1952-GFP InvSCI colony1 7.) CrMATE1952-GFP InvSCI colony2 8.) CrMATE1952-GFP WATII colony1. Each lane was loaded with 10 µg/µl of protein.Membrane was labelled with an anti-GFP antibody.

A GFP expressing yeast strain showed that the localisation of the GFP signal from free GFP, and GFP derived from expressing the CrMATE-GFP fusions was different (Figure 3-28). This is indicative that the GFP derived from the CrMATE-GFP constructs are likely to be due to the transporter protein products being degraded in the lytic vacuole, with the GFP from the fusion accumulating in the vacuolar lumen.

In order to assess whether the cloning strategy and yeast strains utilised in this study are suitable for *C. roseus* membrane transporter expression, a putative purine permease CrPUP_1443 that is expressed in *C. roseus* leaf tissue, was cloned and expressed as a C-terminal GFP fusion in all yeast strains using the vector pYES-DEST52 (Figure 3-28).

In all yeast strains, the CrPUP_1443-GFP fusion localised to a different subcellular compartment to either CrMATE10740-GFP or CrMATE1952-GFP, namely both the plasma membrane as well as internal vesicles. This is reminiscent of the MATE transporters NtJAT1, NtJAT2, NtMATE1, and is indicative that the cloning strategy we have used is suitable for membrane protein expression, and that CrMATE1952 and CrMATE10740 may not be amenable to heterologous expression in *S. cerevisiae.*

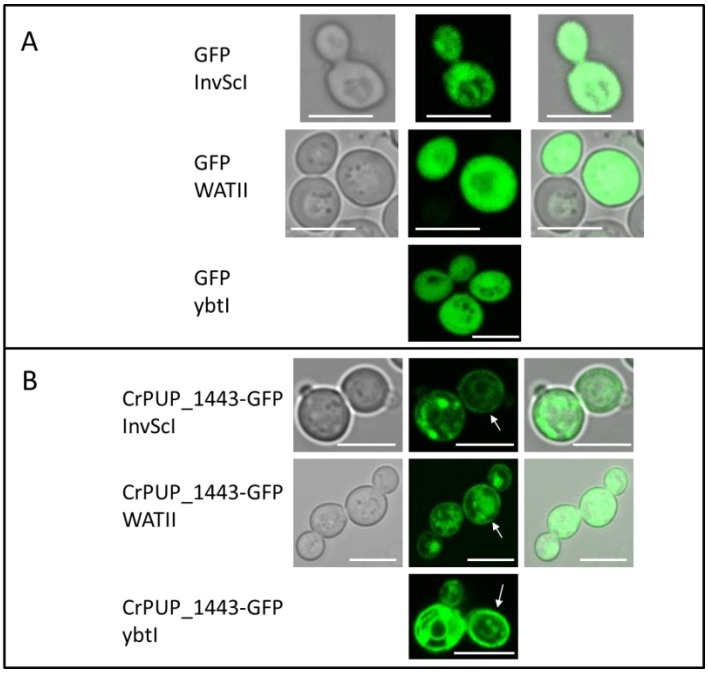


Figure 3-28 Heterologous expression of GFP and CrPUP_1443-GFP in *S. cerevisiae*.
A.) Expression of GFP in three different strains of *S. cerevisiae* WATII, InvSCI and ybtI and visualisation by confocal microscopy. B.) Expression of CrPUP_1443-GFP in three different strains of *S. cerevisiae* WATII, InvSCI and ybtI and visualisation by confocal microscopy. Highlighted by white arrows is the plasma membrane of the yeast to which CrPUP_1443-GFP localised. (l-r) Brightfield, GFP-fluorescence, Merge. Scale bar = 5µm.

Furthermore expression of CrMATE10740-GFP by the group of Vincent Courdavault, University of Tours, in a different yeast strain AD12345678 resulted in the same localisation as that of CrMATE10740-GFP in WAT11 (unpublished results; personal communication) and the group of

Marianna Sottomayer were also unsuccessful in stably expressing CrMATE1952-GFP in *S. cerevisiae* (unpublished results; personal communication).

3.3.10 Heterologous expression in Xenopus laevis oocytes.

An alternative system to test for transport is the *Xenopus* oocyte expression system. In this system, linear cRNA derived from *in vitro* transcription, encoding the transporter to be expressed, is microinjected into the *Xenopus* oocyte for protein expression. The protein is likely expressed at the plasma membrane of the oocyte and transport assays can then be performed relative to water injected controls, to assess whether the transporter can move the substrate (Figure 3-29).

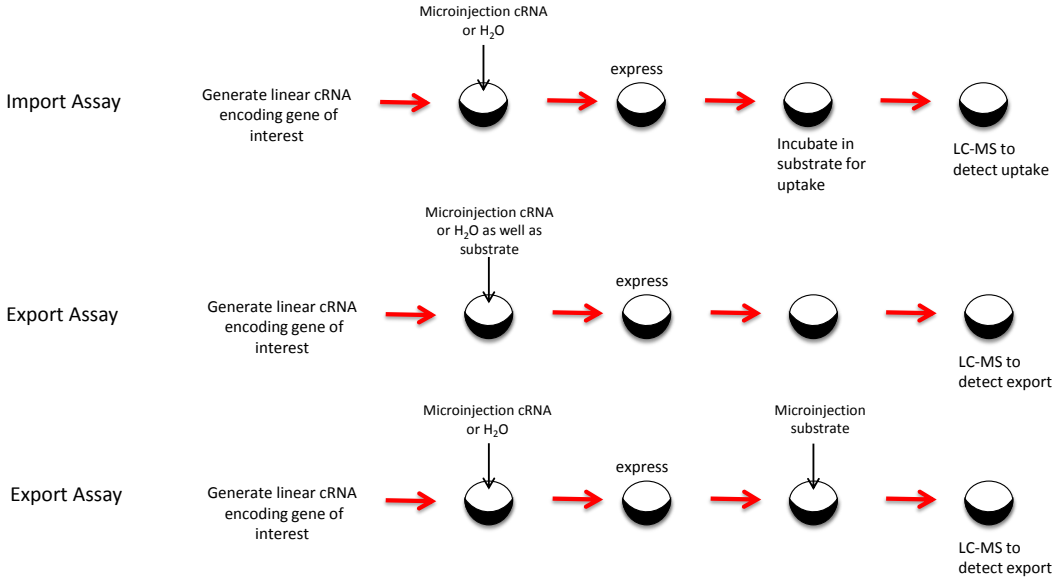


Figure 3-29 Scheme for *Xenopus* oocyte expression.
Xenopus laevis oocytes can be utilised for the heterologous expression of membrane proteins through the microinjection of linear cRNA containing the gene of interest. Depending on the directionality of transport, an import assay or export assay can be performed, and LC-MS utilised to detect the transport of the substrate. The export assay is technically more challenging as it requires either the co-injection of the substrate with the cRNA or the separate injection of substrate.

For the MATE proteins that have been characterised to date, the general directionality of transport is away from the cytoplasm. Therefore, *in planta*, if these transporters are expressed at the tonoplastic membrane, they move substrates into the vacuole, whilst expression at the plasma membrane would result in the movement of the substrate out of the cell.

For the *Xenopus* oocyte based assay in which the MATEs are expressed at the plasma membrane an export assay is required. In this scheme the transport assay requires the microinjection of the

substrate into the oocyte, as well as the cRNA, and subsequent measurement of substrate levels in the oocyte. The assumption is that the oocytes expressing the correct transporter will contain less substrate than the negative controls. This methodology has previously been employed for MATE proteins in the characterisation of organic acid export (Durrett et al., 2007; Furukawa et al., 2007).

3.3.10.1 Microinjection of secologanin into Xenopus oocytes

To assess whether the substrate secologanin is compatible with the *Xenopus* assay system, and whether there are any endogenous transport systems for the substrate that may affect the transport assay, 50 nl of a secologanin stock solution at concentrations of 1 mM, 5 mM and 10 mM was injected into untransformed *Xenopus laevis* oocytes, and incubated in MBS buffered saline for two days. The oocytes were subsequently homogenised and extracted in methanol for 2 hr at 60 ^{o}C, and the methanolic extract monitored by LC-MS.

As can be seen in Figure 3-30 and Figure 3-31 after two days there is no detectable secologanin in the oocytes injected with 1 mM or 5 mM, whilst there is some residual secologanin detected when oocytes were injected with 10 mM.

It appears that the oocytes reduce secologanin to secologanol, since a peak of *m/z* 413 that co-elutes with the secologanol product of VIGS-CrMATE1952 silencing is observed (Figure 3-31). This suggests that the oocyte expresses a non-specific reductase that can convert the aldehyde of secologanin to the alcohol, perhaps as a detoxification mechanism.

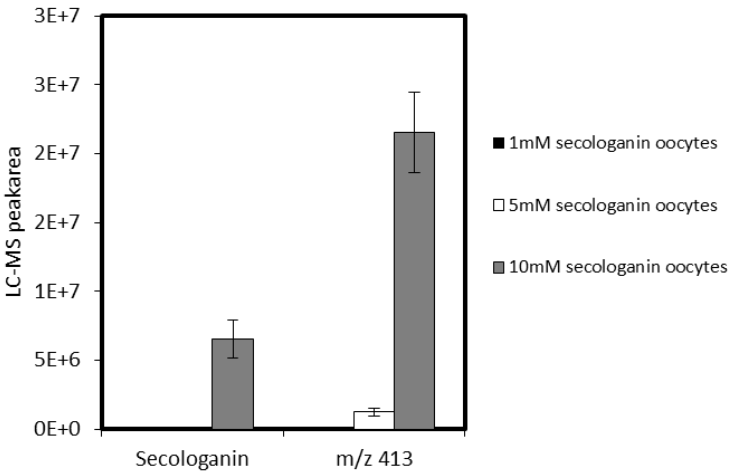


Figure 3-30 Metabolite profile upon microinjection of *Xenopus laevis* oocytes with secologanin.
For each concentration of secologanin 20 *Xenopus laevis* oocytes were microinjected with either 50 nl 1 mM secologanin, 5 mM secologanin and 10 mM secologanin. For each datapoint 5 oocytes were pooled, extracted in methanol and analysed by LC-MS. This therefore resulted in 4 replicates for each injected concentration. No secologanin or secologanol was identified upon injection of 50 nl 1 mM secologanin. All data shown are mean ± SEM.

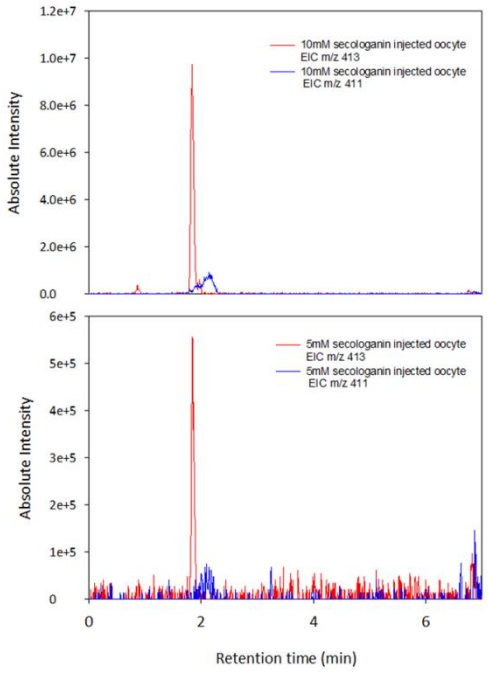


Figure 3-31 Representative Extracted Ion Chromatogram of *m/z* 413 and *m/z* 411 *Xenopus laevis* oocytes injected with secologanin.
Xenopus laevis oocytes were microinjected with either 50 nl 1 mM secologanin, 5 mM secologanin and 10 mM secologanin. For each concentration, 5 oocytes were pooled, extracted in MeOH and run by LC-MS. A.) Extracted ion chromatogram of *m/z* 411 (secologanin) and *m/z* 413 (secologanol) for oocytes injected with 10 mM secologanin. B.) Extracted ion chromatogram of *m/z* 411 (secologanin) and *m/z* 413 (secologanol) for oocytes injected with 5 mM secologanin.

The derivatisation of secologanin will complicate characterisation of CrMATE1952 in this heterologous host.

This provides further evidence for the importance of subcellular compartmentation of this substrate in *C. roseus*, and indicates that its storage and localisation to the plant vacuole may be an integral mechanism to prevent secologanin reduction. An alternative approach to this methodology would be to microinject the *Xenopus* oocytes with substrate immediately prior to performing the oocyte assay. This method, however, would be more technically challenging as it would require two injections of the oocytes.

3.3.11 Overexpression in N. benthamiana

An alternative expression system to *S. cerevisiae* or *X. laevis* oocytes, is through the use of *N. benthamiana* expression. Transient over expression of the two MATE transport proteins CrMATE10740 and CrMATE1952 in *N. benthamiana*, utilising agroinfiltration, together with the purification of vacuoles from this leaf tissue, can be used for transport assays across the tonoplast membrane.

Due to the proposed directionality of transport, in which the CrMATE1952 is hypothesised to result in the import of secologanin into the vacuole, a transport assay for secologanin can be conducted across the tonoplast membrane with the contents of the vacuoles expressing the CrMATE1952 transporter and negative control vacuoles assessed for secologanin uptake by LC-MS.

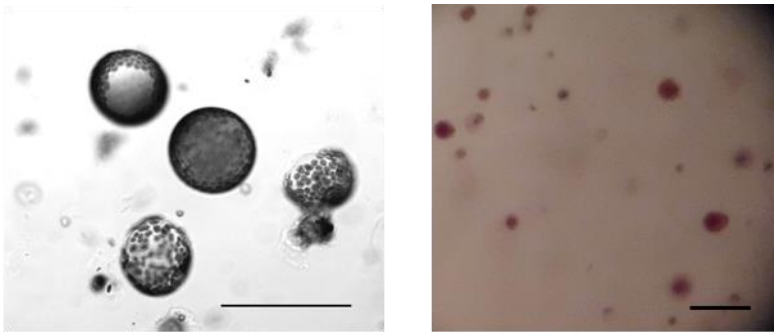
The full length *CrMATE10740* and *CrMATE1952* genes were cloned into the pEAQ-HT DEST1 expression vector and transformed into *A. tumefaciens* strain LBA4404 for transient overexpression by agroinfiltration in *N. benthamiana.*

Overexpression of CrMATE10740 for harvesting of vacuoles was not feasible as the expression of the native protein resulted in the cell death of the infiltrated leaves. This phenomenon occurred irrespective of the optical density used to infiltrate the leaf tissue (Figure 3-32).

A pEAQ-HT CrMATE10740 **B** pEAQ-HT GFP

Figure 3-32 Overexpression of CrMATE10740 in *N. benthamiana* leaf tissue
Agroinfiltration of *N. benthamiana* leaves for expression of the CrMATE10740 protein using the pEAQ-HT system resulted in cell death of the leaf tissue. A.) Expression of pEAQ-HT-CrMATE10740. B.) Expression of pEAQ-HT GFP. Infiltrated leaves are highlighted by white arrows.

In contrast, over expression of the CrMATE1952 leaf tissue did not result in cell death, and, as has been shown previously in this chapter, expression of a C-terminal GFP fusion of CrMATE1952 in *N. benthamiana* epidermal cells resulted in fluorescence in membrane vesicle compartments (Figure 3-6). We have been able to isolate protoplasts and generate vacuoles from *N. benthamiana* plant tissue (Figure 3-33). Therefore, the stage is now set to utilise this expression system to functionally characterise the CrMATE1952 protein.



A Protoplasts *N. benthamiana* **B** Purified vacuoles

Figure 3-33 Purification of plant vacuoles from *N. benthamiana*.
A.) Brightfield image of *N. benthamiana* mesophyll protoplasts. Scale bar = 100 µm B.) Brightfield image of vacuoles purified from *N. benthamiana* mesophyll protoplasts by osmostic lysis and ultracentrifugation on a Ficoll gradient. Neutral red stain was used for visualisation of the vacuoles. Scale bar = 60 µm. Both images were AF6000 stereo brightfield microscope.

3.4 Conclusion

This chapter presents the first example of a MATE transporter directly affecting iridoid glycoside formation in *C. roseus*. Silencing of the CrMATE1952 transporter resulted in a clear metabolic phenotype, with the accumulation of a new compound that, based upon co-elution with known standards, is likely to be the reduced product of the iridoid glycoside secologanin, secologanol.

The localisation of the transporter to the tonoplast membrane, together with the metabolic phenotype observed, suggests that this transporter is likely to be a secologanin importer into the vacuole, and the observed metabolic phenotype occurs because the redistribution of secologanin in the cytoplasm leads to its reduction.

Although silencing of this transporter led to a reproducible metabolic phenotype involved in iridoid glycoside biosynthesis, the effect on the downstream alkaloids is minimal, suggesting that other transporters might also be involved in secologanin import.

Individually silencing other MATE transporters in *C. roseus*, as well as combinatorially silencing these transporters, did not lead to a phenotype that is greater than singly silencing CrMATE1952. Therefore, to date, we only have evidence to suggest that CrMATE1952 is involved in iridoid glycoside transport.

It is possible that other transporter families other than the MATEs are also involved in secologanin import, and therefore the combinatorial silencing of disparate transporter classes may provide a method for generating novel metabolic phenotypes in *C. roseus*.

The work described in this chapter highlights the importance of subcellular compartmentation involved in aldehyde metabolism. Secologanin accumulates in *C. roseus* leaf tissue, and it is possible that its import into the vacuole prevents its derivatisation to the less reactive alcohol secologanol. As has been observed, silencing this tonoplast localised transporter, as well as over-feeding secologanin to wild type plants, is likely to lead to the subcellular redistribution of secologanin. This subsequently results in the accumulation of the reduced product of secologanin, secologanol (*m/z* 391 (H^+ adduct) and *m/z* 413 (Na^+ adduct)).

The importance of subcellular localisation is further emphasised by the fact that other non-plant organisms also possess mechanisms for the reduction of secologanin. Microinjection of secologanin into *Xenopus laevis* oocytes resulted in the reduction of secologanin to a peak of

m/z 413, indicating that an unrelated species possesses a promiscuous reductase for detoxification of this aldehyde.

The localisation of secologanin to the vacuole thereby provides a mechanism by which this substrate is able to accumulate in the plant without its derivatisation, and highlights the importance of the subcellular localisation of this metabolic pathway in controlling the accumulation of reactive intermediates, which are needed as substrates for alkaloid biosynthesis.

Sequestration of intermediates might be especially important in the context of a convergent metabolic network. The formation of strictosidine, the key metabolic branchpoint in the MIA pathway, occurs as a result of condensation of the product of tryptophan, tryptamine and the iridoid glycoside secologanin. If there is a discrepancy in the rate of formation of tryptamine and secologanin, and tryptamine formation becomes limiting relative to secologanin, the storage of secologanin in the vacuole prevents its derivatisation to the alcohol secologanol.

With regards to this, it is interesting that CrMATE1952 is the gene that is both co-regulated at the transcriptional level with known genes in the biosynthetic pathway, as well as being clustered on the genome with tryptophan decarboxylase and strictosidine synthase. It is possible that this spatial proximity on the genome is integral for co-regulation of these genes. The co-regulation and genomic location hints that the two branches of the convergent metabolic pathway to strictosidine synthesis via iridoid glycoside metabolism and the decarboxylation of tryptophan are intricately connected. Further work to pick apart how the pathways of tryptamine synthesis and iridoid glycoside transport are related is still needed.

In the context of metabolite damage, it is known that small molecules are often damaged by side reactions from promiscuous enzymes or through spontaneous chemical reaction, and repair pathways are needed to mitigate the effect of this. Key examples of this include the repair of hydrated NAD(P)H back to NAD(P)H through the action of a fused dehydratase-epimerase (Niehaus et al., 2014) or the repair of 5-formyltetrahydrofolate back to 5,10-methenyltetrahydrofolate by the 5-formyltetrahydrofolate cycloligase (Linster et al., 2013). The transport of secologanin into the vacuole therefore potentially represents a pre-emptive mechanism to prevent the inappropriate reduction to the non-useful side product secologanol.

In vitro evidence for transport of secologanin by CrMATE1952 has not been possible in *S. cerevisiae*, as the protein is not appropriately expressed, or *Xenopus* oocytes, as the endogenous metabolism of secologanin by the oocytes prevents analysis of export. Preliminary evidence

suggests we are able to extract vacuoles from *N. benthamiana* expressing CrMATE1952 and we are now in the position to attempt transport assays for this protein across the tonoplast membrane.

3.5 Future directions

There are a number of future directions and alternative experiments that can be proposed for the characterisation of the MATEs involved in secologanin transport.

1.) As has been described previously, further VIGS experiments are necessary to assess whether other transporters are involved in secologanin vacuolar import. This would utilise the combinatorial VIGS stacking approach that has been developed in this study, and would allow for the silencing of multiple transporter families in one VIGS experiment.

2.) VIGS is a reverse engineering approach that allows relatively fast screening for metabolic phenotypes when compared to other reverse genetic engineering approaches such as RNAi and genetic knock out. Now that a metabolic phenotype has been identified for the MATE transporter CrMATE1952 it could be feasible to knock out the gene in *C. roseus*. A genetic knock out would confirm whether the lack of a phenotype on the downstream alkaloids for this transporter is due to the functional redundancy of this transporter. If the transporter is functionally redundant, we would expect the same metabolic phenotype that we observe by VIGS. If however, the lack of a phenotype for downstream alkaloids is due to the fact that there is enough basal activity of the CrMATE1952 transporter upon VIGS silencing to maintain flux for MIA biosynthesis, a genetic knock out of this transporter should have a greater cumulative effect on the MIA levels.

3.) We need to utilise the *N. benthamiana* vacuole preparation method, coupled to overexpression of CrMATE1952 in this heterologous host, to assess whether these vacuoles can transport secologanin to a greater extent than endogenous *N. benthamiana* vacuoles, to provide further definitive proof that this transporter is involved in the translocation of secologanin.

Iridoid biosynthesis is a crucial part of the biosynthetic pathway for high value anti-cancer agents produced by this plant, and we have uncovered a transporter for this pathway. Altogether this work highlights the importance of a MATE transporter in the distribution of secologanin in *C. roseus* leaf tissue, and how the subcellular distribution of reactive intermediates is integral to prevent their derivatisation in metabolic pathways. This appears to be especially important in

the context of a convergent metabolic pathway where the two convergent branches may not be occurring at the same rate.

4 Characterisation of *C. roseus* NPF transporters

4.1 Introduction to NPF transporters

4.1.1 Background to MFS transporters and the POT/PTR/NPF family of transporters.

The major facilitator superfamily (MFS) is a transporter family that occurs ubiquitously in all living organisms (Pao et al., 1998). Unlike the ABC transporters, which utilize the energy of ATP hydrolysis for movement of small molecules and macromolecules across biological membranes, the MFS transporters are secondary transporters that utilize the chemiosmotic ion gradient for the movement of small molecules, and are involved in the transport of a wide spectrum of substrates (Pao et al., 1998). The MFS proteins encompass both facilitators, which transport a substrate down its concentration gradient, as well as secondary active transporters which utilise the energy of chemiosmotic ion gradients to couple the movement of a solute against its concentration gradient (Reddy et al., 2012).

The major facilitator superfamily is divided into 91 subfamilies, in accordance with the transporter classification database (Saier et al., 2009). This superfamily is broadly characterised as having 12 transmembrane helices with a common structural fold comprising two discretely folded domains, each containing 6 α-helices with pseudo two fold symmetry, and with the N- and C-termini located on the cytoplasmic side of the membrane (Yan, 2015). To date, these transporters are thought to mechanistically transport the substrate across the membrane by an alternating access model. In this mechanism the substrate binding site is alternately exposed to either side of the membrane, allowing for solute movement across the lipid bilayer (Yan, 2013; Yan, 2015).

The proton oligopeptide cotransporter family (POT/PTR/SLCA15) is a subfamily of the MFS proteins with four distinct homologues in humans, PepT1, PepT2, Pht1 and Pht2 (SLC15A1-A4) (Daniel and Kottra, 2004; Smith et al., 2013). This subfamily is present in all kingdoms of life. The bacterial, yeast and human homologues that have been characterised to date demonstrate transport of a range of substrates, including di- and tri-peptides, as well as the β-lactam antibiotics (Newstead et al., 2011). As transport of the di- and tri-peptides show no dependence on substrate sequence, this family therefore displays broad substrate specificity for in excess of 8500 different substrates derived from the 20 proteinogenic L-α-amino acids (Daniel and Kottra, 2004; Smith et al., 2013). The physiological role of these human proteins are dependent on their

tissue specific localisation, with PepT1 primarily localised to the intestine and kidney, and PepT2 localised to the kidney, brain and lung (Daniel and Kottra, 2004; Smith et al., 2013).

In planta the homologues of the bacterial, fungi and mammalian, POT/PTR/ SLC15 family members are known as the Nitrate and Peptide family (NPF) transporters, all of which display a conserved membrane topology with 12 predicted transmembrane helices. In comparison to bacteria, yeast and mammals, the functionality and size of this transporter family is considerably expanded *in planta*, with more than 50 members of this family in *Arabidopsis*.

In bacteria and mammals, the primary function of this transporter family appears to be in nitrogen assimilation from di and tri peptides (Daniel and Kottra, 2004; Newstead et al., 2011; Smith et al., 2013), whilst *in planta* this family has the capacity to transport a wide variety of substrates, and is therefore involved in a diverse array of physiological plant functions, however to date this family has primarily been associated with low affinity nitrate transport (Leran et al., 2014).

A rigorous phylogenetic relationship of the plant NPF transporter family was performed utilising 2398 protein sequences from 33 different plant genomes, and demonstrated that there are eight unambiguous clades in this family in plants, resulting in the characterisation of eight subfamilies NPF1-NPF8 (Leran et al., 2014). It is necessary to note that sequence homologies do not necessarily correlate with substrate specificity, with multiple families being able to transport the same substrates. Additionally some NPF transporters have the capacity to transport more than one substrate, and also act as metabolic sensors.

An overview of the known substrate specificities of the characterised NPF transporters from *Arabidopsis* to date is highlighted in Table 5 (Leran et al., 2014).

Table 5 NPF transporters in *Arabidopsis thaliana*
Table adapted from (Leran et al., 2014).

Previous Name	NPF nomenclature	Substrate
NRT1.11	AtNPF1.2	Unknown
NAXT1	AtNPF2.7	Nitrate
NRT1.9	AtNPF2.9	Nitrate/glucosinolates
GTR1	AtNPF2.10	Nitrate/glucosinolates
GTR2	AtNPF2.11	Nitrate/glucosinolates
NRT1.6	AtNPF2.12	Nitrate
NRT1.7	AtNPF2.13	Nitrate/glucosinolates
Nitr	AtNPF3.1	Nitrate/nitrite
AIT3	AtNPF4.1	Abscisic acid
AIT4	AtNPF4.2	Unknown
NRT1.14	AtNPF4.3	Unknown
NRT1.13	AtNPF4.4	Unknown
AIT2	AtNPF4.5	Unknown
NRT1.2/AIT1	AtNPF4.6	Nitrate/Abscisic acid
PTR3	AtNPF5.2	Dipeptides
NRT1.16	AtNPF5.13	Nitrate
NRT1.15	AtNPF5.14	Nitrate
NRT1.4	AtNPF6.2	Nitrate
NRT1.1	AtNPF6.3	Nitrate/Auxin
NRT1.3	AtNPF6.4	Nitrate
NRT1.8	AtNPF7.2	Nitrate
NRT1.5	AtNPF7.3	Nitrate
PTR1	AtNPF8.1	Dipeptides
PTR5	AtNPF8.2	Dipeptides
PTR2/NTR1	AtNPF8.3	Dipeptides
PTR4	AtNPF8.4	Unknown
PTR6	AtNPF8.5	Unknown

In *Arabidopsis*, the low affinity nitrate transporter family, the NPF transporters, is a large gene family with ~53 members (Leran et al., 2014). The expanded gene family of the NPF transporters suggests that there may be more diverse functional roles of these transporters, other than low affinity nitrate transport, that have not yet been elucidated.

With the exception of the NPF8 subfamily, the majority of the NPF families contain members that have been shown to have at least low affinity nitrate transport (Chiang et al., 2004). An in depth analysis of the role NPF transporters play in nitrogen utilisation *in planta* is beyond the scope of this thesis, this discussion will focus on the diversity in substrate specificity that has been elucidated for this transporter class and the effect this specificity has on plant physiology.

4.1.2 Characterisation of NRT1.1/AtNPF6.3

The first described, and the best characterised, NPF transporter to date is NRT1.1/CHL1/AtNPF6.3. This transporter was first identified using a T-DNA tagged mutant of the low affinity uptake mutant *chl1* (Tsay et al., 1993). Chlorate is a nitrate analogue that is toxic to plants due to its conversion to chlorite by nitrate reductase. Mutants in the *NRT1.1* gene are resistant to chlorate and therefore deficient in nitrate uptake (Tsay et al., 1993).

This transporter is primarily localised to the root, and is involved in both low and high affinity supply of nitrate to the plant from its external environment (Huang et al., 1996; Tsay et al., 1993; Wang et al., 1998). Uniquely this transporter displays both high and low affinity to external nitrate concentration when expressed in *Xenopus* oocytes, displaying biphasic kinetics for nitrate uptake (Huang et al., 1996; Liu et al., 1999; Wang et al., 1998). This biphasic kinetics is due to the specific phosphorylation of threonine residue 101 by a kinase CIPK23. Expression of a T101A mutant version of NRT1.1 in *Xenopus* oocytes displayed only low affinity kinetics, whilst expression of T101D, a phosphomimetic exhibits only high affinity nitrate uptake (Tsay et al., 1993). This transporter, as will be discussed, also has substrate specificity for the plant hormone auxin (Krouk et al., 2010), however is not able to transport nitrite, alanine, sulphate, phosphate or dipeptides (Parker and Newstead, 2014). To date this is the only plant NPF transporter whose structure has been solved by X-ray crystallography (Parker and Newstead, 2014; Sun et al., 2014). Crystal structures for both the wild type protein and the T101D phosphomimetic have been generated to assess the structural difference that allow for low affinity versus high affinity transport. As would be predicted the NRT1.1 adopts a classical major facilitator superfamily fold,

with 12 transmembrane α-helices centred on a pseudo two-fold axis between the two N-terminal and C-terminal domains (Parker and Newstead, 2014; Sun et al., 2014). The central substrate binding site is very similar to the peptide transporters present in bacteria, such as the $PepT_{So}$ from *Shewanella oneidensis* (Newstead et al., 2011). It is known that NRT1.1 utilises a proton gradient for the electrogenic transport of nitrate across cellular membranes, and an EXXER motif that has been shown to be integral for coupling proton movement in bacterial $PepT_{SO}$ is conserved in the *Arabidopsis* NRT1.1. This indicates that the mechanism for coupling proton movement is likely to be conserved irrespective of substrate specificity in these two related transporter classes (Parker and Newstead, 2014).

4.1.3 Role of NPF transporters in transport of nitrate

Although NRT1.1 displays dual affinity for nitrate, the majority of NPF transporters that have been characterised to date are solely low affinity nitrate transporters. As mentioned, NRT1.1/ATNPF6.3 is primarily expressed in the roots. The expression of this gene is inducible by the application of nitrate (Tsay et al., 1993). As well as an inducible system for nitrate uptake in the roots, *Arabidopsis* also has a constitutive low affinity nitrate uptake system, encoded by AtNRT1.2/AtNPF4.6 (Huang et al., 1999).

In *Arabidopsis*, the role of the NPF transporters have primarily been shown to be involved in nitrogen mobilisation and assimilation in different plant tissues, with NPFs from different subfamilies, yet with low affinity nitrate activity, preferentially expressed in different tissues.

Upon uptake of nitrate to the root there is a need to redistribute this nutrient in the plant through root to shoot nitrate translocation, which is achieved through loading of nitrate into the xylem. Mutation in the transporter AtNPF7.3 resulted in a decrease in the nitrate content of xylem sap, indicating this transporters role in nitrate xylem loading (Chen et al., 2012; Lin et al., 2008). In contrast, mutations in transporters AtNPF7.2 and AtNPF2.9 result in an increase in the root to shoot translocation of nitrate, indicating that these transporters play a role in retaining nitrate in the root (Li et al., 2010; Wang and Tsay, 2011). The specific localisation of AtNPF7.2 to xylem parenchyma and AtNPF2.9 to the companion cells of root phloem support this hypothesis (Li et al., 2010; Wang and Tsay, 2011).

Leaf specific expression of NPF genes is essential for nitrogen allocation and sequestration. AtNPF2.13 is involved in the remobilization of nitrate from older to younger leaves, and is

expressed in the phloem of older leaf tissue (Fan et al., 2009). The petiole has a unique function in nitrate homeostasis, and is a site of nitrate storage. The low affinity AtNRT1.4/AtNPF6.2 is only expressed in the petiole, and its mutation reduces the nitrate content of the petiole (Chiu et al., 2004) demonstrating the role of this transporter in localised nitrate sequestration.

Similarly, nitrate transport is integral to the correct development of seeds, and the transporter AtNPF2.12 is specifically localised to the vascular bundle of siliques, and is required for correct embryo development (Almagro et al., 2008). Correct nitrate allocation in the embryo may also require the protein AtNPF5.5 (Léran et al., 2015).

As shown, the distinct cellular localisation of these low affinity nitrate transporters directly impacts nitrogen metabolism in the plant, affecting nitrate reallocation that is integral to plant development, as well as in response to environmental fluctuations in nutrient availability.

4.1.4 The NAXT transporter in Arabidopsis

The AtNPF2.7(NAXT) is unique in its role as a nitrate efflux transporter, as all other characterised nitrate transporters are importers (Segonzac et al., 2007). NAXT1/AtNPF2.7 is localised at the plasma membrane, as has been determined by overexpression of a NAXT1-GFP fusion, together with immunoblotting of plasma membrane purified fractions from *Arabidopsis* root, and is primarily expressed in the mature root cortex. RNAi of the NAXT family did not result in a growth phenotype, however the roots from these plants contained 30% higher nitrate content relative to wild type plants (Segonzac et al., 2007). The *naxt1* T-DNA insertion did not display a strong phenotype, however upon acid load, and acidification of the medium in which the plants were grown, the *naxt1* T-DNA insertion plants displayed lower nitrate efflux activity relative to wild type controls. This efflux capacity is thought to be passive, and therefore thermodynamically downhill (Segonzac et al., 2007). Interestingly the NAXT1 protein does not contain the conserved EXXER motif which in bacterial peptide transporters, and also in the *Arabidopsis* NRT1.1, is conserved and thought to be involved in coupling transport to the proton gradient (Segonzac et al., 2007).

4.1.5 Role of NPF transporters in hormonal transport and plant development

Although the low affinity nitrate activity of this transporter class is the best defined, there is a growing appreciation that the NPF transporters may have other substrate specificities, and be involved in hormone transport.

As well as its role in nitrate transport, the NRT1.1/AtNPF6.3 transporter has an integral role in nitrate sensing, and is the first transceptor identified in plants. Nutrient sensing is critical for the development of a proper lateral root network (Krouk et al., 2010). Nitrate is the main nitrogen source for plants, however is also a key signal molecule for development of lateral roots, which occurs under high nitrogen concentration. In particular NRT1.1 is key for the formation of a nitrate regulated root system by stimulating lateral root growth in response to a local supply of nitrate. This regulation of lateral root growth is achieved by nitrate dependent control of auxin accumulation. In three heterologous expression systems, *Xenopus laevis* oocytes, tobacco BY2 cells and yeast, NRT1.1 was shown to transport auxin (Krouk et al., 2010). In a knockout mutant of *NRT1.1* crossed with an *Arabidopsis* line that contains an auxin inducible GUS reporter system, auxin accumulated in lateral root primordia when grown in a range of nitrate concentrations, which was not observed in wild type, where auxin only accumulated in lateral roots at high nitrate concentration. This is indicative that NRT1.1 represses auxin accumulation at low external nitrate concentration, whilst high external nitrate competes with auxin as a substrate for NRT1.1, consequently leading to auxin accumulation in lateral roots (Krouk et al., 2010).

This demonstrates how transport and nutrient sensing are elegantly intertwined with roles in plant development.

As well as controlling local auxin concentrations, the NPF transporter family has been implicated in the transport of other plant hormones, such as abscisic acid.

As has been mentioned previously abscisic acid is a key hormone involved in the response to abiotc and biotic stresses (Hetherington and Woodward, 2003). Through a yeast functional screen for ABA uptake, a number of NPF transporters were identified as being capable of transporting ABA (Kanno et al., 2012). The yeast screen effectively monitored the internal cellular accumulation of ABA through a yeast two hybrid system. PYR/PYL/RCAR ABA receptors interact with PP2C type protein phosphatases in the presence of abscisic acid. This interaction can be assessed by a yeast two hybrid system utilising a HIS3 selection marker. cDNA libraries in this yeast strain can be screened for survival on -HIS media, with the assumption that sequences

that allow the yeast to survive, increase the concentration of ABA inside the cell, allowing for expression of HIS3, and therefore the cDNA clone is likely to encode for an ABA importer (Kanno et al., 2012). Utilising this screening approach four NPF transporters were identified (AIT 1,2,3,4 or AtNPF4.6, AtNPF4.5, AtNPF4.1 and AtNPF4.2) that had ABA import activity, with AIT1 and AIT3 having highest affinity. As should be noted AIT1/AtNPF4.6 is also the previously identified constitutive low affinity nitrate transporter present in roots. The transport activity was also tested independently by LC-MS and detection of ABA inside the yeast. AIT1 was specific for ABA and could not transport the other plant hormones gibberellin, auxin or jasmonic acid, whilst AIT3 also had gibberellin uptake activity (Kanno et al., 2012). This uptake activity was also observed when the proteins were expressed in *Spodoptera frugiperda* (Sf9) insect cells. A GFP fusion of AIT1 preferentially localised to the plasma membrane, and loss of function of this protein resulted in a decrease in temperature of inflorescence stems, and more open stomata, suggesting a physiologically relevant role for the NPF transporter in ABA transport in planta (Kanno et al., 2012). As has been described previously for NRT1.1/CHL1/AtNPF6.3, the competition of transport between nitrate and auxin allows for NRT1.1 to act as a sensor that is crucial for correct plant development. In comparison, although interactions between ABA signals and nitrate have been speculated, nitrate does not compete with ABA as a substrate for AtNPF4.6 (Kanno et al., 2013).

In *Medicago truncatula* abscisic acid is also transported by members of the NPF family, through a member of the NPF6 subfamily, annotated as MtNPF6.8 (Pellizzaro et al., 2014). Overexpression in *N. benthamiana* as a C-terminal GFP fusion resulted in localisation to the plasma membrane, with this gene expressed in both roots and shoots. The RNAi lines of *MtNPF6.8* did not display any changes in the nitrogen status of the root, however, the knockdown lines were no longer sensitive to nitrate dependent inhibition of primary root length (Pellizzaro et al., 2014). Supplementation with exogenous ABA could restore this inhibition, and expression of MtNPF6.8 in *Xenopus laevis* oocytes demonstrated ABA uptake activity. This is indicative that MtNPF6.8 is a dual affinity transporter for both nitrate and ABA, and is likely to be involved in ABA dependent nutrient sensing in plant development (Pellizzaro et al., 2014).

Similarly another *Medicago truncatula* NPF transporter, MtNPF1.7 has also been implicated in root architecture, with defects in lateral root organ development in *npf1.7* mutants, whose phenotype could be rescued by exogenous application of abscisic acid (Yendrek et al., 2010). This demonstrates that multiple NPF transporter subfamilies may be co-ordinately involved in plant development, and hormone transport.

4.1.6 Alternative substrate specificities for NPF transporters

In *Arabidopsis* the NPF subfamily 8 is unique in that it preferentially transports di- and tri-peptides, like the human SLC15 transporters, and does not transport nitrate (Chiang et al., 2004). The first amino acid and peptide transporter belonging to the NPF 8 class was identified by a heterologous complementation approach in yeast mutants defective for histidine biosynthesis (Frommer et al., 1994). NTR1/AtNPF8.3 was identified as a histidine transporter, however has lower affinity for this amino acid in comparison to amino acid permeases. Furthermore NTR1/AtNPF8.3 was able to complement a yeast strain deficient in oligopeptide transport, and demonstrated high affinity, yet broad specificity, for di- and tri-peptide substrates (Frommer et al., 1994; Rentsch et al., 1995). AtNPF8.1 (AtPTR1), AtNPF8.2 (AtPTR5) and AtNPF8.3 (AtPTR2/NTR1) have also been shown to be specific transporters for di- and tri-peptides through complementation in yeast. AtNPF8.1 and AtNPF8.2 proteins are localised to the plasma membrane in plants, whilst AtNPF8.3 is localised to the tonoplast (Dietrich et al., 2004; Komarova et al., 2008).

Two members of the NPF8 family in *Arabidopsis* AtNPF8.4 (AtPTR4) and AtNPF8.5 (AtPTR6) have been shown to localize to the tonoplast membrane as opposed to the plasma membrane *in planta*, however their functional characterization has not been achieved in *Xenopus laevis* oocytes or in yeast (Weichert et al., 2012) and therefore the function of these transporters remains unclear.

As well as the NPF8 subfamily, one member of the NPF5 subfamily from *Arabidopsis*, AtNPF5.2 (PTR3) has been characterised as a dipeptide transporter through complementation in a yeast mutant defective in oligopeptide transport, and is upregulated in response to salicylic acid (Karim et al., 2007). In comparison, other members of the AtNPF5 family transport nitrate with no affinity for peptides (Léran et al., 2015). This again demonstrates that sequence homology is not necessarily a reliable predictor of substrate specificity (Karim et al., 2007).

NPF transporters have also been implicated in organic acid transport. Alder is a plant species that is able to form a symbiotic relationship with the actinomycete *Frankia*. A nodule specific plant mRNA encoding a NPF transporter was identified, the protein product of which localises to the plasma membrane at the interface between the plant cell and the bacteria. Heterologous expression of the transporter in *Xenopus laevis* oocytes, yeast and *E. coli* demonstrated that this

transporter is specific for dicarboxylates, such as malate and succinate, and is likely to be involved in carbon supply to the bacterium (Jeong et al., 2004).

4.1.7 Role of NPF transporters in secondary metabolism.

The NPF transporter family has been implicated in the transport of secondary metabolites in *Arabidopsis*. The major defence compounds in *Arabidopsis* are a class of secondary metabolites known as glucosinolates (Nour-Eldin et al., 2012). These compounds are present ubiquitously in the plant, however upon senescence, there is a translocation of glucosinolates from leaf to seeds, as has been determined by monitoring ^{14}C-labelled glucosinolates *in planta*. Two NPF transporters, GTR1 and GTR2 (AtNPF2.10 and AtNPF2.11), were identified in a screen of *Arabidopsis* transporters to have uptake activity in *Xenopus* oocytes for the model glucosinolate substrate 4-methylthiobutyl glucosinolate (4MTB) (Nour-Eldin et al., 2012). Biophysical characterisation of these transporters using two voltage electrode clamp in *Xenopus laevis* oocytes demonstrated that transport was dependent on the proton gradient, and that these transporters are capable of utilising a variety of glucosinolates as substrates, however there was no activity towards di- and tri- peptides. The *gtr1/gtr2* double *Arabidopsis* mutant abolishes glucosinolate accumulation in the seeds, with a concomitant increase in methionine derived glucosinolates in leaf tissue, indicating a key role of these transporters in source to sink translocation. Complementation with *GTR1* and *GTR2* resulted in the re-establishment of seed glucosinolate content (Nour-Eldin et al., 2012). Both GTR1 and GTR2 localised to the plasma membrane, and whilst GTR1 is primarily associated with the mesophyll, GTR2 is localised to the vasculature of the plant, indicating a potential role in apoplasmic phloem loading of glucosinolates (Nour-Eldin et al., 2012).

4.1.8 Summary to Introduction

As has been demonstrated, the distinct subcellular and intercellular localisation of NPF transporters, which to date have been primarily shown to be low affinity nitrate transporters, is crucial for both nutrient sensing, and nitrogen allocation throughout the plant, and required for correct plant development.

The recent discovery that the NPF class of transporters is involved in the transport of multiple substrates in *Arabidopsis*, including auxin, glucosinolates, abscisic acid, dicarboxylates, di- and tri-peptides, as well as low affinity nitrate transport, together with the fact that there are ~53 members in the *Arabidopsis* genome, suggests this transporter class may have as of yet

unidentified substrate specificities, and could be involved in the transport of a wide variety of other compounds. As such it would be appealing to assess whether these transporters have any physiological role in MIA transport in *C. roseus*. This would be the first identification that the NPF transporters play a physiological role in secondary specialised metabolism, outside of *Arabidopsis.*

To date, many approaches for the elucidation of NPF transporter function have centred on the *in vitro* characterisation of the transporter, through testing the transporters ability to complement defective yeast strains in oligopeptide or amino acid transport, or overexpression in *Xenopus* oocytes and testing transport on a range of substrates. I would argue to further assess the functionality of these transporters, it is necessary to test these transporters on a wider substrate repertoire, as was demonstrated in the elucidation of glucosinolate transporters (Nour-Eldin et al., 2012) or utilise a strategy whereby knocking down the transporter in a given tissue can be used to screen for a metabolic phenotype that could be indicative of substrate transport, as will be demonstrated in this chapter.

4.2 Aims of this Chapter

1.) To assess whether there are any NPF transport proteins that are involved in MIA metabolism in *C. roseus* using VIGS

2.) Functionally characterise the NPF transporters that display a metabolic phenotype after transient silencing *in planta*.

4.3 Results and Discussion

4.3.1 NPF transporters in C. roseus

There are ~46 genes in *C. roseus* that are annotated as NTR/PTR family members in the transcriptomic databases of *C. roseus* from the MPGR *C. roseus* transcriptome. This is in good agreement with the number found in the Cathacyc transcriptomic resource (Van Moerkercke et al., 2013).

The focus of this chapter will centre on the NPF2 subfamily of transporters present in *C. roseus*. The reason for focusing on this subfamily is two-fold. Firstly the NPF2 subfamily has been demonstrated to be involved in secondary metabolism in *Arabidopsis*, in transporting glucosinolates (Nour-Eldin et al., 2012), and secondly, as will be demonstrated, a homologue of the NPF2 family transcriptionally co-regulates with the biosynthetic genes in the MIA pathway.

There are 10 homologues of the NPF2 subfamily in *C. roseus* that are present in the transcriptome for which full length sequences could be obtained. These homologues are annotated as CrNPF2.1-CrNPF2.10.

The functionally characterised members of the NPF2 subfamily in *Arabidopsis* include the two glucosinolate transporters GTR1 and GTR2 (AtNPF2.10 and AtNPF2.11) (Nour-Eldin et al., 2012), a low affinity nitrate transporter involved in phloem loading for root to shoot nitrate translocation, NRT1.9 (AtNPF2.9) (Wang and Tsay, 2011) and in the reallocation of nitrate from old to young leaves, NRT1.7 (AtNPF2.13) (Fan et al., 2009), a low affinity nitrate transporter for nitrate transport into the developing embryo, NRT1.6 (AtNPF2.12) (Almagro et al., 2008), as well as the NAXT (AtNPF2.7) nitrate efflux transporter (Segonzac et al., 2007).

A phylogenetic analysis of the *C. roseus* homologues (Figure 4-1) demonstrated, as would be expected, the candidates CrNPF2.1-CrNPF2.8 are most similar to the AtNPF2 family members, forming a clade. In comparison, CrNPF2.9 and CrNPF2.10 are in turn most similar to the *Arabidopsis* NAXT/AtNPF2.7 protein, whilst the other characterised proteins from *Arabidopsis* that are present in other subfamilies, NRT1.1 (AtNPF6.3), NRT1.2 (AtNPF4.6), as well as the di- and tri- peptide transporters, AtNPF8.1-AtNPF8.5, form a distinct clade (Figure 4-1).

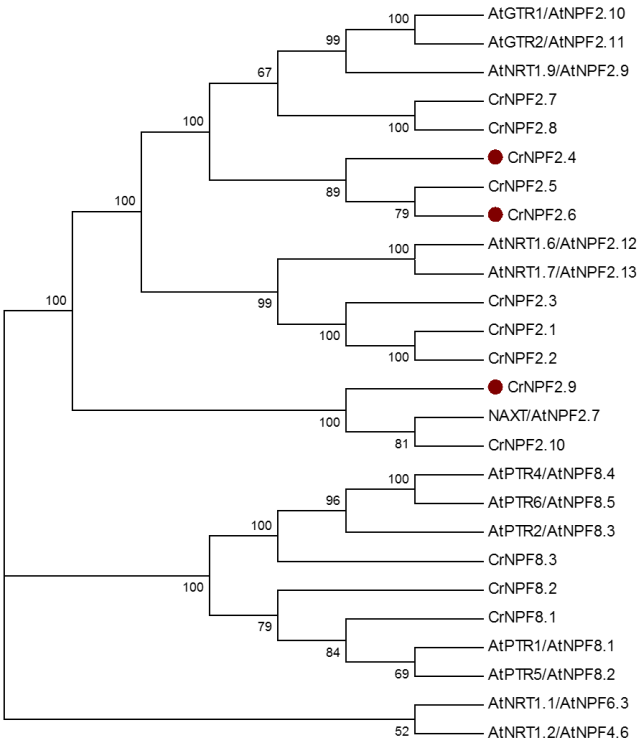


Figure 4-1 Phylogenetic tree of *C. roseus* NPF2 family homologues.
The CrNPF2 homologues are highlighted in this phylogenetic tree together with the AtNPF2 family members, AtNPF2.7, AtNPF2.9, AtNPF2.10, AtNPF2.11, AtNPF2.12, AtNPF2.13, as well as the AtNPF family members AtNPF4.6, AtNPF6.3 and AtNPF8.1-AtNPF8.5. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model, and the bootstrap consensus tree was inferred from 500 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed.

4.3.2 Transcriptomic profiling and candidate selection

Hierarchical clustering and self-organising maps (Chapter II) were utilised to identify candidate transport proteins and enzymes that were co-regulated with the monoterpene indole alkaloid biosynthetic pathway. Together with the two CrMATES that have been described in Chapter III, the other transport protein that was tightly co-expressed with the pathway was the candidate CrNPF2.9, one of the CrNPF2 subfamily members that is closest to the NAXT homologue in *Arabidopsis.* A representative expression profile of the CrNPF2.9 transcript, together with the three biosynthetic genes in MIA metabolism, secologanin synthase, tryptophan synthase and strictosidine synthase, as well as the two MATE proteins identified, is shown in Figure 4-2. CrNPF2.9 is the fourth highest expressed NPF transporter in leaf tissue suggesting that as well

as its strong co-regulation with genes in the MIA pathway, it is also highly expressed in leaf tissue, the major site of MIA metabolism in *C. roseus.*

The full length candidate gene of CrNPF2.9 was cloned from *C. roseus* leaf cDNA into a pCR8GW entry vector and sequenced, confirming the correct assembly of this gene in the transcriptome.

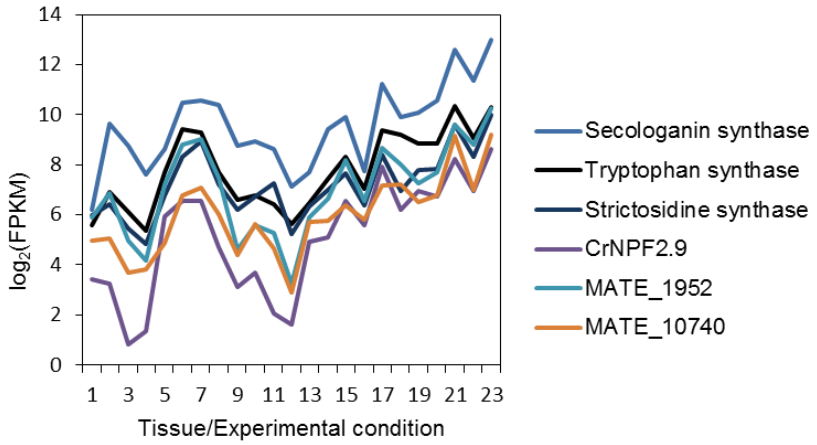


Figure 4-2 Co-regulation of CrNPF2.9 with known genes in the biosynthetic pathway.
Expression profile of CrNPF2.9 measured as the log_2(FPKM) in the 23 different experimental conditions highlighted in Table1. The expression profile of the known co-regulated genes that are involved in MIA biosynthesis; secologanin synthase, tryptophan synthase and strictosidine synthase as well as the two CrMATE1952 and CrMATE10740 are also shown for comparison.

4.3.3 Localisation of CrNPF2.9 to the tonoplast membrane

4.3.3.1 Prediction of tonoplast localisation

CrNPF2.9 is predicted to localise to the tonoplast membrane. The CrNPF2.9 contains an acidic di-leucine motif at its N-terminus, (D/E)X_{3-5}L(L/I), which is known to be involved in the subcellular localisation of NPF transporter proteins to the tonoplast membrane (Komarova et al., 2012). This paper demonstrated that AtPTR2, AtPTR4 and AtPTR6 are localised to the tonoplast whilst AtPTR1 and AtPTR5 are localised to the plasma membrane, and that this acidic di-leucine motif is critical for the correct localisation of AtPTR2, 4 and 6 to the tonoplast membrane. The CrNPF2.9 transporter is the only CrNPF2 subfamily member to contain this motif at its N-terminus (Figure 4-3).

Gene	N-terminal sequence 40 aa
CrNPF2.1	MELEENKMKKQLSSTWFLCCSKCLPVNSSSSSSKVSSPNS
CrNPF2.2	MDSKKTSSSWVSCCTKCLPTKPSNSKQNNGRHDEENQDMV
CrNPF2.3	MESKNEHQMGKSSPWFNICSTKCLPINSSSKVSSPTSSSM
CrNPF2.4	MAEKSSSIDATENGFVKIKEPNHRGVKAMPFIVGNETFEK
CrNPF2.5	MEENQEHKAIKNHEPNYRGIKAMPFVIGNETFEKLGTIGT
CrNPF2.6	MEKNEKLVPNGTKEDGKVAAKNEPNYRGIKAMPFVLGNET
CrNPF2.7	MEKKETMDTKPKEEEEEEPQINYRGIKAMPYIIGNETFEK
CrNPF2.8	MEKNQQKEGAMEEEPVLDVQEKPQIINYKGIKAMPYIIGN
CrNPF2.9	MGDTEAQLLQPGHKQKGGWITFPFILATRTLLTLAVAGFS
CrNPF2.10	MDDAENPKSSSASTKQGGWITFPFIIGTMAGMSLAAGGWV

Figure 4-3 The first 40 amino acids for the CrNPF2 family.
Highlighted in yellow is the (D/E)X_{3-5}L(L/I) present in CrNPF2.9 which does not occur in the other CrNPF2 family members

Additionally it has been reported that vacuolar localised proteins often contain more acidic residues throughout the protein sequence, and therefore have a considerably lower isoelectric point (pI) (Carter et al., 2004).

For membrane proteins localised to the tonoplast there have also been reports that these proteins have a more acidic pI. This has been best described for the plant aquaporin family. The plasma membrane, PIP subfamily, of aquaporins have an isolectric point of ~9 whilst the tonoplast localised aquaporins, TIP subfamily, have an isoelectric point of ~6.0 (Gomes et al., 2009).

Similarly this phenomenon has also been observed for the NPF transporters. The tonoplast localised AtPTR2, 4, 6 have a pI in the range of 5.34-6.57, whilst the NPF transporters localised to the plasma membrane, AtPTR1 and 5, have a pI of 8.3 and 7.87 respectively. Interestingly CrNPF2.9 has a considerably lower pI than the other 9 CrNPF2 transporters in *C. roseus,* with a predicted pI of 6.1, whilst the other CrNPF2 transporters have a pI in the range of 7.53-9.25 (Table 6).

Table 6 Isoelectric points for the CrNPF2 subfamily.
The isoelectric point (pI) of the CrNPF2 subfamily members was performed utilising the ExPASY isoelectric point calculator

Gene Name	Isoelectric point (pI)
CrNPF2.1	7.53
CrNPF2.2	8.74
CrNPF2.3	8.76
CrNPF2.4	9.25
CrNPF2.5	8.95
CrNPF2.6	9.16
CrNPF2.7	9.1
CrNPF2.8	8.81
CrNPF2.9	6.1
CrNPF2.10	9.18

4.3.3.2 NPF transporter localisation and overexpression in N. benthamiana

Overexpression of the CrNPF2.9 gene in *C. roseus* cell suspension culture as a C-terminal YFP fusion resulted in its localisation to the tonoplast membrane, and is consistent with our prediction of tonoplast localisation. This protein co-localised with the tonoplastic marker protein, a potassium channel from *Arabidopsis*, AtTPK1 (Figure 4-4).

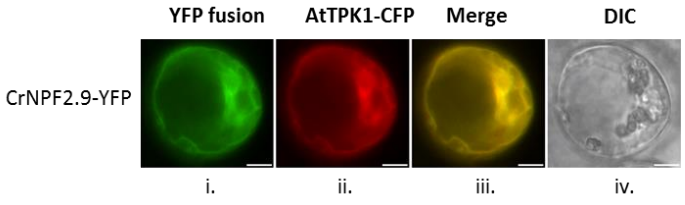


Figure 4-4 Overexpression of CrNPF2.9-YFP in *C. roseus* cell suspension cultures
Co-localisation of CrNPF2.9 with the *Arabidopsi*s potassium channel TPK1 i.) CrNPF2.9-YFP ii.) *Arabidopsis* TPK-CFP iii.) Merge iv.) Differential Interference Contrast. Scale bar = 10 µm *Cloning of the C. roseus CrNPF2.9 gene and Arabidopsis TPK1 gene into pEAQ-HT DEST1 was performed by R.Payne, subcloning and transformation by particle bombardment into C. roseus cell suspension cultures and visualisation by confocal microscopy was performed by Vincent Courdavault, University of Tours.*

Additionally, overexpression of a CrNPF2.9-GFP fusion using the pEAQ-HT system and agroinfiltration in *N. benthamiana* resulted in the localisation to a membrane in *N. benthamiana* epidermal cells, which is likely to be the tonoplast due to the presence of trans-vacuolar strands (Figure 4-5).

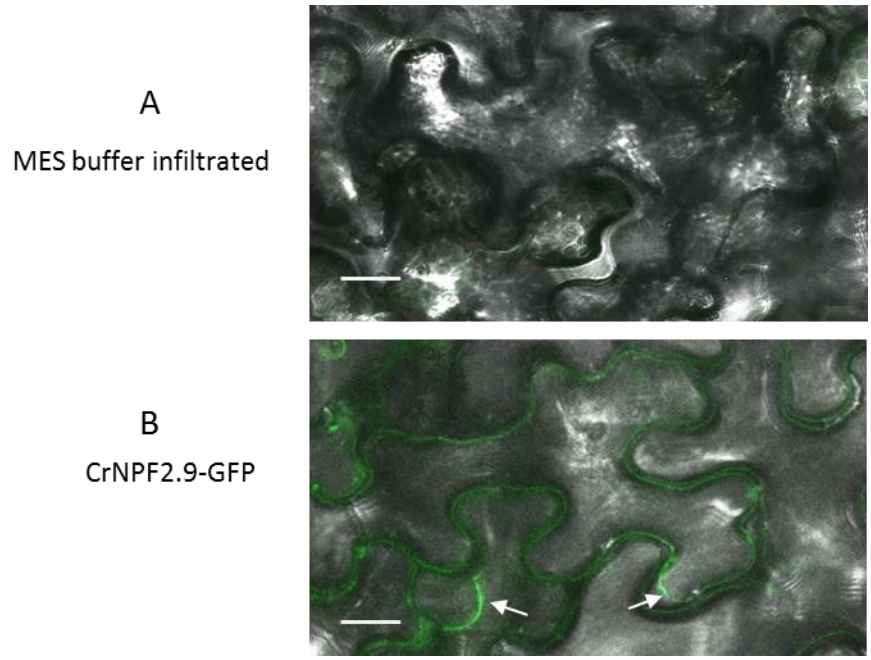


Figure 4-5 Expression of CrMATE1952-GFP in *Nicotiana benthamiana* epidermal cells.
Overexpression of CrNPF2.9-GFP using *Agrobacterium tumefaciens* infiltration of *N. benthamiana* in the pEAQ-HTDEST1 expression vector and visualisation by confocal microscopy. A.) MES buffer infiltrated *N. benthamiana* leaf tissue B.) Overexpression pf CrNPF2.9-GFP using the pEAQ-HTDEST1 expression vector. White arrows indicate the localisation of CrNPF2.9-GFP to trans-vacuolar strands in *N. benthamania* epidermal cells. Scale bar = 20 µm.

4.3.4 VIGS CrNPF2.9

To assess the physiological relevance of CrNPF2.9 in *C. roseus,* a VIGS pTRV2u silencing vector containing ~500 bp of the gene for the NPF-transporter CrNPF2.9 was generated. As can be seen from the phylogenetic tree, the closest homologue of CrNPF2.9 in *C. roseus* is the protein CrNPF2.10.

Although this transporter is the closest homologue to CrNPF2.9, it has less than 50% sequence identity at the nucleotide level and there is no region that has a contiguous sequence of more than 20 bp that would allow for cross-silencing in the VIGS experiment. Similarly a BLAST search of the ~500 bp region of the gene used for VIGS against the *C. roseus* transcriptome did not

result in the identification of other genes with a contiguous stretch of 20 bp indicating a low chance of cross-silencing.

VIGS of this transporter led to a significant metabolic phenotype, resulting in the ~20 fold greater accumulation of the intermediate strictosidine compared to empty vector controls. A significant decrease in the downstream alkaloids, vindoline and catharanthine, was also observed (Figure 4-6). As will be discussed in section 4.3.4.4, silencing also led to a physiological blackening of the leaf tissue.

The VIGS experiment has been independently replicated multiple times, and each time the same metabolic phenotype was observed. Three independent replicates of this silencing experiment are presented in Figure 4-7.

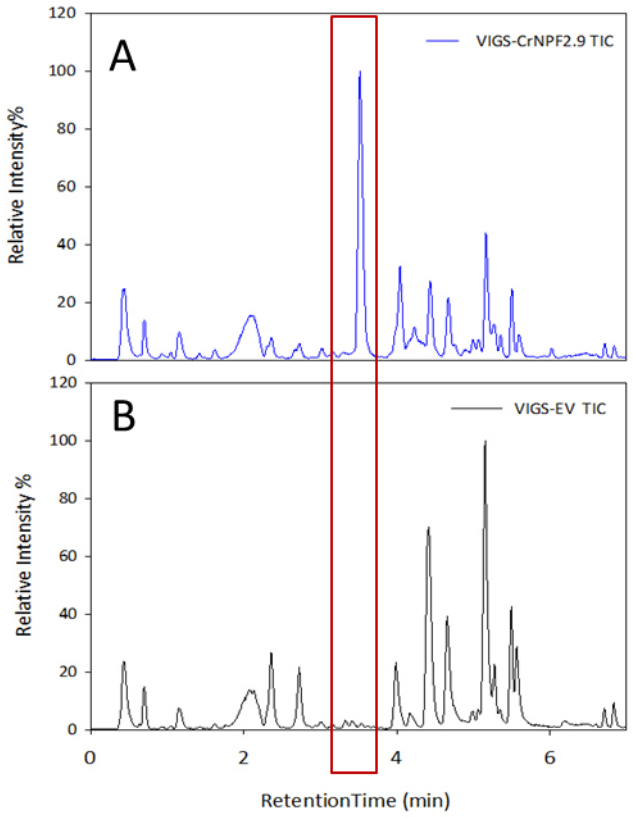


Figure 4-6 Representative Total Ion Chromatogram for VIGS-EV and VIGS-CrNPF2.9
Methanol extracts of leaf tissue from A.) VIGS-CrNPF2.9 and B.) VIGS-EV were analysed by LC-MS. Presented are the total ion chromatograms. Highlighted in the red box is the peak corresponding to strictosidine.

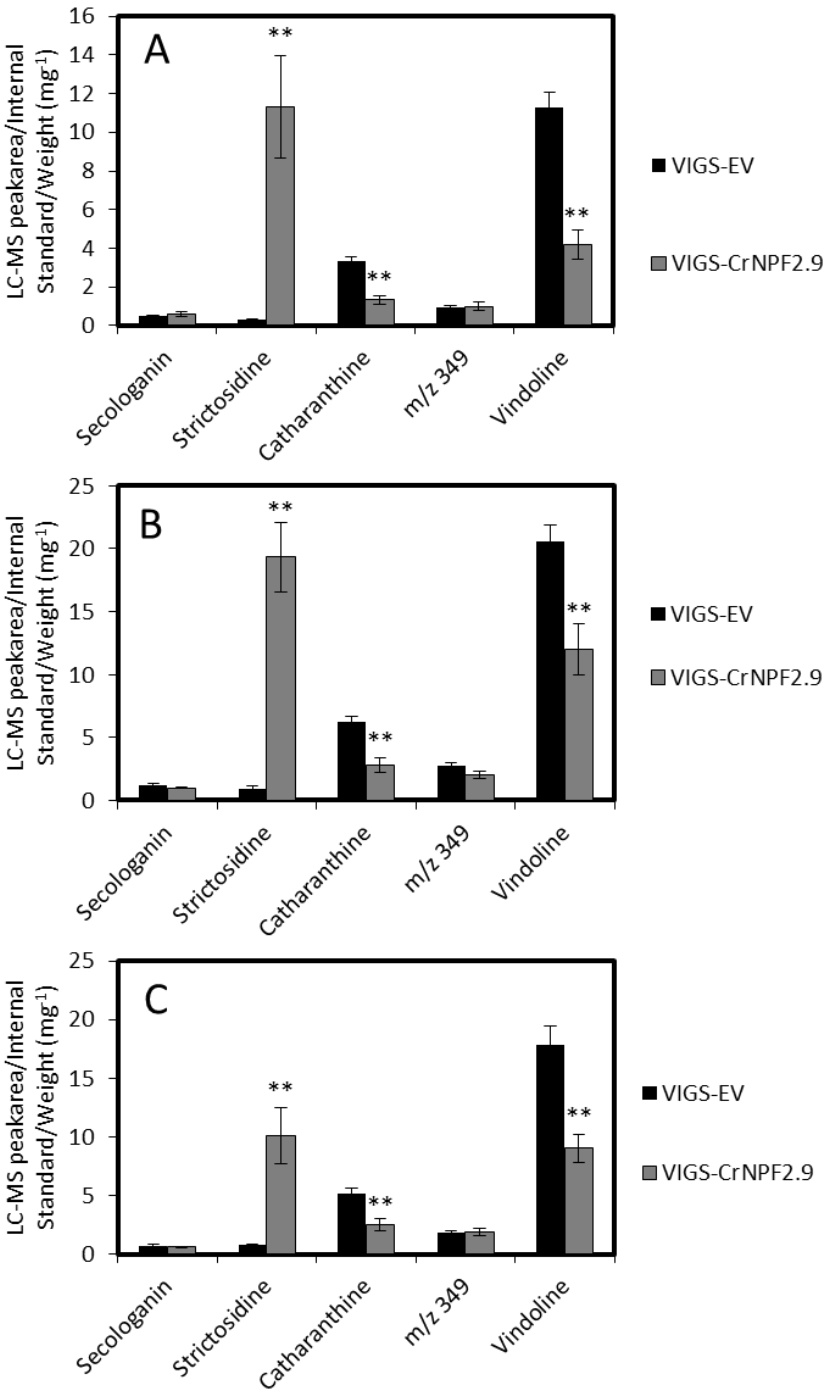


Figure 4-7 Metabolite profile upon VIGS of CrNPF2.9 from three independent VIGS experiments. **A.)** Alkaloid profile for VIGS of CrNPF2.9 relative to empty vector control tissue in Little Bright Eyes VIGS CrNPF2.9 (n=8) VIGS-EV (n=14) ** p<0.01 **B.)** Alkaloid profile for VIGS of CrNPF2.9 relative to empty vector control tissue in Little Bright Eyes VIGS CrNPF2.9 (n=12) VIGS-EV (n=10) ** p<0.01 **C.)**Alkaloid profile for VIGS of CrNPF2.9 relative to empty vector control tissue in Little Bright Eyes VIGS CrNPF2.9 (n=8) VIGS-EV (n=12) ** p<0.01. All data shown is mean ± SEM.

4.3.4.1 Proposed role of CrNPF2.9 in MIA metabolism

The metabolic phenotype observed upon silencing the CrNPF2.9 gene is consistent with the CrNPF2.9 protein being responsible for the transport of strictosidine out of the vacuole. As has been mentioned in Chapter I, strictosidine synthase is localised to the vacuole, whilst the next enzyme in the pathway, strictosidine β-glucosidase, is localised to the nucleus (Guirimand et al., 2010). This therefore requires the transport of strictosidine out of the vacuole for subsequent downstream processing to other MIAs.

The fact that CrNPF2.9 is localised to the tonoplast membrane, and when silenced leads to the accumulation of strictosidine, together with the concomitant decrease in downstream MIAs, vindoline and catharanthine, is indicative that this protein is likely to be the strictosidine exporter from the vacuole.

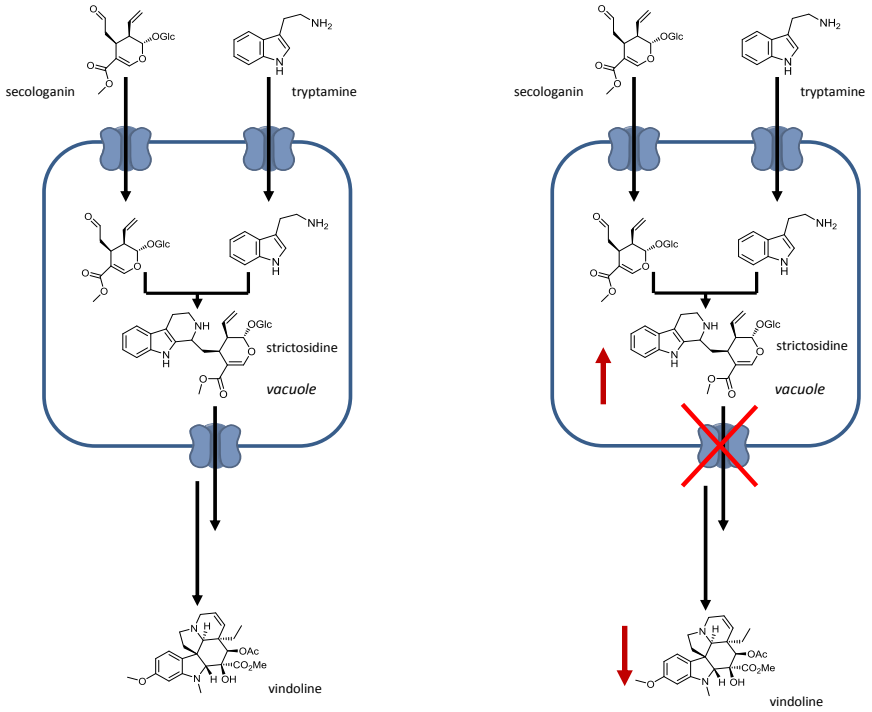


Figure 4-8 Strictosidine is exported from the vacuole after synthesis
A.) Strictosidine is synthesised in the vacuole by strictosidine synthase and exported for production of downstream alkaloids such as vindoline. B.) Silencing of CrNPF2.9 resulted in the accumulation of strictosidine and the decrease in vindoline (highlighted by a red arrow), which is consistent with it being the strictosidine exporter from the vacuole.

This is the first *in vivo* indication that the NPF transporter family is involved in MIA metabolism and is physiologically relevant *in planta*. This is also the first time, outside of *Arabidopsis* that this transporter family has been shown to be involved in secondary metabolism.

4.3.4.2 Untargeted metabolomics VIGS-CrNPF2.9 leaf tissue

As well as the accumulation of strictosidine, a compound with *m/z* 499 also accumulated upon CrNPF2.9 silencing. This compound is present at a low level in the VIGS-EV tissue and this phenotype is reproducible across all VIGS-CrNPF2.9 experiments (Figure 4-9).

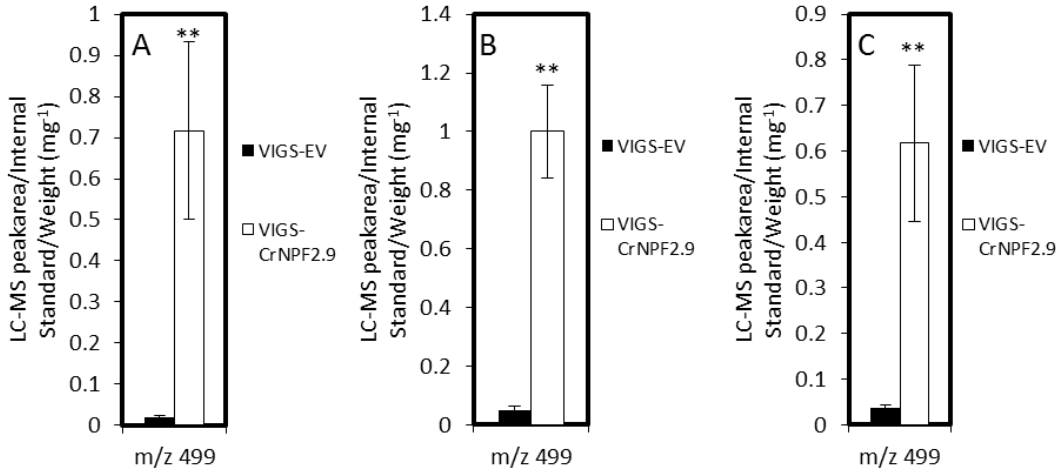


Figure 4-9 Accumulation of m/z 499 upon CrNPF2.9 silencing
This data is from the same three independent VIGS experiments presented in Figure 4-7 A.) *m/z* 499 profile for VIGS of CrNPF2.9 relative to empty vector control tissue in Little Bright Eyes VIGS CrNPF2.9 (n=8) VIGS-EV (n=14) ** p<0.01 B.) *m/z* 499 profile for VIGS of CrNPF2.9 relative to empty vector control tissue in Little Bright Eyes VIGS CrNPF2.9 (n=12) VIGS-EV (n=10) ** p<0.01 C.) *m/z* 499 profile for VIGS of CrNPF2.9 relative to empty vector control tissue in Little Bright Eyes VIGS CrNPF2.9 (n=8) VIGS-EV (n=12) ** p<0.01. All data is mean $\pm$ SEM.

The identity of this compound is unknown, however the *m/z* value is consistent with a derivative of strictosidine, strictosamide. Additionally, the MS^2 fragmentation pattern of this compound is also consistent with this compound (Figure 4-10).

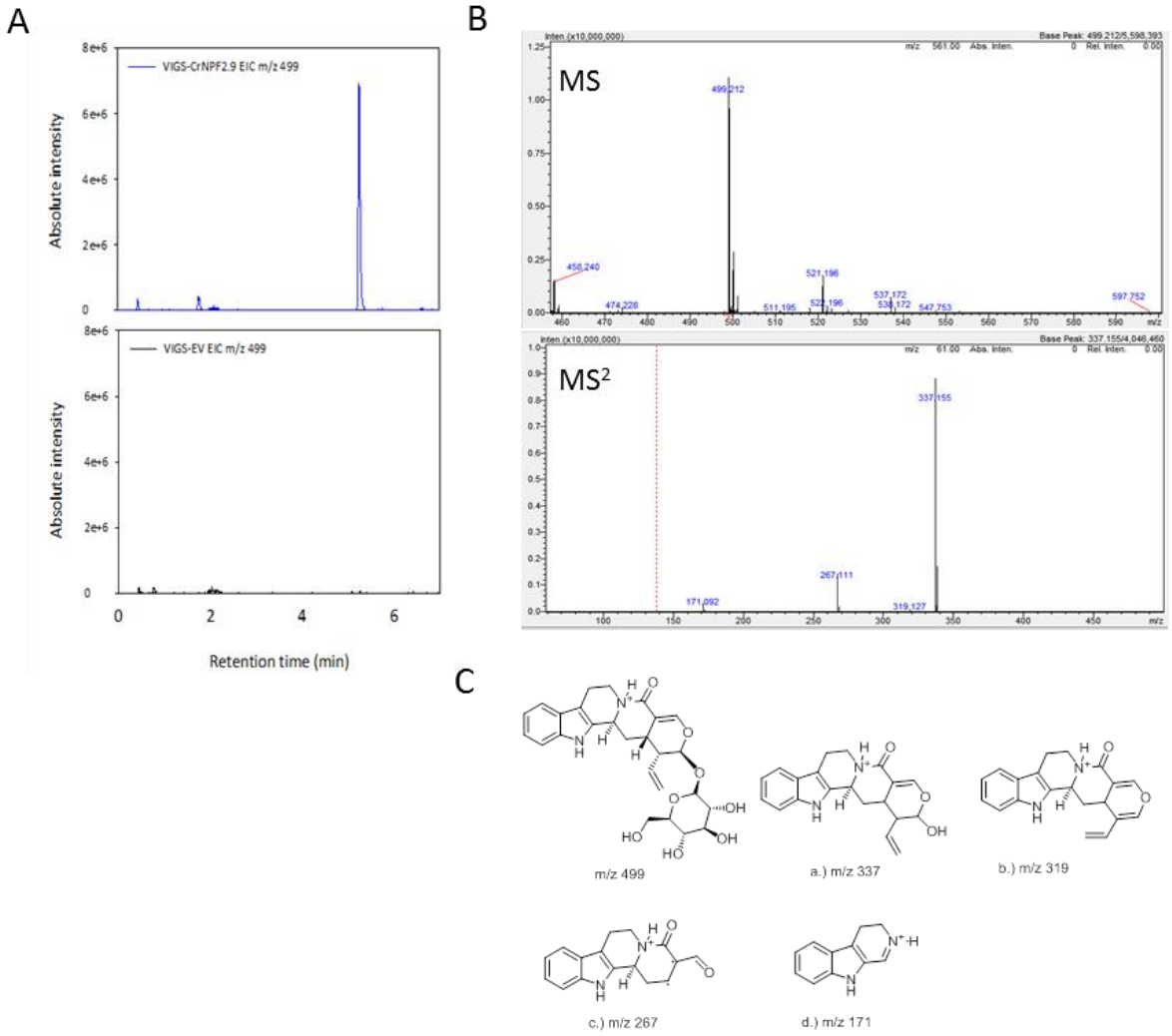


Figure 4-10 Analysis of *m/z* 499 accumulating in VIGS-CrNPF2.9 leaf tissue.
A.) Representative extracted ion chromatograms for *m/z* 499 in methanolic extracts of VIGS-CrNPF2.9 (blue) and VIGS-EV (black) leaf tissue analysed by LC-MS. B.) MS and MS^2 spectra of *m/z* 499 accumulating in VIGS-CrNPF2.9 leaf tissue C.) Proposed structures corresponding to the fragmentations in the MS^2 spectra.

The accumulation of strictosamide would be consistent with the hypothesis that the accumulation of strictosidine in the vacuole could subsequently lead to its derivatisation and intramolecular cyclisation, which is known to happen under non-enzymatic conditions. Alternatively, it is also feasible that the accumulation of this compound is an artefact of the extraction procedure of VIGS-CrNPF2.9 tissue and this cyclisation is promoted due to the higher concentration of strictosidine in the sample.

4.3.4.3 VIGS of CrNPF2.9 stem tissue

Metabolite analysis of the stem tissue from *C. roseus* NPF silenced plants taken from around the wound site from VIGS pinching, indicated that the silencing also affects the stem alkaloid pathway. Unlike in leaf tissue, the dominant alkaloid in stem is the peak with *m/z* 349, which represents a mixture of serpentine and alstonine, as opposed to vindoline *m/z* 457 (Figure 4-11).

As would be expected, in CrNPF2.9 silenced plants there is an increase in the amount of strictosidine present in stem tissue. This indicates biosynthesis of MIAs is not leaf specific, and that this metabolic pathway is active in stem tissue. This also indicates that the CrNPF2.9 transporter is also likely to be involved in MIA metabolism in multiple tissues.

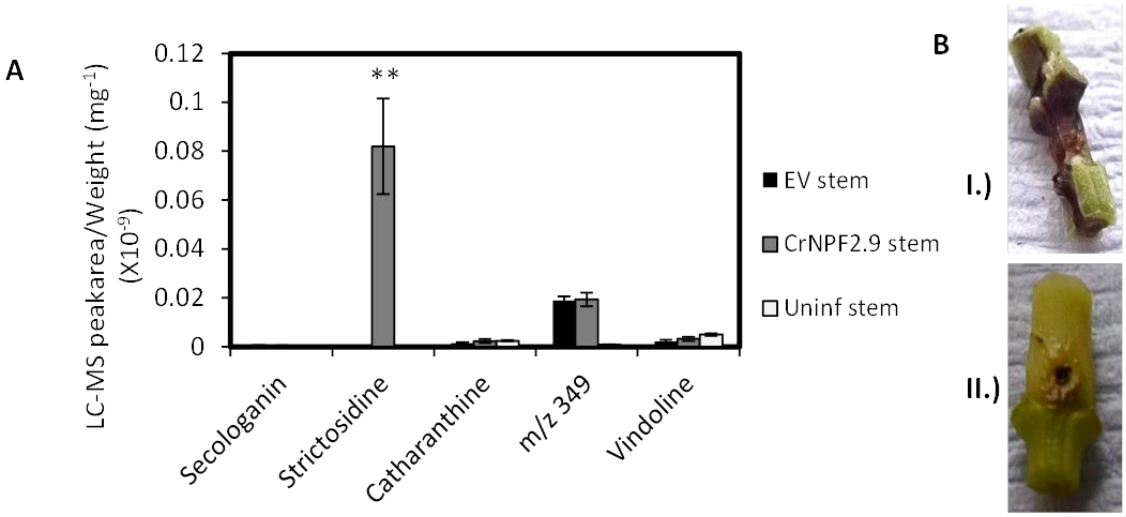


Figure 4-11 Metabolite analysis of stem tissue upon CrNPF2.9 silencing
A.) Alkaloid profile of stem tissue taken from around VIGS wound site for empty vector control (n=4) and CrNPF2.9 silenced plants (n=4) and uninfected stem (n=4) ** $p < 0.01$. All data is mean $\pm$ SEM. B.) Representative photos of stem tissue from I.) VIGS-CrNPF2.9 plants II.) VIGS-EV stem

4.3.4.4 Blackening leaf tissue upon CrNPF2.9 VIGS

The metabolic phenotype observed upon CrNPF2.9 silencing is accompanied by a blackening of the leaf and stem tissue (Figure 4-11, Figure 4-12). This observed phenotype is likely to be caused as a result of cell death, which we think is due to the high levels of strictosidine accumulating in the leaf tissue, which may be toxic.

Figure 4-12 Blackening of VIGS-CrNPF2.9 leaf tissue.
Silencing CrNPF2.9 results in blackening leaves. A.) Three leaves from three separate VIGS-EV plants B.) Three leaves from three separate VIGS-CrNPF2.9 plants displaying the black phenotype. Highlighted by white arrows are the black regions.

Staining the black leaf with 4',6-diamidino-2-phenylindole (DAPI) and propidium iodide and analysis by confocal microscopy indicated that the black tissue is necrotic tissue (Sikorskaite et al., 2013; Truernit and Haseloff, 2008). Propidium iodide staining of the leaf tissue stains the extracellular spaces and the plant cell wall, as is seen in the VIGS-EV control leaves (Figure 4-13). The uptake of the propidium iodide stain internally is indicative of cell death, as is seen with the black leaf upon silencing of the CrNPF2.9 transporter (Figure 4-13). Conversely DAPI staining specifically stains nuclei and double stranded DNA. As is shown for the VIGS-EV control tissue, DAPI staining results in staining of plant nuclei, whilst in the black leaf tissue from CrNPF2.9 silenced plants the dye is diffuse, and does not stain nuclear structures, again indicating that the cells in the blackened area have undergone cell death (Figure 4-13).

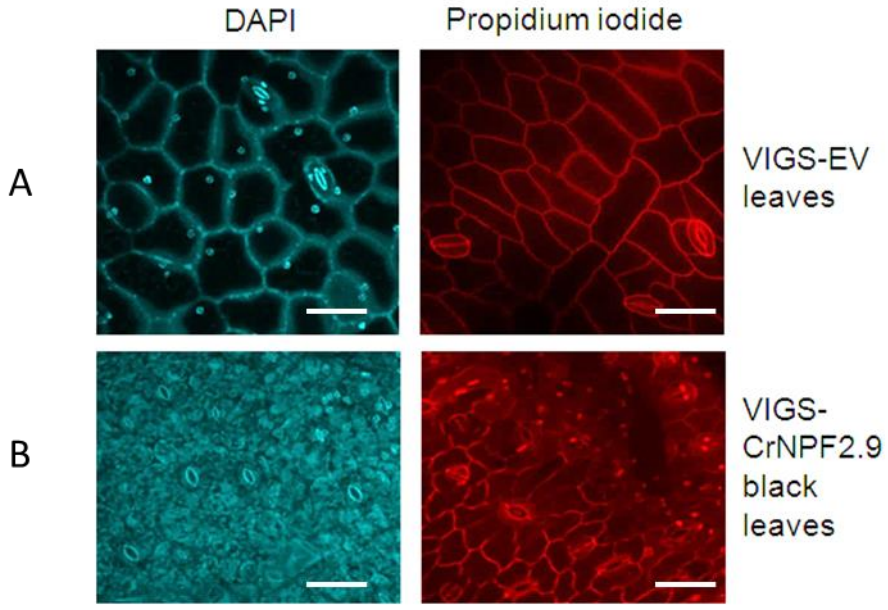


Figure 4-13 Confocal microscopy of VIGS-EV leaf tissue and VIGS-CrNPF2.9 leaf tissue displaying the black phenotype
A.) DAPI and propidium iodide staining of VIGS-EV leaf tissue B.) DAPI and propidium iodide staining of VIGS-CrNPF2.9 leaf tissue. Scale bar = 30 µm.

4.3.4.5 Trypan blue staining VIGS-CrNPF2.9 black leaf tissue

There are a number of potential reasons for the cause of this cell death. Firstly, the accumulation of strictosidine could be toxic to leaf tissue. Secondly, depletion of secondary metabolites upon CrNPF2.9 silencing may make the plant more susceptible to disease. To assess whether the black leaf tissue was due to the presence of a fungal pathogen, leaves from EV *C. roseus* and CrNPF2.9 silenced black leaf tissue were stained with trypan blue, in order to detect fungal hyphae. As can be seen in Figure 4-14 trypan blue staining did not lead to the observation of fungal hyphae in CrNPF2.9 silenced or EV leaf tissue. However, interestingly visualisation of this cleared tissue under UV light resulted in yellow fluorescence which appears to be centred on the regions undergoing cell death.

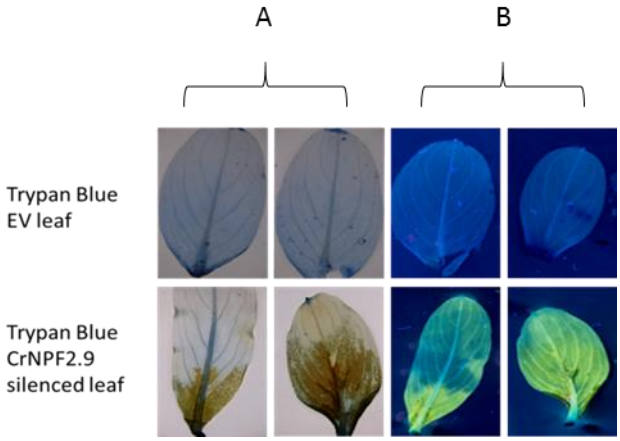


Figure 4-14 UV fluorescence of VIGS-CrNPF2.9 *C. roseus* leaf tissue
A. Trypan blue stained VIGS-EV and VIGS-CrNPF2.9 leaf visualised under standard white light, after removal of chlorophyll. B. Trypan blue stained VIGS-EV and VIGS-CrNPF2.9 leaf visualised under UV light. 2 leaves for VIGS-EV and VIGS-CrNPF2.9 are shown respectively.

Additionally, when the methanolic extract from VIGS-CrNPF2.9 leaf samples is analysed by Thin Layer Chromatography (TLC), additional yellow fluorescence compounds are observed on the TLC plate relative to VIGS-EV controls (Figure 4-15).

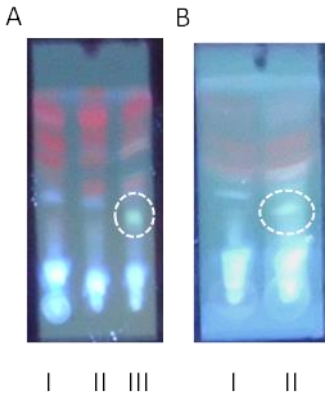


Figure 4-15 UV fluorescence of VIGS-EV and VIGS-CrNPF2.9 leaf tissue methanol extracts from two different cultivars of *C. roseus*
Methanol extracts of VIGS-EV and VIGS-CrNPF2.9 leaf tissue were run by silica TLC in 7:3 Chloroform:MeOH and visualised by 365 nm UV A.) Methanol extracts from cultivar Little Bright Eyes I.) VIGS-EV plant 1 II.) VIGS EV plant 2 III.) VIGS-CrNPF2.9 B.) Methanol extracts Sunstorm Apricot I.) VIGS-EV II.) VIGS-NPF2.9. Yellow fluorescent spots in VIGS-CrNPF2.9 leaf tissue is highlighted by a white dotted circle.

The yellow fluorescent compounds could not be identified after extraction from the TLC silica and analysis by LC-MS, indicating these compounds may not ionise well in either positive or negative ionisation mode by electrospray ionisation.

The reason for this yellow fluorescence specifically in VIGS-CrNPF2.9 leaf tissue, together with the visualisation of yellow fluorescent compounds upon visualisation of the methanolic extract of VIGS-CrPNF2.9 leaf tissue by silica TLC is interesting and requires further investigation.

Yellow autofluorescence of leaf tissue under UV light (365 nm) is used as an indicator of cell stress, and this autofluorescence is often attributable to the accumulation of phenolic compounds (Klement et al., 1990). In the example we present here, the yellow autofluorescence is unlikely to be due to the accumulation of phenolic compounds, as these should be detectable by LC-MS in the methanolic extract of VIGS-CrNPF2.9 tissue.

4.3.4.6 Proposed causes of cell death

Our current hypothesis as to the cause of cell death in VIGS-CrNPF2.9 silenced leaf tissue, is that the accumulation of strictosidine is toxic to the cell. The reason for this toxicity is unclear since strictosidine itself does not contain reactive functional groups, and the accumulation of strictosidine in this VIGS-CrNPF2.9 tissue would be localised to the vacuole and subsequently compartmented away from the cytoplasm.

However the accumulation of strictosidine in the vacuole, an acidic microenvironment, may lead to its spontaneous deglycosylation through acid hydrolysis, or through the action of a non-specific vacuolar glycosidase.

The deglycosylation of strictosidine results in the formation of a reactive dialdehyde species (Figure 4-16). This dialdehyde is unstable and very reactive and has been shown to lead to the crosslinking of proteins and any group with a free amine (Guirimand et al., 2010), thereby having the potential to crosslink proteins, DNA and lipid head groups such as ethanolamine.

As such, accumulation and deglycosylation of strictosidine in the vacuole could potentially result in disruption of vacuole function, resulting in cell death.

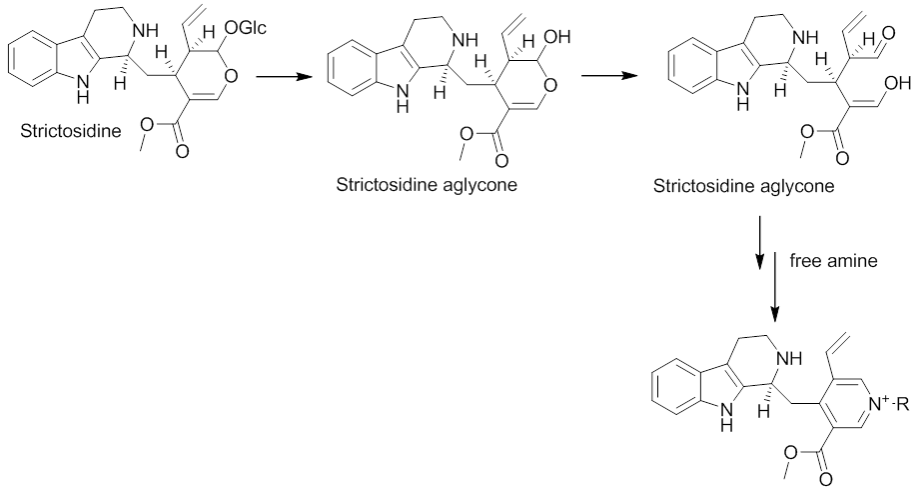


Figure 4-16 Potential mechanism for cell death in strictosidine accumulating plants.
Deglycosylation of strictosidine results in the formation of a reactive dialdehyde that has the potential to crosslink free amines

4.3.4.7 Characterisation of the spatial distribution of metabolism in VIGS tissue

Analysis of methanol extracts of VIGS-CrNPF2.9 leaf tissue by LC-MS demonstrated that there is an accumulation of strictosidine in this silenced tissue relative to VIGS-EV control plants, however it does not provide spatial information as to where the strictosidine is accumulating in the leaf.

To further assess whether the strictosidine accumulation is directly associated with certain areas of the leaf, the VIGS tissue was analysed by Mass Spectrometry Imaging (MSI). There are a number of MSI methods that can be applied to plant tissue utilising a variety of ionisation sources including: secondary ion mass spectrometry (SIMS), laser ablation electrospray ionisation (LAESI), matrix assisted laser desorption ionisation (MALDI) and desorption electrospray ionisation (DESI) (Matros and Mock, 2013). SIMS-MSI and MALDI-MSI have the best spatial resolution for imaging (between 1 µm-10 µm) and have been used on a variety of plant tissues. The MALDI-MSI approach requires application of a matrix to the biological tissue, which assists ionisation of compounds on the surface. Ionisation of the molecular ions then takes place under vacuum. One of the limitations of MALDI-MSI is that only the surface is imaged. If other parts of the tissue need to imaged, the sample needs to be cryosectioned (Brown and He, 2015; Matros and Mock, 2013).

Although DESI-MSI has a lower spatial resolution of ~100 µm the ionisation occurs under atmospheric conditions without the application of a matrix, and therefore sample preparation is less technically intensive (Li et al., 2013). DESI-MSI has been utilised for both the direct and indirect imaging of plant tissues (Müller et al., 2011). Direct imaging results in the electrospray scanned directly over the leaf surface, whilst indirect imaging, by comparison, requires the imprinting of the metabolites from the plant tissue onto either filter paper or microporous Teflon, under high pressure (1 ton), with this imprint subsequently scanned by DESI. Both are able to provide spatial information with regards to the metabolite distribution in leaf tissue, and have been successfully applied to the analysis of the spatial distribution of secondary metabolites in *Hordeum vulgare* (Li et al., 2011), *Hypericum perforatum* (Thunig et al., 2011), *Myristica malabrica* (Ifa et al., 2011), as well as *Catharanthus roseus* (Hemalatha and Pradeep, 2013).

In this study, indirect DESI-MSI of VIGS-EV and VIGS-CrNPF2.9 imprinted leaf tissue was utilised to assess the spatial distribution of metabolites in this leaf tissue. MALDI-MSI was not used for the assessment of the metabolite distribution in VIGS-CrNPF2.9 leaf tissue because the target

compound, strictosidine *m/z* 531, did not ionise well with the standard matrix (α-cyano-4-hydroxy-cinnamic acid). Further optimisation of the matrix for ionisation of this compound is therefore required (Data not shown).

Preliminary analysis of the indirect DESI-MSI of the leaf imprints from VIGS-EV and VIGS-CRNPF2.9 leaf tissue displayed a distinctly different localisation of metabolites. As observed in Figure 4-17 both the downstream alkaloids *m/z* 457 (vindoline) and *m/z* 337 showed a uniform distribution in the VIGS-EV leaf tissue imprint. In comparison in the VIGS-CrNPF2.9 leaf tissue imprints these molecular ions are localised to the tip of the leaf tissue (Figure 4-17).

Strictosidine, *m/z* 531 is not detectable in the VIGS-EV leaf imprint (data not shown), whilst in comparison the strictosidine accumulated in VIGS-CrNPF2.9 in the regions which have lowest *m/z* 457 and *m/z* 337, nearer the leaf petiole (Figure 4-17). This therefore indicates that there is a different localisation of strictosidine and vindoline in the VIGS-CrNPF2.9 leaf tissue, with the highest accumulation of vindoline in the region with the lowest accumulation of strictosidine (Figure 4-17). This is consistent with the standard metabolite profile of the VIGS-CrNPF2.9 leaf tissue, and also demonstrates that the transporter is likely to be involved in the MIA pathway, as it appears silencing the transporter blocks formation of vindoline and leads to the accumulation of strictosidine.

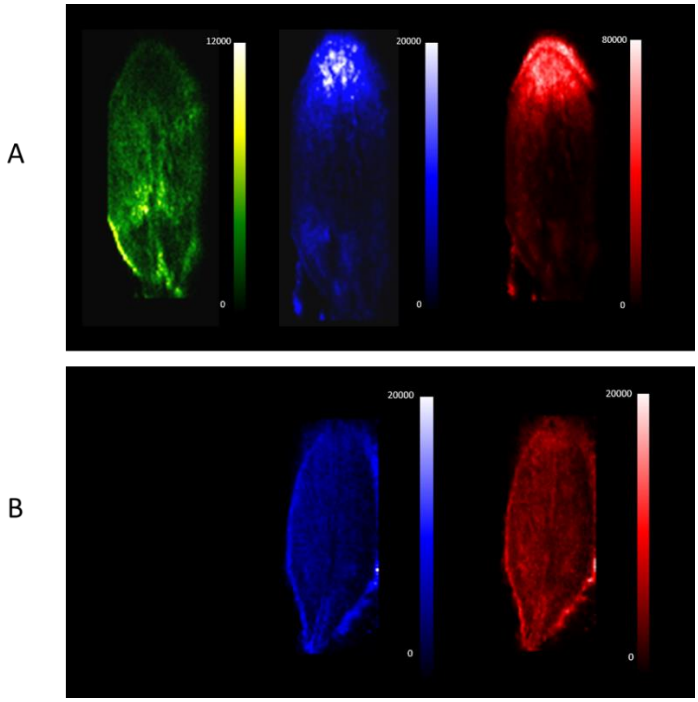


Figure 4-17 DESI-MSI of VIGS-EV and VIGS-CrNPF2.9 leaf tissue
A.) DESI-MSI of VIGS-CrNPF2.9 leaf tissue. (l-r) strictosidine, *m/z* 531, green; *m/z* 337, blue; vindoline, *m/z* 457, red. B.) DESI-MSI of VIGS-EV leaf tissue. (l-r) strictosidine, *m/z* 531, data not shown; *m/z* 337, blue; vindoline, *m/z* 457, red. Scale bar is intensity of ion counts.

Further work and optimisation of this methodology is needed to provide more information about the spatial distribution of metabolism upon silencing other genes in the MIA pathway, however preliminarily it appears that it is a useful tool for assessment of the spatial distribution of metabolites in VIGS tissue, and this is the first example of DESI being applied to VIGS samples in *C. roseus.*

4.3.4.8 Combinatorial silencing of transporters by VIGS

In Chapter III a transporter, CrMATE1952, was identified that produced a VIGS phenotype resulting in the accumulation of a new peak with *m/z* 391 (H^+ adduct) and *m/z* 413 (Na^+ adduct). It was assumed that this phenotype was due to the mislocalisation of secologanin in the cytosol leading to the accumulation of the reduced product secologanol. To further assess whether both CrMATE1952 and CrNPF2.9 are involved in the MIA pathway, and occur sequentially in the pathway, it is necessary to combinatorially silence these two transporters.

Two VIGS fusion pTRV2 vectors were generated, a double fusion that contains ~500 bp of both CrMATE1952 and CrNPF2.9 in the pTRV2u vector, and a triple silencing vector that contains ~500 bp each of CrMATE1952, CrMATE10740, and CrNPF2.9. CrMATE10740 is the other MATE protein apart from CrMATE1952 that also co-regulates with the MIA pathway. These vectors were generated by USER fusion cloning.

Both the double fusion and triple fusion vectors resulted in the same metabolic phenotype. Each of the combinatorial VIGS experiments was performed in parallel with a set of empty vector control plants, as well as individually silenced CrNPF2.9 as a positive control (Figure 4-18, Figure 4-19).

As would be expected, only in vectors containing sequence for silencing CrMATE1952 do we see the accumulation of secologanol, *m/z* 391, and this was observed for both the double and triple silencing vector (Figure 4-18, Figure 4-19). This compound was not present in either the individually silenced VIGS-CrNPF2.9 or VIGS-EV control plants (Figure 4-18, Figure 4-19).

In both the individually silenced CrNPF2.9 plants, as well as in the double and triple VIGS fusions, there was an accumulation of strictosidine relative to empty vector control plants. However the extent of strictosidine accumulation was attenuated in the double and triple silenced plants relative to the individually silenced CrNPF2.9 tissue (Figure 4-18, Figure 4-19).

The fact that both the double and triple silenced plants also show a comparable decrease in the downstream MIAs, vindoline and catharanthine, like the individually silenced CrNPF2.9 plants, is indicative that the two transporters CrMATE1952 and CrNPF2.9 may be working sequentially in the same metabolic pathway.

CrMATE1952 silencing results in the mislocalisation of secologanin leading to the accumulation of *m/z* 391 (H^+ adduct). This decreases the amount of secologanin entering the vacuole, and therefore limits strictosidine synthesis. Any strictosidine that is formed in the vacuole, in turn, accumulates because it cannot exit due to the silencing of CrNPF2.9. The combined effect of silencing both CrMATE1952 and CrNPF2.9 therefore has a comparable effect on vindoline and catharanthine as individually silencing CrNPF2.9 (Figure 4-18, Figure 4-19).

The double and triple silencing plants did not result in the cell death phenotype that is seen upon individually silencing the CrNPF2.9. This correlates with the fact that the strictosidine accumulation in these double and triple silencing sets of plants is attenuated relative to the individually silenced CrNPF2.9 plants.

qPCR analysis on the VIGS tissue for the double silencing and triple silencing sets of plant tissue demonstrated that in both sets, both the CrNPF2.9 gene and the CrMATE1952 are silenced relative to the empty vector control plants (Figure 4-18, Figure 4-19), and that in the triple silenced plants CrMATE10740 is also silenced (Figure 4-19).

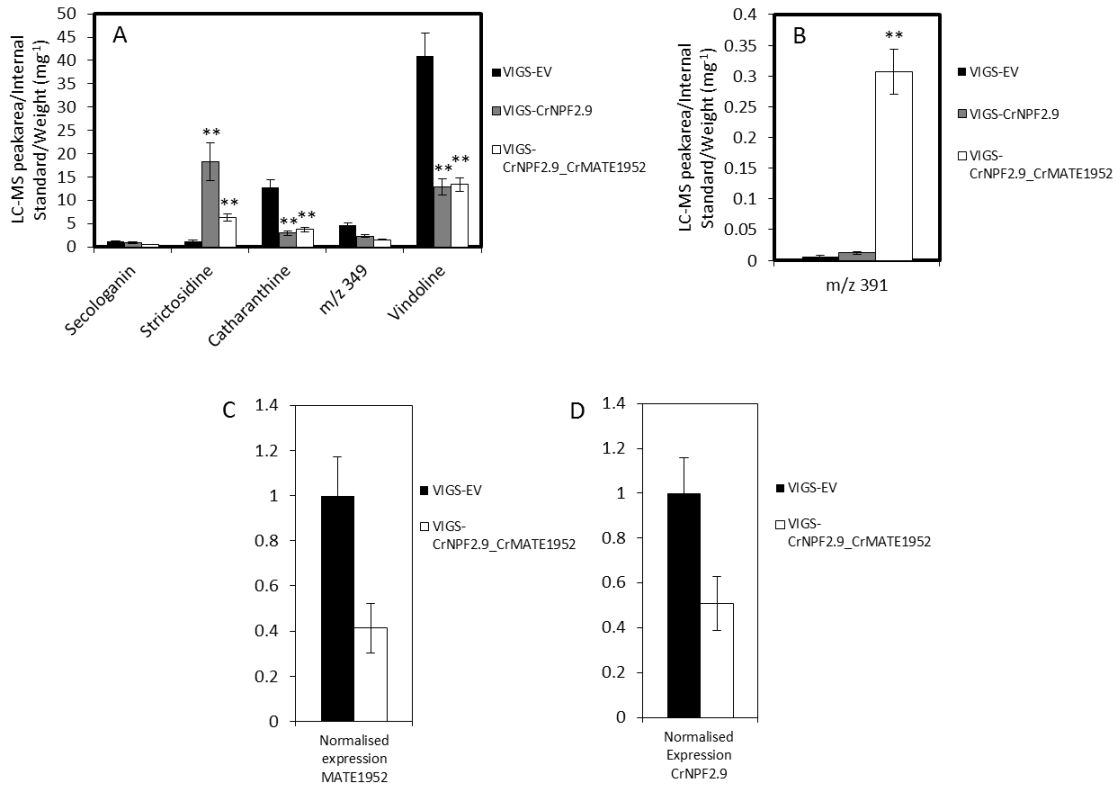


Figure 4-18 Metabolite and qPCR profile for VIGS double fusion vector CrNPF2.9_CrMATE1952
A.) Alkaloid profile for VIGS of empty vector control tissue, individually silenced CrNPF2.9 and the double fusion vector CrNPF2.9_CrMATE1952 in Little Bright Eyes; VIGS CrNPF2.9 (n=10), VIGS-CrNPF2.9_CrMATE1952 (n=15), VIGS-EV (n=9) ** $p<0.01$ B.) *m/z* 391, secologanol (H^+ adduct) accumulation for VIGS of empty vector control tissue, individually silenced CrNPF2.9 and the double fusion vector CrNPF2.9_CrMATE1952 in Little Bright Eyes; VIGS CrNPF2.9 (n=10), VIGS-CrNPF2.9_CrMATE1952 (n=15), VIGS-EV (n=9) ** $p<0.01$ C.) qPCR normalised expression of the CrMATE1952 gene in empty vector control tissue and the double silenced CrNPF2.9_CrMATE1952 silenced tissue; VIGS-CrNPF2.9_CrMATE1952 (n=8) VIGS-EV (n=8). D.) qPCR normalised expression of the CrNPF2.9 gene in empty vector control tissue and the double silenced CrNPF2.9_CrMATE1952 silenced tissue; VIGS-CrNPF2.9_CrMATE1952 (n=8) VIGS-EV (n=8). All data shown is mean ± SEM.

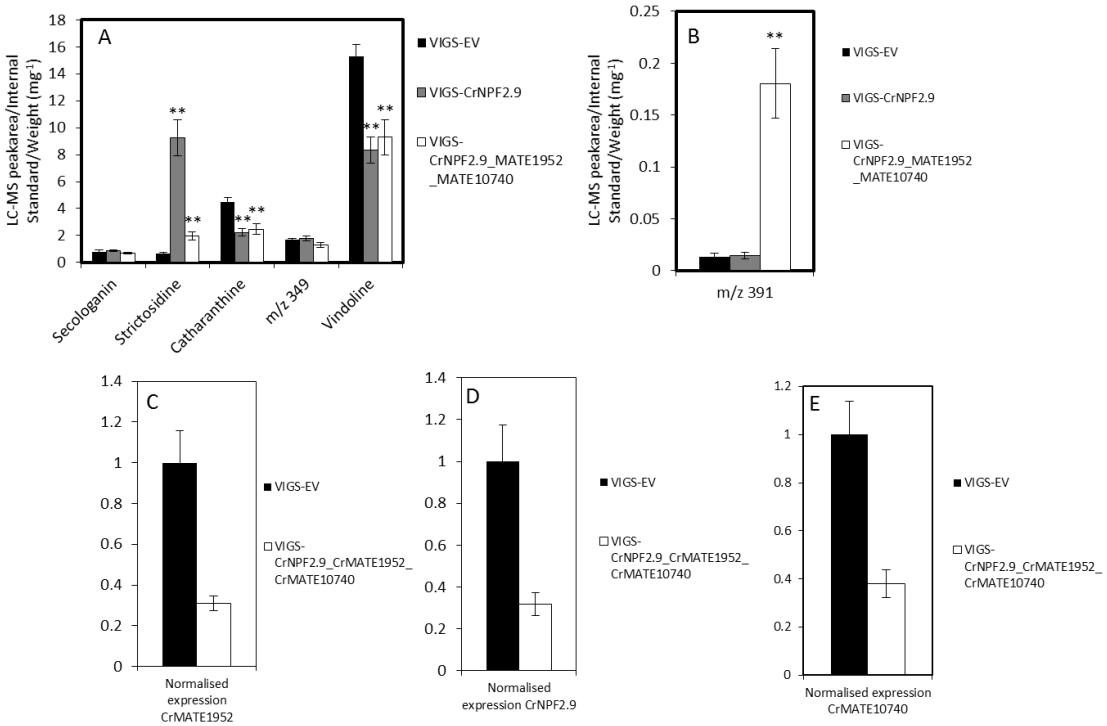


Figure 4-19 Metabolite and qPCR profile for VIGS triple fusion vector CrNPF2.9_CrMATE1952_CrMATE10740

A.) Alkaloid profile for VIGS of empty vector control tissue, individually silenced CrNPF2.9 and the triple fusion vector CrNPF2.9_CrMATE1952_CrMATE10740 in Little Bright Eyes; VIGS CrNPF2.9 (n=12), VIGS-CrNPF2.9_CrMATE1952_CrMATE10740 (n=15), VIGS-EV (n=11) ** $p<0.01$ B.) *m/z* 391, secologanol (H^+ adduct), accumulation for VIGS of empty vector control tissue, individually silenced CrNPF2.9 and the triple fusion vector CrNPF2.9_CrMATE1952_CrMATE10740 in Little Bright Eyes; VIGS CrNPF2.9 (n=12), VIGS-CrNPF2.9_CrMATE1952_CrMATE10740 (n=15), VIGS-EV (n=11) ** $p<0.01$ C.) qPCR normalised expression of the CrMATE1952 gene in empty vector and the triple silenced CrNPF2.9_CrMATE1952_CrMATE10740 silenced tissue; VIGS-CrNPF2.9_CrMATE1952_CrMATE10740 (n=8) VIGS-EV (n=8). D.) qPCR normalised expression of the CrNPF2.9 gene in empty vector and the triple silenced CrNPF2.9_CrMATE1952_CrMATE10740 silenced tissue; VIGS CrNPF2.9_CrMATE1952_CrMATE10740 (n=8) VIGS-EV (n=8). E.) qPCR normalised expression of the CrMATE10740 gene in empty vector and the triple silenced CrNPF2.9_CrMATE1952_CrMATE10740 silenced tissue; VIGS CrNPF2.9_CrMATE1952_CrMATE10740 (n=8) VIGS-EV (n=8). All data shown is the mean ± SEM.

4.3.4.9 qPCR analysis of VIGS-CrNPF2.9 silenced tissue

As has been discussed in Chapter III, qPCR is utilised to assess the relative silencing of genes in a VIGS experiment between the VIGS -EV control plant tissue and the VIGS silenced tissue.

qPCR analysis to assess the level of CrNPF2.9 silencing in the individually silenced CrNPF2.9 VIGS tissue was not possible due to significant variation in the reference housekeeping genes used for normalisation between VIGS-EV and VIGS-CrNPF2.9. The three most commonly utilised reference genes used for qPCR VIGS tissue in *C. roseus* are EIF4α, RPS9 and 60SRPPO (Asada et al., 2013; Geu-Flores et al., 2012; Liscombe and O'Connor, 2011; Salim et al., 2013). Although there have been reports for the selection and validation of other reporter genes for qPCR in *C. roseus* (Pollier et al., 2014) in our hands these genes proved highly variable, and appear to be lowly expressed in leaf tissue (Appendix), in relation to the established standards, EIF4α, RPS9 and 60SRPPO.

The reference genes are used as a method for normalisation of gene expression in qPCR data, and act as internal controls for the qPCR experiment. These genes should be stably expressed in the tissues under consideration in the silencing experiment. VIGS tissue for individually silencing CrNPF2.9 resulted in a twofold lower expression of the reporter genes EIF4α, RPS9 and 60SRPPO relative to VIGS-EV tissue (Figure 4-20). This is reproducible from two separate independent VIGS experiments (Figure 4-20) and is despite RNA quality being comparable between VIGS-EV and VIGS-CrNPF2.9 (Figure 4-20). As shown previously for the double and triple pTRV2 vectors (CrNPF2.9_CrMATE1952 and CRNPF2.9_CrMATE1952_CrMATE10740) the silencing of the CrNPF2.9 gene can be measured in these plants. For plants silenced using the fusion vectors there is no significant difference in the reporter genes for the VIGS-EV and VIGS-fusion expression levels.

The potential reason for this disparity in housekeeping gene expression in VIGS-CrNPF2.9 leaf tissue is likely due to the influence of the cell death VIGS phenotype seen upon CrNPF2.9 silencing.

As cell death is not a usual phenotype that occurs with most VIGS studies, this might explain why this discrepancy in housekeeping gene profile has not been observed before. Cell death is not observed in the CrNPF2.9_CrMATE1952 double silenced plants, or the CrNPF2.9_CrMATE1952_CrMATE10740 triple silenced plants. This may be because, although strictosidine does accumulate, it does not accumulate as highly as in the CrNPF2.9 individually

silenced plants. qPCR was therefore able to be performed on the double silenced, VIGS_CrNPF2.9_CrMATE1952, and triple silenced, VIGS_CrNPF2.9_CrMATE1952_CrMATE10740, as presented in Figure 4-18 and Figure 4-19.

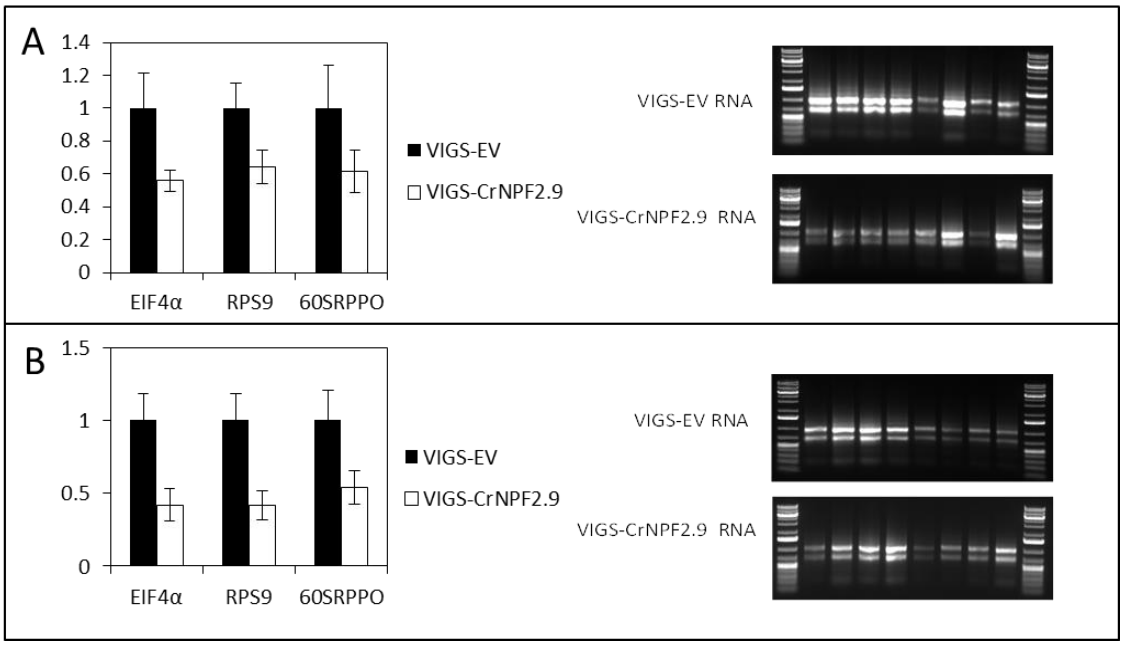


Figure 4-20 Relative expression of housekeeping genes in VIGS-CrNPF2.9 silenced plant tissues.
A.) Relative expression of EIF4α, RPS9 and 60SRPPO in VIGS-CrNPF2.9 silenced leaf tissue to VIGS-EV leaf tissue VIGS-CrNPF2.9 (n=8), VIGS-EV (n=8). RNA extracted from VIGS-EV leaf tissue (n=8) and VIGS-CrNPF2.9 (n=8) used for cDNA synthesis in the qPCR study. B.) Relative expression of EIF4α, RPS9 and 60SRPPO in VIGS-CrNPF2.9 silenced leaf tissue to VIGS-EV leaf tissue VIGS-CrNPF2.9 (n=8), VIGS-EV (n=8). RNA extracted from VIGS-EV leaf tissue (n=8) and VIGS-CrNPF2.9 (n=8) used for cDNA synthesis in the qPCR study. All data shown is the mean ± SEM.

4.3.4.10 VIGS of the CrNPF2 subfamily

As has been shown in this chapter, CrNPF2.9 resulted in a metabolic phenotype upon silencing that is consistent with it being the exporter of strictosidine from the vacuole. There are nine other orthologs of this transporter family in *C. roseus*. The orthologs CrNPF2.4 and CrNPF2.6 are localised to the plasma membrane, and have been shown to have transport activity towards the iridoid glycosides, 7-deoxyloganic acid, loganic acid, loganin and secologanin, when expressed in *Xenopus laevis* oocytes (Larsen et al; submitted). The secoiridoid pathway for the generation of secologanin is inter-cellulary compartmented, and there is a need for the transport of a

mobile intermediate between the IPAP cells and the epidermis. To date this intermediate has been hypothesised to be loganic acid (Miettinen et al., 2014).

To assess whether the *in vitro* activity of CrNPF2.4 and CrNPF2.6 is physiologically relevant these transporters were targeted by VIGS. VIGS of the NPF transporters CrNPF2.4 and CrNPF2.6 did not result in a statistically significant difference in the downstream alkaloids vindoline, catharanthine, strictosidine or serpentine, and there was no reproducible accumulation of any new peaks, when compared to EV control plants. There was also no change in the peak area of secologanin relative to EV control plant tissues. The VIGS experiment was repeated twice in two different cultivars of *C. roseus*, Little Bright Eyes and Sunstorm Apricot (Figure 4-21).

Although the transporters CrNPF2.4 and CrNPF2.6 have the capacity to transport a range of iridoid glycosides *in vitro*, and the work presented in (Larsen et al; submitted) elegantly demonstrates that the NPF transporter family is capable of transporting this class of compounds, thereby expanding the known substrate repertoire of this transporter family, the physiological relevance of these transporters and the role they play in MIA metabolism *in planta* is debatable. It is possible at the plasma membrane that these transporters are functionally redundant, and multiple transporters and transporter families, may be involved in the transport of the iridoid glycosides *in planta*. One way of addressing this is through the use of combinatorial silencing of multiple transporter classes by VIGS and analysing for a metabolic phenotype that results in the accumulation of the iridoid glycosides to be transported.

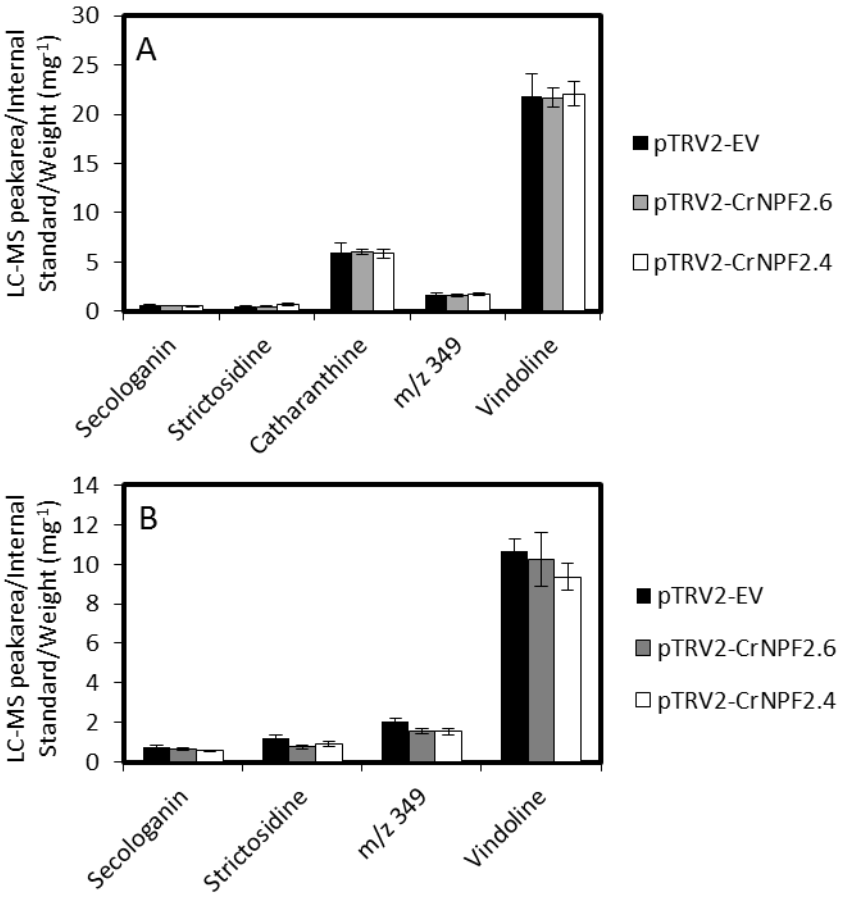


Figure 4-21 Metabolite profile for VIGS for CrNPF2.4 and CrNPF2.6.
A.) VIGS in *C. roseus* cultivar Sunstorm Apricot pTRV2-EV (n=11), pTRV2-CrNPF2.4 (n=11), pTRV2-CrNPF2.6 (n=11), p<0.01**B.) VIGS in *C. roseus* cultivar Little Bright Eyes ; pTRV2-EV (n=8), pTRV2-CrNPF2.4 (n=8), pTRV2-CrNPF2.6 (n=8). p<0.01 **. Untargeted metabolomics using the XCMS software package was performed for both VIGS experiments and did not identify the accumulation of new peaks in the CrNPF silenced lines relative to empty vector control. All data shown is the mean ± SEM.

4.3.5 Heterologous expression of CrNPF2.9 in S. cerevisiae

Due to the predicted directionality of transport for the CrNPF2.9 transporter, in which movement of solute is predicted to occur into the cytoplasm from the vacuole, for a compatible transport assay in *S. cerevisiae*, the CrNPF2.9 transporter would need to be located in the plasma membrane of the yeast. If the CrNPF2.9 transporter localises to the yeast plasma membrane it would be possible to perform an import assay with strictosidine into yeast cells. This whole yeast cell assay has been utilised for the characterisation of the purine permease transporter that transports nicotine in *Arabidopsis*, NUP1 (Hildreth et al., 2011).

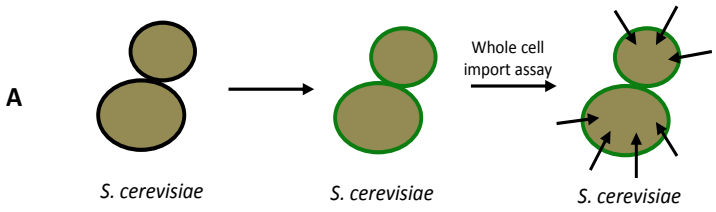


Figure 4-22 Whole cell yeast cell assay
Due to the predicted directionality of the NPF transporter a whole cell yeast import assay would be feasible if the transporter is expressed at the plasma membrane

If the CrNPF2.9 localises to internal membranes, such as the yeast vacuolar membrane, the assay becomes technically challenging as it then becomes necessary to pre-load the yeast vacuole with substrate prior to the assay and detect export from the vacuole.

To assess whether the CrNPF2.9 transporter is expressed at the plasma membrane, it was over expressed as a C-terminal GFP fusion in *S. cerevisiae* under control of a galactose inducible promoter in the yeast expression vector pYES-DEST52.

Over expression of CrNPF2.9 in *S. cerevisiae* as a C-terminal GFP fusion resulted in a similar localisation in the yeast as the two MATE proteins expressed in Chapter III, with localisation to internal membranes in the yeast strains InvSCI and ybtI or the yeast lytic vacuole for WATII (Figure 4-23). Although there is localisation to what appears to be the yeast plasma membrane when expressing the protein in the ybt1 strain, this localisation is not consistent between individual yeast cells.

This is in contrast to the *C. roseus* purine permease, CrPUP_1443, whose expression resulted in localisation to the yeast plasma membrane, as well as in internal membranes for all three yeast strains, or soluble GFP that localises to the yeast cytoplasm (Chapter III).

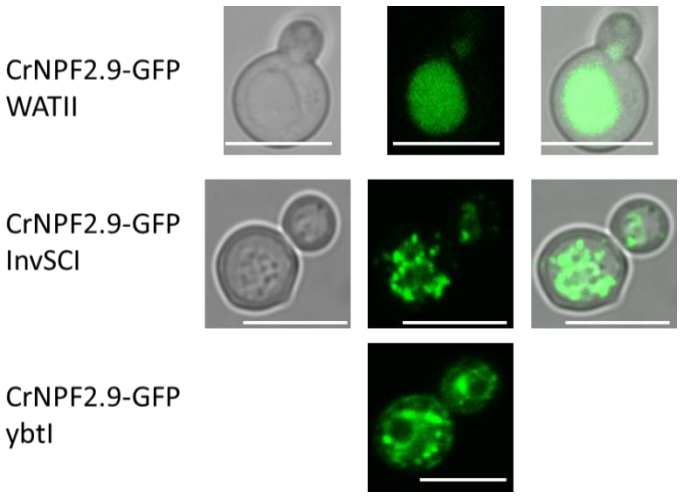


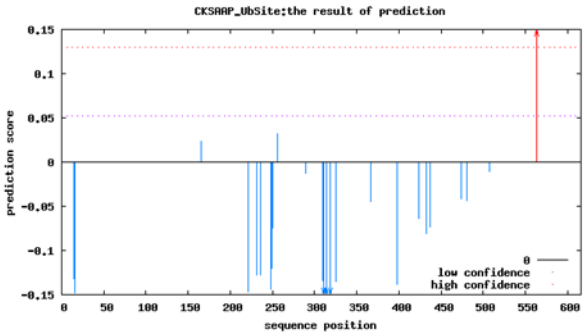
Figure 4-23 Expression of CrNPF2.9-GFP in *S. cerevisiae.*
CrNPF2.9-GFP was cloned into the yeast expression vector pYES-DEST52, expressed in three separate yeast strains WATII, InVScI and ybt1, and imaged by confocal microscopy. (l-r) Brightfield, GFP-fluorescence, Merge. Scale bar = 5 µm.

4.3.5.1 Site directed mutagenesis of the CrNPF2.9

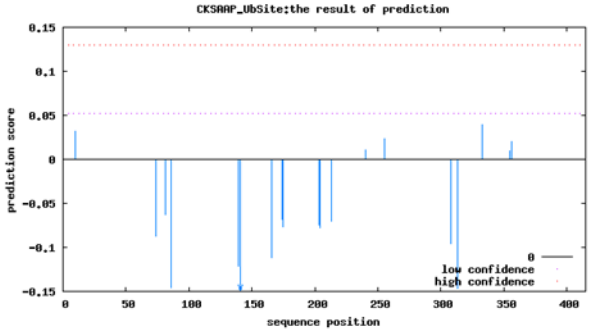
A variant of the CrNPF2.9-GFP construct was made to assess whether the length of the linker between the membrane protein and GFP affects protein stability and localisation. The linker between the C-terminus of the membrane protein utilised a three amino acid glycine-serine-glycine linker, and this was extended to an eight amino acid linker GSGMGGSG.

Additionally, it was postulated that the ubiquitination state of the membrane transporter might affect its turnover rate at the plasma or vacuolar yeast membrane. The CrNPF2.9 transporter has a predicted C-terminal yeast ubiquitination site (Figure 4-24), whilst the *C. roseus* putative purine permease, CrPUP_1443, which was shown to express at the yeast plasma membrane as a C-terminal GFP fusion in Chapter III does not (Figure 4-24).

A



B



C

Gene name	C-terminal 20 aa
CrNPF2.9	YEYSSVMEDEDETNDNKIYM
CrNPF2.9K>R	YEYSSVMEDEDETNDNRIYM

Figure 4-24 Ubiquitination site prediction for the *C. roseus* purine permease and CrNPF2.9
A.) The CrNPF2.9 contains a predicted yeast ubiquitination site at the C-terminus of the protein, highlighted by a red arrow in accordance with the CKSAAP_UBsite prediction server. B.) The purine permease CrPUP_1443 does not contain a predicted ubiquitination sequence C.) The C-terminal 20 amino acids of the CrNPF2.9 protein showing the lysine to arginine mutation generated, highlighted in yellow, to prevent putative monoubiquitination of the transporter.

Ubiquitin is a 76-amino acid protein that is expressed in all eukaryotic cells and is used as a post-translational modification for a diverse array of physiological functions; including signalling for proteolytic degradation by the proteasome, and altering cellular localisation (Horák, 2003; MacGurn et al., 2012). Ubiquitination occurs on lysine residues via the formation of an isopeptide bond. In yeast ubiquitination is a key modification for the promotion of plasma membrane protein turnover and ubiquitin mediated endocytosis. The ubiquitin modification acts as a targeting signal for endocytosis of plasma membrane proteins and their subsequent degradation in the lytic vacuole (Horák, 2003; MacGurn et al., 2012).

This mechanism has been shown for the endogenous yeast uracil permease (Fur4p) (Galan et al., 1996), the ABC transporter Ste6p (Kölling and Hollenberg, 1994), the yeast general amino acid permease (Gap1p) (Springael et al., 1999), the yeast multidrug transporter Pdr5p (Egner and Kuchler, 1996), as well as for the other membrane spanning receptors such as the G-protein

coupled receptor Ste2p (Hicke and Riezman, 1996). An overview of this protein degradation pathway is presented in Figure 4-25.

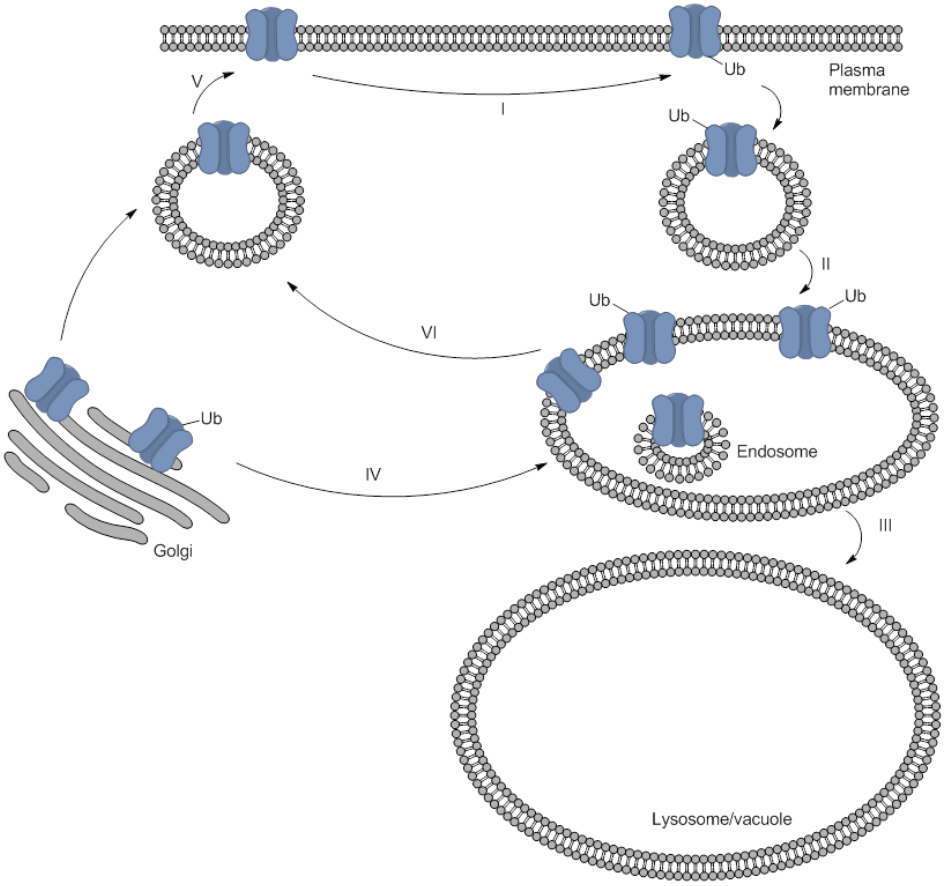


Figure 4-25 Overview of ubiquitin mediated endocytosis and vacuolar degradation of plasma membrane proteins in yeast
I.) Ubiquitination of the membrane protein II.) Ubiquitination mediated endocytosis III.) Degradation in the vacuole IV.) Golgi to endosome traffic V.) exocytosis VI.) Endosomal recycling to cell surface. Figure adapted from (MacGurn et al., 2012).

The CrNPF2.9 putative ubiquitination site was mutated from a lysine to an arginine (CrNPF2.9 K>R) utilising a reverse primer that contained the point mutation, and cloned into the pYES-DEST52 yeast expression vector. Both the short and extended C-terminal GFP linker versions of the CrNPF2.9 K>R protein were generated and expressed in the *S. cerevisiae* yeast strain WAT11. WAT11 was selected as the yeast strain, as previous GFP fusion proteins showed distinct localisation to the lytic vacuole in this strain, and it would be easier to monitor redistribution of the protein to a different membrane compartment.

Overexpression of the CrNPF2.9 with the extended linker to GFP, or with the mutated ubiquitination site, led to a minor relocalisation, however not to the plasma membrane (Figure

4-26). As such, *S. cerevisiae* as a platform for heterologous expression of this transporter was abandoned.

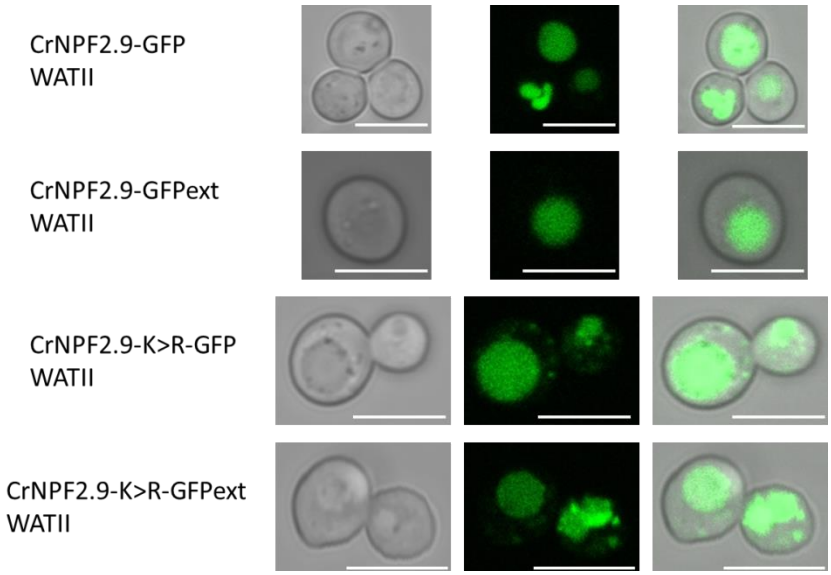


Figure 4-26 Localisation of CrNPF2.9-GFP in *S. cerevisiae*
Extension of the GFP linker between CrNPF2.9 and GFP in the fusion protein, together with mutation of a predicted ubiquitination site was performed to assess whether there is a change in subcellular localisation of the protein when heterologously expressed in *S. cerevisiae*. In all cases the CrNPF2.9-GFP fusions resulted in localisation to the lytic vacuole of the yeast (l-r) Brightfield, GFP-fluorescence, Merge. Scale bar = 5 µm.

4.3.6 Heterologous expression of CrNPF2.9 in Xenopus laevis oocytes

An alternative expression system for CrNPF2.9 is the *Xenopus* oocyte expression system. In this system, linear cRNA derived from *in vitro* transcription, encoding the CrNPF2.9 to be expressed, is microinjected into the *Xenopus* oocyte for protein expression. The protein is likely to be expressed at the plasma membrane of the oocyte and transport assays can then be performed relative to water injected controls, to assess whether the transporter can uptake the substrate (Figure 4-27).

Due to the predicted directionality of transport, if the CrNPF2.9 is localised to the plasma membrane of the *Xenopus laevis* oocyte, it is feasible to perform an import assay into the *Xenopus laevis* oocyte, and detect uptake of the substrate by LC-MS (Figure 4-27).

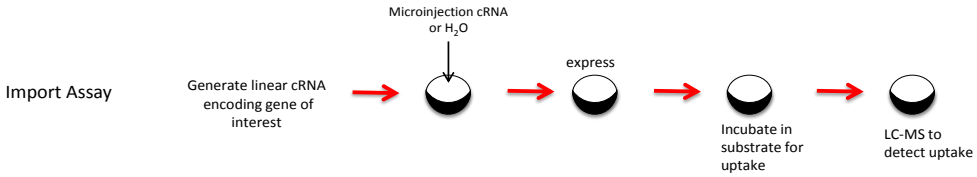


Figure 4-27 Overview of the *Xenopus laevis* oocyte uptake assay
cRNA generated by *in vitro* transcription is microinjected into the oocyte for expression of the protein. After 3 days incubation in MBS buffer the oocytes are incubated at pH 5 containing the substrate strictosidine for 1 h. The oocytes are subsequently washed and extracted in MeOH for analysis of strictosidine uptake by LC-MS.

The full length CrNPF2.9 transporter was cloned into two separate *Xenopus laevis* oocyte expression vectors pT7TS and pLIFE0016, and complementary cRNA was generated by *in vitro* transcription. This cRNA was used for the microinjection of *Xenopus laevis* oocytes, which were utilised to assess strictosidine uptake relative to water injected controls. The strictosidine used in this study was produced enzymatically and purified by preparative HPLC.

Although there are some clear examples where import is greater in the CrNPF2.9 injected *Xenopus* oocytes relative to water injected controls, as can be seen from Figure 4-28 there is batch to batch variability in the oocytes when the assay is replicated between different weeks. As such it is not possible to definitively state that we are observing transport of strictosidine into the oocyte.

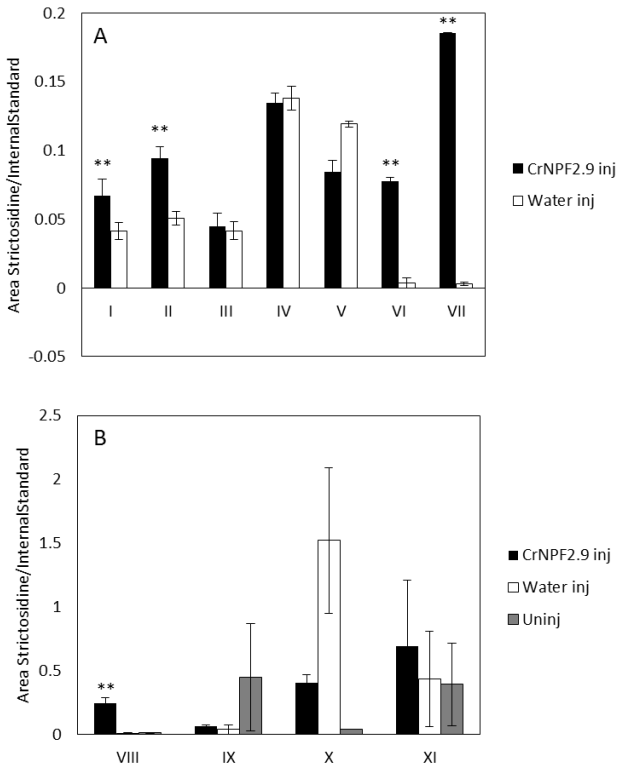


Figure 4-28 Strictosidine uptake assays in *Xenopus laevis* oocytes
To assess uptake of strictosidine in *Xenopus laevis* oocytes, ~20 oocytes were microinjected with 50 nl 1 µM cRNA, and ~20 oocytes microinjected with H_2O. Each batch of oocytes was incubated in 750 µl of 100 µM strictosidine. For each batch of 20 oocytes, oocytes were pooled in batches of 4 for extraction in MeOH and analysis by LC-MS, resulting in 5 replicates per batch. A.) Microinjection with cRNA derived from CrNPF2.9-pT7TS. Roman numerals I-VII represent separate weeks. B.) Microinjection with cRNA derived from CrNPF2.9-pLIFE0016. Roman numerals VIII-XI represent separate weeks on which the assay was performed. For uptake the CrNPF2.9 injected bars should be higher than water injected and uninjected controls $p<0.01$**. All data shown is mean $\pm$ SEM.

There are a number of potential limitations in this *Xenopus* oocyte uptake study. Firstly, as has been highlighted in Colman and Drummond, 1986, batch to batch variation in oocyte quality can greatly affect the experiment. As can be seen in the data in Figure 4-28 repetition of the experiment varies between weeks, indicating that different batches of oocytes may have different transport capacity for the substrate strictosidine. Sourcing oocytes from a commercial supplier may eliminate this problem.

A second limitation of the study in oocytes is the fact that CrNPF2.9 is predicted to be a facilitative transporter, and not an active transporter, of strictosidine.

The structure of the *Arabidopsis* nitrate transporter NRT1.1 (Parker and Newstead, 2014) implicates a conserved motif that is integral for coupling of solute to the proton gradient. It has been shown that an EXXER motif in the structure of the peptide transporter *Shewanella*

oneidensis $PepT_{So}$ is necessary for proton binding to the transporter for peptide transport (Newstead et al., 2011). This motif, EXXE(R/K), is conserved in the related NPF transporter family and is present in the glucosinolate transporters AtGTR1/AtGTR2, and suggests that the coupling of proton movement for structural changes in this class of transporters is conserved and substrate independent, for the active transport of solute (Nour-Eldin et al., 2012).

Interestingly, the *Arabidopsis* NAXT transport protein, which has been thought to lead to nitrate export (Segonzac et al., 2007), is a passive transporter that is not thought to utilise the proton gradient for transport, but instead transports nitrate down its concentration gradient. This transporter does not contain the conserved EXXE(R/K) motif for proton coupling (Figure 4-29).

CrNPF2.9 has the closest sequence identity to the NAXT transporter from *Arabidopsis* and also does not contain the conserved EXXE(R/K) sequence (Figure 4-29).

```
AtNAXT          ---------------------------------------MASSVTGDAETAISADSSTKR 21
CrNPF2.9        -------------------------------------------MGDTEAQLLQPGHKQK 16
AtGTR1          MKSRVILNHRDRRDKNHNNNNTNHYTQVDTMERKPLEVEPSTTTTNTDVVDSFEEEQRKI 60
AtNRT1.1        ----------------------------------MSLPETKSDDILLDAWDFQGRPADRS 26
                                                                         :

AtNAXT          RGGGWITFPFMIATLLGLTIAAWGWLLNLIVYLIEEFNVKSIAAAQIANIVSGCICMVPA 81
CrNPF2.9        --GGWITFPFILATRTLLTLAVAGFSSNLIVYLINEFNVNRIDSAQIYNVVNGCMALFPL 74
AtGTR1          VYRGWKVMPFIIGNETFEKLGIIGTLSNLLVYLTSVFNLKSYTAATIINAFSGTINFGTF 120
AtNRT1.1        KTGGWASAAMILCIEAVERLTTLGIGVNLVTYLTGTMHLGNATAANTVTNFLGTSFMLCL 86
                   **   .:::       :   *   **:.**   :::    :*   . . *   :
```

Figure 4-29 N-terminal alignment of the NPF transporters
As can be seen a conserved EXXE(R/K) motif is present in the AtGTR1 and AtNRT1.1 transporters highlighted in yellow

Based on this we would predict that therefore CrNPF2.9 is a facilitative transporter, that moves strictosidine down its concentration gradient, but does not couple this movement to the proton gradient (Figure 4-30).

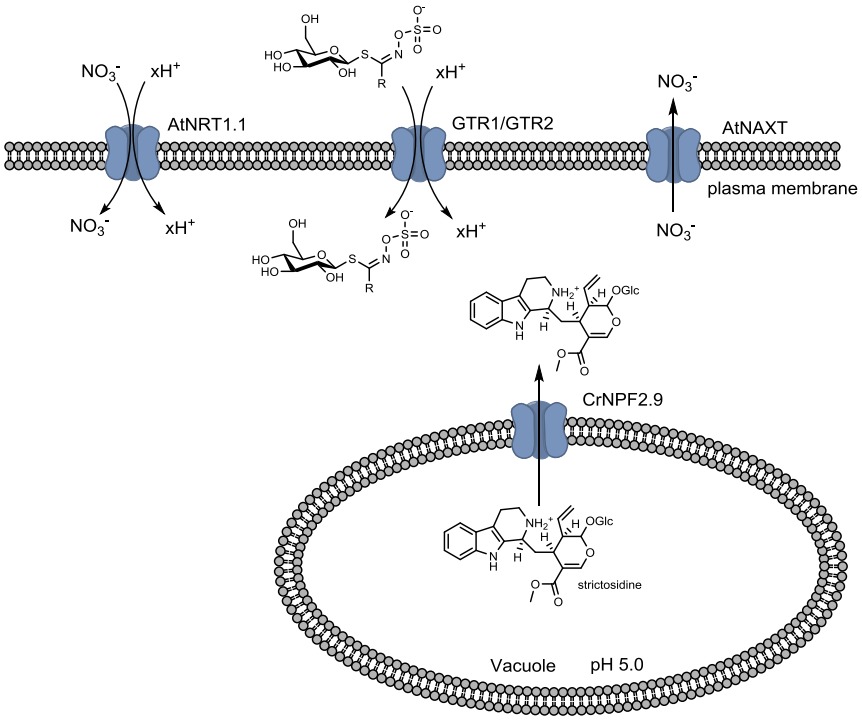


Figure 4-30 Overview of the predicted transport of NPF transporters
AtNRT1.1 couples the movement of nitrate against its concentration gradient utilising the electrochemical potential of the proton gradient. Similarly AtGTR1/AtGTR2 utilise the proton gradient for coupling of glucosinolate transport into the cytoplasm. Based on sequence identity and the absence of an EXXE(R/K) motif in both AtNAXT and CrNPF2.9 both transporters are thought to be facilitative transporters. AtNAXT has been shown to passively transport nitrate, whilst CrNPF2.9 is predicted to be a facilitative strictosidine exporter.

For the NPF transporters characterised to date, that have utilised an LC-MS approach for detection of transport, such as for the glucosinolate/H^+ co-transporters AtGTR1 and AtGTR2, the transporters are active transporters, which utilise the electrochemical gradient of across the tonoplast membrane to move the substrate against the concentration gradient (Nour-Eldin et al., 2012).

This makes the assay potentially easier to interpret, as the nature of active transport means the concentration of substrate can significantly build up in the oocyte relative to the external medium.

Secondly, the AtGTR1 and AtGTR2 are amenable for analysis by electrophysiology (Nour-Eldin et al., 2012). It was shown that AtGTR1 and AtGTR2 result in negative currents when incubated in the substrate glucosinolates. As the glucosinolate is negatively charged, yet a negative current is seen upon incubation of the oocyte in the substrate, this suggests that the transporters couple

the charge across the tonoplast membrane with a stoichiometric ratio of $n_{H+} > n_{glucosinolate}$ (Nour-Eldin et al., 2012).

Similarly, the canonical nitrate transporter from *Arabidopsis* NRT1.1 results in the electrogenic transport, coupling the symport of NO_3^- to the proton gradient, and therefore the ratio of proton: solute movement across the membrane is at least 2:1 (Parker and Newstead, 2014).

At the pH of the vacuole (ca. pH 5), strictosidine is likely to be positively charged, and therefore movement of strictosidine into the cytoplasm by CrNPF2.9 would be energetically downhill and dissipate the electrochemical potential across the membrane. However the theoretical concentrative capacity of the transporter is likely to be dependent on the concentration gradient of strictosidine across the tonoplast membrane. Therefore it would not be possible to detect transport into the oocyte, greater than the concentration of the external medium, which may also limit detection of transport in our assay system. Therefore further experiments are needed with higher quality oocytes in order to definitively assess *in vitro* transport for CrNPF2.9.

4.4 Conclusion

This chapter highlights the identification of an NPF transport protein that is directly involved in MIA metabolism in *C. roseus*. There are ten homologues of the CrNPF2 subfamily in *C. roseus*. It has been shown in (Larsen et al; submitted), that two of these proteins, CrNPF2.4 and CrNPF2.6 are able to transport a range of iridoid glycosides when expressed in *Xenopus laevis* oocytes. VIGS of these transporters, however, did not result in a metabolic phenotype, indicating that these two proteins might be functionally redundant. Future work would be to combinatorially silence these transporters and assess whether they compensate for each other's activity. It is also possible at the plasma membrane that there might be other transporter families that could be contributing to transport of the iridoid glycosides as well, and a combinatorial reverse genetics screen with other transport families should be employed to pick apart this potential functional redundancy.

To date, despite the *in vitro* characterisation of the two NPF transporters, CrNPF2.4 and CrNPF2.6 as iridoid glycoside transporters, which further expands the substrate repertoire of this transporter family as presented in (Larsen et al; submitted), the physiological importance of these transporters in MIA metabolism is still open for debate.

CrNPF2.9 is a transport protein whose closest homologue is the NAXT transporter in *Arabidopsis*. CrNPF2.9 strongly co-regulates with known genes in the MIA biosynthetic pathway, and this protein has been shown to localise to the tonoplast membrane when overexpressed in two heterologous systems, *C. roseus* cell suspension cultures, and *N. benthamiana*.

Silencing of this transporter resulted in a strong metabolic phenotype, resulting in the over accumulation of strictosidine, the central metabolic branch point in MIA metabolism, suggesting this transporter is the exporter of strictosidine from the vacuole.

This is the first demonstration that an NPF transporter is physiologically relevant to MIA metabolism. Co-silencing of CrNPF2.9 with the MATE proteins identified in Chapter III further indicate that both proteins are likely to be involved in MIA biosynthesis.

The metabolic phenotype is accompanied by cell death, which is attributable to the toxic accumulation of strictosidine in the tissue, as determined by confocal microscopy and preliminary analysis by DESI mass spectrometry imaging which indicated that strictosidine is accumulating in the region surrounding the dying tissue.

To date, heterologous expression of CrNPF2.9 has proven elusive for characterisation of strictosidine transport in both *S. cerevisiae* as well as *Xenopus* oocytes. Based on sequence analysis of the transporter, as with the NAXT protein from *Arabidopsis*, the CrNPF2.9 lacks the EXXE(R/K) motif present in NPF transporters that is thought to couple transport to proton gradient and allow for active transport, therefore it is predicted that CrNPF2.9 is likely to act as a facilitative transporter.

The strong metabolic phenotype, resulting in the over accumulation of strictosidine, poses a number of intriguing questions over the regulation of the MIA metabolic pathway and the fact that this accumulation of strictosidine can result in cell death of the plant tissue is indicative that there is a need to limit the accumulation of strictosidine in plant tissue.

4.5 Future directions

There are a number of experiments that are needed to further assess the involvement of the NPF transporter class in MIA metabolism.

1.) As has been highlighted, although CrNPF2.4 and CrNPF2.6 are capable of transporting iridoid glycosides *in vitro*, there is a need to assess whether this is physiologically relevant to MIA metabolism. Individually silencing these two transporters did not result in a metabolic phenotype, and therefore combinatorially silencing these transporters is needed to assess physiological involvement in the pathway.

2.) There are seven other NPF homologues present in the *C. roseus* transcriptome, and individually silencing these other transporters may highlight an involvement in the MIA pathway. These other transporters do not co-regulate transcriptionally and are predicted to localise to the plasma membrane, therefore if they are involved in the MIA pathway it would be predicted that they are involved in the later stages of MIA biosynthesis.

3.) To date CrNPF2.9, despite resulting in a strong VIGS phenotype, has proven recalcitrant to assessment of transport in a heterologous system. Further work is needed to demonstrate that this transporter is able to transport strictosidine, and assess the mechanistic detail concerning whether this transport is a facilitative transporter, as is predicted form the sequence.

Altogether the work in this chapter has identified a tonoplast transporter, CrNF2.9, that, when silenced, leads to the accumulation of strictosidine. This phenotype is consistent with it being the exporter of strictosidine from the vacuole. Furthermore, this metabolic phenotype is accompanied by cell death, indicating that the export of strictosidine from the vacuole, and maintenance of a low level of strictosidine in the cell is likely to be fundamental for control of the MIA pathway.

5 Characterisation of Reticuline Oxidase Like enzymes in *C. roseus*

5.1 Introduction to the Reticuline Oxidase family

5.1.1 Flavoproteins

Flavin dependent enzymes are characterised by the binding of the cofactors flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) (Heuts et al., 2009). These cofactors can be associated with the apoprotein either non-covalently or by covalent attachment, and use the redox-active isoalloxazine ring system for catalysis. More than 90% of the FAD-dependent enzymes are oxidoreductases, however a few FAD-dependent proteins are also able to catalyse non-redox based reactions (Heuts et al., 2009).

FAD is a cofactor derived from riboflavin (Vitamin B_2), and can be synthesised *de novo* by plants, bacteria and fungi (Figure 5-1) (De Colibus and Mattevi, 2006). Because of the chemical properties of the isoalloxazine ring, FAD can exist in a number of different redox states, with both the one-electron reduced semiquinone and two-electron reduced dihydroflavin, being catalytically relevant entities (Walsh and Wencewicz, 2013) (Figure 5-1). The tight binding of FAD to the enzyme prevents its dissociation and subsequent uncontrolled autoxidation outside of the enzyme. This means that during catalysis of the oxidation reduction reaction, the FAD in the enzyme undergoes a catalytic cycle through two half reactions, to allow FAD to be regenerated (Walsh and Wencewicz, 2013).

The first reductive half reaction of FAD results in the oxidation of the substrate with concomitant reduction of the FAD, whilst the second re-oxidative half reaction results in the reoxidation of the reduced flavin (Figure 5-1) (Walsh and Wencewicz, 2013). The redox potential of free riboflavin is ~-200 mV, though the enzyme microenvironment in which the FAD is situated is able to tune this redox potential. The redox potential of the FAD in different flavoproteins has been measured at values ranging from -400 mV to +200 mV (Heuts et al., 2009).

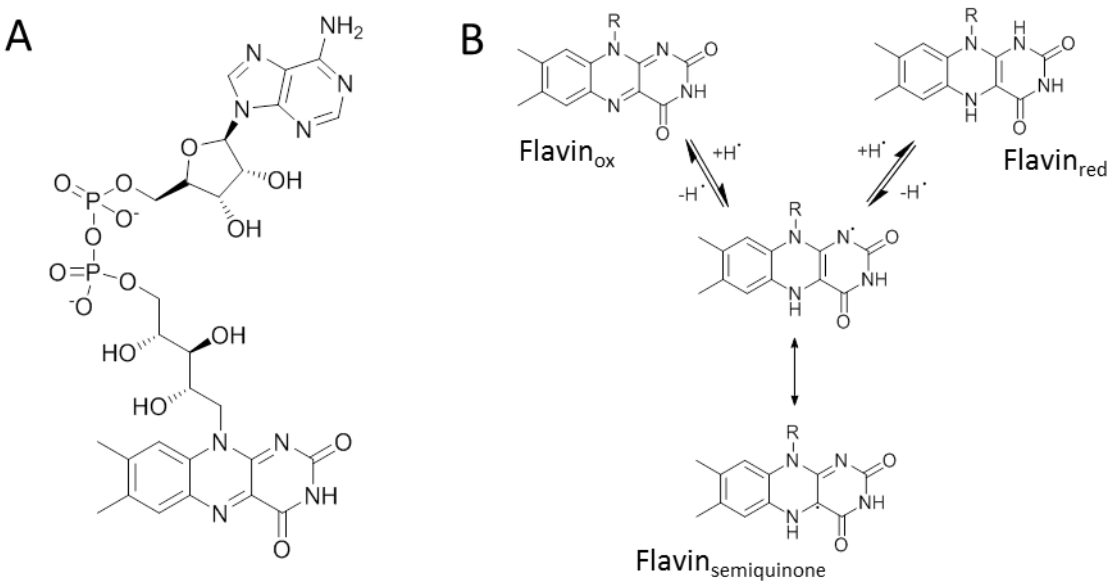


Figure 5-1 Flavin Adenine Dinucleotide
A.) Structure of Flavin Adenine Dinucleotide (FAD) B.) The three biologically accessible redox states of the isoalloxazine ring of FAD are oxidised Flavin, reduced Flavin and the semiquinone. Figure adapted from (Walsh and Wencewicz, 2013)

Flavin dependent enzymes can react with molecular oxygen in different ways. Flavin dependent oxidases utilise molecular oxygen as the final electron acceptor to reoxidise the reduced FAD, subsequently releasing hydrogen peroxide (Mattevi, 2006). Flavin dependent mono-oxygenases form a C4 α-(hydro)peroxide adduct of the flavin that allows insertion of an oxygen atom into the substrate, with release of H_2O (Mattevi, 2006). Flavin dependent dehydrogenases, do not react readily with oxygen, and utilise an alternative molecule, such as a quinone or redox protein, as the final electron acceptor to reoxidise FAD (Mattevi, 2006).

The diversity in the redox states of the isoalloxazine moiety of FAD, the fact the redox potential can be tuned, together with the differential nature by which FAD can interact with molecular oxygen, has ensured that flavoproteins are ubiquitous in nature and their chemical versatility has meant they are involved in a wide range of biological processes, act on a wide variety of substrates, and catalyse novel chemistry (Mattevi, 2006; Walsh and Wencewicz, 2013).

Sequence structure analysis of FAD-containing proteins suggests that there are four main structural folds; the glutathione reductase (GR) fold, ferredoxin reductase (FR) fold, the pyruvate oxidase (PO) fold, and the *p*-cresol methylhydroxylase (PCMH) or vanillyl alcohol oxidase (VAO)

fold (Dym and Eisenberg, 2001). Reticuline oxidases, the subject of this chapter, are part of the VAO family.

The VAO fold is characterised by two subdomains consisting of both α helices and β sheets. The members of the VAO family share a conserved FAD binding domain, in which the FAD cofactor is accommodated between the two sub domains, adopting an elongated conformation, with the enzyme active site at this interface (Dym and Eisenberg, 2001). Members of the VAO superfamily have been identified in all kingdoms of life and are involved in a diverse range of metabolic processes, although these enzymes primarily act as either oxidases or dehydrogenases.

Although 90% of FAD associated proteins bind FAD in a non-covalent manner, the VAO flavoprotein family has a relatively higher propensity for covalent binding, with one in four VAO family members predicted, based upon genome mining, to contain a conserved histidine residue that allows for histidyl-FAD attachment (Heuts et al., 2009). There are four main types of covalent attachment characterised in this enzyme family, monocovalent FAD linkage via either a histidine or tyrosine residue generating an 8α-N^3-histidyl-FAD, 8α-N^1-histidyl-FAD or 8α-*O*-tyrosyl-FAD attachment, as well as bicovalent linkage through attachment to a histidine and cysteine resulting in an 8α-N^1-histidyl-6-*S*-cysteinyl-FAD linkage (Figure 5-2). In plant enzymes, covalent attachment has only been shown via the monocovalent, 8α-N^1-histidyl-FAD and bicovalent 8α-N^1-histidyl-6-*S*-cysteinyl-FAD attachments, with the linking histidine present in the N-terminal part of the protein (Heuts et al., 2009).

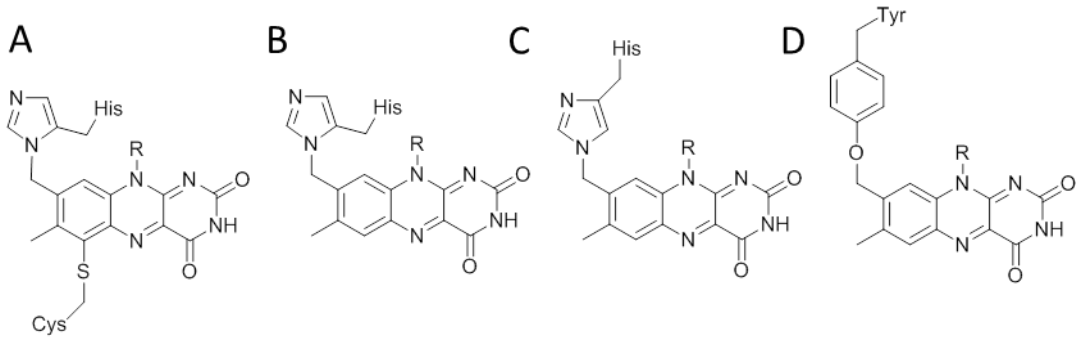


Figure 5-2 Types of covalent linkage for FAD
There are four main types of covalent linkage in FAD dependent enzymes A.) 8α-N^1-histidyl-6-*S*-cysteinyl-FAD B.) 8α-N^1-histidyl-FAD C.) 8α-N^3-histidyl-FAD D.) 8α-*O*-tyrosyl-FAD. Only 8α-N^1-histidyl-6-*S*-cysteinyl-FAD and 8α-N^1-histidyl-FAD have been seen *in planta.* (Conventional numbering has been used in this scheme).

5.1.1.1 The role of the Vanillyl Alcohol Oxidase Superfamily in planta

In planta the VAO superfamily is involved in a wide variety of biological processes, including catalysing novel oxidative reactions that lead to structural heterogeneity in secondary metabolites, developmental processes through controlling the catabolism and anabolism of plant hormones, and in plant defence through the production of hydrogen peroxide (Leferink et al., 2008a). There are three primary sub families of VAO protein *in planta* that have been characterised and shown to have both a physiological function and/or catalytic activity *in vitro*; the aldonolactone oxidoreductase family, the cytokinin dehydrogenase family, and the berberine bridge enzyme family (Leferink et al., 2008a).

The only plant VAO type enzyme to be characterised that does not contain a covalent linkage to FAD is L-galactono-1,4-lactone dehydrogenase (GALDH), which is involved in the biosynthesis of ascorbic acid in plants, oxidising L-galactono-1,4-lactone to L-ascorbate, with the concomitant reduction of cytochrome c (Figure 5-3) (Leferink et al., 2008b). This enzyme is an example of an aldonolactone dehydrogenase. Recombinant GALDH is also able to catalyse the conversion of gulono-1,4-lactone to L-ascorbate, however at a lower catalytic efficiency than what is thought to be its natural substrate L-galactono-1,4-lactone (Leferink et al., 2008b). This enzyme is integral to plant metabolism and its silencing by RNAi in tomato results in numerous plant and fruit developmental defects thought to be related to changes in ascorbate redox state, and mitochondrial function (Alhagdow et al., 2007).

Cytokinin dehydrogenase is an enzyme that contains a monocovalent histidyl-FAD and is involved in the degradation of the plant hormone family, the cytokinins, to adenine or adenosine by the oxidative cleavage of their sidechains (Werner et al., 2003). There are multiple families of cytokinin dehydrogenases that have different subcellular localisation, are expressed at different times in development, and have different substrate specificities for the cytokinin side chain. Since cytokinin dehydrogenases control cytokinin catabolism, these enzymes are integral in controlling the local cytokinin level in plant tissue and, consequently, the action of the hormone (Schmülling et al., 2003; Werner et al., 2003). Transgenic *Arabidopsis* overexpressing cytokinin dehydrogenases have a decreased cytokinin content, and display the associated deficiencies in shoot and flower development, coupled with deficiencies in leaf vasculature, and enhanced root growth and formation (Werner et al., 2003). The best studied of these enzymes is CKX1 from maize for which a crystal structure has been elucidated, revealing the VAO protein architecture (Malito et al., 2004).

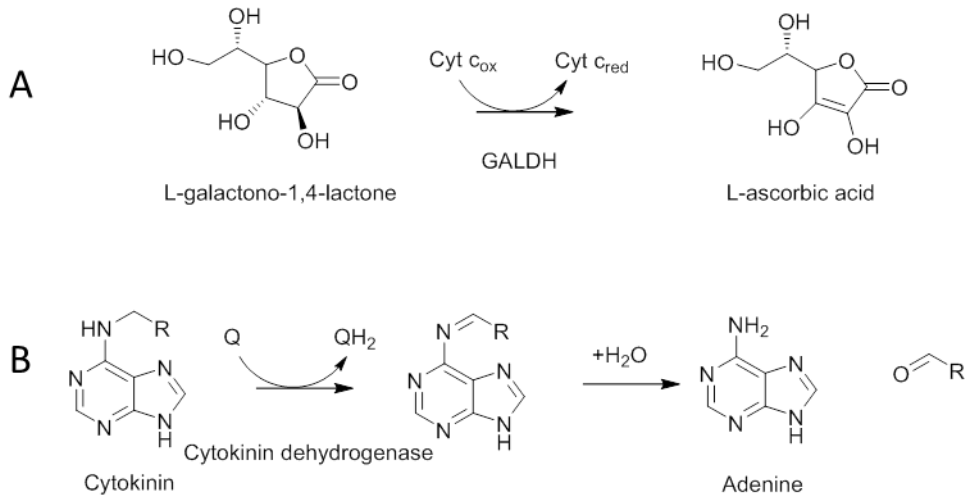


Figure 5-3 Role of VAO enzymes in plant metabolism
A.) GALDH catalyses the oxidation of L-galactono-1,4-lactone to L-ascorbic acid with reduction of cytochrome c B.) The cytokinin dehydrogenase family catalyses the degradation of cytokinins to adenine with concomitant reduction of an electron acceptor that is thought to be a quinone.

5.1.2 The Berberine Bridge Enzyme (BBE)-like family

The berberine bridge enzyme (BBE)-like family is a large subfamily of the VAO superfamily of FAD dependent oxidoreductases and these enzymes have been shown to have a diverse array of metabolic functions *in planta.*

The first role to be discovered for a member of the BBE enzyme family was the oxidation of reticuline to a berberine precursor in benzoisoquinoline alkaloid (BIA) biosynthesis (Dittrich and Kutchan, 1991). This class of alkaloids are present in species of the Papaveraceae, Ranunculaceae, Berberidaceae, Fumariaceae, and Menispermaceae and all are derived from two molecules of L-tyrosine (Figure 5-4) (Beaudoin and Facchini, 2014). This alkaloid family is diverse, with more than 2,500 structures described and contains a number of pharmaceutically relevant compounds, including the antimicrobial agent, sanguinarine, the cough suppressant, noscapine, as well as the narcotic analgesics, morphine and codeine (Beaudoin and Facchini, 2014). (*S*)-reticuline is a central metabolic intermediate that acts as a branchpoint towards the morphinan, protoberberine, phthalideisoquinoline and benzo[c]phenanthridine structural scaffolds found in the BIA family.

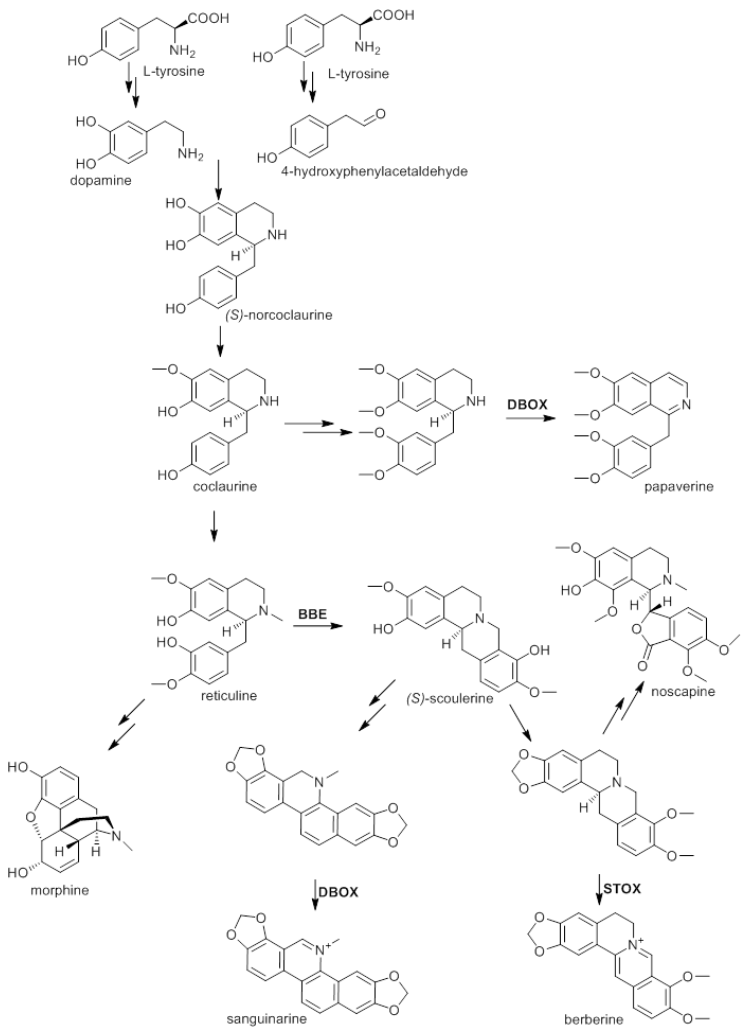


Figure 5-4 Overview of benzylisoquinoline alkaloid biosynthesis
Figure adapted from Beaudoin and Facchini, 2014.

The berberine bridge enzyme, also called reticuline oxidase, converts (*S*)-reticuline to the berberine precursor (*S*)-scoulerine by catalysing the oxidative cyclisation between the N-methyl moiety of (*S*)-reticuline into the berberine bridge carbon C-8 of (*S*)-scoulerine with concomitant release of H_2O_2 (Figure 5-5). This is the first step in the pathway towards the protoberberine, phthalideisoquinoline and benzo[c]phenanthridine alkaloids (Beaudoin and Facchini, 2014). The berberine bridge enzyme was originally discovered and purified to homogeneity from the cell culture of *Berberis beania* and was shown to turnover (*S*)-reticuline, (*S*)-protosinomenine and (*S*)-laudanosaline with stoichiometric release of H_2O_2 (Steffens et al., 1985). The gene for BBE was first cloned from *Eschscholtzia californica* cell suspension cultures, where it was shown that the level of the gene in the cell culture was induced by treatment with a fungal elicitor, suggesting that upregulation of this pathway enzyme, is integral to plant defence (Dittrich and

Kutchan, 1991). In this study the gene was cloned and heterologously expressed in *S. cerevisiae* where the enzyme resulted in the stereospecific conversion of (*S*)-reticuline to (*S*)-scoulerine (Dittrich and Kutchan, 1991). *In vitro* the enzyme has been tested with 23 isoquinoline alkaloids, and the enzyme was able to form the berberine bridge carbon with five compounds: (*S*)-reticuline, (*S*)-protosinomenine, (*R,S*)-crassifoline, (*R,S*)-laudanosoline and (*R,S*)-6-O-methyllaudanosoline (Figure 5-5) (Kutchan and Dittrich, 1995). As well as C-C bond formation to form the berberine bridge in tetrahydrobenzylisoquinoline alkaloids, BBE also catalyses the *N*-demethylation of *N*-methylcoclaurine, implicating a likely methylene imminium intermediate in the reaction mechanism (Figure 5-5). This is further highlighted by the fact BBE was shown to have oxidative activity against some tetrahydroprotoberberine alkaloids, namely (*S*)-coreximine and (*S*)-norseponine, forming a double bond between the nitrogen and the berberine bridge C-8 carbon (Kutchan and Dittrich, 1995). This activity has also been observed for the tetrahydroprotoberberine alkaloid (*S*)-scoulerine which is thought to be the natural product of BBE (Winkler et al., 2009b) (Figure 5-5).

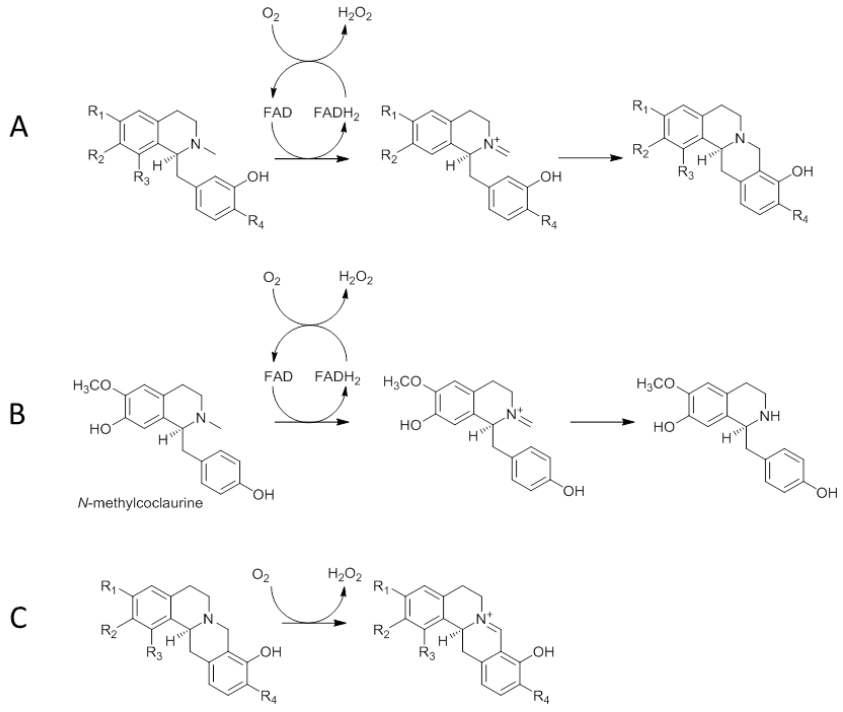


Figure 5-5 Reactions catalysed by the canonical BBE from *E. californica* and *P. somniferum*
A.) The canonical formation of the carbon carbon bond for formation of the berberine bridge B.) BBE catalyses the *N*-demethylation of *N*-methylcoclaurine, implicating a likely methylene iminium intermediate in the reaction mechanism C.) Oxidation of tetrahydroprotoberberine alkaloids

The homologue of BBE from opium poppy (*Papaver somniferum*) has also been cloned (Facchini et al., 1996), and it has 70% identity to that found in *E. californica*. As in *E. californica* cell

suspension cultures, this gene was upregulated by treatment of *Papaver somniferum* cell suspension cultures with a fungal elicitor. The BBE gene also showed a distinct spatial distribution, being primarily localised to the root and stems of mature opium poppy, which is consistent with its involvement in sanguinarine biosynthesis, which is primarily localised to the root (Facchini et al., 1996). The BBE from *P. somniferum* has been shown to contain an N-terminal signal sequence that targets BBE to the endoplasmic reticulum, as well as an adjacent vacuolar targeting sequence that leads to its subsequent translocation to this compartment (Bird and Facchini, 2001). BBE from opium poppy has been shown to have the same activity against (*S*)-reticuline as the homologue found in *E. californica* (Hagel et al., 2012). There are five other full length BBE like proteins present in the transcriptome of opium poppy, one of which has 94% identity to the opium poppy BBE and is assumed to have the same biological activity, as it can catalyse the same reactions *in vitro* (Hagel et al., 2012).

There have been a number of reports of the role of the BBE in planta and whether it's silencing leads to an altered metabolic phenotype (Frick et al., 2004; Hagel et al., 2012; Park et al., 2002; Park et al., 2003). In cell suspension cultures of *E. californica*, cell lines expressing antisense BBE mRNA transcripts resulted in a reduction in BBE mRNA levels, together with a decrease in BBE activity. This decrease in BBE activity was accompanied by a decrease in benzophenanthridine alkaloids, an alteration in the free amino acid pools in the cell culture, together with a reduction in the growth rate of the cell culture, however there is no concomitant increase in pathway intermediates preceding BBE in the antisense lines for this enzyme (Park et al., 2002). A similar phenomenon was seen in transgenic root cultures of *E. californica*. Antisense BBE expression again led to a decrease in BBE activity, together with a decrease in benzylphenanthridine alkaloids, a change in the size of the free amino acid pool and altered growth rate of the culture. In comparison overexpression of BBE resulted in a two fold increase in BBE activity, together with a five fold increase in benzylphenanthridine alklaoids (Park et al., 2003). In whole plants of opium poppy, the stable transformation of antisense BBE results in an increase in reticuline, laudanine, laudasonine, dehydroreticuline, salutaridine, and (*S*)-scoulerine and altered the ratio of morphinan and tetrahydrobenzylisoquinonline alkaloids in latex, however no change in benzophenanthridine alkaloids, such as sanguinarine, occurred in roots (Frick et al., 2004). This is interesting, as a decrease in sanguinarine might be expected as BBE catalyses the branchpoint reaction to (*S*)-scoulerine. VIGS of BBE in opium poppy using a silencing construct that targets both BBE homologues resulted in a decrease in sanguinarine

and dihydrosanguinarine, together with an increase in the substrate reticuline (Hagel et al., 2012). The combined evidence from cell suspension cultures and whole plants is indicative that BBE plays a role in benzylisoquinoline biosynthesis in both California and opium poppy.

In comparison overexpression of BBE resulted in a two fold increase in BBE activity, together with a five fold increase in benzylphenanthridine alklaoids (Park et al., 2003).

As well as in the conversion of (*S*)-reticuline to (*S*)-scoulerine and the unusual formation of the berberine bridge carbon, BBE-like enzymes have also been found to be involved in other aspects of benzylisoquinoline alkaloid biosynthesis.

There are six full length homologues of BBE present in opium poppy. Two are thought to be involved in performing the canonical BBE reaction, conversion of (*S*)-reticuline to (*S*)-scoulerine (Hagel et al., 2012). Three of the remaining homologues could not be expressed in *Pichia pastoris*, and when silenced in opium poppy no change in the metabolic phenotype was observed. The remaining homologue displayed novel enzymatic activity in the benzylisoquinoline pathway, catalysing the oxidation of dihydrosanguinarine to sanguinarine, and was therefore named dihydrobenzophenanthridine oxidase (DBOX) (Figure 5-6) (Hagel et al., 2012). This enzyme also similarly oxidised a number of protoberberine and benzylisoquinoline alkaloids, including scoulerine, tetrahydropapaverine tetrahydrocolumbamine, tetrahydropalmitine, canadine and stylopine, and therefore displayed (*S*)-tetrahydroprotoberberine oxidase (STOX) activity, suggesting this enzyme is a multifunctional oxidase. VIGS of this enzyme resulted in a decrease in sanguinarine, dihydrosanguinarine and papaverine, suggesting it is an integral enzyme in BIA metabolism (Hagel et al., 2012).

Other BIA producing plants also contain a third type of BBE-like protein that has protoberberine dehydrogenase activity, (*S*)-tetrahydroprotoberberine oxidase, (STOX). This enzyme activity was originally identified in suspension cultures of *Berberis* (Amann et al., 1984). The gene was cloned and expressed in Sf9 insect cell cultures, together with a homologue from *Argemone mexicana* (Gesell et al., 2011). Both BwSTOX and AmSTOX were able to act as oxidases on a number of different protoberberine scaffolds, and showed differential substrate specificity. Both acted on (*S*)-tetrahydropalmatine, generating two double bonds to form palmatine, whilst BwSTOX showed activity towards (*S*)-scoulerine, and converted (*S*)-canadine to berberine, and AmSTOX demonstrated *in vitro* activity towards (*S*)-corexime generating a new alkaloid product that does not have a conjugated ring system (Figure 5-6). It has been shown that BBE enzyme from *E.*

californica also has STOX activity towards its end product (*S*)-scoulerine, forming dehydroscoulerine, however this reaction occurs at a rate around 3000 times lower than that of its primary reaction (Winkler et al., 2009b).

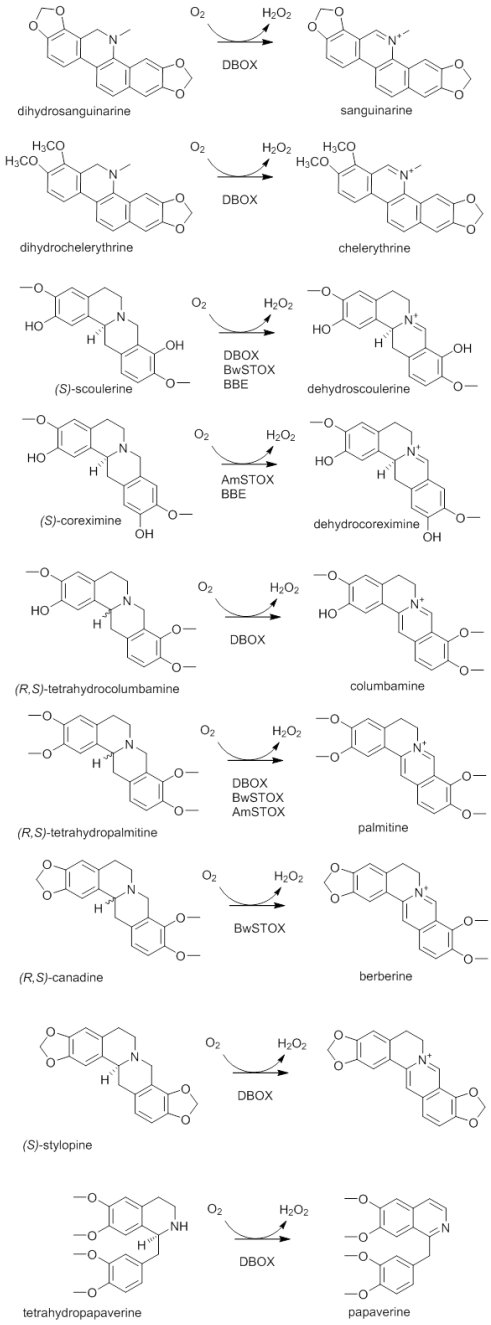


Figure 5-6 Proposed roles of Berberine Bridge Like enzymes in benzylisoquinoline alkaloid biosynthesis
Each reaction involves the oxidation of a carbon-nitrogen bond

5.1.2.1 The role of berberine bridge enzyme family in pyridine alkaloid metabolism

BBE-like genes have also been implicated in the biosynthesis of the pyridine alkaloid, nicotine, in tobacco (*Nicotiana tabacum*) (Kajikawa et al., 2011). Four full length BBE-like genes were identified in the tobacco genome (*BBLa*, *b*, *c*, *d*), and their expression was shown to be under the transcriptional control of the *NIC* regulatory locus, induced by MeJA and highly expressed in roots, which is similar to the known enzymes involved in nicotine biosynthesis. Silencing these *BBL* genes either constitutively or inducibly, by RNA interference in tobacco hairy roots resulted in decreased nicotine levels, together with the accumulation of a novel metabolite identified as dihydrometanicotine (DMN) (Kajikawa et al., 2011). In transgenic tobacco plants in which the *BBL* genes were silenced, nicotine levels were lower in the leaves and roots compared to wild type controls, whilst DMN was the most abundant alkaloid in roots in silenced lines. Silencing of the *BBL* genes in tobacco BY2 cell suspension cultures resulted in a decrease in the accumulation of the pyridine alkaloids anatabine, anabasine, though DMN did not accumulate (Kajikawa et al., 2011).

This work was further validated in field grown tobacco in which transgenic tobacco carrying an RNAi construct to reduce *BBL* expression, also resulted in a decrease in nicotine levels (Lewis et al., 2015). Additionally, an EMS mutant screen for disruption of each of the BBL homologues, *BBLa*, *b* and *c*, combined with sexual crossing to generate all possible homozygous combinations demonstrated that *BBLa* had the biggest single effect on alkaloid content, however the greatest decrease in nicotine and total alkaloid levels occurred in the homozygous line that disrupted all three *BBLa*, *b*, *c* genes (Lewis et al., 2015).

Collectively, these data strongly implicate the BBL-genes in nicotine biosynthesis in *Nicotiana tabacum*. The data suggest that the BBL proteins oxidise DMN to form nicotine (Figure 5-7). As with BBE from *E. californica*, the tobacco BBL proteins contain an N-terminal signal peptide that targets the enzymes for localisation to the vacuole. Recombinant expression of BBLa in *Pichia pastoris* resulted in the expression of a heavily glycosylated protein that was shown to bind FAD, though incubation of the enzyme with DMN did not lead to the predicted conversion to nicotine, and therefore the exact enzymatic reaction catalysed by these tobacco BBL proteins is still not known (Kajikawa et al., 2011).

5.1.2.2 The role of berberine bridge enzyme family in terpenophenolic metabolism

BBE-like proteins have also been implicated in the formation of both Δ9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) synthesis in *Cannabis sativa* (Sirikantaramas et al., 2004; Taura et al., 2007). Marijuana contains unique terpenophenolics called cannabinoids that are psychoactive, however also possess analgesic and anti-inflammatory properties. Psychoactive *C. sativa* contains a higher level of THCA, whilst fiber-type *Cannabis* has a higher proportion of CBDA. These molecules are derived from the same precursor, cannabigerolic acid (CBGA), and therefore the relative activity of THCA synthase to CBDA synthase is important in determining the psychoactivity of the *Cannabis* strain (Onofri et al., 2015; Taura et al., 2007). THCA synthase, a BBE-like enzyme, was shown to be a flavinylated oxidase that could catalyse the oxidative cyclisation of CBGA to THCA (Figure 5-7) forming a carbon oxygen bond and releasing H_2O_2, when expressed recombinantly in tobacco hairy roots or insect cell culture (Sirikantaramas et al., 2004; Taura et al., 1996; Taura et al., 1995). It contains a 28 amino acid signal peptide that likely targets it to the secretory pathway.

Recombinant THCA synthase expressed in insect cells or tobacco BY2 cells is secreted to the culture medium, whilst overexpression of THCA synthase as a GFP fusion in transgenic *N. tabacum* resulted in its localisation to the storage cavity of glandular trichomes. THCA synthase activity was detected in endogenous *Cannabis sativa* trichomes, and suggested that the storage cavity is both the site of cannabinoid accumulation and biosynthesis (Sirikantaramas et al., 2005). Interestingly CBDA synthase is also a BBE-like enzyme that has 84% identity to THCA synthase. Both of these enzymes catalyse the oxidative cyclisation of CBGA to either THCA or CBDA respectively, and are thought to proceed through a similar mechanism (Taura et al., 1996; Taura et al., 2007). The distinct catalytic activity of THCA synthase or CBDA synthase therefore plays a role in controlling the chemotype of the *C. sativa* variety (Taura et al., 2007). Sequence heterogeneity of CBDA and THCA synthases in *C. sativa* is a key determinant of chemical phenotype and determining whether a strain is psychoactive. With increasing availability of sequences for *C. sativa* transcriptomes and genomes, there is a growing appreciation that there are multiple CBDA- and THCA like synthases expressed (Onofri et al., 2015). Analysis of different *C. sativa* lines with different mutations in CBDA and THCA homologues, along with analysis of the corresponding chemical phenotype, suggested that many mutations that impair functionality of these enzymes in *C. sativa* led to an increase in the precursor CBGA as would be expected (Onofri et al., 2015; Shoyama et al., 2012).

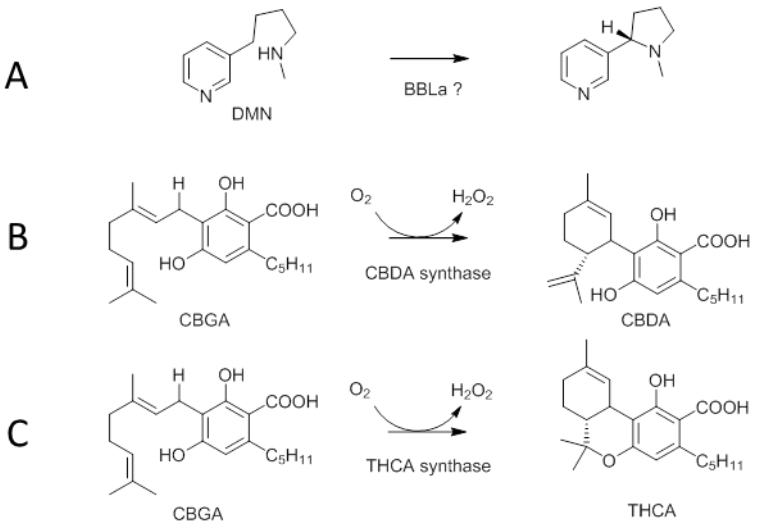


Figure 5-7 The role of BBE like enzymes in specialised plant metabolism
Outside of benzylisoquinoline biosynthesis BBE like enzymes are involved in pyridine and terpenophenolic metabolism. A.) Based on silencing experiments in tobacco, BBLa is thought to lead to the conversion of dihydrometanicotine (DMN) to nicotine, however this has not yet been shown *in vitro* B.) The enzyme CBDA synthase catalyses the conversion of cannaigeriolic acid (CBGA) to cannabidiolic acid (CBDA) C.) The enzyme THCA synthase catalyses the conversion of cannabigeriolic acid (CBGA) to tetrahydrocannabinolic acid (THCA).

5.1.2.3 The role of the berberine bridge enzyme family in primary metabolism

BBE-like enzymes are also involved in a diverse range of metabolic processes outside of secondary metabolism. In *Arabidopsis,* a plant that does not contain benzylisoquinoline, pyridine alkaloids, or cannabinoids, there are 28 different members of this enzyme family (Daniel et al., 2015). This family of enzymes has been shown to be expressed during certain developmental stages as well as in response to environmental stresses. Two of these genes AtBBE-like 13 and 15, have been functionally characterised. AtBBE-like 15 is located in the cell wall and gene disruption of this enzyme resulted in defects in female gametophytic development. Heterologous expression in *Pichia pastoris*, and screening against a targeted library of compounds demonstrated that both AtBBE-like 15 and AtBBE-like 13, showed catalytic activity in which cinnamyl-, *p*-coumaryl, -coniferul and sinapyl alcohol, are converted to the corresponding aldehydes, and suggested that these enzymes may act as monolignol oxidoreductases (Daniel et al., 2015). The localisation of the AtBBE-like 15 to the apoplastic space is consistent with a physiological role for this activity, as monolignols are thought to be exported in their alcohol forms to the apoplast, where they act as substrates for cell wall polymerization (Daniel et al., 2015). Furthermore, this enzyme co-expressed with known genes involved in monolignol biosynthesis, including phenylalanine ammonia lyase, an enzyme involved in entry into the phenylpropanoid pathway, as well as monolignol glucoside hydrolases.

Structural analysis of the active site residues in the AtBBE-like 13 and 15 enzymes, identified important structural sequences that are involved in substrate recognition, and based on alignments to the other 28 members of this family implicated a further 12 of the AtBBE-like proteins as acting as alcohol oxidases, however this needs to be experimentally verified (Daniel et al., 2015).

BBE-like proteins in *Helianthus annus* (Ha-CHOX), and *Lactuca sativa* (Ls-CHOX) were shown to be carbohydrate oxidases, with the ability to oxidise a range of monosaccharide and disaccharide substrates and led to the formation of H_2O_2 (Custers et al., 2004). With glucose as a substrate, Ha-CHOX forms δ-glucono-1,5-lactone. Despite homology to BBE-like enzymes, it is unable to catalyse the conversion of (*S*)-reticuline to (*S*)-scoulerine and form the berberine bridge carbon. Transcription of the *Ha-CHOX* gene from sunflower was shown to be upregulated in response to salicylic acid, as well as in response to the plant pathogens *Sclerotinia sclerotiorum, Botyrtis cinerea* and *Diaporthe helianthi* and transgenic tobacco overexpressing Ha-CHOX was shown to be more resistant to *Pectobacterium carotovorum* (Custers et al., 2004). It is thought therefore that Ha-CHOX is an active part of the plant defence response, as it is upregulated under pathogen attack, and contributes to formation of H_2O_2 as a result of carbohydrate oxidation (Custers et al., 2004).

Similarly in ornamental tobacco, a BBE-like protein called Nectarin V is also thought to be involved in plant defence (Carter and Thornburg, 2004). Nectarin V was purified from the nectar of ornamental tobacco, subjected to N-terminal sequencing, and the cDNA encoding the full length protein was elucidated. The temporal expression of *Nec5* suggested it was only expressed during maturation of the nectary, and spatially is only localised to this organ. Plant nectar in tobacco does not contain alkaloids, however does contain a range of other mono- and di-saccharides, amino acids, organic acid vitamins and oils. Nectarin V was shown to display carbohydrate oxidase activity, and in particular was active against glucose. This carbohydrate oxidase activity is again thought to lead to the high concentration of H_2O_2 found in nectar, together with the action of Nectarin I which is a superoxide dismutase (Carter and Thornburg, 2004).

The timothy grass pollen allergen Phlp4, protein has been shown to be homologous to BBE and is a contributory factor in respiratory allergies (DeWitt et al., 2006). Enzymatically it catalyses the oxidation of D-glucose, however unusually for a bicovalently linked FAD BBE-like protein, it acts as a dehydrogenase as opposed to an oxidase (Zafred et al., 2015).

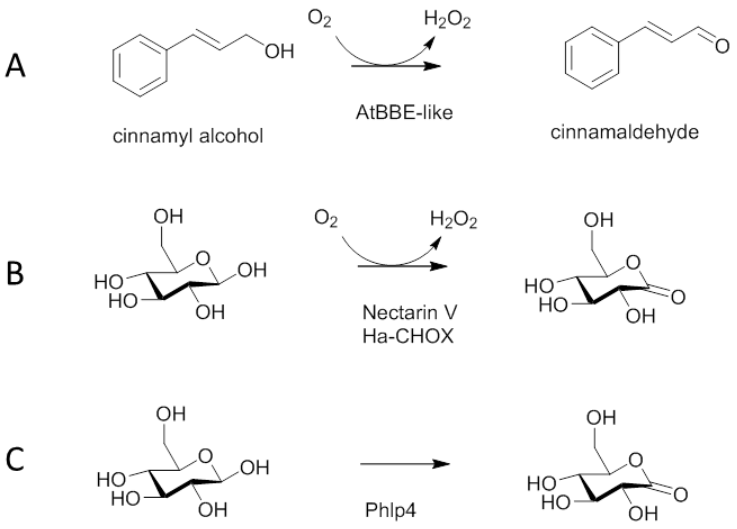


Figure 5-8 The role of BBE-Like enzymes in primary metabolism
A.) AtBBE-like proteins have been shown to be involved in the generation of monolignol aldehydes involved in cell wall biosynthesis. B.) Nectarin V and Ha-CHOX catalysed the oxidation of D-glucose to glucono 1,5-lactone, with the release of H_2O_2 which is thought to contribute to plant defence. C.) Similarly the pollen allergen Phlp4 catalyses the same reaction, however is a dehydrogenase and doesn't use oxygen as a final electron acceptor. The identity of the endogenous electron acceptor is not known.

The fact that BBE-like proteins are ubiquitous in the plant kingdom, and that many species contain multiple BBE-like homologues, is suggestive that this class of enzymes has a general role in plant metabolism. The aforementioned specialised roles of BBE-like enzymes in alkaloid and cannabinoid biosynthesis may have arisen by gene duplication and mutational diversification from primitive oxidoreductases (Daniel et al., 2015).

5.1.3 The process of autoflavinylation and the role of covalent flavinylation

With the exception of GALDH, all VAO family proteins from plants have been shown to contain either monocovalent or bicovalent attachment of FAD and there has been debate as to the role covalent attachment plays in catalysis.

In all VAO proteins studied to date, flavinylation occurs autocatalytically, and for bi-covalently associated FAD, the 8α-*N*[1]-histidyl and 6-*S*-cysteinyl-FAD covalent attachment in the enzyme occurs independently. This has been shown for BBE from *E. californica*, THCA synthase from *C. sativa* as well as in glucooligosaccharide oxidase (GOOX) from *Acremonium strictum* and chitooligosaccharide oxidase (ChitO) from *Fusarium graminearum*. In all cases single mutations of either the histidine or cysteine FAD linkages in the enzymes still resulted in FAD covalently

linked to the enzyme, suggesting each covalent attachment is formed independently (Heuts et al., 2008; Huang et al., 2008; Sirikantaramas et al., 2004; Winkler et al., 2007).

There are a number of sequence determinants for FAD covalently binding to VAO type enzymes. GALDH, a VAO protein from *A. thaliana* that binds FAD non-covalently, has a leucine in the position that is usually a conserved histidine for covalent attachment. Mutagenesis of the leucine to an isoleucine, phenylalanine or histidine destabilised the protein, and all mutants bound FAD non-covalently. This therefore suggested that despite the presence of histidine in the L56H mutant, other residues are required for autoflavinylation (Leferink et al., 2008b). This has been further shown for the VAO protein from *Penicillium simplicissimum*. The mutation H61T prevented the formation of a covalent link between H422 and FAD, suggesting this H61 is important for activating H422 to covalently react with FAD (Tahallah et al., 2002).

The redox potential of flavins can be influenced by the surrounding chemical environment, and therefore the protein environment of the enzymatic active site may have a direct effect on its redox potential (Heuts et al., 2009). It has been shown that covalent coupling of flavin increases the midpoint potential of FAD, with bicovalent attachment having the greatest effect (Heuts et al., 2009). An increase in the midpoint potential of the FAD allows the enzyme to oxidise the substrate more efficiently. This has been elegantly shown for the BBE from *E. californica.* This enzyme has a bicovalent attachment to FAD at H104 and C166 (Winkler et al., 2006; Winkler et al., 2008; Winkler et al., 2009a). The wild type protein has a high midpoint potential of +132 mV. Mutagenesis of the amino acids H104 and C166 demonstrated that covalent attachment is integral to achieving this high midpoint potential (Winkler et al., 2007). Mutation of the histidine to alanine or threonine resulted in low yields of protein, suggesting the histidyl-FAD covalent attachment might be integral to the stability of the protein. In comparison mutating the C166 to an alanine resulted in good yield of protein as it still had a single covalent link to H104, however it resulted in a decrease in the redox potential to +53 mV (Winkler et al., 2007). Additionally, the steady state turnover of the enzyme decreased, with a switch in the rate limiting step in catalysis from re-oxidation of the FAD, to the oxidation of the (Winkler et al., 2007). This suggested the bicovalent attachment fine tunes the FAD redox potential for improved catalytic efficiency. In comparison, for THCA synthase, the mutation of the H114 still resulted in soluble protein, however this mutation abolished enzyme activity suggesting the histidyl-FAD is integral to catalysis (Sirikantaramas et al., 2004). This has also been shown for bicovalent VAO proteins outside of plants. Glucooligosaccharide oxidase from *Acremonium strictum*, has been crystallised and was the first enzyme identified to show bicovalent attachment of FAD (Huang

et al., 2005). In each of the single mutants H70A and C130A, there was a decrease in the midpoint potential of the FAD from +126 mV to +69 mV and +61 mV respectively, which correlated with a decrease in the activity of the enzyme. The double mutant H70A/C130A, which did not contain FAD, was inactive even upon supplementation of the assay with exogenous FAD (Huang et al., 2008). In the context of GOOX, the covalent attachment appears to be essential for the structural integrity of a well ordered FAD binding cavity in the enzyme active site (Huang et al., 2008). Conversely in Chito-oligosaccahride oxidase (ChitO) from *Fusarium graminearum*, another bicovalent VAO type protein, mutagenesis of C154A reduced the redox potential from +131 mV to +70 mV, whilst mutation of H94A resulted in an increased redox potential, though both mutations resulted in a decrease in enzymatic activity (Heuts et al., 2008). This suggests that although C154A is likely involved in tuning the redox potential of the FAD, the bicovalent nature of the FAD linkage may be important for the correct positioning of the cofactor in the active site of the enzyme (Heuts et al., 2008).

A key proposed role of covalent flavinylation, as well as its impact on redox potential, has been its implied involvement in the structural integrity of the protein. In many bicovalently linked proteins, mutations that led to the mono-covalent attachment of FAD were still viable and soluble but had lower catalytic efficiency. Mutations that completely abolish FAD binding usually cannot be expressed, as has been shown for BBE from *E. californica* (Winkler et al., 2007) and ChitO from *F. graminearum* (Heuts et al., 2008). For GOOX, from *Acremonium strictum,* although the double mutant H70A/C130A, was expressed, it did not contain FAD, suggesting that removal of both covalent linkages to FAD from the protein alters the structure of the protein and prevents FAD binding non-covalently, demonstrating that the bicovalent linkage is required for correct FAD positioning as well as in catalysis (Huang et al., 2008). However for VAO, as mentioned, mutagenesis of the histidyl-FAD covalent link leads to a reduction in catalytic efficiency, though the mutants are all able to tightly bind FAD non-covalently, demonstrating that the covalent linkage is not needed to form a suitable FAD binding site (Tahallah et al., 2002). Similarly in the bicovalently linked protein GilR, an oxidoreductase involved in gilovarcin biosynthesis the double mutant H65A/C125A, despite having no enzymatic activity, was appropriately folded and bound FAD non-covalently (Noinaj et al., 2011). This therefore suggests that the role of covalent flavinylation on protein structural integrity is protein dependent.

The majority of covalently linked VAO proteins are oxidases, both *in planta* and in other kingdoms of life, and are able to utilise molecular oxygen as the final electron acceptor. One of the postulated reasons for this is that covalent flavinylation increases the redox potential of the

FAD and therefore limits the redox partners that can be used in electron transfer (Heuts et al., 2009). However there are exceptions where proteins that have a covalent-FAD linkage act as dehydrogenases. The archetypal example of this is the cytokinin dehydrogenase family of enzymes, which have been shown to react sluggishly with molecular oxygen and instead have a preference for a range of quinones that act as the final electron acceptor to oxidise FAD (Frébortová et al., 2004; Frébortová et al., 2010). Additionally, it has been shown that bicovalent plant allergen Phlp 4, despite a very high redox potential of +200 mV, does not react well with oxygen, and therefore likely utilises an alternative electron acceptor for re-oxidation of FAD. Mutagenesis studies of the BBE, Phlp4 and GALDH enzymes have identified two key residues termed 'the gatekeeper' (Leferink et al., 2009) and 'alternative gatekeeper' (Zafred et al., 2015). These residues are involved in forming an oxygen pocket that influences the reactivity of the enzyme towards molecular oxygen. In particular, through mutagenesis, it was shown that it is possible to increase the oxygen reactivity of the dehydrogenase Phlp4 (Zafred et al., 2015), to convert the dehydrogenase GALDH, that binds FAD non-covalently, into an oxidase (Leferink et al., 2009), and to decrease the reactivity of BBE towards molecular oxygen (Zafred et al., 2015). This work increases the predictive power through sequence analysis, of whether a VAO enzyme is likely to react with molecular oxygen as the final electron acceptor. However it should be noted that, to date, all VAO type enzymes that bind FAD non-covalently act as dehydrogenases.

5.1.4 Summary for Introduction

The BBE family, which is conserved throughout the plant kingdom, has been shown to catalyse varied and unique oxidation reactions on a plethora of structurally diverse substrates. Members of this protein family have been implicated in the metabolism of three different families of natural products, the benzylisoquinoine alkaloids, pyridine alkaloids and terpenophenolics. Therefore, this family is an attractive target to search for biosynthetic gene candidates in MIA metabolism in *C. roseus*.

5.2 Aims of this Chapter

1.) Assess whether any VAO enzymes are involved in MIA biosynthesis in *C. roseus* using an *in planta* loss of function screen.

2.) Functionally characterise the VAO enzymes that display a metabolic phenotype.

5.3 Results

5.3.1 The reticuline oxidase family in Catharanthus roseus

As discussed in Chapter II the reticuline oxidase like protein annotated as locus_5608 in the *C. roseus* transcriptome, co-expresses with a number of genes that have previously been shown to be involved in MIA metabolism in *C. roseus*, including strictosidine synthase, secologanin synthase, tryptophan decarboxylase, tryptophan synthase as well as the transporters characterised in Chapter III and Chapter IV (Figure 5-9). Due to this co-expression it was therefore assessed to be a good candidate for its potential involvement in MIA metabolism, and will hereafter be referred to as CrRO.

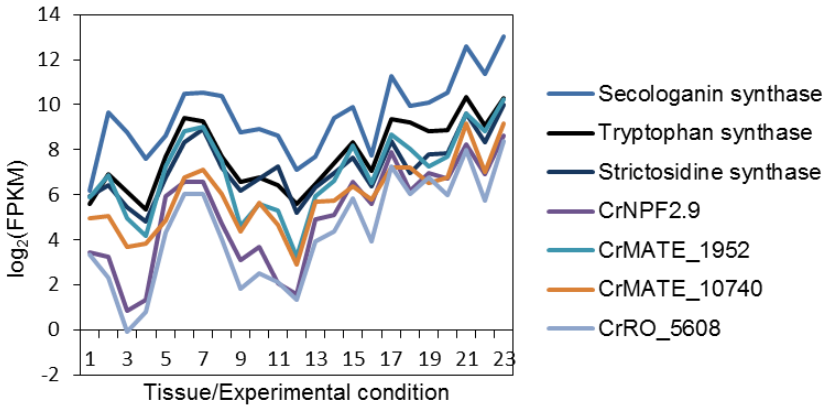


Figure 5-9 Transcriptomic profile of CrRO
Analysis of the expression profile of CrRO across 23 different tissues and experimental conditions, which have previously been described in Chapter II demonstrates that it has a similar expression pattern as known genes in the MIA biosynthetic pathway, as well as a similar expression pattern to the transporters CrNPF2.9, CrMATE1952 and CrMATE10740 that have been demonstrated to result in a VIGS phenotype earlier in this thesis.

CrRO encodes for a protein 530 amino acids in length and the full length gene was cloned from *C. roseus* leaf cDNA into a pCR8GW gateway compatible entry vector and sequenced. The sequence was consistent with that found in the MPGR *C. roseus* transcriptome, and genome, demonstrating that the gene is real, is expressed in leaf tissue and that the gene was correctly assembled in the transcriptome.

The *C. roseus* genome shows that *CrRO* does not contain any introns and exists as one contiguous exon, which is consistent with reticuline oxidases from other plant species (Facchini et al., 1996). There is one homologue of *CrRO* present on the *C. roseus* genome that has a greater than 70%

identity at the amino acid level and encodes for a 550 amino acid protein. Interestingly these proteins are spatially close on the genome and are both present on the same 23,195 bp genomic contig. This is suggestive that these two genes could have arisen from a gene duplication event. The homologue of CrRO has the same genomic organisation as CrRO, and again occurs as one contiguous exon (Figure 5-10).

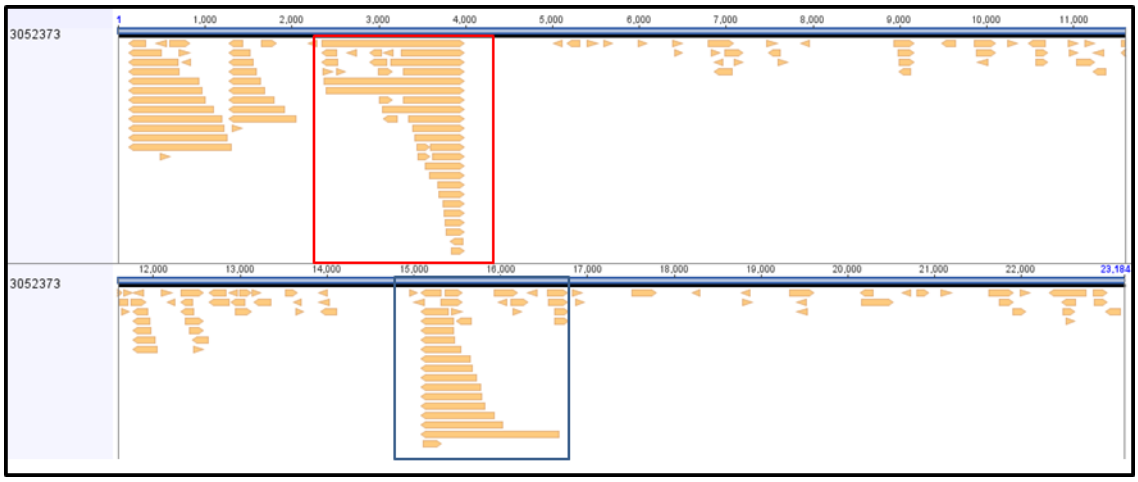


Figure 5-10 Genomic organisation of the CrRO
There are two reticuline oxidases, CrRO and CrRO homologue, spatially close on a genomic contig ~23,000 bp in length. Open reading frames are highlighted by orange arrows. As with other characterised reticuline oxidases each gene occurs as one contiguous exon. The CrRO is highlighted by a blue box, and the CrRO homologue highlighted by a red box.

A basic local alignment using BLAST of the CrRO homologue against the *C. roseus* transcriptome did not identify a transcript that had a complete open reading frame, unlike CrRO, but instead had greatest sequence identity to a transcript annotated as locus_15353 that was misassembled. This misassembled transcript was expressed at low levels in the *C. roseus* transcriptome in all tissues. A BLAST of the CrRO homologue against two other openly available *C. roseus* transcriptomes, Cathacyc (Van Moerkercke et al., 2013), and Phytometasyn (Xiao et al., 2013) also failed to retrieve a full length transcript of the CrRO homologue.

To assess whether the CrRO homologue gene is actually expressed in leaf tissue, primers were designed to amplify the full length gene of the CrRO homologue from *C. roseus* leaf cDNA using the sequence from the genome, to amplify a truncation of the CrRO homologue that aligns better with the start codon of the CrRO, and to amplify two internal regions of the CrRO homologue gene 100 bp in length. As a positive control the full length CrRO and 100 bp of the CrRO were amplified.

In all cases no band of the correct size could be amplified from leaf cDNA for the CrRO homologue (Figure 5-11) indicating that despite having significant sequence identity to CrRO, the CrRO homologue does not appear to be expressed in leaf tissue and therefore is unlikely to have a role in leaf MIA metabolism, and may be a pseudogene.

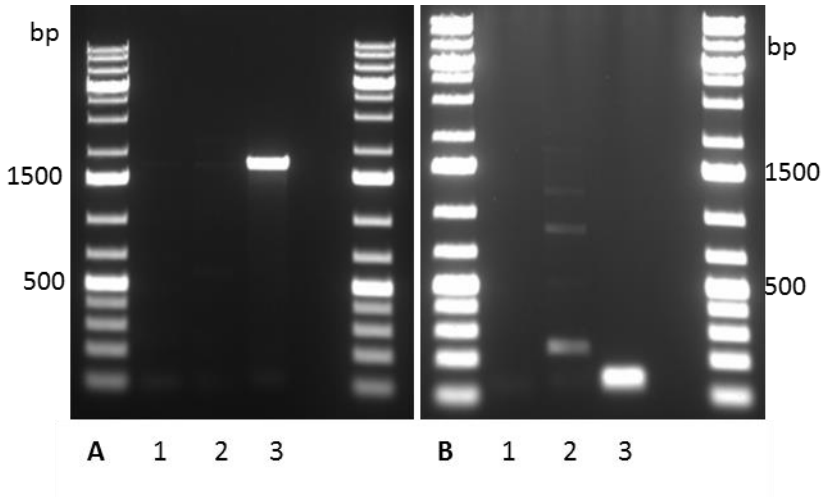


Figure 5-11 PCR amplification of CrRO and CrRO Homologue from *C. roseus* cDNA
A.) Full length PCR amplification of the CrRO homologue from *C. roseus* leaf cDNA. 1.) Amplification full length CrRO homologue 2.) Amplification N-terminal truncation of CrRO homologue 3.) Amplification CrRO gene. **B.)** Amplification of 100 bp regions of CrRO homologue 1.) Amplification section1 100 bp CrRO homologue 2.) Amplification section 2 100 bp CrRO homologue 3.) Amplification 100bp of CrRO

5.3.1.1 Assessment of the CrRO sequence

CrRO does not contain the conserved sequence required for the bicovalent attachment of FAD. The conserved histidine residue that ordinarily coordinates attachment to FAD, as has been found in all BBE-like enzymes *in planta,* is not present, and instead has an alanine at this position. Additionally the cysteine that is conserved and allows for bicovalent attachment of FAD is also not present and is substituted with a serine (Figure 5-12, Figure 5-13). This suggests that FAD is likely to bind non-covalently in CrRO. Although more than 90% of enzymes that bind FAD have a non-covalent attachment, the enzymes that have been characterised to date in the BBE like enzyme subfamily all contain a conserved motif that allows for at least monocovalent attachment to FAD.

It is therefore the first protein to my knowledge that has sequence identity to the berberine bridge enzyme scaffold, but is predicted to bind non-covalently to FAD (Figure 5-12).

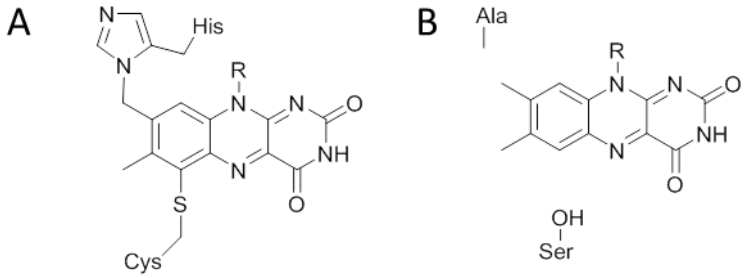


Figure 5-12 The predicted residues for covalent FAD binding in CrRO are not present.
A.) BBE contains a histidine and cysteine that allows for bicovalent attachment of FAD by an 8α-*N*[1]-histidyl-6-*S*-cysteinyl-FAD attachment B.) In CrRO the equivalent residues are an alanine and a serine.

As has been mentioned previously, all VAO type proteins identified to date that do not bind FAD covalently act as dehydrogenases (Leferink et al., 2008a). It is therefore tempting to speculate that the CrRO would not use molecular oxygen, but another electron acceptor instead in its catalytic cycle for re-oxidation of FAD.

However based on the work in (Leferink et al., 2009; Leferink et al., 2008b; Zafred et al., 2015) this CrRO enzyme is predicted to be able to bind molecular oxygen. CrRO has a glycine at the 'gatekeeper' residue and a valine at the 'alternative gatekeeper', (Figure 5-13, Figure 5-14), which is consistent with the oxidases such as BBE, as opposed to the dehydrogenases such as Phlp4 or GALDH. Whether this enzyme acts as an oxidase or dehydrogenase is therefore still an open question that requires experimental validation, and cannot be determined by analysis of the sequence alone.

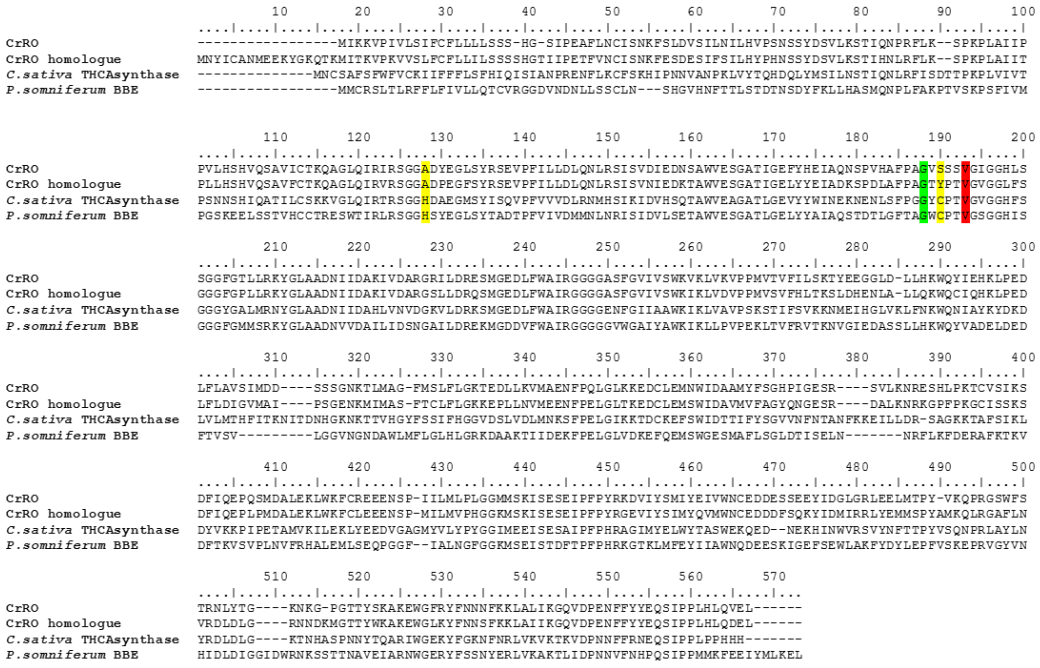


Figure 5-13 Protein alignment of CrRO.

Alignment of CrRO against the CrRO homologue present on the *C. roseus* genome, the *Cannabis sativa* THCA synthase and the canonical BBE from *Papaver somniferum*. Highlighted in yellow are the histidine and cysteine that are involved in bicovalent attachment of FAD in *C. sativa* THCA synthase and *P. somniferum* BBE, which are in turn an alanine, serine or tyrosine in CrRO and CrRO homologue. Highlighted in green is the gatekeeper residue that is a glycine in all sequences, and highlighted in red is the alternative gatekeeper residue, which is valine in all sequences indicating that all have the predicted capacity to bind molecular oxygen.

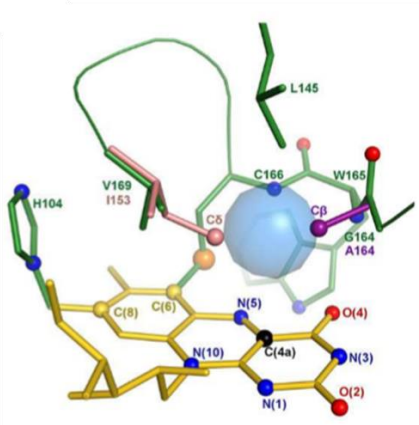


Figure 5-14 Active site of *E. californica* BBE and Phl p 4

Aligned structures of the wild type E. californica BBE (green) and Phl p 4 (light pink) with the FAD highlighted in yellow. A pocket of space (blue) allows for oxygen binding in the oxidase BBE. Phl p 4 has a residue I153 (alternative gatekeeper) that prevents oxygen binding, whilst when G164 (the gatekeeper) of the BBE is mutated to A164 (dark pink) this also prevents oxygen binding. CrRO has a glycine at the gatekeeper residue and a valine at the alternative gatekeeper and is therefore predicted to bind molecular oxygen. *Figure sourced from Zafred et al., 2015.*

The Apocynaceae family of plants produces a wide variety of MIAs that have the aspidosperma, iboga and corynanthe alkaloid scaffolds. If the CrRO is integral to MIA biosynthesis, in a part of the metabolic pathway that is common between different plant species, we may expect other species to contain a homologous protein to CrRO.

A BLAST search of this protein against the transcriptomes of other members of the *Catharanthus* genus, *Catharanthus ovalis* and *Catharanthus longifolius*, using the Phytometasyn server, demonstrated that this CrRO is conserved with 99.3% sequence identity in these species. A homologous protein was also conserved in more distant family members, such as *Vinca minor* and *Tabernaemontana elegans* each protein having 72% identity and 67.9% identity respectively. In comparison, *Rauwolfia serpentina* and *Camptotheca acuminata*, despite also producing MIAs, do not contain a homologue that is mutated for bicovalent attachment.

Interestingly the homologues from *Catharanthus ovalis, Catharanthus longifolius, Vinca minor* and *Tabernaemontana elegans*, like the CrRO from *C. roseus*, do not contain the histidine or cysteine that allows for bicovalent attachment of FAD, with the histidine replaced by an alanine in *C. ovalis* and *C. longifolius*, whilst replaced by a serine in *V. minor* and *T. elegans*, and the cysteine replaced by a either a tyrosine or serine in each of these plant species.

This demonstrated that although the CrRO protein is unusual in having a berberine bridge scaffold, but is predicted to bind FAD non-covalently, this reticuline oxidase like protein is conserved in a number of species of MIA producing plants.

The alkaloid content of the different plant species may provide clues as to which specific MIA metabolic pathway CrRO might be involved in. The fact that *Rauwolfia serpentina* and *Camptotheca acuminata* contain a homologue that has lower sequence identity and also appears to bind FAD differently, suggests the CrRO from *C. roseus* may have a unique function that is not is involved in ajmalan, yohimbine, ajmalicine or quinoline type alkaloid biosynthesis which are the predominant alkaloid families found in *Rauwolfia serpentina* and *Camptotheca acuminata*. As well as ajmalicine type indole alkaloids, *C. roseus* also contains aspidosperma and iboga type alkaloids, which are highlighted by the MIAs, vindoline and catharanthine. The iboga and aspidosperma alkaloid families are thought to be synthesised from a common chemical intermediate, preakuammicine. Similarly *Tabernaemontana elegans* contains iboga based alkaloids, whilst *Vinca minor* produces vincamine, a hunteria type alkaloid that based on feeding studies (Ahmad et al., 1983), has been shown to be derived from the aspidosperma alkaloid tabersonine. The conservation of a reticuline oxidase like protein that is predicted to bind FAD

non-covalently in the *Catharanthus* genus*, Tabernaemontana elegans* and *Vinca minor*, whilst such a protein is absent in *Rauwolfia serpentina* or *Camptotheca acuminata*, may suggest that this protein is involved in the biosynthesis of iboga and aspidosperma based MIAs.

As more transcriptomic and genomic data becomes available for plant species that produce MIAs it will be interesting to ascertain whether this particular reticuline oxidase like protein is conserved in these plant species, as has been shown so far in *C. roseus*, *V. minor* and *T. elegans.*

There are nine contigs in the *C. roseus* transcriptome that are annotated as either berberine bridge enzymes or reticuline oxidases that are also expressed in leaf tissue. Furthermore a BLAST search of the *P. somniferum* BBE against the *C. roseus* transcriptome identified a further seven genes that are annotated as either an FAD-dependent oxidoreductase, a CPRD2 protein, carbohydrate oxidase, lactate dehydrogenase, or misannotated as a gene of unknown function, resulting in a gene family of ~16 genes. With the exception of CrRO and the CrRO homologue, all of these proteins have the predicted conserved histidine present for monocovalent attachment of FAD, and 10/16 are predicted to also bind in a bicovalent manner due to the presence of a conserved cysteine residue. All of the BBE-like enzymes contain a glycine residue at the 'gatekeeper' position, however 3 out of the 16 have an isoleucine or leucine instead of a valine at the 'alternative gatekeeper' residue suggesting three of the proteins may act as dehydrogenases.

Only CrRO was shown to cluster with regards to its expression profile with genes shown to be involved in the MIA pathway (Figure 5-9). Additionally this CrRO is also highly expressed in leaf tissue and has the third highest expression of this class of enzyme in leaf tissue in *C. roseus*.

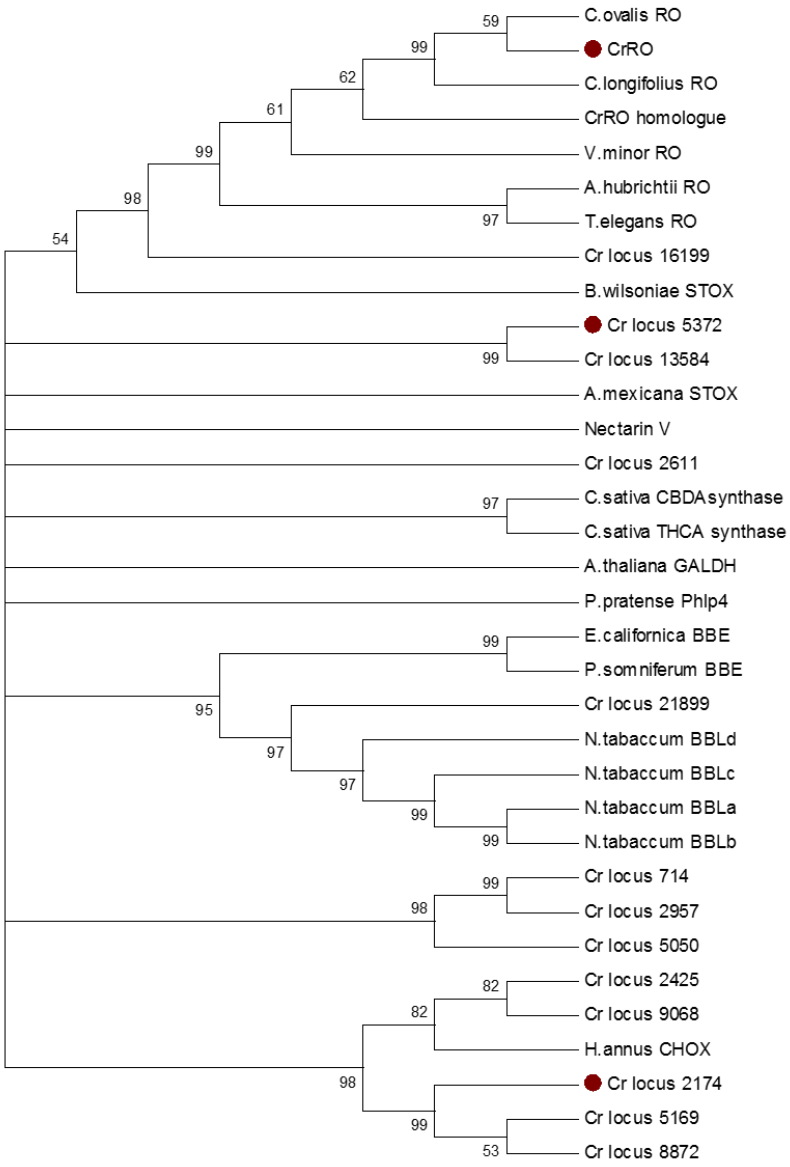


Figure 5-15 Phylogenetic tree of BBE-like enzymes

A phylogenetic tree showing sequences for all of the BBE-like enzymes present in *C. roseus*, the homologues of CrRO present in *C. longifoilus, C. ovalis, V. minor, T. elegans*, as well as the characterised BBE like enzymes, BBE from *P. somniferum* and *E. californica*, NtBBLa-d from *N. tabacum*, THCA synthase and CBDA synthase from *C. sativa*, AmSTOX from *A. mexicana*, BwSTOX from *B. wilsoniae*, Ha-CHOX from *H. annus*, Nectarin V from ornamental tobacco and Phlp4 from *P. pratense are* presented. Highlighted by the red circles are the *C. roseus* reticuline oxidase like enzymes silenced in this thesis. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model, and the bootstrap consensus tree was inferred from 500 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed.

5.3.2 VIGS of Reticuline Oxidase-Like Enzymes in C. roseus

5.3.2.1 VIGS of CrRO

To assess whether CrRO is involved in MIA metabolism, it was silenced by VIGS. Two regions of the reticuline oxidase like gene (CrRO) from the N-terminus (500 bp in length) and C-terminus (281 bp in length) were selected for the construction of two pTRV2u silencing vectors. An N- and C-terminal region of the gene were selected for silencing to ensure that any phenotype observed was due to silencing of this specific RO and not as a result of cross-silencing other members of the reticuline oxidase like family in *C. roseus.* A BLAST against the transcriptome for each of the regions from the N- and C-terminus suggested that the VIGS fragments selected do not contain regions of homology that has significant overlap to other genes in the transcriptome that could cause potential cross-silencing.

VIGS of the CrRO using both either the N-terminal or C-terminal fragment in the VIGS vector resulted in a novel metabolic phenotype, showing the accumulation of two new peaks with *m/z* values of 355 and 397, relative to the empty vector control VIGS tissue (Figure 5-18). These peaks are not found, or found at very low levels in the EV control, in comparison to the CrRO silenced leaf tissue (Figure 5-18). The fact both N-terminal and C-terminal VIGS vectors generated the same metabolic phenotype is indicative that silencing of this gene, CrRO, is the likely cause of the accumulation of these peaks.

The whole VIGS experiment has been replicated multiple times, generating the same metabolic phenotype. The results from the whole VIGS experiment are presented in (Figure 5-16) utilising the N-terminal VIGS vector, as well as the VIGS experiment in which silencing of the CrRO with the N-terminal and C-terminal VIGS vector are performed in parallel (Figure 5-17). Further replicates are presented in the Appendix. This is the first evidence that the reticuline oxidase (and more broadly the VAO family of enzymes) is fundamentally involved in metabolism in *C. roseus.* qPCR to confirm the silencing of the CrRO gene was performed for multiple replicates of the VIGS experiments (Figure 5-16, Figure 5-17).

The qPCR demonstrated that use of both the N- and C terminally targeted VIGS vector results in efficient CrRO silencing (Figure 5-17).

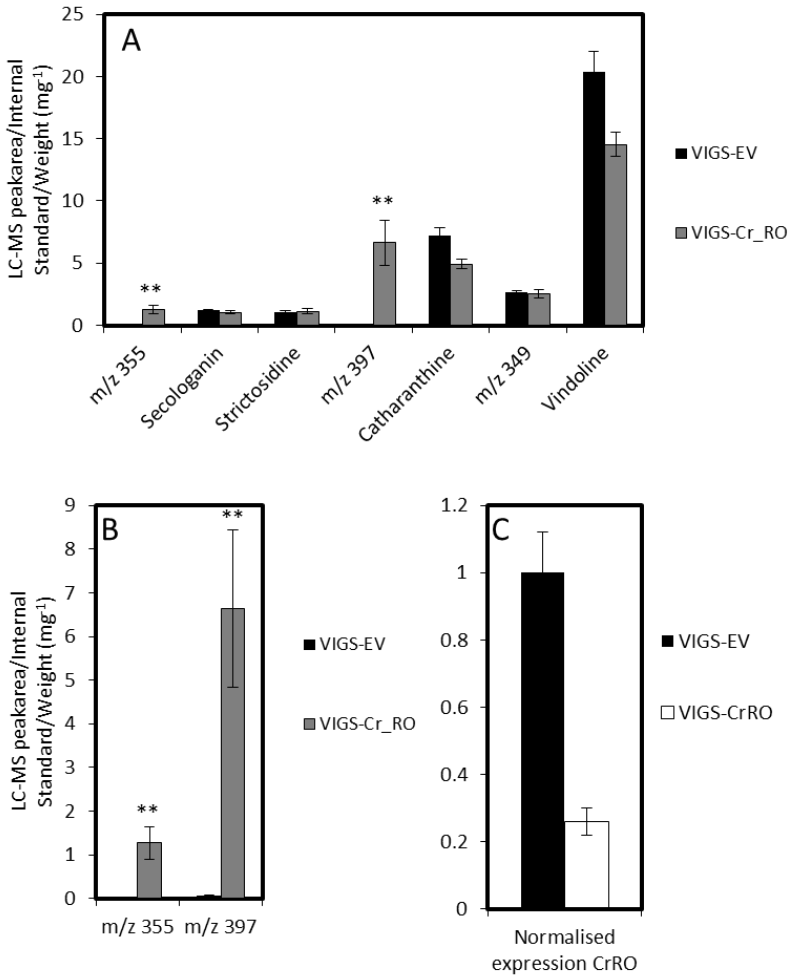


Figure 5-16 Metabolite and qPCR profile upon VIGS of CrRO using the N-terminal VIGS vector
A.) Alkaloid profile for VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes; VIGS-CrRO (n=8) VIGS-EV (n=8) ** p<0.01 B.) *m/z* 355 and *m/z* 397 profile for VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes; VIGS- CrRO(n=8) VIGS-EV (n=8) ** p<0.01 C.) qPCR normalised expression of the CrRO gene in empty vector and CrRO silenced plants; VIGS- CrRO (n=8) VIGS-EV (n=8). All data presented is the mean ± SEM.

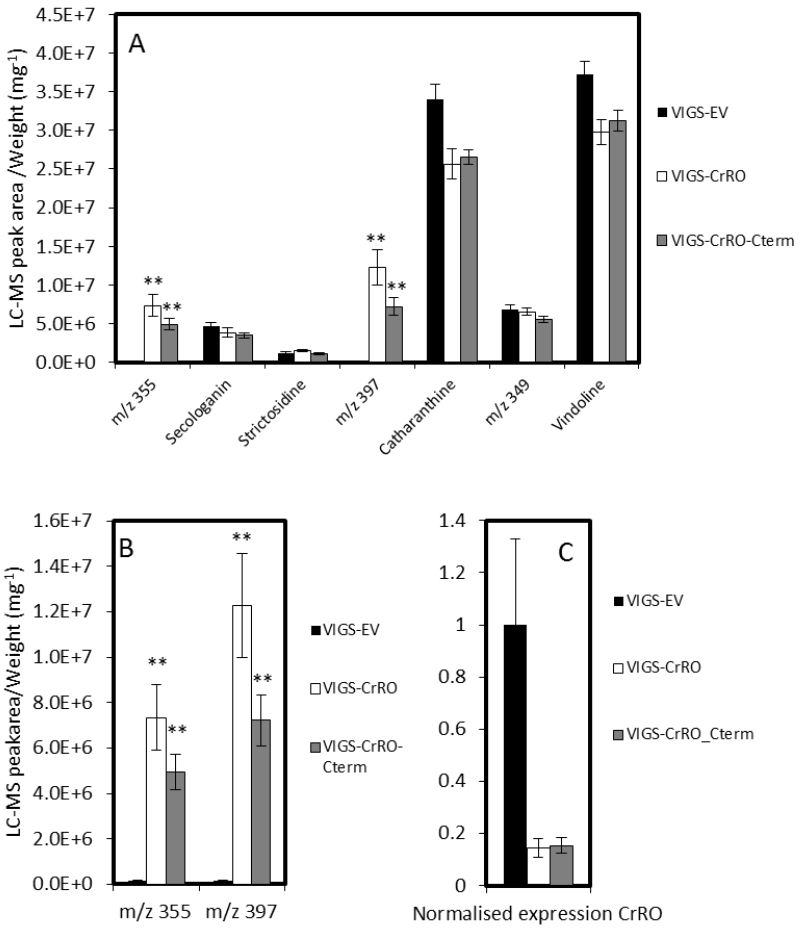


Figure 5-17 Metabolite and qPCR profile upon VIGS of CrRO using the N- and C-terminal VIGS vector
A.) Alkaloid profile for VIGS of CrRO and CrRO_Cterm relative to empty vector control tissue in Little Bright Eyes VIGS-CrRO (n=10), VIGS-CrRO_Cterm (n=8), VIGS-EV (n=10) ** p<0.01 B.) *m/z* 355 and *m/z* 397 profile for VIGS of CrRO and CrRO_Cterm relative to empty vector control tissue in Little Bright Eyes; VIG-CrRO(n=10), VIGS-CrRO_Cterm (n=8), VIGS-EV (n=10) ** p<0.01 C.) qPCR normalised expression of the CrRO gene in empty vector and CrRO silenced plant; CrRO (n=8), VIGS-CrRO_Cterm (n=8), VIGS-EV (n=8). All data presented is the mean ± SEM.

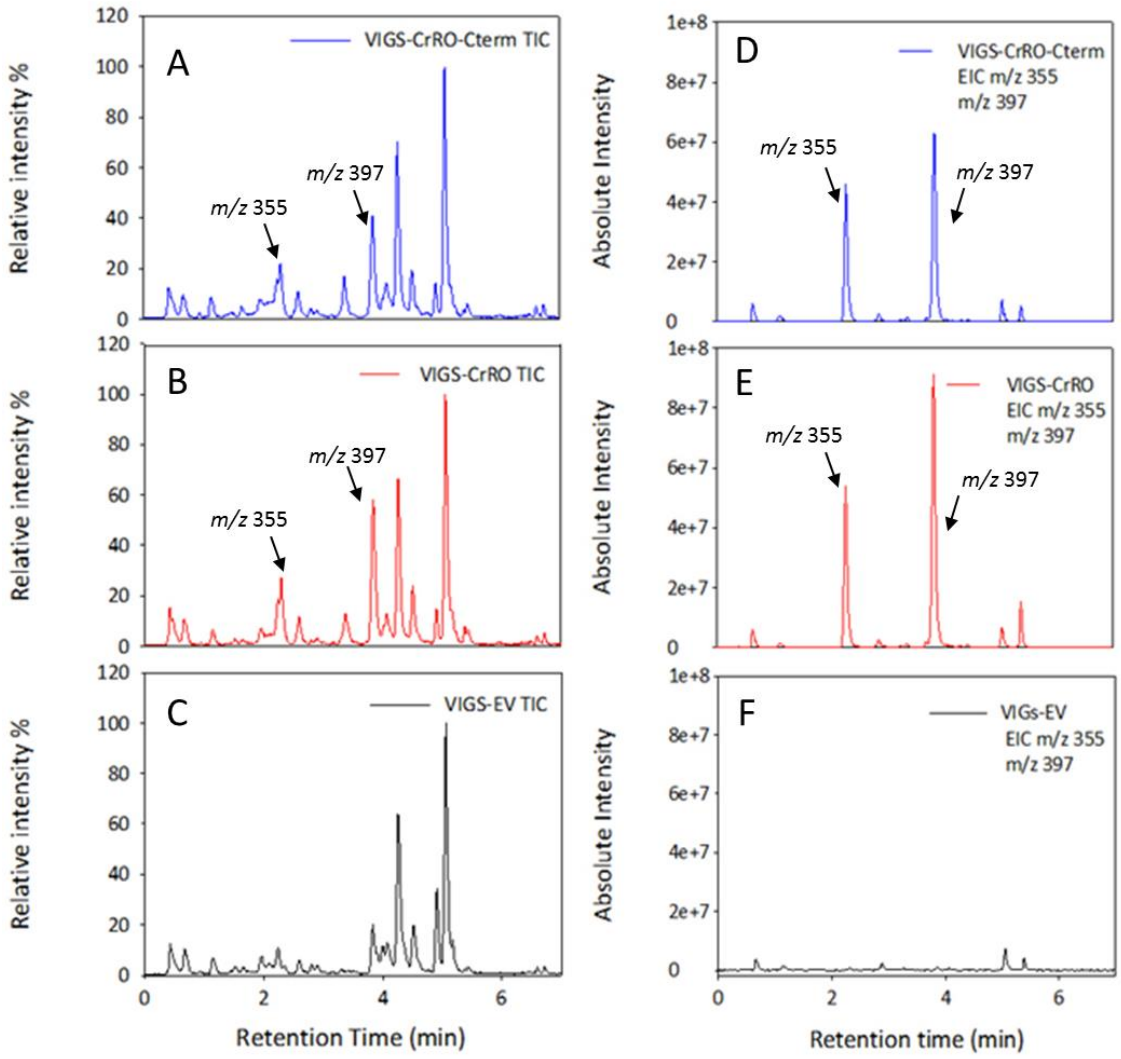


Figure 5-18 Representative Total Ion Chromatogram and Extracted Ion Chromatograms for VIGS-EV, VIGS-CrRO and VIGS-CrRO Cterm leaf tissue.

Methanol extracts of leaf tissue from VIGS-CrRO_Cterm, VIGS-CrRO and VIGS-EV leaf tissue were analysed by LC-MS. Total ion chromatograms of A.) VIGS-CrRO_Cterm B.) VIGS-CrRO C.) VIGS-EV as well as Extracted ion chromatograms of *m/z* 355 and *m/z* 397 for D.) VIGS-CrRO_Cterm, E.) VIGS-CrRO and F.) VIGS-EV are presented. The *m/z* 355 and *m/z* 397 do not accumulate in VIGS-EV leaf tissue.

5.3.2.2 VIGS of RO homologues

As mentioned, there are multiple reticuline oxidase like genes in the *C. roseus* transcriptome. CrRO is the third most abundantly expressed in leaf tissue and the only one whose expression profile clusters with known genes in the MIA metabolic pathway. The most abundant RO-like protein (locus_2957) present in leaf tissue was silenced using VIGS by a colleague in the O'Connor group and did not result in a metabolic phenotype (Franziska Kellner; personal communication). Two other reticuline oxidases from *C. roseus* were selected for silencing from disparate regions of the phylogenetic tree, locus_5372 and locus_2174. These genes were the most highly expressed in leaf tissue on their respective branches of the phylogenetic tree, however did not co-regulate with the MIA pathway.

pTRV2u silencing constructs were generated to individually silence the locus_5372 and locus_2174. The individual silencing constructs did not lead to a metabolic phenotype in *C. roseus.* There was no accumulation of new compounds and there was no statistically significant change in the peak area of vindoline, catharanthine, secologanin or strictosidine (Figure 5-19).

Furthermore plants in which the locus_5372 or locus_2174 is silenced do not accumulate compounds with an *m/z* value of 355 or 397. This suggested that the role of CrRO in the plant is unique to this enzyme (Figure 5-19, Figure 5-20).

A triple silencing vector to allow for the combinatorial silencing of CrRO, locus_5372 and locus_2174 was generated by USER fusion cloning. If CrRO is functionally redundant and compensating for the activity of locus_5372 and locus_2174 then combinatorial silencing of all three genes should lead to either a novel metabolic phenotype or an enhanced metabolic phenotypic greater than that observed through individual silencing.

VIGS using this triple silencing construct resulted in the same metabolic phenotype as individually silencing the CrRO gene (Figure 5-19, Figure 5-20). This suggested that there was no cumulative silencing effect upon silencing all three RO genes at the same time, and therefore the evidence to date only implicates CrRO as being physiologically relevant to *C. roseus* metabolism.

qPCR confirmed that all three genes CrRO, locus_5372 and locus_2174 in the triple silenced lines are silenced in these experiments. Additionally, in CrRO individually silenced plants there appeared to be a minor amount of cross-silencing of locus_5372 and no cross-silencing with locus_2174. The fact that the locus_5372 individually silenced plants do not show a phenotype

suggested that the cross-silencing observed is unlikely to be contributory to the metabolic phenotype (Figure 5-19, Figure 5-20).

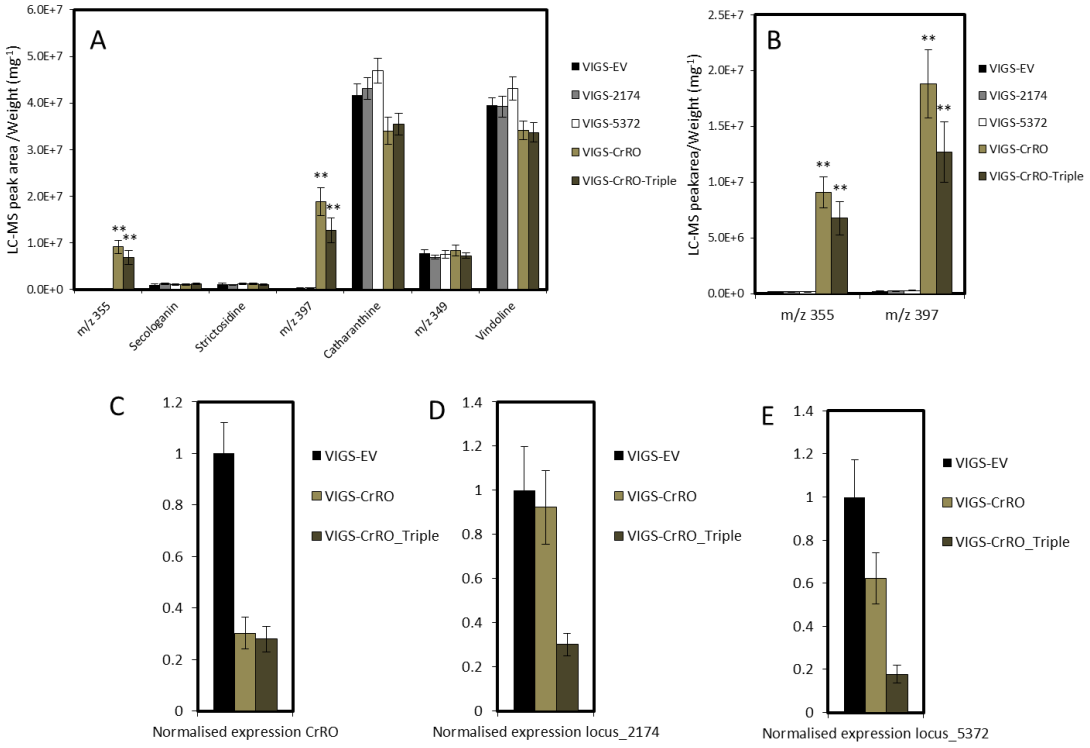


Figure 5-19 Metabolite and qPCR profile upon VIGS of CrRO, locus_5372, locus_2174 and the CrRO_Triple silencing fusion vector.
A.) Alkaloid profile for VIGS of CrRO, locus_5372, locus_2174 and a triple silencing vector of all three, relative to empty vector control tissue in Little Bright Eyes: VIGS-CrRO (n=11), VIGS-2174 (n=11), VIGS-5372 (n=9), VIGS-CrRO-Triple (n=10), VIGS-EV (n=12) ** p<0.01 B.) *m/z* 355 and *m/z* 397 profile for VIGS of CrRO , locus_5372, locus_2174 and a triple silencing vector of all three relative to empty vector control tissue in Little Bright Eyes VIGS-CrRO (n=11), VIGS-2174 (n=11), VIGS-5372 (n=9), VIGS-CrRO-Triple (n=10), VIGS-EV (n=12) ** p<0.01. All data presented is the mean ± SEM. C.) Normalised expression of the gene CrRO in VIGS-EV (n=8), VIGS-CrRO (n=8) and VIGS-CrRO_Triple tissue (n=8) D.) Normalised expression of the gene locus_2174 in VIGS-EV (n=8), VIGS-CrRO (n=8) and VIGS-CrRO_Triple tissue (n=8) E.) Normalised expression of the gene locus_5372 in VIGS-EV (n=8), VIGS-CrRO (n=8) and VIGS-CrRO_Triple tissue (n=8).

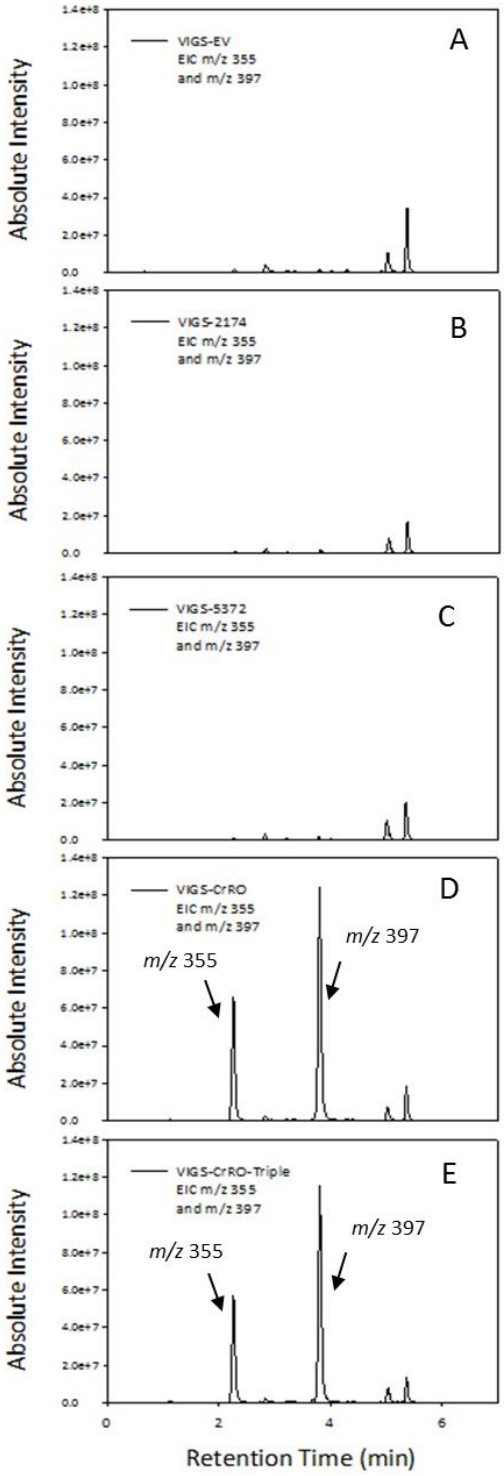


Figure 5-20 Representative Extracted Ion Chromatogram of *m/z* 355 and *m/z* 397 for VIGS-EV, VIGS-2174, VIGS-5372, VIGS-CrRO and VIGS-CrRO-Triple leaf tissue.
Methanol extracts of leaf tissue from A.) VIGS-EV B.) VIGS-2174 C.) VIGS-5372 D.) VIGS-CrRO E.) VIGS-CrRO-Triple leaf tissue were analysed by LC-MS. Presented are the extracted ion chromatograms for the *m/z* 355 and *m/z* 397 which accumulated in the VIGS-CrRO and VIGS-CrRO-Triple tissue but not in the other tissue samples.

5.3.2.3 Analysis of Stem Tissue for CrRO silenced plants

Silencing of the CrNPF2.9 transporter in Chapter IV demonstrated that the VIGS can affect the metabolic phenotype occurring in both leaf and stem. Metabolite analysis of VIGS-EV stem and stem from VIGS-CrRO silenced plants demonstrated that there was no significant difference in the peaks present in these tissues and that the compounds represented by *m/z* 355 and 397 did not accumulate in stem (Figure 5-21). This may be indicative that the physiological role of CrRO is leaf specific. As demonstrated previously (Chapter IV), the peak with value *m/z* 349 is the dominant peak in stem as opposed to vindoline in leaves.

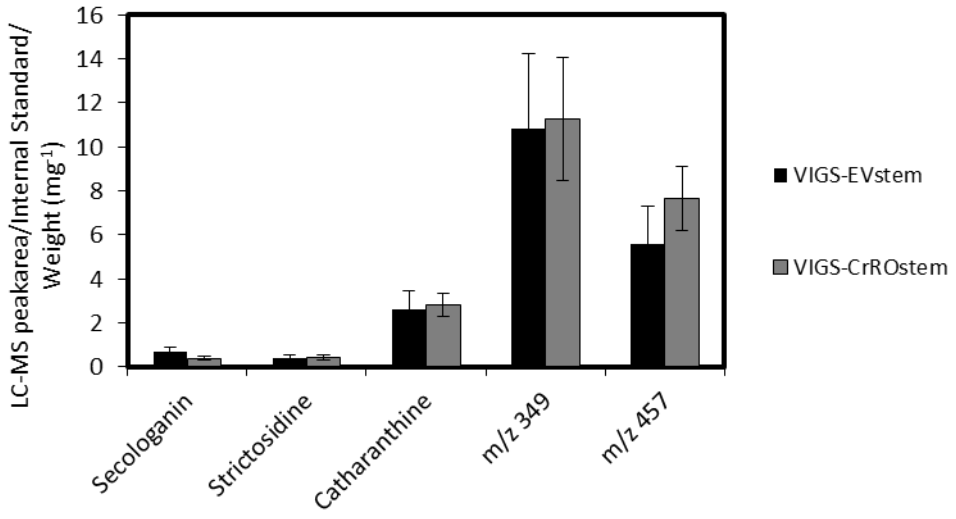


Figure 5-21 Metabolite analysis of stem tissue upon CrRO silencing
Alkaloid profile of stem tissue taken from around VIGS wound site for Empty vector control (n=8) and CrRO silenced plants (n=8) ** $p < 0.01$**.** All data presented is the mean ± SEM

5.3.2.4 Isotopic labelling of MIAs in VIGS-CrRO leaf tissue

We do not know the identity of the compounds with *m/z* 355 and *m/z* 397 that are accumulating in the CrRO silenced VIGS experiments. An exact mass for the *m/z* 397 peak was obtained, with a value of 396.2048 (Gerhard Saalbach, John Innes Centre). The best molecular formula for this exact monoisotopic mass is $C_{23}H_{28}N_2O_4$. This is consistent with this compound being a MIA, due to the presence of two nitrogen atoms in the molecular formula.

To further demonstrate that these compounds of *m/z* 355 and *m/z* 397 are MIAs, and that the CrRO is involved directly in MIA biosynthesis, CrRO VIGS tissue was fed with deuterated (d_4)-

tryptamine. As MIAs are derived from tryptamine we predict we would observe *m/z* increase by 4 for all MIA compounds. Individual leaves from CrRO plants were cut at the petiole and fed 5mM d_4-tryptamine over the course of 4 days.

Label was incorporated into the known alkaloids, strictosidine and vindoline (Figure 5-23, Figure 5-24). As has been observed in other feeding experiments (Chapter III), there is an accumulation of labelled strictosidine, with the strictosidine pool in leaf tissue labelled to ~ 75% (Table 7). This is in contrast to the label incorporation into downstream alkaloids, potentially implicating that either the export of strictosidine from the vacuole or its downstream utilisation is a metabolic bottleneck.

For CrRO VIGS tissue the label was also incorporated into the peaks with *m/z* value of 355 and 397, resulting in peaks of *m/z* 359 and 401 respectively. This is indicative that these peaks are derived from tryptamine, and therefore likely to be MIAs (Figure 5-23, Figure 5-24). The amount of label incorporation into the *m/z* 355 and *m/z* 397 peak is not as great as the label incorporated into strictosidine, however comparable to that observed in the downstream MIAs vindoline. Vindoline results in a ~2% label enrichment, whilst the label enrichment for *m/z* 355 and *m/z* 397 is between 8-11% (Table 7).

Although we do not know the role of CrRO, this labelling data is indicative that the two peaks that are accumulating do not have an oxidised β-carboline ring (Figure 5-22). The fact that labelling results in *m/z* + 4 for both *m/z* 355 and *m/z* 397 is indicative that there is no loss of deuterium. Therefore although we do not know the specific catalytic role of CrRO, this labelling data is indicative that the two peaks that are accumulating do not have an oxidised β-carboline ring (Figure 5-22).

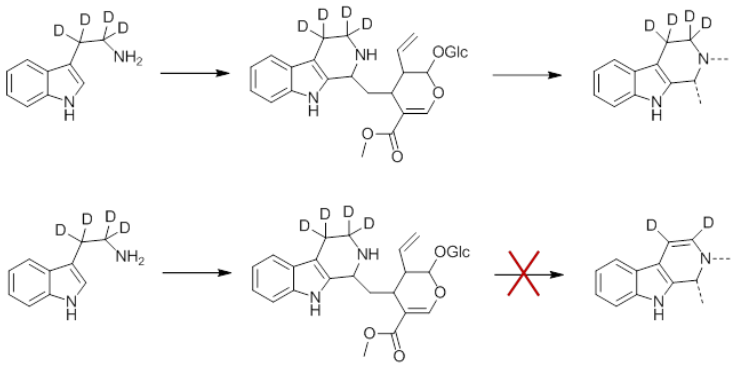


Figure 5-22 d_4-Tryptamine feeding results in deuteration of β-carboline ring.
Tryptamine labelling results in the deuteration of the β-carboline ring. The fact deuterated label is incorporated into the *m/z* 359 and *m/z* 397 peaks is indicative that these compounds contain a non-oxidised β-carboline ring system as the label is still present.

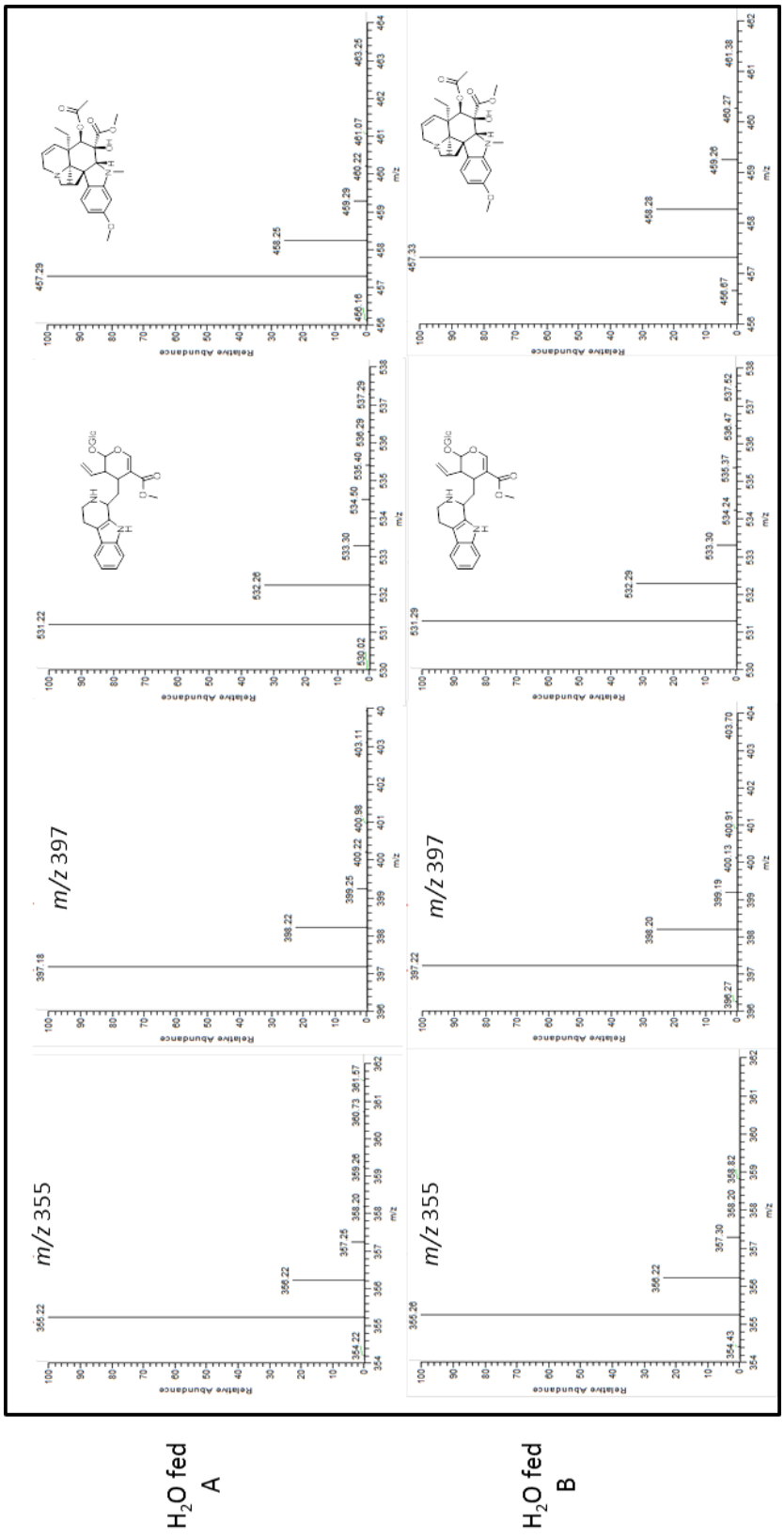


Figure 5-23 Mass Spectra for H_2O fed *C. roseus* leaf tissue upon VIGS-CrRO silencing
Mass spectra of m/z 355, m/z 397, strictosidine and vindolinefor two separate replicates of *C. roseus* VIGS-CrRO leaf tissue fed H_2O;H_2O fed A and H_2O fed B. As would be expected no label is present in the mass spectra for H_2Ofed tissue.

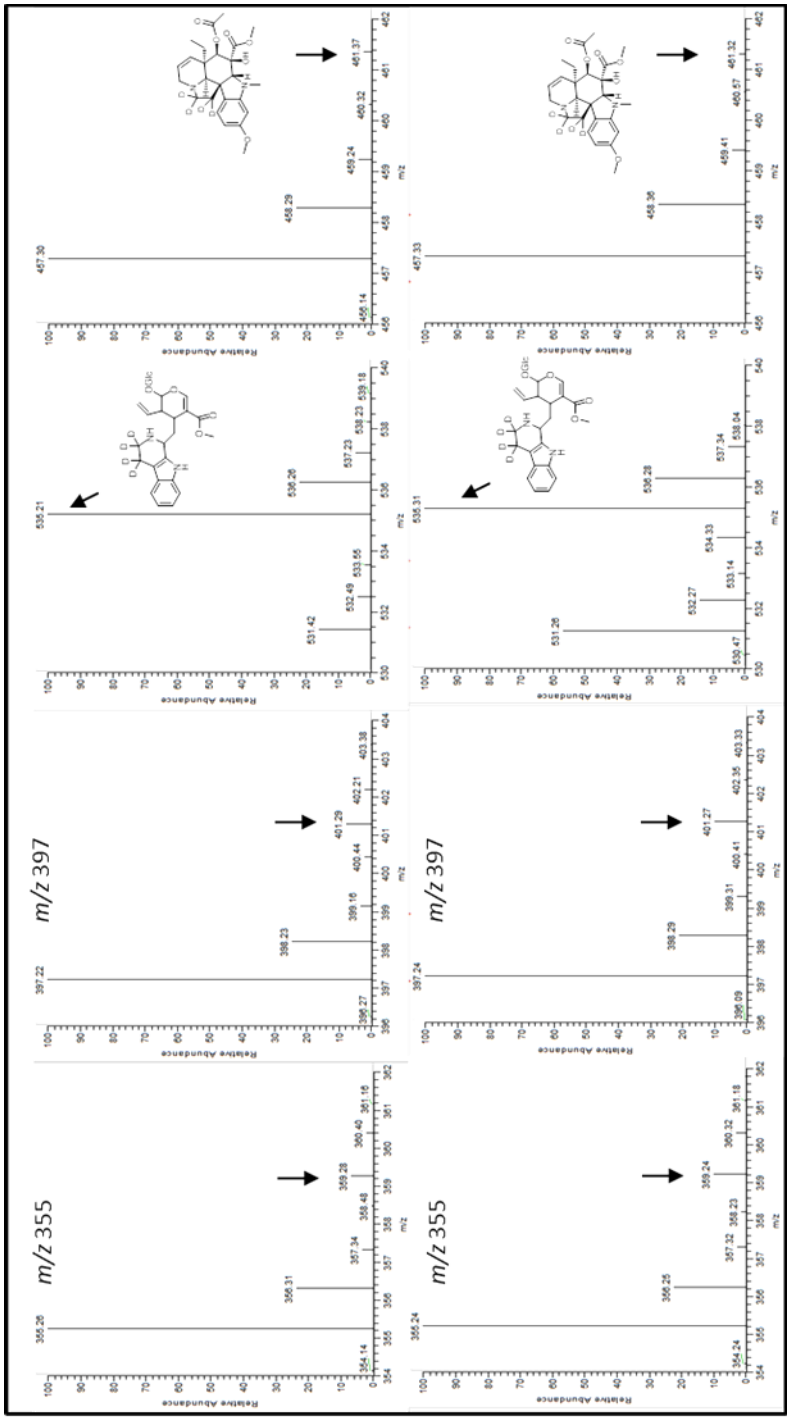


Figure 5-24 Mass Spectra for d_4-tryptamine fed *C. roseus* leaf tissue upon VIGS-CrRO silencing
Mass spectra of m/z 355, m/z 397, strictosidine and vindoline for two separate replicates, A and B, of *C. roseus* VIGS-CrRO leaf tissue fed d_4-tryptamine. Label incorporation *m* + 4 as a result of d_4-tryptamine feeding is highlighted by a black arrow.

	m/z 355	*m/z* 359	*m/z* 397	*m/z* 401	strictosidine	d_4-strictosidine	vindoline	d_4-vindoline
VIGS-CrRO H_2Ofed A	99.91	0.09	99.96	0.04	99.03	0.97	99.99	0.01
VIGS-CrRO H_2Ofed B	99.97	0.03	99.86	0.14	98.79	1.21	99.89	0.11
VIGS-CrRO d_4-tryptamine fed A	89.63	10.37	92.73	7.27	13.72	86.28	97.78	2.22
VIGS-CrRO d_4-tryptamine fed B	91.04	8.96	91.03	8.97	36.22	63.78	97.45	2.55

Table 7 Enrichment of labelling upon d4-tryptamine feeding to VIGS-CrRO leaf tissue
The relative enrichment of the *m/z* +4 relative to *m/z* is presented for each compound. Enrichment values were calculated as a percentage of the combined peak area of *m/z* and *m/z* + 4; (peak area (*m/z* + 4))/(∑(peak area (*m/z* + 4) + peak area (*m/z*)). The presence of deuterated compounds in the water fed controls can be explained by baseline noise and natural isotope abundance.

5.3.2.5 Deacetylation of VIGS-CrRO leaf extracts

The fact that two new peaks, *m/z* 355 and *m/z* 397, accumulated in the VIGS tissue is indicative that either two separate metabolic pathways are being affected by silencing of the CrRO, or that the two compounds are related. It is possible that silencing of the CrRO results in the accumulation of one compound in the plant, which in turn is then detoxified or derivatised by the endogenous metabolism of *C. roseus*.

The mass difference of 42 between 397 and 355 suggests that *m/z* 397 might be the acetylated derivative of *m/z* 355.

Potassium carbonate (K_2CO_3) can be used for the deacetylation of compounds in aqueous methanol (Wuts and Greene, 2006). Treatment of a methanolic extract of VIGS-CrRO tissue with K_2CO_3 resulted in the depletion of the *m/z* 397 and the accumulation of *m/z* 355, relative to its non-treated control (Figure 5-25).

In comparison treatment of a VIGS-EV methanolic extract with potassium carbonate did not lead to the significant accumulation of a compound with *m/z* 355 relative to untreated controls, indicating that the increase in *m/z* 355 upon treatment with K_2CO_3 in the VIGS-CrRO tissue is likely to be due to its conversion as a result of the deactylation of *m/z* 397 (Figure 5-25).

The base-dependent conversion of the compound *m/z* 397 to *m/z* 355, strongly suggested that the *m/z* 397 compound is likely to be the acetylated derivative of *m/z* 355. Therefore these compounds are directly related to each other, and CrRO is acting on one branch of the metabolic pathway.

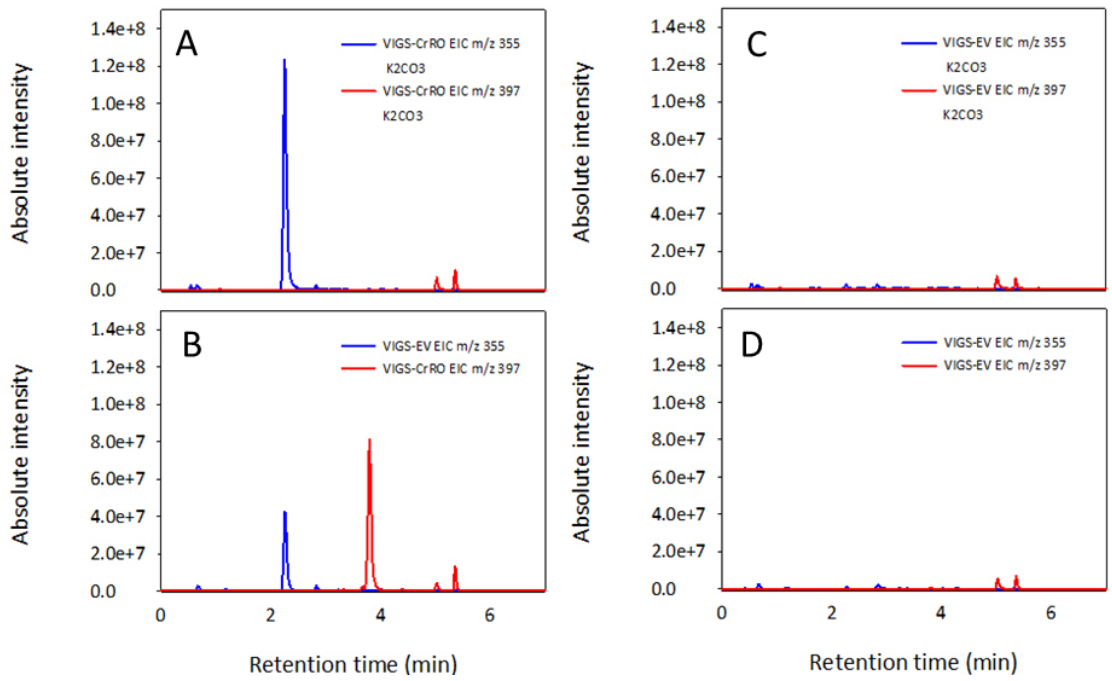


Figure 5-25 Extracted ion chromatograms for VIGS-CrRO and VIGS-EV extract treated with K_2CO_3 in aqueous MeOH for *m/z* 355 and *m/z* 397.
Methanolic extracts of the VIGS-CrRO treated A.) with and B.) without potassium carbonate and methanolic extracts of the VIGS-EV treated C.) with and D.) without potassium carbonate. Plotted are the extracted ion chromatograms for *m/z* 355 and *m/z* 397

It is likely that the *m/z* 355 that is accumulating in VIGS-CrRO leaf tissue is acetylated on a free hydroxyl group, since the deacetylation reaction (K_2CO_3, 1 hour at room temperature in MeOH) was performed under conditions too mild to effect the hydrolysis of an amide.

Additionally, vindoline and catharanthine were still present in the sample upon treatment with K_2CO_3 indicating that the reaction conditions were not strong enough for the liberation of the methoxy group from methyl-esters (Figure 5-26).

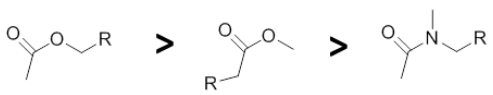


Figure 5-26 Reactivity of acetylated groups to deacetylation by K_2CO_3
Esters are more reactive towards cleavage of the acetyl group with K_2CO_3 than either methyl esters, which would liberate a methoxy group or amides, which would liberate the amine.

Acylation of oxygen- and nitrogen containing secondary metabolites in plants is a common reaction resulting in the formation of esters and amides respectively (Schmidt et al., 2015). This

reaction is often catalysed by the BAHD acyltransferase superfamily. In *C.roseus*, the terminal step in vindoline biosynthesis is catalysed by deactylvindoline acetyltransferase (DAT), resulting in the addition of an acetyl group to a hydroxyl moiety, forming vindoline (St-Pierre et al., 1998). This family of enzymes also catalyses the transfer of the acyl donors of malonate, hydroxycinnamate, tiglate, anthranilate and benzoate moieties to various alcohol substrates. A key recent example is the esterification of benzoyl-CoA and methylecgonine in the biosynthesis of cocaine (Schmidt et al., 2015). As well as the formation of esters there a number of examples of BAHD acyltransferases involved in the formation of amides, a key example of this is agmatine coumaroyltransferase, which results in the formation of *N*-(4-guanidinobutyl)-4-hydroxycinnamamide (Burhenne et al., 2003).

5.3.2.6 Proposed MIA structures for the m/z 355 and m/z 397

There are a number of MIAs that have an *m/z* ratio of 355, with a predicted molecular formula of $C_{21}H_{26}N_2O_3$, together with a non-substituted β-carboline ring and a free hydroxyl group. These include the compounds vincamine and yohimbine, two MIAs of the eburnamine and corynanthine classes respectively, as well as the early pathway intermediates isositsirikine, together with its isomer sitsirikine, and stemmadenine.

Although not present in *C. roseus*, vincamine is found in a number of related species such as *Vinca minor*. It has recently been shown that vincamine can be formed from the structural rearrangement of tabersonine (Kellner et al., 2015a), a key intermediate in the formation of vindoline. It is hypothetically feasible that if tabersonine is inappropriately rearranged to vincamine in *C. roseus*, a repair pathway back to tabersonine might be necessary, which would encompass a series of oxidation and reduction reactions (Figure 5-27). If CrRO is involved in this repair pathway, vincamine (*m/z* 355) might be expected to accumulate upon CrRO silencing. Although feasible, this scenario is unlikely, however, as *V. minor* also contains a homologue of CrRO.

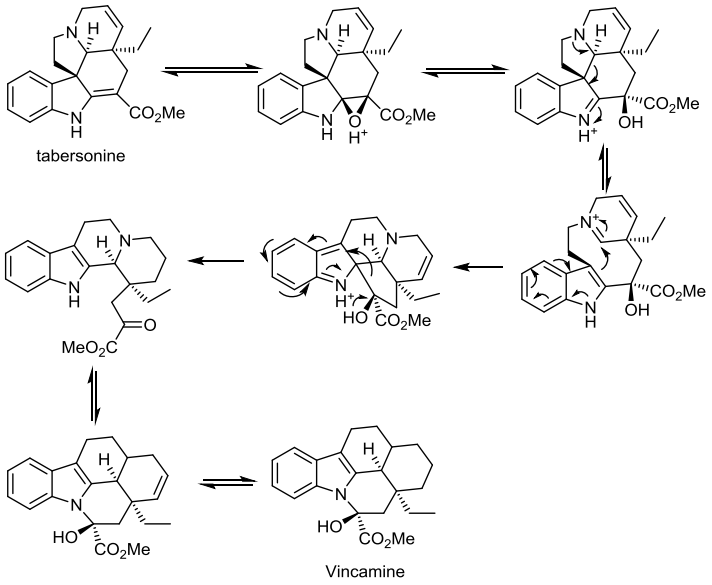


Figure 5-27 Tabersonine can rearrange to vincamine
The rearrangement of tabersonine to vincamine encompasses a number of oxidation and reduction reactions. Figure adapted from (Kellner et al., 2015a).

The yohimbine class of alkaloids are a group of corynanthine alkaloids that are structurally related to the *C. roseus* heteroyohimbine alkaloids ajmalicine, epi-ajmalicine and tetrahydroalstonine. The yohimbine alkaloids yohimbine, rauwolscine, and corynanthine all have an *m/z* 355. These compounds have not been found in *C. roseus* leaf tissue, however are present in the related plant species *Rauwolfia serpentina* (Figure 5-28). These compounds fit the criteria for the *m/z* 355 for the compound accumulating in the VIGS-CrRO silenced tissue, however the hypothetical reason for their accumulation upon VIGS-CrRO silencing and the role CrRO plays in their metabolism is unknown.

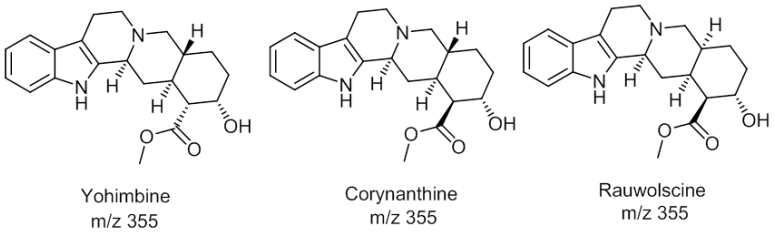


Figure 5-28 Diastereoisomers of the yohimban alkaloids.
Three known diasteroisomers of yohimban alkaloids accumulate in *Rauwolfia serpentina,* yohimbine, corynanthine and rauwolscine

Isositsirikine is a compound that is derived from the deglycosylated product of strictosidine and is the reduced form of geissoschizine (Figure 5-29). It could be envisaged that if geissoschizine is

inappropriately reduced to isositsirikine, it would need to be reoxidized to its enol form. If CrRO was the oxidase responsible for this, its silencing would result in the accumulation of *m/z* 355, and then this molecule could then be acetylated to *m/z* 397.

If this is the case, then CrRO would be acting on an off-pathway intermediate, effectively acting to recycle isositsirikine back to geissoschizine and this might explain why CrRO silencing does not lead to a reproducible statistically significant decrease in vindoline or catharanthine, as it is not a direct pathway enzyme, but is instead involved in metabolite repair (Figure 5-29).

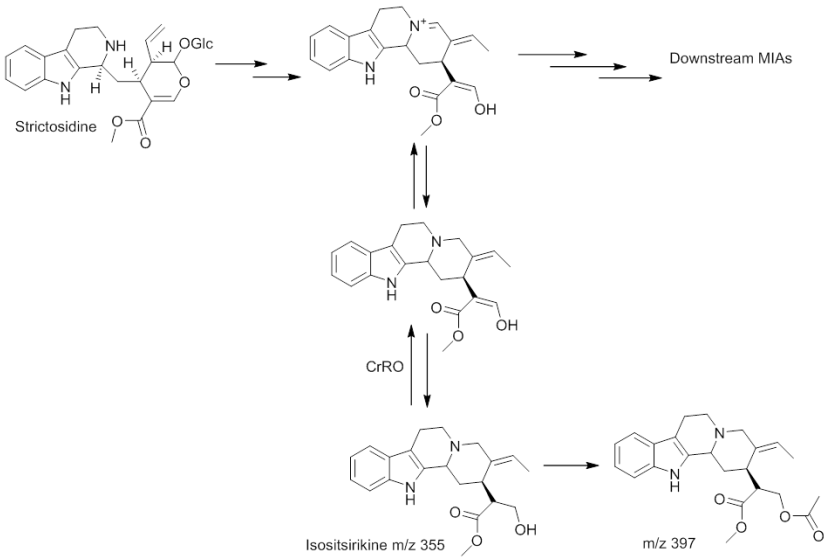


Figure 5-29 Proposed Scheme for regeneration of geissoschizine from isositsirikine utilising CrRO

Similarly stemmadenine, a molecule that also has an *m/z* 355, is thought to be an integral intermediate in the metabolic pathway to vindoline and catharanthine biosynthesis, however could be converted back to the substrate preakuammicine by a hypothetical oxidation reaction catalysed by CrRO. This would result in the formation of a carbon-nitrogen bond, as has been observed in many reactions catalysed by the BBE-like enzymes in benzylisoquinoline alkaloid biosynthesis (Figure 5-30).

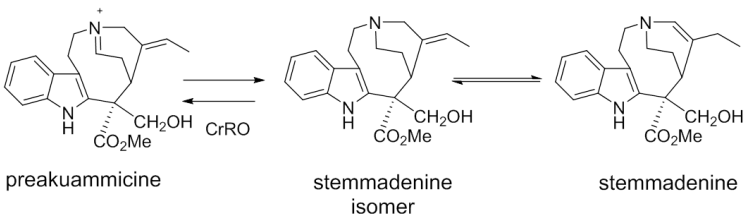


Figure 5-30 Proposed generation of preakuammicine from stemmadenine

5.3.3 LC-MS Analysis of plant extracts and known standards

To assess whether there are any compounds that could account for either the *m/z* 355 and *m/z* 397 peaks that accumulated upon CrRO silencing, methanolic extracts of known MIA producing plants, as well as known standards that have an *m/z* 355 were analysed by LC-MS to see if any compounds co-elute with the new compounds that occur in VIGS-CrRO leaf tissue.

5.3.3.1 LC-MS Analysis of plant extracts

For the crude methanolic extracts of plant tissue, *C. roseus* root, as well as *V. minor* and *V. major* leaf was extracted in methanol. *C. roseus* root has a different alkaloid profile to the leaf tissue. It does not contain the aspidosperma or iboga alkaloids that occur in leaf, and primarily accumulates serpentine. In comparison *V. minor* and *V. major* are known to accumulate vincamine as well as a number of other MIAs.

Methanolic extracts of *V. minor*, *V. major,* as well as extracts from other *C. roseus* tissues such as roots, contain compounds that have *m/z* values of 355 and 397. However none of these compounds co-eluted with the *m/z* 355 and *m/z* 397 present in the *C. roseus* CrRO VIGS tissue (Figure 5-31).

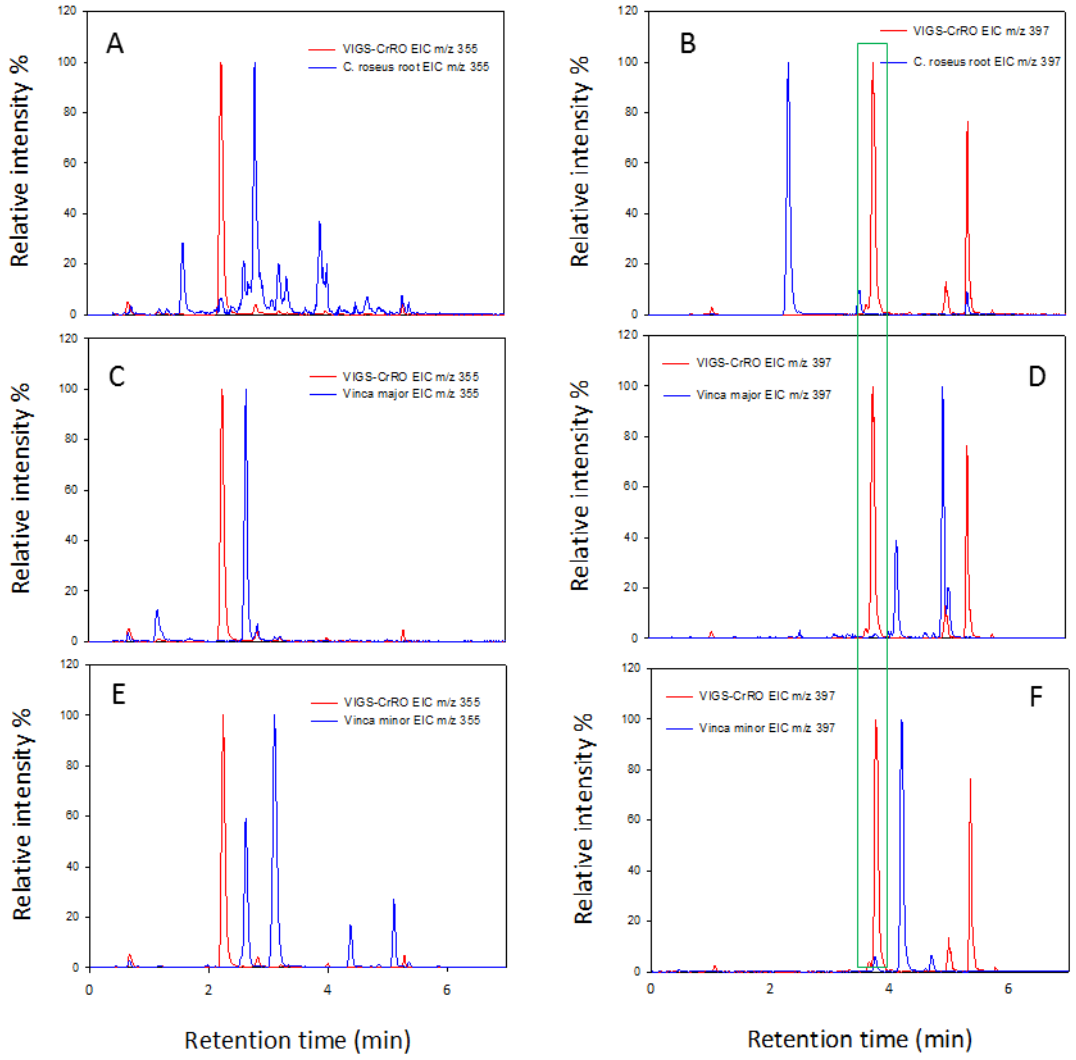


Figure 5-31 Mass Chromatogram for VIGS-CrRO, *V. minor*, *V. major, and C. roseus* root EIC *m/z* 355 and *m/z* 397

Methanolic extracts of VIGS-CrRO leaf tissue as well as *C. roseus* root*, V. minor* and *V. major* were run by LC-MS and the extracted ion chromatogram (EIC) with *m/z* 355 and *m/z* 397 plotted for each sample. *C. roseus* root *m/z* 355 (A) *m/z* 397 (B). *V. minor m/z* 355 (C) *m/z* 397 (D*), V. major m/z* 355 (E) *m/z* 397 (F). Highlighted in the green box is the *m/z* 397 that accumulated in the VIGS-CrRO leaf tissue.

5.3.3.2 LC-MS standards of known monoterpene indole alkaloids

The MIAs isositsirikine, sitsirikine and stemmadenine are not commercially available and therefore could not be analysed in this study. Vincamine, rauwolscine, corynanthine and yohimbine are commercially available and have been run as standards to compare the retention time of the *m/z* 355 from VIGS-CrRO extracts (Figure 5-32). Yohimbine, corynanthine, and vincamine do not co-elute with the *m/z* 355 that accumulated in VIGS-CrRO tissue, indicating that they are different compounds. In comparison, the yohimban alkaloid rauwolscine co-eluted with the same retention time as that of *m/z* 355 that accumulated in VIGS-CrRO leaf tissue (Figure 5-32).

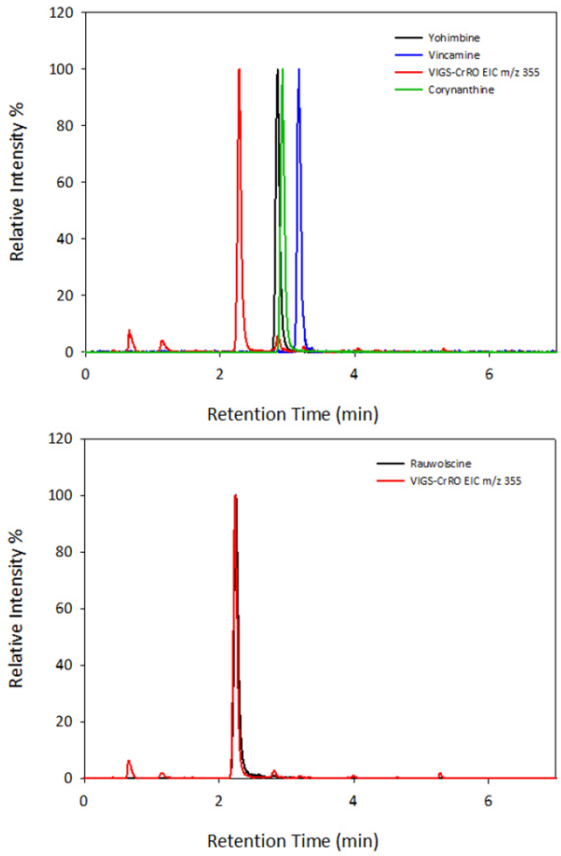


Figure 5-32 Mass Chromatogram for VIGS-CrRO, yohimbine, vincamine, rauwolscine and corynanthine EIC *m/z* 355

Methanolic extracts of the VIGS-CrRO, yohimbine, corynanthine, rauwolscine and vincamine were analysed by LC-MS and the extracted ion chromatogram (EIC) with *m/z* 355 plotted for each sample. Vincamine, yohimbine, corynanthine and the *m/z* 355 peak from VIGS-CrRO tissue do not co-elute and are therefore different compounds, whilst rauwolscine and the *m/z* 355 peak from VIGS-CrRO co-elute.

The co-elution of the *m/z* 355 from VIGS-CrRO leaf tissue and rauwolscine, however, appeared to be an artefact of the chromatography used in the LC-MS. Despite co-elution and the fact the molecule has an *m/z* 355, these two compounds are likely to be different. MS/MS fragmentation of the yohimban alkaloids yohimbine, corynanthine and rauwolscine resulted in the dominant ions on the MS^2 spectra of 144, 212 and 224 (Figure 5-33).

In comparison the *m/z* 355 and *m/z* 397 compounds accumulating in VIGS-CrRO tissue preferentially fragment with ions of 337 and 230 (Figure 5-33). Therefore the *m/z* 355 from VIGS-CrRO and rauwolscine are different compounds.

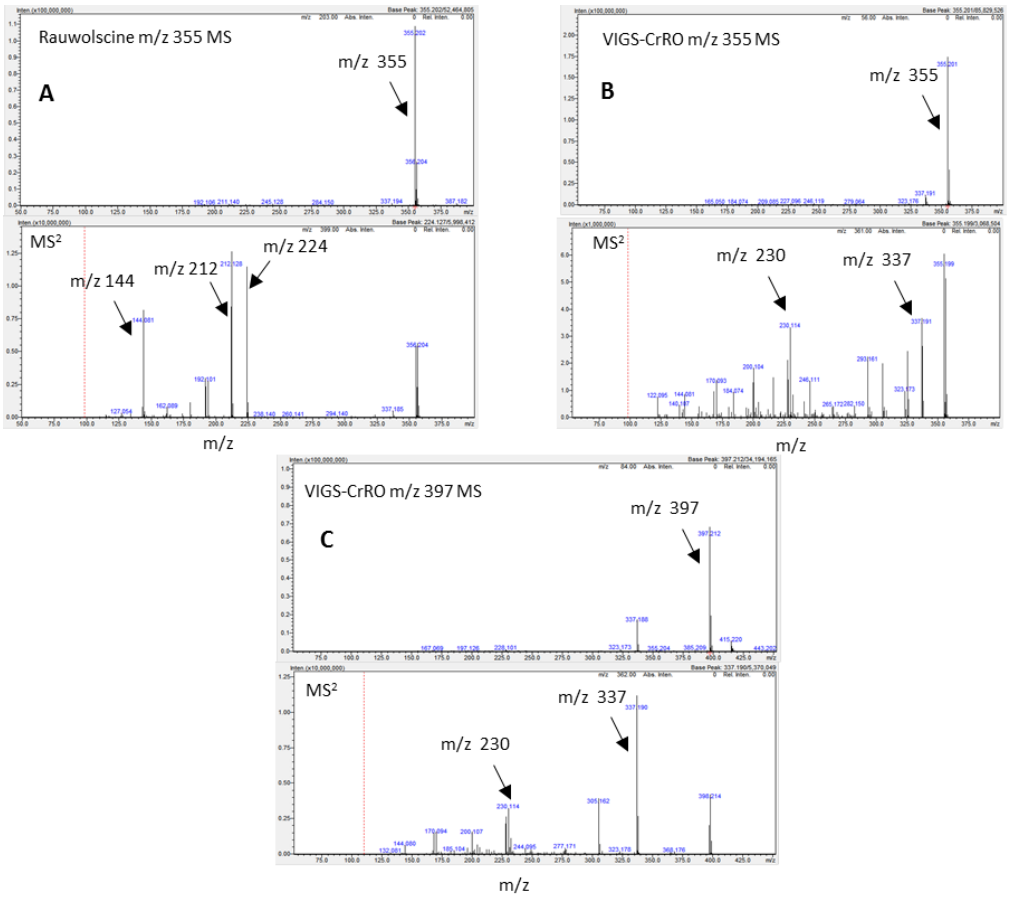


Figure 5-33 MS^n fragmentation of Rauwolscine and the *m/z* 355 and *m/z* 397 compounds from VIGS-CrRO

A.) MS and MS^2 fragmentation of rauwolscine *m/z* 355 B.) MS and MS^2 fragmentation of VIGS-CrRO *m/z* 355 C.) MS and MS^2 fragmentation of VIGS-CrRO *m/z* 397.

In summary, we know that the two compounds that accumulated upon silencing of the CrRO are MIAs and that these two compounds are related to each other, with the *m/z* 397 the acetylated derivative of *m/z* 355.

To date we have not identified a compound that co-elutes with either the *m/z* 355 or *m/z* 397 present in VIGS-CrRO tissue and that also has the same MS^n fragmentation pattern, from analysis of known standards or plant tissues known to contain MIAs.

Altogether, further VIGS experiments are therefore needed to obtain enough silenced leaf tissue so that the compounds *m/z* 355 and *m/z* 397 can be isolated for structural elucidation by NMR, in order to confirm the identity of these compounds.

5.3.4 Heterologous expression of CrRO

Our assumption is that the compounds of *m/z* 355 and *m/z* 397 are accumulating because they are the substrate or are a derivatised version of the substrate for CrRO.

To test this hypothesis, it is necessary to heterologously express CrRO and incubate it with a protein free extract of the silenced tissue that contains the alkaloid products. Consumption of the *m/z* 355 and *m/z* 397 compounds will demonstrate that one or both of these compounds is a substrate for CrRO.

Furthermore the identity of any new compound that accumulates will provide further information about the structure of *m/z* 355 and *m/z* 397.

Three separate heterologous expression systems have been utilised in this study; *S. cerevisiae, E. coli* and Sf9 insect cells.

To date the heterologous expression and characterisation of plant berberine bridge like enzymes has primarily utilised: *E. coli*, *S. cerevisiae, Pichia pastoris*, and Sf9 insect cells. The expression systems used for the characterisation of these known plant enzymes that has been published is summarised in Table 8.

Of the results in the published literature for these enzymes, the most reproducible expression for this enzyme family is Sf9 insect cell culture and *Pichia pastoris*.

Table 8 Expression systems for RO proteins published

Gene	Plant species	*E. coli*	*S. cerevisiae*	*P. pastoris*	Sf9 insect cells	Reference
Berberine Bridge Enzyme	*E. californica*		✓	✓	✓	Kutchan 1991, Kutchan 1995, Winkler 2006, 2007, 2009, Wallner 2012
Berberine Bridge Enzyme	*P. somniferum*		✓	✓		Hawkins and Smolke2008, Hagel 2012
Berberine Bridge Enzyme	*C. japonica*		✓			Minami 2008
(*S*)-tetrahydroprotoberberine oxidase (STOX)	*B. wilsoniae*				✓	Gesell 2011
(*S*)-tetrahydroprotoberberine oxidase (STOX)	*A. mexicana*				✓	Gesell 2011
dihydrobenzophenanthridine oxidase (DBOX)	*P. somniferum*			✓		Hagel 2012
tetrahydrocannabinolic acid synthase (THCA synthase)	*C. sativa*		✓	✓	✓	Taura 2007, Zirpel 2015, Sirikantaramas 2004
cannabidiolic acid synthase (CBDA synthase)	*C. sativa*				✓	Taura 2007
BBLa, b, c, d	*N. tabacum*			✓		Kajikawa 2011
Phl p 4	*P. pratense*	✓		✓		DeWitt 2006, Zafred 2015
Carbohydrate oxidase (Ha-CHOX)	*H. annuus*	✓				Custers 2004
AtBBE-like 13	*A. thaliana*			✓		Daniel 2015
AtBBE-like 15	*A. thaliana*			✓		Daniel 2015

5.3.4.1 Heterologous expression in S. cerevisiae

There is literature precedent for successful expression of some reticuline oxidase like proteins in *S. cerevisiae*. The BBE from *E. californica* was first cloned and shown to be catalytically active when heterologously expressed in *S. cerevisiae* (Dittrich and Kutchan, 1991). Under a galactose inducible promoter system, utilising the pYES2 yeast expression vector, the BBE from both *Coptis japonica* or *P. somniferum* was also shown to be capable of converting (*S*)-reticuline to (*S*)-scoulerine, when expressed in *S. cerevisiae* yeast culture (Hawkins and Smolke, 2008; Minami et al., 2008). However, attempts to reconstitute the berberine derived alkaloid dihydrosanguinarine, which utilises the product of the berberine bridge enzyme, (*S*)-scoulerine, as an intermediate, demonstrated that the BBE from *P. somniferum* was limiting flux to the downstream in *S. cerevisiae* (Fossati et al., 2014).

CrRO was cloned into the inducible yeast expression vector pYES-DEST52, and transformed into the yeast strain ybt1.

Liquid cultures started from two separate yeast colonies containing the CrRO pYESDEST52 vector were incubated with lyophilised CrRO silenced VIGS extract containing the compounds *m/z* 355 and *m/z* 397. As a negative control, the uninduced CrRO yeast colonies, two GFP expressing ybt1

yeast strains, and two uninduced GFP expressing ybt1 yeast strains were also incubated with CrRO silenced VIGS extract. After 10 hours incubation there was no turnover of either the *m/z* 355 or *m/z* 397, measured by LC-MS from either the yeast pellet or from the yeast media, and untargeted metabolomics by XCMS did not demonstrate the accumulation of any other new compounds (Figure 5-34). While there are several possible reasons why turnover was not observed, the most likely is that the protein is misfolded and non-functional.

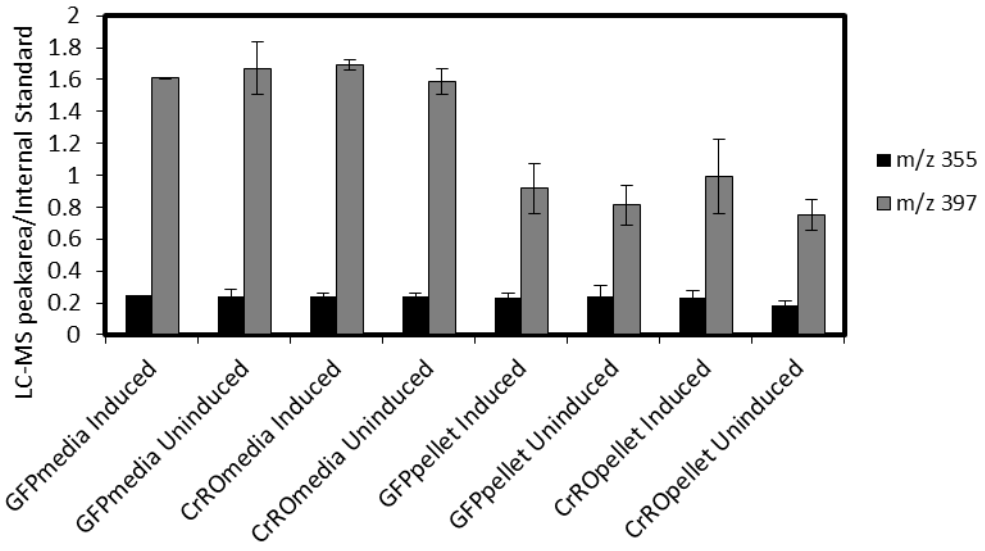


Figure 5-34 *S. cerevisiae* feeding study
Two separate yeast colonies were induced for expression of CrRO, together with a GFP control, and the media supplemented with the aqueous fraction of the VIGS-CrRO extract, and incubated for 10 hours. The yeast media and yeast pellet were extracted in methanol and these samples run on the LC-MS to detect whether the peaks of *m/z* 355 and *m/z* 397 had been consumed.

5.3.4.2 Heterologous expression of CrRO in E. coli

For expression of CrRO in both *E. coli* (and Sf9 insect cells, in Section 5.3.4.3) a suite of vectors was generated utilising the pOPIN cloning series of vectors (Berrow et al., 2007) that encompassed a series of signal sequences and tags that may aid with solubility of the protein. This is summarised in Table 9.

Additionally variants of CrRO lacking a putative N-terminal signal peptide as predicted by the SignalP 4.1 server (Petersen et al., 2011), or had a C-terminal truncation that aligned better with the consensus sequence of both *C. sativa* THCA synthase and *P. somniferum* BBE were cloned

(Figure 5-35). These variants were cloned into vectors that contained either an N- or C-terminal His tag, as well as a vector that contained an insect cell secretion signal and C-terminal His tag.

Table 9 Variants of CrRO was cloned into a suite of pOPIN vectors that are compatible with expression in *E. coli* and Sf9 insect cells.

Gene name	Nterm amino acid number	Cterm amino acid number	Vector	Tag	Comment	E.coli compatible	Sf9 compatible	Variant number
CrRO	1	529	pOPINF	N-term His tag	Full length protein	✓	✓	1
CrRO	24	529	pOPINF	N-term His tag	N-term truncation	✓	✓	2
CrRO	1	524	pOPINF	N-term His tag	C-term truncation	✓	✓	3
CrRO	24	524	pOPINF	N-term His tag	N- and C-term truncation	✓	✓	4
CrRO	1	529	pOPINE	C-term His tag	Full length protein	✓	✓	5
CrRO	24	529	pOPINE	C-term His tag	N-term truncation	✓	✓	6
CrRO	1	524	pOPINE	C-term His tag	C-term truncation	✓	✓	7
CrRO	24	524	pOPINE	C-term His tag	N- and C-term truncation	✓	✓	8
CrRO	1	529	pOPING	Signal sequence and C-term His tag	Full length protein	X	✓	9
CrRO	24	529	pOPING	Signal sequence and C-term His tag	N-term truncation	X	✓	10
CrRO	1	524	pOPING	Signal sequence and C-term His tag	C-term truncation	X	✓	11
CrRO	24	524	pOPING	Signal sequence and C-term His tag	N- and C-term truncation	X	✓	12

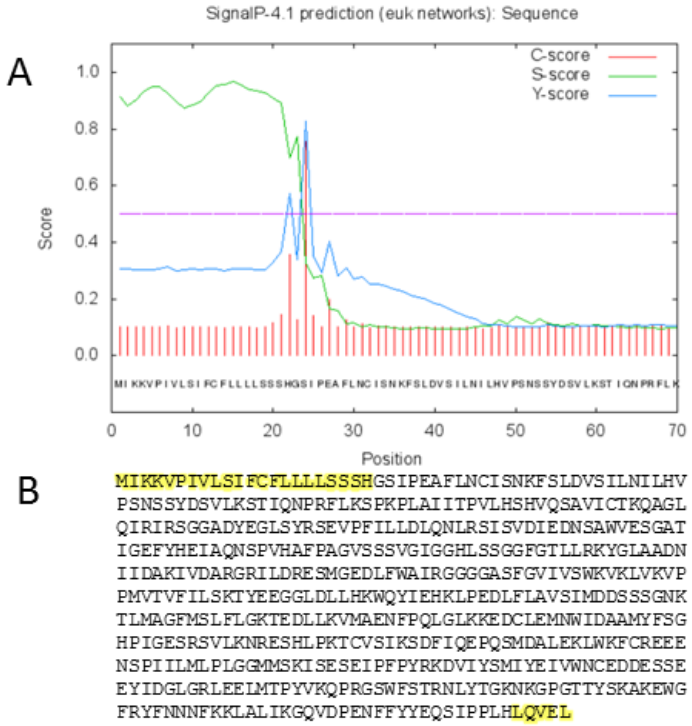


Figure 5-35 Protein sequence of CrRO

A.) Using the secretion signal prediction software SignalP 4.1, a signal sequence was predicted at the N-terminus of the CrRO protein B.) Amino acid sequence of CrRO. Highlighted in yellow at the N-terminus is the predicted signal sequence. The protein variants that were cloned include the full length protein, together with variants that have truncations at the N- and C-terminus of the protein. Amino acid residues removed in the truncated variants at the N- and C-terminus are highlighted in yellow.

The full length CrRO, together with the N-terminal truncation that eliminated the putative secretion signal peptide, were expressed as N-terminal His tagged proteins in the *E. coli* strain SoluBL21 (Figure 5-36). A positive control for flavoprotein expression in *E. coli* from *Pisum sativum* was also expressed (Construct donated by Lorenzo Caputi, John Innes Centre).

As can be seen in Figure 5-36, the flavoprotein from pea was expressed in soluble form, whilst the N-terminal truncated CrRO protein was insoluble. In comparison the full length protein of CrRO, from two separate colonies, did not lead to protein expression, indicating *E. coli* utilising an N-terminal histidine tag is not a suitable system for expression of this protein (Figure 5-36).

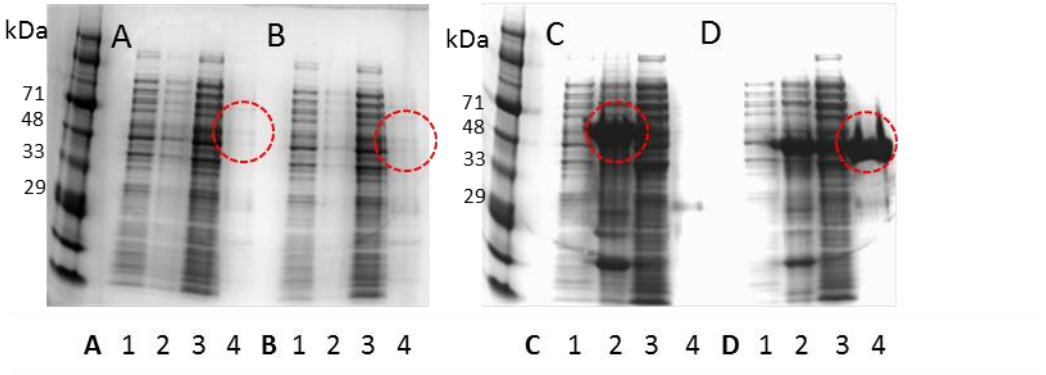


Figure 5-36 *E. coli* expression of CrRO and pea flavoprotein
Expression test of CrRO in pOPINF with an N-terminal histidine tag in *E. coli* strain SoluBL21. 1.) Uninduced cell culture 2.) Insoluble pellet 3.) Flow through 4.) Elution from Ni-NTA resin. A.) Full length CrRO pOPINF colony 1 red circle shows no soluble protein expression B.) Full length CrRO pOPINF colony 2 red circle shows no soluble protein expression C.) N-term truncation CrRO red circle shows insoluble protein expression D.) Flavoprotein from *Pisum sativum* red circle shows soluble protein expression.

5.3.4.3 Expression Screen of CrRO Sf9 Insect Cells

All twelve cloned variants of the CrRO were screened for expression in Sf9 insect cell culture (Figure 5-37). A summary of successful soluble expression from both the secreted and intracellular fraction of Sf9 insect cells is presented in Table 10.

For the Sf9 insect cells, as with *E.coli* expression, an N-terminal histidine tagged variant of CrRO did not result in soluble protein, irrespective of N- or C-terminal truncation.

However, the full length CrRO protein, along with the C-terminal truncated version of CrRO, which both contained a C-terminal histidine tagged expressed at good levels in soluble form. Additionally good expression was also obtained from the pOPING vectors that had either the N-terminal truncation or both N- and C-terminal truncation (Figure 5-37).

This result therefore emphasises that the secretory signal is required for the correct folding of the CrRO protein. An N-terminal histidine tag could likely interfere with the correct sorting of the CrRO protein, and no CrRO protein with an N-terminal tag is expressed in insect cells or *E. coli*. Only the full length and C-terminal truncations with a C-terminal His tag led to soluble protein production in insect cells. The N-terminal truncated variants of CrRO expressed as C-terminal His-tagged proteins, and therefore do not contain the endogenous secretory signal, are lowly expressed in insect cells.

Notably, the variants of CrRO in which the endogenous secretion signal is replaced by the insect cell secretion signal, are abundantly expressed. The CrRO variants that contain both the

endogenous secretion signal and the insect cell secretion signal in tandem are also expressed in insect cells though in lower levels than the variants containing only one secretion signal (Table 10).

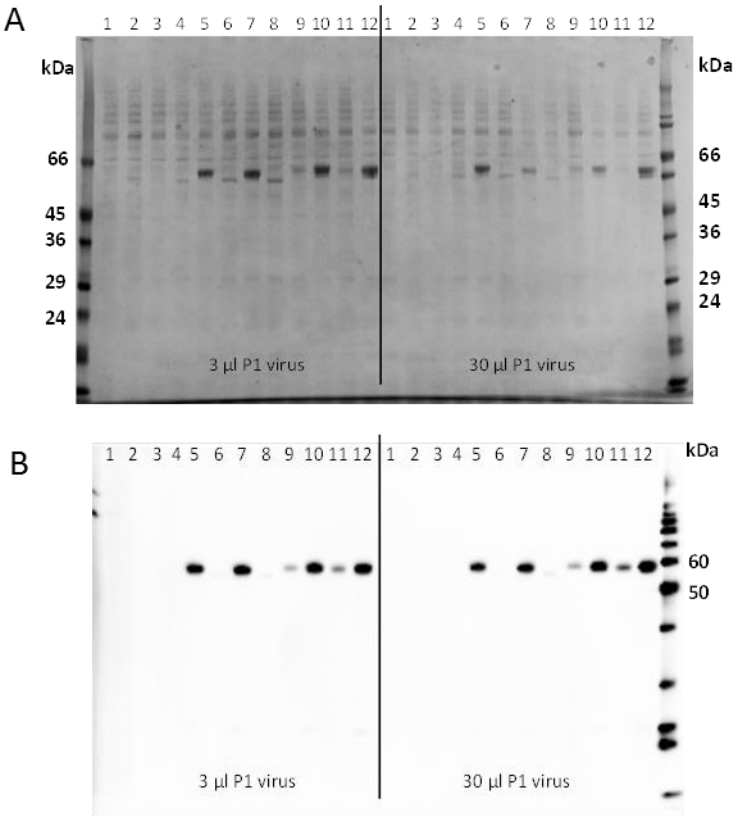


Figure 5-37 Expression screening of pOPIN CrRO constructs in Sf9 Insect cells.
All twelve variants of the CrRO gene were screened for expression in Sf9 insect cells A.) SDS-PAGE gel with coomassie staining for the intracellular protein fraction of the insect cells after purification by Ni-NTA. B.) Western blot of the intracellular insect cell fraction with an anti-His antibody. Lane numbers 1.)CrRO-pOPINF full length 2.)CrRO-pOPINF N-terminal truncation 3.) CrRO-pOPINF C-terminal truncation 4.) CrRO-pOPINF N- and C-terminal truncation 5.) CrRO-pOPINE full length 6.) CrRO-pOPINE N-terminal truncation 7.) CrRO-pOPINE C-terminal truncation 8.) CrRO-pOPINE N- and C-terminal truncation 9.) CrRO-pOPING full length 10.) CrRO-pOPING N-terminal truncation 11.) CrRO-pOPING C-terminal truncation 12.) CrRO-pOPING N- and C-terminal truncation. Screening of insect cell expression was performed by Joanne Nettleship at the Oxford Protein Production Facility.

There is literature precedent for the requirement for a secretion signal for correct folding of BBE-like enzymes in both Sf9 insect cells and *Pichia pastoris*, with soluble protein expressed from either the full length variants of these enzymes or with the endogenous signal sequence replaced by an alternative targeting sequence. For the BBE from *E. californica,* active enzyme was generated in *P. pastoris* from both a construct expressing the native full length BBE sequence, or from a culture where the endogenous signal peptide from BBE was replaced by the

S. cerevisiae α-mating signal sequence (Winkler et al., 2006). Similarly, BBE from *P. somniferum* and DBOX , dihydrobenzophenanthridine oxidase (Hagel et al., 2012), AtBBE-like 15, AtBBE-like 13 (Daniel et al., 2015) and the BBLa, b, c and d proteins from *N. tabacum* (Kajikawa et al., 2011) were functionally expressed in *P. pastoris*, by replacing the endogenous signal peptide with the *S. cerevisiae* α-mating prepro signal sequence.

Protein that is secreted from the insect cells should be correctly folded as it has passed through the secretory pathway. Insect cell media culturing the pOPINE constructs as well as the two highest expressing pOPING constructs, were assessed for secreted protein (Figure 5-38). As would be expected the constructs that resulted in the secreted protein were the CrRO variants that contained either the endogenous signal sequence, or the endogenous signal sequence replaced by the insect cell secretion signal (Figure 5-38). Both the intracellular protein and secreted protein will be assessed for enzymatic activity.

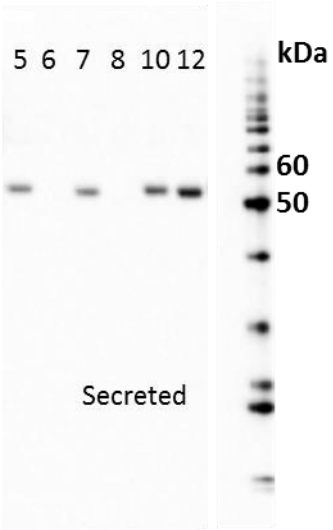


Figure 5-38 Expression screening of pOPIN CrRO constructs in Sf9 Insect cells for secreted expression
All four pOPINE vectors v5.) CrRO-pOPINE full length 6.) CrRO-pOPINE N-terminal truncation 7.) CrRO-pOPINE C-terminal truncation 8.) CrRO-pOPINE N- and C-terminal truncation, together with the best expressing pOPING vectors 10.) CrRO-pOPING N-terminal truncation 12.) CrRO-pOPING N- and C-terminal truncation were assessed for protein expression in both the secreted fraction. Presented is a western blot targeting the histidine tag of the secreted protein. Screening of insect cell expression was performed by Joanne Nettleship at the Oxford Protein Production Facility.

Table 10 Overview of soluble expression of CrRO enzyme variants in Sf9 insect cells.
All twelve variants of the CrRO were screened for soluble expression in Sf9 insect cells. X- no expression, ✓expression, ✓✓good expression.

Gene name	Vector	Tag	Comment	Well No.	Intracellular expression	Secreted expression
CrRO	**pOPINF**	**N-term His tag**	**Full length protein**	**1**	**X**	**X**
CrRO	pOPINF	N-term His tag	N-term truncation	2	X	X
CrRO	**pOPINF**	**N-term His tag**	**C-term truncation**	**3**	**X**	**X**
CrRO	pOPINF	N-term His tag	N- and C-term truncation	4	X	X
CrRO	**pOPINE**	**C-term His tag**	**Full length protein**	**5**	**✓✓**	**✓✓**
CrRO	pOPINE	C-term His tag	N-term truncation	6	✓	X
CrRO	**pOPINE**	**C-term His tag**	**C-term truncation**	**7**	**✓✓**	**✓✓**
CrRO	pOPINE	C-term His tag	N- and C-term truncation	8	✓	X
CrRO	**pOPING**	**Signal sequence and C-term His tag**	**Full length protein**	**9**	**✓**	**✓**
CrRO	pOPING	Signal sequence and C-term His tag	N-term truncation	10	✓✓	✓✓
CrRO	**pOPING**	**Signal sequence and C-term His tag**	**C-term truncation**	**11**	**✓**	**✓**
CrRO	pOPING	Signal sequence and C-term His tag	N- and C-term truncation	12	✓✓	✓✓

As has been discussed previously, CrRO is the first naturally occurring variant of the reticuline oxidase family that does not contain a putative covalent binding site for FAD utilising either a monocovalent histidine attachment, or a bicovalent 8-α*N*[1]-histidine and 6-*S*-cysteine linkage. For the BBE from *E. californica*, site directed mutagenesis of the covalent linkage between FAD and the enzyme at both the conserved histidine and cysteine residue resulted in very low expression in *Pichia pastoris* relative to wild type expression levels (Winkler et al., 2007). This implies that the covalent linkage of FAD to the enzyme is structurally important for proper protein folding.

Similarly in other VAO type proteins the covalent attachment of FAD by the bicovalent linkage has been shown to be integral for protein stability. A H94A/C154A double mutant of the ChitO protein in *Fusarium graminearum* results in no expression of the enzyme (Heuts et al., 2008), and for the glucooligosaccharide oxidase from *Acremonium strictum*, the double mutant H70A/C130A, is expressed but does not bind FAD (Huang et al., 2008). In comparison the GilR oxidoreductase, the double mutant H65A/C125A binds FAD non-covalently, however this mutation abolishes enzymatic activity (Noinaj et al., 2011).

As well as in analysis of the structural integrity of the protein, the potential for the enzyme to bind non-covalently with FAD is also interesting in the context of the redox potential of the FAD. It has been shown previously that covalent binding of FAD increases the redox potential of the enzyme, which in turn limits the final electron acceptor for regeneration of the FAD, hence most BBE enzymes utilise molecular oxygen (Heuts et al., 2009; Leferink et al., 2009; Zafred et al., 2015). The fact that CrRO is predicted to bind non-covalently potentially suggests this enzyme may not utilise molecular oxygen as a final electron acceptor.

We have been able to obtain soluble protein from insect cells of a naturally occurring variant of a BBE like enzyme that is predicted to bind FAD non-covalently, and utilising this expression system we now intend to functionally characterise the protein using the *m/z* 355 from VIGS-CrRO tissue as its putative substrate.

5.4 Conclusion

This chapter presents the first identification of an FAD dependent oxidase that has homology to the BBE family to be physiologically relevant to MIA biosynthesis in *C. roseus*.

Upon silencing of the CrRO gene with two separate pTRV2u vectors that target distinct regions of the gene, we observed the accumulation of two new compounds which have an *m/z* 355 and *m/z* 397. These compounds are not present in VIGS-EV control leaf tissue or in other *C. roseus* tissues such as stem or root. These compounds do not accumulate upon individually silencing other reticuline oxidase like genes from *C. roseus* indicating this CrRO is involved in the biosynthesis of these compounds. Isotopic labelling of these compounds suggested that the compounds with *m/z* 355 and *m/z* 397 are MIAs. Chemical derivatisation suggested that the two compounds are structurally related, with *m/z* 397 being the acetylated derivative of *m/z* 355.
To date the structural elucidation of these alkaloids has not been determined, however we know that these compounds are not yohimban derived alkaloids, or known alkaloids that accumulate in *C. roseus, V. minor* or *V. major*, that the β-carboline ring is not substituted, and that the *m/z* 355 is likely to contain a free hydroxyl group. A number of intermediates have been proposed, and to date the most likely candidates for *m/z* 355 are isositsirikine/sitsirikine and stemmadenine, with the *m/z* 397 being the acetylated derivative of these compounds.

The protein is the first example of a BBE that is predicted to lack a covalent FAD binding site. There are homologues of this protein present in other MIA producing species that also have the putative covalent binding site mutated, including *T. elegans* and *V. minor,* indicating that this reticuline oxidase may work in a region of the MIA pathway that is conserved in these species. Similarly, based on the expression profile of this gene in the *C. roseus* transcriptome, the fact it co-expresses with secologanin synthase, strictosidine synthase, and the three transporters that have been characterised in Chapters III and IV indicates that it is likely to be working in an early to middle part of the MIA pathway.
Three heterologous expression systems were utilised to obtain soluble protein of this CrRO. Enzymatic activity was not observed upon overexpression of the enzyme in *S. cerevisiae*, and soluble protein could not be produced in *E. coli*. This is consistent with the literature preference for characterisation of this family of proteins in Sf9 insect cells or *P. pastoris*.
Soluble protein has been obtained through expression screening in Sf9 insect cells, with the best variants expressed being either the full length protein with a C-terminal histidine tag or a variant

of the enzyme that has the endogenous secretory signal replaced with an insect cell secretion signal with a C-terminal histidine tag. This is consistent with previous literature reports of the need for a secretion signal for correct protein folding. This enzyme is the first report of a naturally occurring variant of the berberine bridge like enzymes identified that is predicted to bind FAD non-covalently, and future work will assess the importance of this putative non-covalent linkage for enzymatic activity.

5.5 Future directions

There are a number of future directions to continue the characterisation of this reticuline oxidase described in this chapter.

1.) We have highlighted the importance of this enzyme in *C. roseus* metabolism by silencing using VIGS. Feeding studies with isotopically labelled tryptamine suggested that the compound is an MIA. It is necessary to increase the amount of VIGS material in order to obtain enough material for the structural characterisation of the MIA compounds that are accumulating in the plant by NMR.

2.) We have silenced four homologues of the reticuline oxidase family in *C. roseus* that do not appear to impact the metabolism of *C. roseus* leaves. However silencing of additional members, as well as combinatorial silencing, of this protein family may identify novel enzymatic activity.

3.) We hypothesise that the accumulating products of *m/z* 355 and *m/z* 397, that occur by silencing the CrRO are likely to be the substrate for the enzyme. We intend to assess whether the CrRO is able to turnover these compounds *in vitro* and further characterise this enzymatic reaction. Since we have demonstrated that the enzyme can be expressed in Sf9 insect cells the stage has been set to carry out these experiments.

4.) The unique property of this enzyme, in that it does not contain a predicted covalent FAD binding site needs to be addressed. It is feasible that either this enzyme binds FAD non-covalently or that it utilises a novel covalent linkage, and it will be interesting to address whether this is integral for enzymatic activity. The protein can be denatured to determine whether the flavin is in fact non-covalently bound, and further mutational analysis can provide insight into how this novel oxidase functions.

Altogether this study expands the substrate repertoire of the BBE-like enzyme family and provides the first evidence that this enzyme family is physiologically involved in MIA metabolism in *C. roseus*.

6 Conclusion

Despite the intensive effort to elucidate the MIA pathway in *C. roseus* in the last decade, there are still a large number of unknown parts to this metabolic pathway. In particular the utilisation of the branchpoint intermediate strictosidine, and how its deglycosylated product is shuttled into different alkaloid structural scaffolds in *C. roseus*, together with how secondary metabolite transporters integrate the subcellular localisation of this metabolic pathway, has largely remained undiscovered.

This thesis has aimed to emphasise the use of a reverse genetics strategy in the identification of a number of genes in the MIA pathway. Chapter II highlighted the use of a transcriptomic mining strategy to identify gene candidates involved in the MIA pathway utilising the assumption that genes involved in metabolic pathways are likely to be co-regulated.

Chapter II further emphasised a number of the pitfalls of using a reverse genetics strategy. While a negative metabolic phenotype could be attributed to the fact the gene is not involved in the pathway, the lack of phenotype could be due to other reasons, for example functional redundancy in the gene of interest being silenced, whereby other homologous genes could compensate for activity, or inefficiency in the ability to silence the gene. As has been shown, a number of genes upon individual silencing do not lead to a phenotype. To date, we have assumed that this is indicative that there is no evidence that these genes are involved in the MIA pathway. However, further functional screening of these genes is necessary to assess with certainty whether the genes do not have a functional role in the MIA pathway. If further functional screening demonstrates activity towards MIA substrates, it potentially suggests that the lack of a physiological phenotype upon silencing could be due to functional redundancy, and therefore a combinatorial VIGS approach, in which multiple genes are silenced in the same experiment, is necessary to pick apart the complexity of this gene redundancy. Development of such a combinatorial silencing approach has been highlighted in Chapters III, IV and V.

However, in comparison, a positive VIGS phenotype clearly provides evidence that the gene is involved in the *C. roseus* MIA pathway, and examples of such positive outcomes are highlighted in Chapters III, IV and V.

Chapter III and IV demonstrated that the individual silencing of two genes encoding tonoplast localised transporters, a MATE protein, CrMATE1952, and an NPF transporter, CrNPF2.9 result in a metabolic phenotype that perturbs MIA metabolism in *C. roseus.* Silencing of CrMATE1952

resulted in the accumulation of an iridoid glycoside derivative, which is likely to be the reduced product of secologanin, secologanol. This product formation is likely to be due to the mislocalisation of secologanin to the cytosol as a result of its inability to enter the vacuole. This product is also produced upon feeding secologanin to wild type leaf tissue, as well as upon reducing secologanin with $NaBH_4$. Despite a reproducible metabolic phenotype affecting iridoid glycoside formation upon silencing, the CrMATE1952 tissue does not lead to a consistent decrease in downstream MIAs. This could be because other transporters are also involved in the translocation of secologanin, or because secologanin formation is not the limiting step in MIA biosynthesis. The formation of strictosidine is a convergent metabolic pathway between secologanin and tryptamine biosynthesis. If there is enough secologanin flux into the vacuole, even despite an 80% knockdown in the VIGS experiment, relative to tryptamine production, then a decrease in downstream MIAs would not be expected.

Silencing of the homologues of CrMATE1952 did not lead to a comparative decrease in MIAs, indicating, to date, there is only evidence that CrMATE1952 is involved in secologanin translocation. Furthermore feeding tryptamine to wild type leaf tissue resulted in the accumulation of strictosidine relative to water fed control tissue, indicating that it is likely that tryptamine biosynthesis, as opposed to secologanin biosynthesis, is limiting strictosidine formation in leaf tissue.

The silencing of CrNPF2.9 resulted in the accumulation of strictosidine in leaf tissue. This phenotype is consistent with this transporter being the exporter of strictosidine from the vacuole, and was accompanied by a decrease in the downstream MIAs such as vindoline and catharanthine. Surprisingly this phenotype also resulted in cell death, which we currently believe is due to the strictosidine accumulation being toxic to the cell, however this requires further investigation.

The combinatorial silencing of both CrNPF2.9 and CrMATE1952 resulted in a complementary phenotype to that observed upon individual silencing, with the accumulation of both secologanol, and strictosidine, together with the decrease in the MIAs vindoline and catharanthine.

The current model for the tonoplastic transport of secologanin and strictosidine based on the work in this thesis is presented in Figure 6-1.

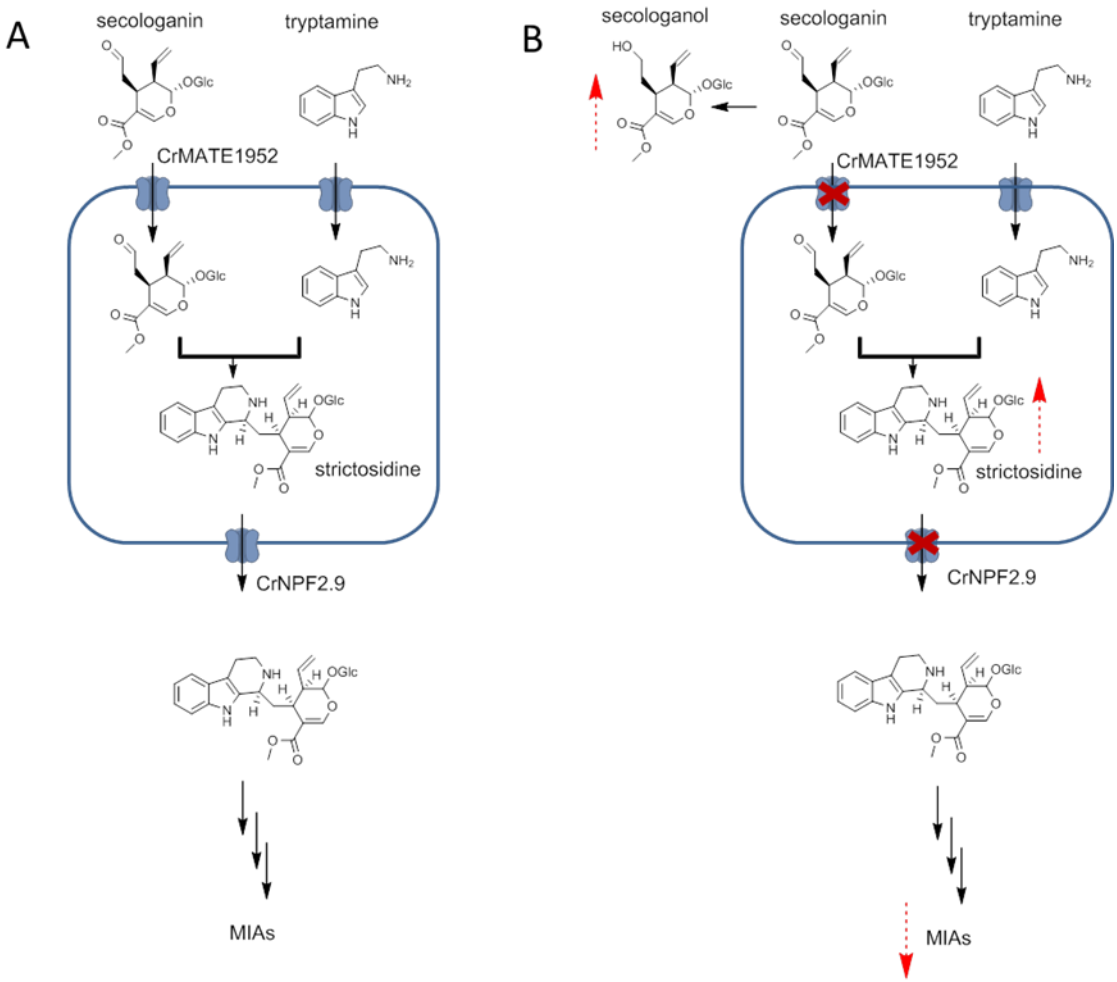


Figure 6-1 The tonoplast transport of secologanin and strictosidine.
A.) VIGS of CrNPF2.9 and CrMATE1952 result in metabolic phenotypes that are consistent with them being the strictosidine exporter from the vacuole and the secologanin importer into the vacuole respectively.B.) Combinatorial silencing of both CrMATE1952 and CrNPF2.9 resulted in the accumulation of secologanol, the accumulation of strictosidine, and the decrease in the downstream alkaloids, vindoline and catharanthine. Red dashed arrows indicate accumulation or depletion of the compound upon VIGS of CrNPF2.9 and CrMATE1952.

Further work on the heterologous expression of these transporters for functional characterisation *in vitro* is ongoing. The result from this reverse genetics approach allows for the targeted design for transport assays, with the predicted substrates for CrNPF2.9 and CrMATE1952, strictosidine and secologanin respectively.

In Chapter V the silencing of a gene encoding a reticuline oxidase like protein, CrRO, resulted in a distinct, reproducible metabolic phenotype, with the accumulation of two new compounds, *m/z* 355 and *m/z* 397. These compounds are likely to be MIAs based upon isotopic labelling studies with d_4-tryptamine, and it has also been demonstrated that the *m/z* 397 compound is likely to be the acetylated product of *m/z* 355. These compounds do not co-elute or have the

same MS^2 fragmentation pattern as known alkaloid species that are commercially available and have the same mass-charge ratio.

This is the first identification that a member of the reticuline oxidase family of proteins is involved in MIA metabolism. Moreover, it is an unusual protein as it has homology to the reticuline oxidases, yet is predicted to lack the conserved covalent linkages to FAD.

Reverse genetics screens, coupled to untargeted metabolomics, has the advantage over the use of large scale screening for enzymatic activity, as it has the potential to discover novel aspects of metabolic pathways. Screening a particular substrate against enzyme libraries inherently biases the screen towards the identification of a preconceived reaction. Although this is incredibly useful for the identification of enzymes that have activity towards known substrates, such screening approaches are not useful for substrates that are expensive or not commercially available and therefore cannot be screened on a large scale, or for the identification of steps in the metabolic pathway that would not be inherently predictable.

As demonstrated with the CrRO, it is unlikely that this enzyme would have been identified through a functional screen, as the substrate for the enzyme is unknown, however this reverse genetics approach allows us to interrogate the effect these genes have on MIA metabolism directly.

Metabolism has the potential to be inherently complex, and for the elucidation of gene discovery in unknown metabolic pathways, it is not necessarily useful or justified to only focus on predicted steps in the metabolic pathway. Such predicted steps in metabolic pathways might be wrong, and it is highly possible that that the metabolic pathway may be non-linear and contain unforeseen reaction steps that significantly contribute to metabolic flux, that may be unpredictable.

Chapter V, and the example of CrRO, therefore highlights the potential role of reverse genetic screens for uncovering non-predictable elements in pathways or for demonstrating the physiological relevance of genes whose substrates are not commercially available. It is through the use of this reverse genetics approach that we now have a platform to identify the compounds accumulating in the CrRO silenced lines, and functionally interrogate this enzyme to ascertain its relevance in MIA metabolism in *C. roseus.*

6.1 Perspectives and Outlook

6.1.1 Novel mining strategies for genes of interest

As has been presented in this thesis, there are a number of mining strategies available for gene discovery. The wealth of sequence information that is becoming available for medicinal plants (Góngora-Castillo et al., 2012; Kellner et al., 2015b; Van Moerkercke et al., 2013; Xiao et al., 2013) allows for direct interrogation and identification of enzymes that could be involved in natural product metabolic pathways. The mining approaches demonstrated in this thesis use the concept that genes in a metabolic pathway are likely to be co-regulated at the transcriptional level, and clustering of the expression data by both hierarchical clustering and self-organising maps, has allowed for the identification of gene candidates that could be involved in the MIA pathway in *C. roseus.* These approaches offer the opportunity for the identification of large numbers of gene candidates. As has been shown in this thesis, however, it is necessary to experimentally verify their involvement in the pathway of interest by either functional characterisation or through reverse genetics to demonstrate their physiological relevance to the metabolic pathway. As demonstrated in Chapter II, while some genes have a highly similar expression profile to known genes in the MIA pathway, no evidence to date implicates these genes in the MIA pathway by the reverse genetics approach used in this thesis.

As has also been presented in Chapter I, not all enzymes involved in the MIA pathway in *C. roseus* have been identified, with many of the downstream enzymes after strictosidine deglycosylation remaining unelucidated. To date all identified genes that act after the deglycosylation of strictosidine do not cluster at the expression level with each other or with other known genes in the pathway. This therefore suggests that alternative mining strategies that are not biased towards co-expression analysis are needed for the elucidation of these steps.

There have been a number of methods proposed for the identification of new genes in this and other secondary metabolic pathways. As highlighted in Kellner et al., 2015b, the *C. roseus* genome has been published, which gives spatial information with regards to the proximity of genes on the genome. There is a growing body of evidence that in some secondary metabolic pathways, genes involved in the same pathway are clustered on the genome (Nützmann and Osbourn, 2014; Osbourn, 2010). This offers an alternative approach to novel gene identification, whereby genes spatially close to known genes in the pathway can be tested by either a reverse genetics approach to assess for physiological relevance to the MIA pathway in the plant, or

through a functional screen of enzymatic activity. This approach has been successful in the identification of tetrahydroalstonine synthase (THAS) (Stavrinides et al., 2015) and further work in this area is ongoing, as more clusters in the MIA pathway are elucidated.

Alternatively it is possible to take genes with known predicted functionality, such as genes identified as glucosyltransferases of natural products, and functionally screen against substrates for which this activity is expected to take place.

The limitations of this approach arise when the novel chemistry of the reactions in the pathway cannot be predicted, or for which the enzymology is novel. A key example was observed with the progesterone-5β reductase family, which has been shown to catalyse novel cyclisation of 8-oxogeranial. From gene annotation alone this functionality could not be predicted, demonstrating that searches of natural product enzymes should not be limited to known families that have been previously shown to be involved in secondary metabolic pathways. Instead it is necessary to look into other gene families, which may not have previously been analysed for natural product enzymatic activity, or which may be annotated incorrectly.

In particular this type of analysis is necessary for enzymatic reactions which are potentially unprecedented *in planta*. In the *C. roseus* pathway, the conversion of dehydrosecodine to either tabersonine or catharanthine, which is a main branchpoint in *C. roseus* to the aspidosperma and iboga classes of alkaloids, is predicted to be a net Diels-Alder reaction (Figure 6-2). In plants this type of enzyme has not yet been identified, and therefore prediction of the enzyme class that can perform this type of reaction is difficult.

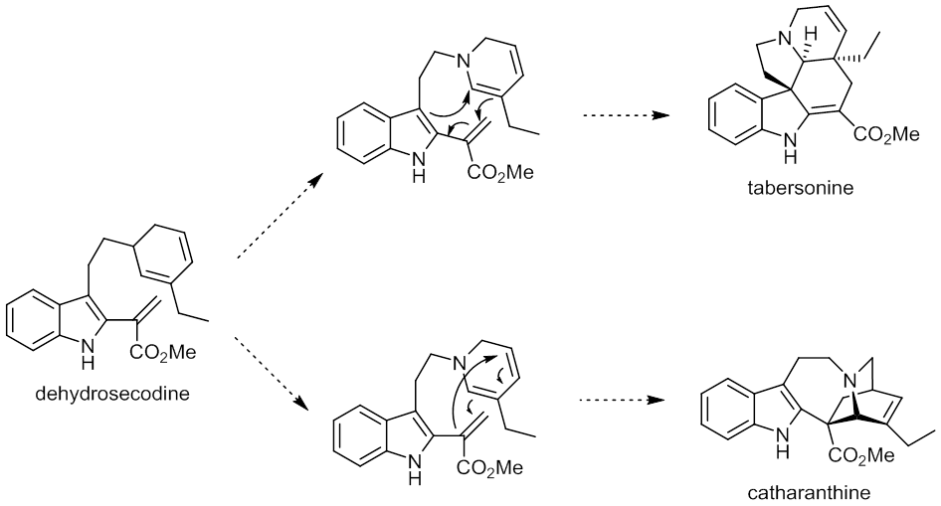


Figure 6-2 Diels-Alder like enzymes in MIA biosynthesis
Dehydrosecodine is the intermediate for both tabersonine and catharanthine and is predicted to form these molecules by a Diels-Alder like mechanism. To date this type of reaction has not been demonstrated to be catalysed by an enzyme *in planta.*

As well as genes that are mis-annotated in terms of function, yet contain domains that can help to predict function, such as co-factor binding sites, there are also genes in plants that have no clear function at all. These genes of unknown function account for up to 40% of genes in plant genomes (de Crécy-Lagard and Hanson, 2007) (Bradbury et al., 2013) and could therefore also be the key source of novel enzymatic activity. Further work by both reverse genetics and functional screening for enzymatic activity is needed to analyse this potential untapped resource.

In *C. roseus*, there are a number of methods by which it is possible to filter these unknown genes to enable more effective unbiased searches. Genes can be filtered on tissue level expression data, to filter genes based upon the localisation to the known biosynthetic site for metabolite synthesis. Other methods include comparative genomics to assess which genes are conserved between different plant species. With the current availability of transcriptomic resources in a number of MIA producing plant species, including *Catharanthus longifolius, Catharanthus ovalis, Tabernaemontana elegans, Vinca minor, Vinca major, Rauwolfia serpentina* and *Camptotheca acuminata,* it is possible that the enzymatic activity for the generation of the iboga and aspidosperma based alkaloids is conserved amongst these species. Homology based searches for genes of unknown function between species that produce similar MIA structural scaffolds may be a method for identifying this enzymatic activity, and this approach could also be utilised for other unknown metabolic pathways that involve novel biochemistry.

As has been observed in Chapter V, the CrRO that has been shown to be involved in MIA biosynthesis in *C. roseus* also has a homologue in both *V. minor* and *T. elegans.* In both of these species, as with the homologue for *C. roseus*, the putative FAD covalent binding site is mutated. Further dissection of these homologues is necessary to assess whether they are playing the same function in MIA metabolism in these species as in *C. roseus*, or whether they catalyse other functions. It is only through comparing this sequence data between species that produce similar compounds that we are able to predict that this reticuline oxidase like enzyme may have common functionality in these medicinal plants.

Therefore the development of novel mining strategies, utilising both transcriptomic and comparative genomic tools, should be utilised for analysis of the genes of unknown function present in plants, to assess whether they are involved in secondary natural product formation, and able to catalyse novel chemistry.

6.1.2 Elucidation of secondary metabolite transporters

While the identification of new enzymes in metabolic pathways is essential, there is a growing precedence and need for the identification of transport steps that might also be involved in the MIA pathway. Despite the need for transport in the MIA pathway, to date only one transporter CrTPT2 has been functionally characterised, and is involved in catharanthine export (Yu and De Luca, 2013).

This thesis has further identified two physiologically relevant tonoplast localised transport proteins, CrMATE1952 and CrNPF2.9 that give phenotypes upon silencing that perturb MIA biosynthesis, indicating that a reverse genetics screen can be successful in transporter identification.

As with enzyme families, the functional annotation of current transporters in plant transcriptomic databases may not truly reflect the physiological substrate that they are transporting.

As more natural product transporter classes are identified we will gain a better understanding of the potential diversity for secondary metabolite transport. The two largest classes of natural product transporters characterised in plants to date are the ABC transporters and the MATEs, however since 2012 there has been a growing appreciation that other transporter families are also involved in natural product transport (Jelesko, 2012; Nour-Eldin et al., 2012).

For example, the NPF transporters are a large class of the MFS proteins found in plants, and in *Arabidopsis* these have been shown to transport glucosinolates (Nour-Eldin et al., 2012).. Similarly the purine permeases from both *Arabidopsis* and *N. tabacum,* which are part of the DMT superfamily, have been shown to transport other substrates such as the pyridine alkaloid nicotine (Jelesko, 2012). The work in this thesis further highlights that the MATE transporters are likely to have a role in the transport of glycosylated iridoids, and secondly that the NPF transporters have an expanded repertoire of substrates, as in *C. roseus* there is strong evidence that this transporter family is involved in strictosidine export from the vacuole.

The fact that there are ~50 NPF transporters in many plant species, suggests that this family is a good starting point to assess novel transport functions for other secondary metabolites, in other medicinal plants. To date, evidence for the role of NPF transporters in secondary metabolism has been limited to *Arabidopsis*, and the work presented in this thesis further expands the physiological relevance of this transporter class to the medicinal plant *C. roseus.*

This also highlights one of the key points in a recent review, addressing the difficulty in identifying the transporters for specialized metabolites (Nour-Eldin and Halkier, 2013) that there is a need to not limit homology searches for secondary metabolite transporters to the MATEs and ABCs, but instead to analyse other transporter families, by either a functional screen or reverse genetics approach, for their involvement in the movement of natural products. This is needed to address whether the neo-functionalisation of primary metabolite transporter families, for specific roles in natural product transport, is a common occurrence in plant metabolism.

There is currently limited information on natural product transport *in planta*, and there is a burgeoning need for further elucidation of natural product transporters in medicinal plants.

6.1.3 The role of spatial organisation in metabolic pathways and its applications to pathway engineering

The role of transport in metabolic pathways in eukaryotic systems is still cryptic. The need for transport of metabolites between organelles in eukaryotic cells is likely to be due to a complicated mix of the evolutionary history of the organelle in which the metabolic pathway is localised, together with the physiological relevance and selection for that metabolic pathway to be localised to that subcellular compartment (Sweetlove and Fernie, 2013).

This study further highlights the role of spatial distribution of the metabolic pathway in the MIA pathway in *C. roseus*. As demonstrated with the silencing of the MATE protein CrMATE1952, there is an increase in the derivatised product of secologanin, secologanol, which is likely to be due to the mis-localisation of secologanin. The role of the spatial distribution of the metabolic pathway could therefore be to act as a pre-emptive mechanism to prevent the derivatisation of intermediates that are needed in the metabolic pathway that have the potential to be reactive. This is especially true in the context of convergent metabolic pathways where the flux to the branchpoint enzyme may not be equal, leading to the accumulation of one of the substrates along the branchpoint.

In the context of the MIA pathway, the build-up of secologanin is likely to be due the fact that the synthesis of tryptamine is limiting relative to the synthesis of the iridoid secologanin.

With the silencing of the NPF transporter CrNPF2.9 the accumulation of strictosidine in the vacuole is toxic to the cell, potentially indicating that there is a need to limit the accumulation of this intermediate in the cell.

The spatial organisation of metabolism needs to be addressed in engineering efforts for the MIA pathway in heterologous systems. For the heterologous production of strictosidine in *S. cerevisiae*, enzymes that were targeted to different subcellular organelles in plants had the signal peptide removed (Brown et al., 2015). In the context of this heterologous pathway in yeast the titers of strictosidine are relatively low, and it was mentioned that the generation of the iridoid secologanin was the limiting step in this pathway (Brown et al., 2015). This is in contrast to the situation in plants in which the limiting step appears to be the supply of tryptamine to strictosidine biosynthesis.

If such yeast strains are engineered to increase the yield of secologanin, whereby tryptamine becomes the limiting step as opposed to iridoid biosynthesis, it may be necessary to partition secologanin biosynthesis to prevent its derivitisation to secologanol.

Secondly, as observed *in planta* upon VIGS of the CrNPF2.9 transporter, the accumulation of strictosidine appears to be toxic. It may therefore be necessary to limit the accumulation of strictosidine in such yeast strains, by ensuring that steady state concentrations of strictosidine remain low. This would require balancing flux from secologanin and tryptamine synthesis, and the downstream enzymes after strictosidine is generated.

As observed in this thesis, the utilisation of plant transporters for engineering purposes in yeast is likely to be limited by the ability to express these transporters efficiently in the heterologous

host. Alternative methods for dealing with the spatial organisation of these metabolic pathways, such as the need to limit secologanin and strictosidine accumulation to prevent derivatisation and toxicity respectively are needed for the generation of higher titers of downstream alkaloids in heterologous systems.

It may be possible that endogenous yeast transporters, or transporters from other species may also be capable of transporting strictosidine, and therefore the plant transporter CrNPF2.9 that has been identified as a strictosidine exporter may not be needed for engineering purposes. The use of transporter engineering in yeast has been successful in the production of high value alkanes, which have the potential for use as next-generation biofuels (Chen et al., 2013). These compounds are cytotoxic, however the expression of two ABC transporters ABC2 and ABC3 from *Yarrowia lipolytica* that have been demonstrated to export alkanes, increases the tolerance in this yeast strain to these products (Chen et al., 2013). Additionally transporters could be evolved by directed evolution to select for alternative substrates, or for increased stability with regards to protein expression in the heterologous hosts.

For the engineering of MIA biosynthesis, *in planta*, in a heterologous host such as *N. benthamiana*, the opportunity for transporter engineering may be easier. As demonstrated in the reconstitution of secologanin biosynthesis in *N. benthamiana* (Miettinen et al., 2014) although secologanin is generated, a number of the terpene intermediates in this pathway are derivatized and acetylated by endogenous metabolism. Although there are no reports that the secologanin is converted to secologanol in *N. benthamiana*, it will be necessary to assess whether the secologanin generated in this engineered system is also derivatised to other products, and whether partitioning biosynthesis to other sub-cellular compartments could increase efficiency to minimise secologanin derivatisation. The reconstitution of strictosidine biosynthesis in *N. benthamiana* resulted in the formation of strictosidine (Miettinen et al., 2014), indicating that *N. benthamiana* has an endogenous transport capacity for secologanin, as the strictosidine synthase utilised in this study would be localised to the vacuole. Expression of the CrMATE1952 may increase the efficiency with which secologanin can enter the vacuole in this hetereolgous host, however, and further heterologous expression of the other transport steps are necessary to assess whether they have any influence on the flux to engineered strictosidine or the downstream MIAs.

Further work is therefore needed to firstly identify novel secondary metabolite transporters in *C. roseus* and other plant species to further expand our understanding of transporter substrate

specificity, and secondly to utilise these transport steps in engineering strategies to assess the relevance of the spatial organisation of metabolic pathways to carbon flux.

6.2 Final Remarks

Altogether this study has highlighted the benefits of utilising a reverse genetics strategy for the identification of physiologically relevant enzymes and transporters in MIA metabolism in *C. roseus.* The study has provided key new insights into the role the subcellular localisation of this metabolic pathway plays in metabolic organisation. This work further expands the repertoire of MATE transporters in their involvement in secondary metabolic processes, and provides the first demonstration in *C. roseus* that NPF transporters are physiologically relevant to MIA metabolism. Moreover this reverse genetics screen has identified an enzyme CrRO that is involved in MIA metabolism, and provides the first evidence that the reticuline oxidase like enzyme family is involved in this metabolic pathway.

In each of these three examples where a positive metabolic phenotype is observed upon silencing the gene, it is unlikely that these genes would have been identified without a reverse genetics method. The two transporters, CrMATE1952 and CrNPF2.9, have proven recalcitrant to efficient heterologous expression to test for *in vitro* activity, and may therefore not have been detected by a functional screen for transport. Secondly the enzyme CrRO, in all likelihood, utilises a substrate that is not commercially available and therefore screening for enzymatic activity would be potentially futile. This work therefore highlights the need for complementary approaches through both functional screening and reverse genetics for gene discovery in medicinal plants.

7 Materials and Methods

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich and all restriction enzymes from New England Biolabs.

7.1 Methods for Bioinformatics

7.1.1 MeV v4.7 software

The transcriptomic data for *C. roseus* was downloaded from http://medicinalplantgenomics.msu.edu together with the expression data in the form of the $\log_2$(FPKM) value for each unigene across the different transcriptome libraries (Góngora-Castillo et al., 2012). This data was filtered based on the expression level in leaf tissue, and upon induction by MeJA. This filtered dataset was input into the MeV 4.7 software (Eisen et al., 1998; Saeed et al., 2003) for hierarchical clustering analysis. Multiple hierarchical clustering dendograms were generated using this dataset using different statistical distance metrics for the clustering algorithm. Known genes in the pathway were highlighted on the dendograms and this was used as a basis for candidate selection for VIGS experiments.

7.1.2 Self-Organising Maps.

This method for application to the C. roseus transcriptome was developed by Marc Jones (John Innes Centre)

The data from the transcriptome dataset in the form of $\log_2$(FPKM) was mean centred and given a unit of variance. Self-organising maps were implemented and visualised in R (Team, 2015) using the `kohonen` package (Wehrens and Buydens, 2007). The map sizes were chosen to give approximately forty contigs per node, with the map shape selected such that the ratio between the two edge lengths was the same as the ratio between the two largest eigenvalues of the data (Vesanto and Alhoniemi, 2000). A toroidal map was used so that every node had the same number of neighbours. This avoided boundary effects occurring when the neighbourhood distance metric was calculated. The known genes in the MIA pathway were mapped to the self-organising maps, and the genes that were also present in the same and neighbouring nodes considered candidates for VIGS. These gene lists were compared to those generated by hierarchical clustering using the MeV v4.7 software. To assess how well the generated self-

organising map fitted the data, two quality metrics were analysed. The first was the within-node distance, which is defined as the mean distance from the weight vector of a node to the samples mapped to it. Therefore, the smaller the within-node distance, the more accurately the node's weight vector represents the samples mapped to the node. The other quality metric used was the inter-nodal distance, defined as the sum of the distances from a node's weight vector to the weight vectors of its neighbouring nodes. The smaller the value, the more similar a node's weight vector is to its surrounding nodes' weight vectors. In order for a node to be classed as high quality in this analysis, both of the described quality metrics for that node had to be in the lowest quartile compared to all nodes. Quality nodes are highlighted green on the self organising map.

7.1.3 Phylogenetic guide trees

MEGA7 software (Tamura et al., 2011) and ClustalW (Larkin et al., 2007) were used for alignment and analysis of phylogenetic relationships. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992), and a bootstrap consensus tree was inferred from 500 replicates.

7.1.4 Prediction Servers

The protein isoelectric point (pI) was calculated using the ExPASy bioinformatics resource portal (Artimo et al., 2012). Signal peptide prediction was performed using the SignalP 4.1 server (Petersen et al., 2011). Prediction of membrane protein ubiquitination state was performed using the CKSAAP_UbSite predictor (Chen et al., 2011).

7.1.5 Geneious

All sequence data, including *C. roseus* transcriptome and genomic data, was analysed in Geneious (Kearse et al., 2012) and was also utilised for all alignments, sequence manipulation and basic local alignment search tool (BLAST) against publicly available sequence databases.

7.2 Methods for General Molecular Biology

7.2.1 Sequencing

All sequencing was performed by SourceBioscience, Cambridge using plasmid DNA as template and both vector specific and gene specific primers for the sequencing reaction.

7.2.2 Primers

All primers were supplied by either Sigma-Aldrich or Integrated DNA Technologies (IDT).

7.2.3 Polymerase Chain Reaction (PCR)

All PCR primers were designed to have a melting temperature of 58 ^{o}C. Unless otherwise stated, the conditions used for the PCR that utilised Uracil specific excision reagent (USER) compatible primers per 25 µl reaction was 5 µl 10X HFBuffer, 2 µl dNTPs, 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 0.2 µl X7 polymerase, 0.2 µl template DNA and 15.3 µl water. For non-USER based PCR the conditions for the PCR per 25 µl reaction was 2.5 µl 10X HFbuffer, 1.5 µl $MgSO_4$, 2.5 µl dNTPs, 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 1 µl KOD polymerase and 15 µl water.

For the PCR, the stepwise gradient consisted of a melting temperature: 95 ^{o}C for 2 min, annealing temperature: 53 ^{o}C for 20 s, and extension temperature: 72 ^{o}C. If the PCR product <1 kbp the extension temperature was 1 min, whilst for PCR product >1 kbp the extension temperature was set at 1 min per kbp. This was repeated for between 28 and 31 cycles.

7.2.4 Agarose Gel Electrophoresis

Agarose gels (1% w/v) were made and run in TAE buffer (40mM Tris(hydroxymethyl)aminomethane (Tris), 20 mM acetic acid, and 1 mM Ethylenediaminetetraacetic acid (EDTA)). GelRed™(Biotium cat. no. 41010) and gel loading dye (NEB) were added to samples before loading and the gels were run at 100 V until the dye reached the lower portion of the gel.

7.2.5 Gel and PCR purification

All PCR products used for downstream cloning were gel purified from the 1% agarose gel using a Nucleospin Gel and PCR clean-up kit (Machery-Nagel cat no. 740609.50) in accordance with the manufacturer's instructions. All DNA was eluted from the column in 40 µl nuclease free H_2O and stored at -20 ^oC.

7.2.6 Common Media used for Molecular Biology

Table 11 Media recipes for molecular biology

Media	Contents	Microorganism cultivated
Lysogeny Broth (LB)	1%(w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, ddH_2O	*E. coli, A. tumefaciens*
SOC	2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose	*E. coli, A. tumefaciens*
YPD	1% yeast extract, 2% peptone, 2% D-glucose	*S. cerevisiae*
SC-Uracil selection medium	0.67% yeast nitrogen base (without amino acids, with NH_4SO_4), Formedium complete synthetic dropout without uracil, 2% glucose	*S. cerevisiae*
SC-Uracil induction medium	0.67% yeast nitrogen base (without amino acids, with NH_4SO_4), Formedium complete synthetic dropout without uracil, 2% galactose, 2% raffinose.	*S. cerevisiae*

7.2.7 Antibiotic selection

For the selection of *E. coli* and *A. tumefaciens* bacterial strains containing plasmids with a specific antibiotic resistance, LB liquid media and agar plates containing the specific antibiotic were used for bacterial growth. The antibiotics used in this study are presented in Table 12.

Table 12 Antibiotics used for bacterial selection

Antibiotic	Final concentration (µg/ml)
Carbenicillin	100
Kanamycin	50
Rifampicin	100
Gentamycin	50
Streptomycin	100

7.2.8 Extraction of RNA from plant tissue for cloning

RNA from both mature and immature of *C. roseus* and *A. thaliana* leaves was extracted and purified with on column DNase digestion with the Qiagen RNeasy Plant Mini Kit (cat no: 74904) using the manufacturer's protocol, for downstream cloning. <100 mg of leaf tissue was ground under liquid nitrogen and used for RNA extraction. The quality of the RNA was checked by gel electrophoresis on a 1% agarose gel and the RNA concentration determined on the nanodrop (NanoDrop ND-1000)

7.2.9 cDNA synthesis from plant RNA for cloning

cDNA was synthesised from the plant RNA extracted using Superscript III reverse transcriptase with RNaseH digestion (Invitrogen Superscript III reverse transcriptase cat no: 18080-044, Invitrogen RNaseOUT inhibitor cat no:10777-013 and Invitrogen oligod$(T)_{12\text{-}18}$ primers cat no: 18418-012) for first strand cDNA synthesis using the manufacturers protocol. Briefly up to 5 µg total RNA was incubated together with 5 µM oligod(T)20 primer and 1 mM dNTPs in a total volume of 10 µl and incubated at 65 ^{o}C for 5 min, and placed on ice. To the RNA, oligodT primer, dNTP solution, 10 µl of cDNA synthesis mix (consisting of RT buffer, 10 mM $MgCl_2$, 20 mM dithiothreitol (DTT), 40 units RNaseOUT and 200 units Superscript III reverse transcriptase) was added. To instigate first strand cDNA synthesis, the mixture was incubated at 50 ^{o}C for 50 min and the reaction was terminated by incubation at 85 ^{o}C for 5 min. To each reaction 1 µl RNaseH was added and incubated at 37 ^{o}C for 20 min. The resulting cDNA was stored at -20 ^{o}C.

OligodT primers were used for cDNA generation as they hybridise to the 3' poly(A) tail of eukaryotic mRNAs. The oligod(T) primers are therefore a more specific priming method, and

decrease the complexity of the cDNA generated, as it only consists of cDNA from expressed mRNA, which is representative of only 1-2% total RNA.

7.3 General Methods for *Escherichia coli*

Table 13 *E. coli* strains used in this thesis

E. coli strain	Source	Function
One Shot® TOP10	Invitrogen	Plasmid propogation
SoluBL21™	Amsbio	Protein expression

7.3.1 Preparation of chemically competent E. coli

A 5 ml culture of the appropriate *E. coli* strain was grown in LB overnight at 37 ^{o}C. This was used as a starter culture, added to 500 ml SOC medium and incubated at 30 ^{o}C until the absorbance at 600 nm (OD_{600}) was between 0.4-0.6. This culture was chilled on ice for 10 min and all remaining steps were kept on ice. The 500 ml cell suspension was centrifuged at 6,000 x g at 4 ^{o}C, this pellet was resuspended in 100 ml ice cold TB buffer (10 mM HEPES pH 6.7, 15 mM $CaCl_2$, 55 mM $MnCl_2$, 250 mM KCl), and this suspension incubated on ice for 10 min before centrifugation at 4000 x g at 4 ^{o}C. The pellet was resupsended in 18.6 ml ice cold TB buffer and 1.4 ml dimethyl sulfoxide (DMSO). The suspension was aliquoted into an appropriate volume, snap frozen in liquid N_2 and stored at -80 ^{o}C until further use.

7.3.2 Chemical transformation of E. coli

2 µl of plasmid was added to 50 µl chemically competent *E. coli* cells, mixed gently and placed on ice. Chemical transformation of the plasmid into the *E. coli* cells was induced by heat-shock by incubating the mixture at 42 ^{o}C for 45 s. After heat shock the cells were placed on ice. The cells were incubated in 150 µl SOC medium immediately following transformation, at 37 ^{o}C for 1 h, before being spread on agar plates containing the appropriate antibiotic selection, and grown overnight at 37 ^{o}C.

7.3.3 Plasmid extraction

Bacterial colonies were grown in 5 ml of LB + appropriate antibiotic selection and the plasmid extracted using a Qiagen Miniprep kit (cat no: 27104) in accordance with the manufacturer's instructions**.** This procedure utilises alkaline lysis together with adsorption of the DNA on a silica

membrane. DNA was eluted from the silica membrane using 40 µl ddH_2O and stored at -20 ^oC for future use.

7.4 General Methods for *Agrobacterium tumefaciens*

Table 14 *A. tumefaciens* strains used in this thesis

A. tumefaciens **strain**	**Function**
GV3101	Virus Induced Gene Silencing
LBA4404	Protein expression in *N. benthamiana*

7.4.1 Generation of electrocompetent A. tumefaciens

Electrocompetent *A. tumefaciens* strains GV3101 and LBA4404 were generated using the protocol for the Biorad Micropulser electroporation apparatus (Catalog no 165-2100). Briefly a 5 ml overnight culture of *A. tumefaciens* was used to inoculate 1 litre of LB broth, and this was grown at 30 ^oC with shaking to an OD of ~5x10^7 cells/ml. All subsequent steps were kept at 4 ^oC. The culture was centrifuged at 3000 x g, the supernatant discarded, and the cell pellet washed with 500 ml ice cold 10% glycerol. This wash step was repeated twice with a volume of 500 ml 10% glycerol and once with a volume of 5 ml 10% glycerol. The final cell pellet was resuspended in 0.5 ml sterile ice cold 1 M sorbitol, and dispensed in 200 µl aliquots for long term storage at -80 ^oC.

7.4.2 Transformation into A. tumefaciens

100-300 ng of plasmid DNA was mixed with 45 µl of electrocompetent *A. tumefaciens* GV3101 or *A. tumefaciens* LBA4404, put into a 2 mm electroporation cuvette and placed on ice for 5 min. Using a Biorad micropulser (165-2100) this mixture was electroporated at 2.2 kV. Immediately after electroporation 150 µl of SOC was added to the cuvette, placed into a 1.5 ml eppendorf, and incubated at 28 ^oC for 2 h. This mixture was then spread onto agar plates containing rifampicin, gentamycin and the plasmid specific antibiotic selection marker for *A. tumefaciens* GV3101, and rifampicin, streptomycin and the plasmid specific antibiotic selection marker for *A. tumefaciens* LBA4404 . The plates were incubated for 2 days at 28 ^oC until individual colonies were visible.

Between 3-5 colonies were picked and grown in 5 ml liquid LB media containing antibiotic selection. Plasmid DNA was extracted and used as a template for PCR. Gene specific primers were used to test for the presence of the insert in the *A. tumefaciens* strain. For colonies testing positive for the insert, the liquid culture was mixed 1:1 with 25% glycerol and stored at -80 ^{o}C for future use.

7.5 General Methods for *Saccharomyces cerevisiae*

Table 15 *S. cerevisiae* strains used in this thesis

***S. cerevisiae* strain**	**Function**
InvSC1	Protein expression
Ybt1	Protein expression
WAT11	Protein expression

7.5.1 Generation of electrocompetent yeast

Electrocompetent *S. cerevisiae* was generated using the protocol for the Biorad Micropulser electroporation apparatus (Catalog no 165-2100). Briefly, 500 ml YPD media was inoculated with an overnight culture of *S. cerevisiae*. The culture was incubated at 30 ^{o}C to a cell density of ~10^{8} cells/ml. The cells were chilled on ice to prevent further growth, and all subsequent steps were performed on ice or at 4 ^{o}C. The culture was centrifuged at 3000 x g for 10 min and the supernatant discarded. The pellet was washed twice with 250 ml ice cold water, and subsequently washed with 20 ml sterile ice cold 1 M sorbitol. The solution was centrifuged, the supernatant discarded and the cell pellet resuspended in 0.5 ml sterile ice cold 1 M sorbitol, which was then dispensed into 200 µl aliquots for storage at -80 ^{o}C.

7.5.2 Transformation into yeast by electroporation

100-300 ng of plasmid DNA was mixed with 45 µl of electrocompetent *S. cerevisiae*, put into a 2 mm electroporation cuvette and placed on ice for 5 min. Using a Biorad micropulser this mixture was electroporated at 1.5 kV. Immediately after electroporation 1 ml of 1 M ice cold sorbitol was added to the cuvette, placed into a 1.5 ml eppendorf. This solution was spread onto SC agar plates without uracil (SC-URA) for selection, with colonies appearing after 2 days of growth at 30 ^{o}C.

7.6 General Methods for Proteins

7.6.1 SDS-PAGE

Protein samples were run using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), using pre-cast 12 well RunBlue SDS-PAGE gels (Expedeon). Protein samples were incubated with LDS-loading dye and 4% β-mercaptoethanol, and heated for 5 min prior to gel loading. The gels were run at 180 V for 1 h in 1 x SDS running buffer (24.8 mM Tris, 192 mM glycine, 0.1% SDS). Protein gels were visualised using InstantBlue (Expedeon).

7.6.2 Western blots

For transfer of the proteins to a nitrocellulose membrane the SDS-PAGE gel was soaked in transfer buffer (25 mM Tris, 190 mM Glycine, 20% (v/v) MeOH, H_2O), before being sandwiched between a nitrocellulose membrane, two pieces of filter paper and two sponges, and placed between a –ve and +ve electrode. The transfer was performed on ice or at 4 ^{o}C at 100 mA for 1-2 hr, until the protein ladder was transferred to the nitrocellulose membrane.

The nitrocellulose membrane was blocked overnight in a solution of TBS (50 mM Tris, 150 mM NaCl, pH 7.5) and 5% milk powder (Marvel dried milk powder). The primary antibody was diluted 1:10,000 in TBS and 5% milk powder and the nitrocellulose membrane incubated in this solution for 2 h at room temperature with gentle agitation. The nitrocellulose membrane was then washed 5 times in TBS containing 0.05% Tween-20 over the course of 45 min. The secondary antibody was diluted 1:5,000 in TBS and 5% milk powder, before incubation with the nitrocellulose membrane for 1 h and then subsequently washed 5 times in TBS and 0.05% Tween-20. The secondary antibody is conjugated to horse radish peroxidase (HRP) and was visualised using a Pierce™ ECL Western Blotting Substrate (Thermo Scientific) using the manufacturer's instructions. This is an enhanced luminol based chemiluminescent substrate for detection of HRP on western blots, and is non-radioactive. The western blot was imaged using the auto exposure on an ImageQuant LAS500.

7.6.3 Bicinchoninic acid (BCA) assay

Protein samples were quantified using a BCA assay, which is a colorimetric detection method for protein concentration. This method utilises the colorimetric detection of Cu^{1+} by BCA upon the reduction of Cu^{2+} to Cu^{1+} by protein in an alkaline medium, resulting in a purple coloured reaction

product with strong absorbance at 562 nm that is linear with protein concentration. A standard curve for absorbance relative to protein concentration was generated using known concentrations of bovine serum albumin (BSA) in the range of 62.5 µg/ml to 1 mg/ml, with each concentration measured in triplicate, and the R^2=0.99 for the standard curve. Protein samples were measured for absorbance at 562 nm and the concentration determined by comparison to the standard curve.

7.7 General Methods Analytical Chemistry

7.7.1 Thin Layer Chromatography (TLC)

Thin layer chromatography was performed using glass backed silica plates (Millipore TLC silica gel 60 F254) cut to 5 cm x 2.5 cm. Around 10 µl of substrate was spotted onto the TLC plate, and the development conditions for each TLC are specified in the text.

7.7.2 Mass Spectrometry

High resolution mass spectrometry using electrospray ionisation (ESI), was performed by Gerhard Saalbach (John Innes Centre) using a Synapt G2 HDMS (Waters).

7.7.2.1 Liquid Chromatography Mass Spectrometry (LC-MS)

All samples were run on either a Thermo-Finnigan instrument equipped with a Deca XP ion trap detector or a Shimadzu LCMS-IT-TOF Mass Spectrometer.
For both instruments the column used for chromatographic separation was a Phenomenex Kinetix 5µ C18 100A (100 × 2.10 mm, 5 µm), and the binary solvent system consisted of acetonitrile (ACN) and 0.1% formic acid in water. The elution program was the following: 1 min isocratic at 12% ACN, 3.5 min gradient up to 25% ACN, 2.5 min gradient up to 50% ACN, 1 min gradient up to 100% ACN, 6 min isocratic at 100% ACN, 1 min gradient down to 12% ACN, and 2.5 min isocratic at 12% ACN.

For the Shimadzu LCMS-IT-TOF mass spectrometer MS/MS fragmentation was induced by collision induced dissociation (CID), with collision energy 50%, collision gas 50% and isolation width of *m/z* 3.0.

For the Thermo-Finnigan instrument peak areas were calculated using the ICIS algorithm in Finnigan's Xcalibur, and for the Shimadzu LCMS-IT-TOF Mass Spectrometer peak areas were calculated using the Shimadzu Profiling Solution software. All representative mass chromatograms presented in this thesis are from samples analysed on the Shimadzu LCMS-IT-TOF Mass Spectrometer. The chromatogram data was extracted and plotted using SigmaPlot version 12.5.

7.8 Methods specific for Virus Induced Gene Silencing

7.8.1 Cloning for VIGS pTRV2 Vectors

7.8.1.1 USER cloning into pTRV2u for individual silencing vectors

Uracil excision-based cloning is a ligation-independent cloning technique. It utilises PCR primers that contain a single deoxyuridine residue near the 5' end. Treatment of the PCR products with USER enzyme leads to the generation of 8 nucleotide 3' overhangs, which can complement the ends of a USER compatible linearised vector (Geu-Flores et al., 2007; Nour-Eldin et al., 2010). The stable hybridisation product that occurs due to the complementarity between the overhanging sequences and the linearised destination vector allows for transformation of *E. coli* without prior ligation (Figure 7-1).

A pTRV2u plasmid containing a USER cassette in the multiple cloning site of pTRV2 vector was designed by Fernando Geu-Flores (Geu-Flores et al., 2012). For USER compatible cloning, the pTRV2u plasmid was linearised by incubation with the restriction enzymes AsisI and Nt.BbvC1 overnight at 37 ^{o}C and the linearised plasmid subsequently gel purified from a 0.7% agarose gel. The linearised plasmid can be stored at -20 ^{o}C until further use.

A 500 bp region of each gene to be transiently silenced was selected and cloned into a non-structural ORF of the pTRV2u vector. In order for selective knockdown of the gene of interest the 500 bp region of the gene is selected to be that which minimises the potential for off target knockdown of homologous gene candidates from the same gene family, based upon a BLAST search of the region of interest against the *C. roseus* transcriptome.

The 500 bp region of the gene of interest for silencing was amplified using PCR from *C. roseus* leaf cDNA with gene specific primers containing overhanging sequences that complement the

overhangs on the pTRV2u linearised vector. These overhanging sequences end in a single deoxyuridine residue upstream of the gene specific region of the primer, that upon treatment with USER enzyme result in long complementary overhangs that can anneal to the overhanging sequences of the pTRV2u linearised vector. The resulting PCR product was gel purified from a 1% agarose gel.

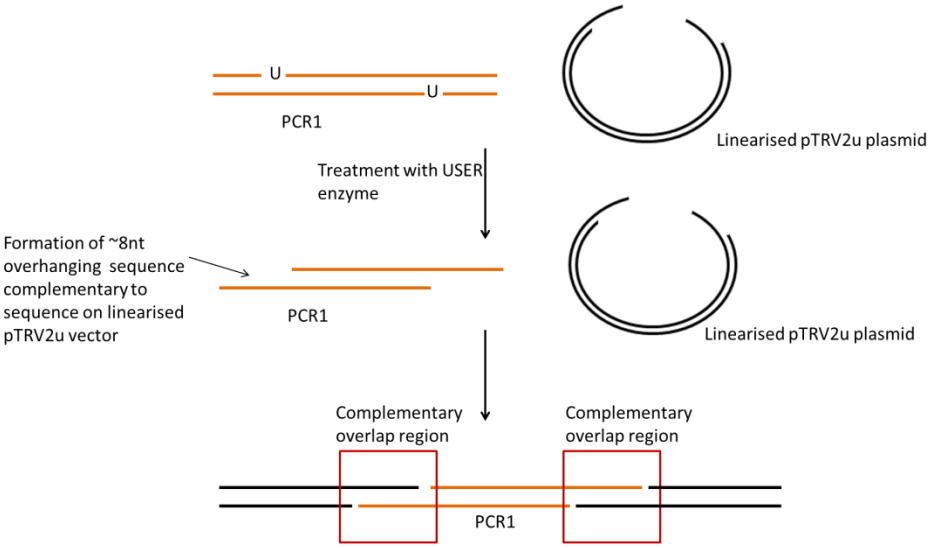


Figure 7-1 Overview of USER-cloning into the pTRV2u plasmid

The 500 bp PCR products, linearised pTRV2u vector, and USER enzyme were incubated together in a reaction volume of 10 µl at 37 °C for 40 min, and subsequently kept on ice for 15 min. This reaction mixture was used to transform 40 µl TOP10 *E. coli* by heat shock without prior ligation. The *E. coli* was recovered and grown in 200 µl SOC media at 37 °C for 40 min, and this was then spread onto kanamycin selective agar plates and incubated at 37 °C overnight. Three colonies from each plate were grown up in 5 ml overnight cultures. Plasmid DNA was extracted by a Qiagen Miniprep kit and tested for the 500 bp insert by PCR using gene specific primers and positive plasmids were sent for sequencing. pTRV2u constructs of the correct sequence were subsequently used for transformation into *Agrobacterium tumefaciens* strain GV3101.

Table 16 pTRV2 Vectors for single gene silencing and primers for VIGS vectors Chapter II

VIGS vectors	Name	VIGS target	Parent plasmid		Primer Sequence (5'-3')	Generated in this study
Transporter Vectors	1770_CrAA1_VIGS	Amino acid transporter	pTRV2u	Fwrd	GGCGCGAUAGCATGAGGAGGATGCTGGG	✓
				Rev	GGTTGCGAUATGTGAACCTCAAAGAATCAATGCGCTTTA	
	24210_CrWAT1_VIGS	DMT family	pTRV2u	Fwrd	GGCGCGAU GGT TGA CTT CAT GTT TAA CAG AAG CTT TGC	✓
				Rev	GGTTGCGAU ATC TTT TCA ATG AGG ATT CAT CTC CTG AAT ACC	
	1443_CrPUP_VIGS	Purine permease	pTRV2u	Fwrd	GGCGCGAUATCAAGTTAGGAGCAGGAGCAAGG	✓
				Rev	GGTTGCGAUTACTTCCTTGCAAGCCCAAAACAA	
	1763_CrABC_VIGS	ABC transporter	pTRV2u	Fwrd	GGCGCGAUCAGATTCCCTAGTTATCGGATTATCATACTTGGTT	✓
				Rev	GGTTGCGAUAAGGATCCAAATTTGGAACATAGAAGAGCAG	
Glutathione s transferase vectors	7197_CrGST1_VIGS	Glutathione s transferase	pTRV2u	Fwrd	GGCGCGAU GTGCTTATAGTATAAGAGTAGATTTAGCACTTAATGCA	✓
				Rev	GGTTGCGAU GCT TTC CAT CTT CAA AAA GAT CTT TGG CAT	
	6502_CrGST2_VIGS	Glutathione s transferase	pTRV2u	Fwrd	GGCGCGAU AAC AAG AAA AGG CCT TTT CCG TTT ACT C	✓
				Rev	GGTTGCGAU GAG CAG CGA CCT TCT GGA G	
	2878_CrGST3_VIGS	Glutathione s transferase	pTRV2u	Fwrd	GGCGCGAU ATT ACG AGT ACA AAG AAG AGA ATT TGG GCA	✓
				Rev	GGTTGCGAU CCA CAC TGT CCC TTT GCA TGC	
Vectors from VIB Ghent	BIS1_VIGS	bHLH transcription factor	pTRV2		Construct was generated by Alex Van Moerkercke	*
	CrP5βR5_VIGS	Progesterone b reductase	pTRV2		Construct was generated by Alex Van Moerkercke	*
	CrP5βR4_VIGS	Progesterone b reductase	pTRV2		Construct was generated by Alex Van Moerkercke	*
	CrP5βR2_VIGS	Progesterone b reductase	pTRV2		Construct was generated by Alex Van Moerkercke	*

Table 17 pTRV2 Vectors for single gene silencing and primers for VIGS vectors Chapter III

VIGS vectors	Name	VIGS target	Parent plasmid		Primer Sequence (5'-3')	Generated in this study
MATE vectors	CrMATE1952_VIGS	MATE	pTRV2u	Fwrd	GGCGCGAU CCAAACAAAACTATGAAATAAACCAACCACTGT	✓
				Rev	GGTTGCGAU GTTTTGAGAATATGTAAACAAACATAATAACAATGCCTGTTA	
	CrMATE10740_VIGS	MATE	pTRV2u	Fwrd	GGCGCGAU TTGGACAGACCTTGAAGCTGG	✓
				Rev	GGTTGCGAU ATCCCCTGGCAACACCTG	
	2720_CrMATE_VIGS	MATE	pTRV2u	Fwrd	GGCGCGAUGCTTGAAAGATGCTGATGTTGCAGTT	✓
				Rev	GGTTGCGAUGCATTGCCCACCATATGCCAT	

Table 18 pTRV2 Vectors for single gene silencing and primers for VIGS vectors Chapter IV

VIGS vectors	Name	VIGS target	Parent plasmid		Primer Sequence (5'-3')	Generated in this study
NPF **vectors**	CrNPF2.9_VIGS	NPF transporter	pTRV2u	Fwrd	GGCGCGAUCGCTTGCGGTAGCTGGATTTA	✓
				Rev	GGTTGCGAU CCCATGACCAACTCACATTATCTTGAAC	
Vectors from VIB Ghent	CrNPF2.4_VIGS	NPF transporter	pTRV2		Construct was generated by Alex Van Moerkercke	*
	CrNPF2.6_VIGS	NPF transporter	pTRV2		Construct was generated by Alex Van Moerkercke	*

Table 19 pTRV2 Vectors for single gene silencing and primers for VIGS vectors Chapter V
Additionally primers for the amplification of the CrRO homologue are also highlighted

VIGS vectors	Name	VIGS target	Parent plasmid		Primer Sequence (5'-3')	Generated in this study
Reticuline oxidase vectors	CrRO_VIGS	Reticuline oxidase like protein	pTRV2u	Fwrd	GGCGCGAU TCTCTTCCTCTGTTGGAATTGGC	✓
				Rev	GGTTGCGAU TCCAATTCATTTCTAAGCAATCTTCCTTTTTCA	
	CrRO_Cterm_VIGS	Reticuline oxidase like protein	pTRV2u	Fwrd	GGCGCGAU ATCAGAATCAGAAATCCCATTTCCTTACAGAA	✓
				Rev	GGTTGCGAU GGGCCAACTTTTTGAAATTATTATTAAAATACCG	
	locus_5372_VIGS	Reticuline oxidase like protein	pTRV2u	Fwrd	GGCGCGAU TTG TGT TTC TTC TGT TAT CAT CTT CAT TTC TGA TGA	✓
				Rev	GGTTGCGAU CAG TGG CAC CGG CCT	
	locus_2174_VIGS	Reticuline oxidase like protein	pTRV2u	Fwrd	GGCGCGAU TCC CTC CTC CTC TGT TTT TGC CTT CTA A	✓
				Rev	GGTTGCGAU CCA AGG GTA GCA CCA GAT TGG	
	locus_16199_VIGS	Reticuline oxidase like protein	pTRV2u	Fwrd	GGCGCGAU ACT ATG AAG GCG TGT CTT ACA GAT CC	✓
				Rev	GGTTGCGAU TCC TAT AAA TTG CCA TTT GTG AAT AAG ATT TAT GGC	
		CrRO homologue full length		Fwrd	AGGAGATATACCATG AAC TAT ATA TGT GCG AAT ATG GAA GAA AAG TAT GGC	
				Rev	GTGATGGTGATGTTT GAG TTC ATC TTG TAA ATG AAG GGG CGG	
		CrRO homologue truncated		Fwrd	AGGAGATATACCATG ATC ACG AAA GTC CCA AAA GTT GTT TCC	
				Rev	GTGATGGTGATGTTT GAG TTC ATC TTG TAA ATG AAG GGG CGG	
		CrRO homologue 100bp fragment 1		Fwrd	ATA ATG TAC CAA GTA ATG TGG AAT TGT GAA GAC G	
				Rev	TTC ATA GCA TAA GGA CTC ATC ATC TCA TAT AGC C	
		CrRO homologue 100bp fragment 2		Fwrd	CTT CTT ATA CTC TCA TCA TCA TCC CAT GGC	
				Rev	ATA ATG AAG AAT GCT GAA AAT GGA TTC ATC TGA TTC A	

7.8.1.2 USER fusion cloning for generation of combinatorial VIGS vectors

USER fusion cloning is a versatile method that allows for the ligation independent fusion of multiple PCR products into a vector, and was used for the generation of multiple VIGS fusion constructs, allowing for the stacking of ~500 bp regions of genes into one pTRV2u plasmid to allow for the combinatorial silencing of multiple genes in one VIGS experiment.

As with the standard USER cloning the versatility of this method is due to the ability to generate long custom designed complementary overhanging sequences that allow for the formation of a stable hybridisation product that can be transformed into *E. coli* without prior ligation (Figure 7-2). Custom overhanging sequences can be designed to generate a series of PCR products that have complementary overhangs that upon treatment with USER enzyme will result in the formation of a stable hybridisation product that can be transformed into *E. coli.*

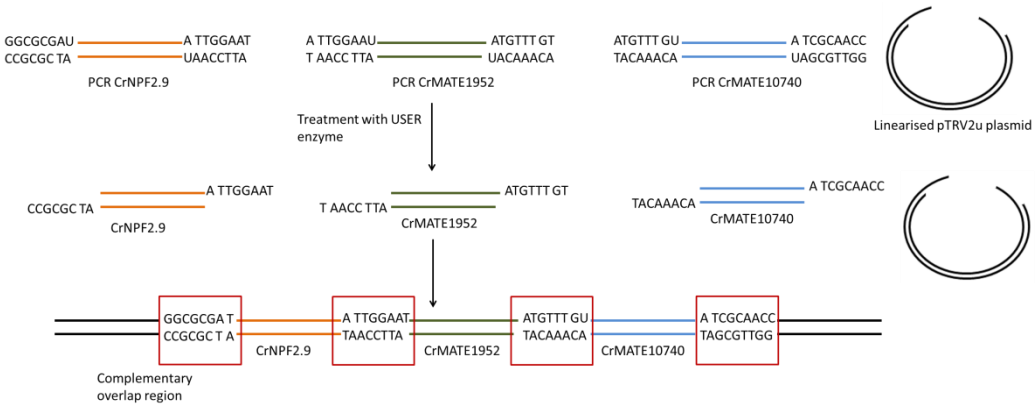


Figure 7-2 Overview of USER-cloning into the pTRV2u plasmid for the generation of combinatorial fusion vectors
The example presented in this schematic resulted in the generation of the triple silencing vector pTRV2u-CrNPF2.9_CrMATE1952_CrMATE10740 which was used in Chapter IV.

For the formation of the fusion vectors each of the 500 bp gel purified PCR products for formation of the fusion, linearised pTRV2u vector, and USER enzyme were incubated together in a reaction volume of 10 µl at 37 °C for 40 min, and subsequently kept on ice for 15 min. This reaction mixture was used to transform 40 µl TOP10 *E.coli* by heat shock without prior ligation. The *E. coli* was recovered and grown in 200 µl SOC media at 37 °C for 40 min, and this was then spread onto kanamycin selective agar plates and incubated at 37 °C overnight. Three colonies from each plate were grown up in 5 ml overnight cultures. Plasmid DNA was extracted by a Qiagen Miniprep kit and tested for the 1000-1500 bp insert by PCR using gene specific primers and positive plasmids were sent for sequencing. pTRV2u constructs of the correct sequence were subsequently used for transformation into *Agrobacterium tumefaciens* strain GV3101.

Table 20 Primers for the generation of VIGS fusion vectors

Primer Name	Primer Sequence (5'-3')	Primer ID
CrNPF2.9_forward_VIGS	GGCGCGAUCGCTTGCGGTAGCTGGATTTA	1
CrNPF2.9_reverse_VIGS_fusion	ATTCCAAU GCCACTGATTACTGAAGCAGCATT	2
CrNPF2.9_CrMATE1952_forward_VIGS_fusion	ATTGGAAUCCAAACAAAACTATGAAATAAACCAACCACTG	3
CrMATE1952_forward_VIGS	GGCGCGAU CCAAACAAAACTATGAAATAAACCAACCACTGT	4
CrMATE1952_reverse_VIGS	GGTTGCGAU GTTTTGAGAATATGTAAACAAACATAATAACAATGCCTGTTA	5
CrMATE1952_reverse_VIGS_fusion	ACAAACAU AAT AAC AAT GCC TGT TAG GGT TAG AAG AAT TG	6
CrMATE1952_CrMATE10740_forward_VIGS_fusion	ATGTTTGU TGG ACA GAC CTT GAA GCT GG	7
CrMATE_10740_reverse_VIGS	GGTTGCGAU ATCCCCTGGCAACACCTG	8
1952_3327_forward_VIGS_fusion	ATGTTTGU GCT GCT TCC TCC CTT GGC AAT AAT	9
3327_reverse_VIGS_fusion	AGCTACU G ATA TAT ATG CAC TTG GAT TAA TAA TG CTC	10
3327_8527_forward_VIGS_fusion	AGTAGCU CTG CCT CCT TAG GCA ACC AG	11
8527_reverse_VIGS	GGTTGCGAU CAA TGA AAT GTA GGC ACT TGG CTT CAC TAT A	12
5608_CrRO_forward_VIGS	GGCGCGAU TCTCTTCCTCTGTTGGAATTGGC	13
5608_CrRO_reverse_VIGS_fusion	ATT CAT TTC UAA GCA ATC TTC CTT TTT CAA TCC AAG	14
5608_2174_forward_VIGS_fusion	AGAAATGAAU TCC CTC CTC TGT TTT TGC CTT CTA A	15
2174_CrRO_reverse_VIGS_fusion	AGC ACC AGA UTG GAC CCA AGC	16
2174_5372_forward_VIGS_fusion	ATCTGGTGCU TTG TGT TTC TTC TGT TAT CAT CTT CAT TTC TGA TGA	17
5372_CrRO_reverse_VIGS	GGTTGCGAU CAG TGG CAC CGG CCT	18

Presented in Table 21 are the combinations of primers used andPCR products used for the generation of the combinatorial VIGS vectors utilised in this thesis.

An overview of all VIGS vectors utilised in this thesis is presented in Table 22.

Table 21 Combinations of primers and PCR products for the VIGS fusion vectors using primer ID in Table 20

VIGS fusion Vectors	VIGS target	Parent plasmid	Primers	PCR product 1	PCR product 2	PCR product 3	Chapter
CrNPF2.9_CrMATE1952_Double_VIGS_fusion	CrNPF2.9 and CrMATE1952	pTRV2u	Fwrd	1	3	X	III
			Rev	2	5	X	
CrMATE1952_CrMATE10740_Double_VIGS_fusion	CrMATE1952 and CrMATE10740	pTRV2u	Fwrd	4	7	X	III
			Rev	6	8	X	
CrNPF2.9_CrMATE1952_CrMATE10740_Triple_VIGS_fusion	NPF transporter and MATE	pTRV2u	Fwrd	1	3	7	IV
			Rev	2	6	8	
CrMATE1952_CrMATE3327_CrMATE8527_Triple_VIGS_fusion	CrMATE1952 homologues	pTRV2u	Fwrd	4	9	11	III
			Rev	6	10	12	
CrRO5608_5372_2174_Triple_RO_VIGS_fusion	Reticuline oxidase like protein	pTRV2u	Fwrd	13	15	17	V
			Rev	14	16	18	

Table 22 Overview of all VIGS vectors generate and used in this thesis

VIGS vectors	Name	VIGS target	Parent plasmid	Chapter
Transporter vectors	CrNPF2.9_VIGS	NPF transporter	pTRV2u	IV
	CrMATE1952_VIGS	MATE	pTRV2u	III
	CrMATE10740_VIGS	MATE	pTRV2u	III
	CrNPF2.9_CrMATE1952_Double_VIGS_fusion	NPF transporter and MATE	pTRV2u	IV
	CrMATE1952_CrMATE10740_Double_VIGS_fusion	MATE	pTRV2u	III
	CrNPF2.9_CrMATE1952_CrMATE10740_Triple_VIGS_fusion	NPF transporter and MATE	pTRV2u	IV
	CrMATE1952_CrMATE3327_CrMATE8527_Triple_VIGS_fusion	MATE homologues Triple silencing	pTRV2u	III
	2720_CrMATE_VIGS	MATE	pTRV2u	III
	1770_CrAA1_VIGS	Amino acid transporter	pTRV2u	Appendix
	24210_CrWAT1_VIGS	Auxin Induced 5NG4	pTRV2u	II
	1443_CrPUP_VIGS	Purine permease	pTRV2u	II
	1763_CrABC_VIGS	ABC transporter	pTRV2u	II
Reticuline oxidase vectors	5608_CrRO_VIGS	Reticuline oxidase like protein	pTRV2u	V
	5608_CrRO_Cterm_VIGS	Reticuline oxidase like protein	pTRV2u	V
	5372_CrRO_VIGS	Reticuline oxidase like protein	pTRV2u	V
	2174_CrRO_VIGS	Reticuline oxidase like protein	pTRV2u	V
	5608_5372_2174_Triple_RO_VIGS_fusion	Reticuline oxidase like protein	pTRV2u	V
Glutathione s transferase vectors	7197_CrGST1_VIGS	Glutathione s transferase	pTRV2u	II
	6502_CrGST2_VIGS	Glutathione s transferase	pTRV2u	II
	2878_CrGST3_VIGS	Glutathione s transferase	pTRV2u	II
Vectors from VIB Ghent	CrNPF2.4_VIGS	NPF transporter	pTRV2	IV
	CrNPF2.6_VIGS	NPF transporter	pTRV2	IV
	BIS1-VIGS	bHLH transcription factor	pTRV2	II
	CrPOR5_VIGS	Progesterone b reductase	pTRV2	II
	CrPOR4_VIGS	Progesterone b reductase	pTRV2	II
	CrPOR2_VIGS	Progesterone b reductase	pTRV2	II

7.8.2 Virus Induced Gene Silencing

7.8.2.1 C. roseus plant growth conditions

Two cultivars of *C. roseus* have been used for VIGS in this study, cv. Little Bright Eyes (purchased from B and T world seeds) and cv. Sunstorm Apricot (Syngenta). All plants were grown in a walk in growth chamber at 25 ^{o}C under 12 h days using the John Innes compost mix No.2 (peat based).

7.8.2.2 VIGS Experiment

For any given VIGS experiment, *Agrobacterium* containing the plasmids pTRV1, pTRV2-Empty Vector, pTRV2-Magnesium chelatase, and the pTRV2 constructs generated containing the 500 bp of the gene of interest to be transiently knocked down, were grown in 5 ml LB liquid media containing rifampicin, gentamycin and kanamycin overnight. The cultures were centrifuged, and resuspended in *Agrobacterium* infiltration buffer (10 mM MES, 10 mM $MgCl_2$, 100 µM acetosyringone, pH 5.5) to an optical density of 0.7 at 600 nm, and incubated for 2 hr at room temperature. Each pTRV2 construct was then mixed 1:1 with the pTRV1 *Agrobacterium* strain, and this mixture was used for pinching.

For each pTRV2 construct a minimum of eight eight week old *C. roseus* plants were pinched on the stem using a pair of bent forceps, just below the meristem, and a bubble of *Agrobacterium* was placed over the wound site. The pTRV2-MgChl construct, which silences magnesium chelatase, and therefore chlorophyll biosynthesis, was used as a positive visual control to ensure the silencing experiment has worked. Between 21-25 days post pinching the MgChl plants turn yellow as a result of downregulation of chlorophyll biosynthesis and the VIGS experiment was harvested.

For each plant pinched with the pTRV2-EV construct and the pTRV2-gene of interest constructs, the newest leaves emerging after the wound site were harvested. The leaf tissue was ground using a retsch ball mill and stored at -80 ^{o}C. This tissue was used for both analysis by LC-MS to assess the metabolite profile upon gene silencing, as well as qPCR to detect the change in expression level of the gene being silenced.

7.8.2.3 LC-MS of C. roseus tissue

Between 10-20 mg of ground tissue of each plant was weighed, collected into 200 µl methanol containing 40 µM caffeine as an internal standard and incubated at 60 ^{o}C for 2 h. After a 30-min centrifugation step at 5,000 x *g*, an aliquot of the supernatant (50 µl) was mixed with an equal volume of water and analysed by LC-MS. VIGS experiments that gave a positive metabolic phenotype were run and quantified on both the Thermo-Finnigan instrument and the Shimadzu-IT-TOF to ensure reproducibility of the VIGS experiment. One difference that needs to be highlighted is that for the iridoid glycoside secologanin, the dominant ion is different between the two instruments. For the Thermo-Finnigan instrument it runs as the proton adduct *m/z* 389, whilst for the Shimadzu-IT-TOF it runs as the sodium adduct *m/z* 411. Additionally the new peak accumulating in VIGS-CrMATE1952 that we believe is secologanol runs as the proton adduct *m/z* 391 on the Thermo-Finnigan instrument, and as the sodium adduct *m/z* 413 on the Shimadzu-IT-TOF.

7.8.2.4 Untargeted metabolomics using XCMS

LC-MS metabolite data from VIGS samples was converted to mzXML files for analysis by the programme XCMS (Smith et al., 2006; Team, 2015). This can be downloaded from http://www.bioconductor.org/ and run in the statistical programme R by running the command `library(xcms)`. The VIGS data in mzXML format consists of two treatments, VIGS-EV and VIGS-gene of interest, which are separated into subdirectories. A peak list for each individual file is generated by the function `xcmsSet()`. A grouping algorithm implemented by the command `group()` cross references the peaks in individual files and creates a common peak list with similar mass and retention time. Using peaks common to all samples, XCMS can estimate how the retention times have changed between runs, and re-calculate the retention time of the peaks using the `retcor()` method. This new dataset can subsequently be re-grouped with higher stringency. Some peaks do not exist in all files and the command `fillPeaks ()` integrates the baseline in samples where no peak was found to prevent a peak list with 0 values. The peak list of individual files, which contains the peak area, together with retention time correction, was exported into Excel. The peak areas for each individual sample were corrected for the peak area of caffeine, and the weight of the leaf. This normalised peak area for each mass was used for comparison between VIGS-EV and VIGS-gene of interest treatments, with the average peak area for each treatment, the fold change between VIGS-gene of interest and VIGS-EV controls, as well as a t-test calculated between the two sets. To correct

the p-value for multiple testing, the peaks present in the XCMS dataset were ranked based on their p-value, and the Benjamini-Hochberg method for False Discovery Rate (FDR) used to correct the p-value for the occurrence of false positives (Benjamini and Hochberg, 1995).

7.8.2.5 qPCR VIGS tissue

7.8.2.5.1 RNA extraction

RNA needs to be extracted for in order to address the expression level of the gene in the plant. Less than 100 mg of plant tissue was used for extraction of RNA from each replicate in a VIGS experiment using a Qiagen RNeasy plant mini kit in accordance with the manufacturer's instructions. For each set of pTRV2u plants, a minimum of eight plants were selected, that gave the best metabolic phenotype, for RNA extraction and qPCR. RNA quality was assessed on a 1% agarose gel and the concentration was measured on the nanodrop (ND-1000).

7.8.2.5.2 cDNA synthesis for qPCR

cDNA for qPCR from each replicate was synthesised using the Biorad iScript cDNA synthesis kit (catalog no 170-8891). Briefly the 5x iScript reaction mix, iScript reverse transcriptase, nuclease free H_2O and 1 µg RNA from each VIGS sample was mixed in a total volume of 20 µl and incubated for 5 min at 25 ^oC, 30 min at 42 ^oC, and 5 min 85 ^oC , resulting in the formation of the cDNA which was used for qPCR.

7.8.2.5.3 qPCR conditions

Real-time qPCR, to assess gene expression of the genes silenced by VIGS, was performed on a CFX96 touch Real-Time PCR system (Biorad) using the SYBR Green I technology. Each reaction was performed in a reaction volume of 25 µl and consisted of a normalised concentration of cDNA, 0.2 mM forward and reverse primers and the SsoAdvanced SYBR Green Supermix (Biorad catalog no: 1725271), which contains dNTPs, Sso7d fusion polymerase, $MgCl_2$, SYBR® Green I and ROX normalisation dyes. The qPCR reaction was initiated by a denaturation step at 95 ^oC for 10 min followed by 40 cycles at 95 ^oC for 15 s and 60 ^oC for 1 min.

For each qPCR experiment, to assess the relative expression level of genes of interest that have been silenced by VIGS, cDNA made from eight replicates of pTRV2-EV and eight replicates of pTRV2-gene of interest were used, and the qPCR reaction performed in technical duplicate. The

relative quantification of gene expression was performed using the ΔΔCq cycle method using either the 40S ribosomal protein 9 (Rps9), 60S acidic ribosomal protein (RPP0C) or elongation initiation factor 4a (EIF4a) as reference genes for normalisation. These reference genes have previously been published for use in VIGS qPCR experiments in *C. roseus* (Asada et al., 2013; Geu-Flores et al., 2012; Liscombe and O'Connor, 2011; Salim et al., 2014). Although there are reports of other qPCR targets being validated for *C. roseus* (Pollier et al., 2014), the primers that were published for these genes, annotated as CrN227, in our hands, did not give reliable data (Appendix) and therefore we utilised the previously published primers for normalisation in the VIGS experiment.

7.8.2.5.4 Primer optimisation

Prior to performing the qPCR reaction on the VIGS experiment, the qPCR primers for each gene of interest to be measured were optimised. Each primer pair was used in a standard PCR reaction against *C. roseus* cDNA to ensure only one target was amplified.

For each primer pair a standard curve was generated to ensure the primer pair amplification efficiency had a linear relationship with cDNA concentration. cDNA from the VIGS-EV and VIGS-pTRV2 constructs was pooled, and a serial dilution of this pooled cDNA was generated at concentrations of 1:5, 1:10, 1:20, 1:80, 1:320. A qPCR reaction for each primer pair was performed in technical triplicate at each concentration, and the quantitation cycle (Cq) plotted against log[cDNA]. Only primer pairs that gave a linear regression (R^2) value of 0.99 were used in this study. This standard curve was also used to assess the appropriate starting concentration of cDNA to use for the qPCR reaction and ensure the Cq value for each gene to be analysed in the experiment occurred between 20-30 cycles in the qPCR reaction.

Together with this, the melting curves generated were used to ensure that the primer pair is specific in its amplification and that it has only one amplified target in the cDNA. The primer efficiency values generated in this study were between 99-101%.

Table 23 qPCR primers used in this study

	Primer Name	Target	Sequence (5'-3')
qPCR primers	CrNPF2.9_qPCR_Forward	NPF transporter	AAG AAG CTC TCC TTG TCA AAC AAA GAA GAA G
	CrNPF2.9_qPCR_Reverse	NPF transporter	CTT GAA ACT TTC ATC AGG TAC AGT AGC AGG
	CrMATE1952_qPCR_Forward	MATE	CCG GTG TTG CTG TTG GAT GT
	CrMATE1952_qPCR_Reverse	MATE	CAA GTT TGA AGT AGA ACC CAA GAA GTG CA
	CrMATE10740_qPCR_Forward	MATE	CTT GAA CTT GCC GCT GCA AAT CTT
	CrMATE10740_qPCR_Reverse	MATE	CCT TGT CCA CAC AAA CTT TCC AGT GC
	CrRO_qPCR_Forward	Reticuline oxidase like protein	CCG AGA ATC TCA TCT TCC AAA GAC ATG
	CrRO_qPCR_Forward	Reticuline oxidase like protein	GCG TTT CGA TCA AAT CAG ACT TTA TTC AAG AAC
	locus_5372_qPCR_Forward	Reticuline oxidase like protein	GAT AAA ACA ATT GTA ACA GCC TAT AAC GCG TTG
	locus_5372_qPCR_Reverse	Reticuline oxidase like protein	TAC AAT CTT TTC GAG TCA ATC CCA GTT CG
	locus_2174_qPCR_Forward	Reticuline oxidase like protein	CTC GAT GCC CAA GTC GTT GAT G
	locus_2174_qPCR_Reverse	Reticuline oxidase like protein	ATC CCA AAA TAA CAC CAA AAC TAG CAC C
	Rps9_qCR_Forward	40S ribosomal protein 9	TTG AGC CGT ATC AGA AAT GC
	Rps9_qCR_Reverse	40S ribosomal protein 9	CCC TCA TCA AGC AGA CCA TA
	EIF4a_qPCR_Forward	Elongation initiation factor 4A	TCA GGA GGC TCT TCC TGG TGA
	EIF4a_qPCR_Reverse	Elongation initiation factor 4A	AGC TCC CTT GGC AGG GTC AT
	RPP0C_qPCR_Forward	60S acidic ribosomal protein-0C	CAAGGTTGGAGCCCCTGCTCGTGTT
	RPP0C_qPCR_Reverse	60S acidic ribosomal protein-0C	TCTTAGTTGGAATGTTCAGCACCTG

7.8.2.6 Confocal microscopy of VIGS leaf tissue

7.8.2.6.1 DAPI staining

C. roseus leaves were vacuum-infiltrated and fixed in a solution of ethanol/acetic acid (3:1 v/v) for 1-2 h and then washed in phosphate buffered saline (PBS) pH 7.0.

Tissue was cut into small pieces (approx 3 mm x 3 mm) that included both blackened regions and normal looking tissue. Material was incubated in a 1 µg/ml solution of DAPI (4', 6-diamidino-2-phenyl-indole) in dH_2O for 1h, then washed in dH_2O and mounted in Vectashield (Vector laboratories) and imaged on a Zeiss LSM780 confocal microscope using a x40 oil immersion lens. For visualisation of DAPI staining, excitation was at 405 nm and detected at 410-585 nm. Z-stacks with 1 µm spacing were used to generate Maximum projection images using ImageJ software (http://rsb.info.nih.gov/ij/). This work was performed in collaboration with Alison Pendle (John Innes Centre).

7.8.2.6.2 Propidium iodide staining

Fresh unfixed tissue was used for propidium iodide staining. Tissue pieces were mounted in a 10 µg/ml solution of propidium iodide in dH_2O and viewed on a Zeiss LSM780 confocal microscope using a x40 oil immersion lens.

For visualisation of propidium iodide, excitation was at 488 nm and the signal detected at wavelengths of 535-660 nm. Z-stacks with 1 µm spacing were used to generate maximum projection images using ImageJ software (http://rsb.info.nih.gov/ij/). This work was performed in collaboration with Alison Pendle (John Innes Centre).

7.8.2.6.3 Trypan blue staining and visualisation by UV

C. roseus leaf tissue was placed in a clearing solution consisting of (300 ml 95% ethanol, 125 ml 90% lactic acid, 800 g chloral hydrate in 1 l chloroform) until the leaves were cleared of chlorophyll, and stored in 2.5 g/ml chloral hydrate. The leaves were stained in a solution of 0.1% trypan blue and lactoglycerol (lactic acid:glycerol:H_2O 1:1:1) for 24 hours, and subsequently destained in solution of 2.5 g/ml chloral hydrate. The stained *C. roseus* leaves were analysed by an AF6000 stereo brightfield microscope using a 1x/0.02 objective lens.

Photos of trypan blue stained leaves under UV light were taken by Andrew Davies (JIC photography department).

7.8.2.7 Desorption Electrospray Ionisation Mass Spectrometry Imaging (DESI-MS)

The freshly collected silenced leaves were imprinted either on Whatman grade 1 paper (VIGS-CrNPF2.9) or polytetrafluroethylene (PTFE) sheets (VIGS-EV). For imprinting on Whatman paper, the leaf was placed between two layers of paper and manually imprinted using a hydraulic press under 1 tonne pressure. For PTFE imprinting, the abaxial side of the leaf was laid on the PTFE sheet and the upper side of the leaf was then covered with filter paper. The leaf imprints were mounted onto a microscope glass slide using double-sided tape.

DESI experiments were performed using a Synapt G2 HDMS mass spectrometer (Waters) operated in positive mode with a scan time of 1.5 s in the mass range of 50-1200 *m/z*. Capillary voltage was 1.5 kV, cone voltage 40 V, source temperature 150 ^{o}C. DESI-MSI experiments were carried out using a Prosolia Inc. DESI 2D source. The leaf imprints were scanned in horizontal rows separated by 200 µm vertical steps until the entire tissue sample had been analysed. The

lines were scanned at a constant velocity of 200 µm/s, and mass spectra were recorded over the *m/z* 50-1200 range.

The spray voltage was set at 1.5 kV. Nitrogen gas (100 psi) was used for nebulisation. The spray solvent, 0.11% formic acid in acetonitrile/water (70:30), was sprayed at a constant flow rate of 1.5 µl/min. Mass spectra were recorded for a few seconds during electrospray contact with the leaf surface.

Non-commercial software designed by Waters was used to convert the MassLynx mass spectra files (.raw) into a format compatible with HDI, which was used to process the mass spectral data to generate two dimensional spatially accurate ion images. This work was performed in collaboration with Lorenzo Caputi and Gerhard Saalbach (John Innes Centre)

7.8.2.8 Chemical derivitisation of VIGS samples

7.8.2.8.1 Treatment with $NaBH_4$

The chemical reduction of methanolic extracts of VIGS-EV *C. roseus* leaf tissue was achieved by incubating a 600 µl methanol VIGS-EV extract with 1 mg $NaBH_4$. 20 µl of this sample was mixed with 30 µl MeOH and 50 µl H_2O, and analysed directly by LC-MS.

7.8.2.8.2 Treatment with potassium carbonate

Deacetylation of compounds in VIGS-EV and VIGS-CrRO leaf tissue was achieved by incubating a 600 µl methanol extract of each leaf tissue in K_2CO_3 (250 mM). This solution was incubated for 1 h at room temperature and analysed by LC-MS.

7.8.2.9 In vivo C. roseus leaf feeding experiments

7.8.2.9.1 Secologanin feeding experiments

2-3 cm long *C. roseus* leaves of wild type *C. roseus* leaf tissue were cut at the petiole and placed in 100 µl of either 5 mM secologanin or water. The leaves were then fed daily 100 µl of the same compound over the course of 4-5 days. After feeding, the leaves were blotted to remove excess liquid, and weighed. Each leaf was extracted in 200-400 µl methanol for 2 h at 65 ^{o}C. 100 µl of the methanolic extract was removed, and centrifuged at 14000 rpm for 40 min. The supernatant was mixed 1:1 with H_2O and analysed by LC-MS.

7.8.2.9.2 Isotopic Labelling Experiments

2-3 cm long *C. roseus* leaves of wild type *C. roseus* leaf tissue, VIGS-EV leaf tissue, or VIGS-CrRO leaf tissue were cut at the petiole and placed in 100 µl of either 5 mM d_4-tryptamine, 5 mM d_5-tryptophan, or water. Unless otherwise stated the leaves were then fed daily 100 µl of the same compound over the course of 4-5 days. After feeding, the leaves were blotted to remove excess liquid, and weighed. Each leaf was extracted in 200-400 µl methanol for 2 h at 65 °C. 100 µl of the methanolic extract was removed, and centrifuged at 14000 rpm for 40 min. The supernatant was mixed 1:1 with H_2O and analysed by LC-MS.

7.9 Methods specific for heterologous expression

7.9.1 USER cloning into pCR8-GW

7.9.1.1 pCR8-GW

The Multisite Gateway® technology allows for movement of genes of interest in an entry vector into multiple vector systems for functional characterisation of proteins of interest. It is a universal cloning method that utilises the site-specific recombination properties of bacteriophage lambda (Walhout et al., 1999). pCR8-GW is a Gateway® compatible entry vector that allows for the recombination based transfer of a gene of interest in the entry vector into a wide variety of downstream destination vectors through use of the LR clonase reaction. The entry clone contains an *att*L1 and *att*L2 sites in the vector between which the gene of interest is cloned. This allows for recombination mediated transfer into destination vectors that contain two *att*R sites which contain a *ccd*B gene between them for negative selection, with recombination between the *att*L and *att*R sites resulting in the replacement of the *ccd*B gene with the gene of interest in the destination vector. Full length *C. roseus* genes were therefore cloned into pCR8-GW, sequenced and then this entry vector was subsequently used for downstream cloning into plant, yeast and *Xenopus* oocyte specific destination vectors.

A USER-compatible version of the pCR8-GW vector, pCR8-GWu, was generated by Fernando Geu-Flores, which as described previously, is a ligation independent cloning technique that is facilitated by the generation of custom designed overhanging sequences that, when treated with USER enzyme, allow for the generation of a stable hybridisation product that can be

transformed into *E. coli* without prior ligation. pCR8-GWu was linearised through treatment with the enzymes PacI and NtBbvcI overnight at 37 °C, and gel purified on a 0.7% agarose gel.

7.9.1.2 Cloning of full length gene sequences into pCR8-GW

The full length genes of interest in this thesis were PCR amplified from *C. roseus* cDNA using gene specific primers and the PCR product was gel purified on a 1% agarose gel. The PCR products, linearised pCR8-GWu vector, and USER enzyme were incubated together in a reaction volume of 10 µl at 37 °C for 40 min, and subsequently kept on ice for 15 min. This reaction mixture was used to transform 40 µl TOP10 *E.coli* by heat shock without prior ligation and grown on spectinomycin. Three colonies from each plate were grown up in 5 ml overnight cultures. Plasmid DNA was extracted by a Qiagen Miniprep kit and tested for the insert by PCR using gene specific primers and positive plasmids were sent for sequencing.

7.9.1.3 Cloning of C-terminal GFP fusions into pCR8-GW

C-terminal GFP fusions of the full length genes of interest were generated by USER fusion cloning, and contained a GSG linker or an extended GSGMGGSG linker between the C-terminus of the protein and the N-terminus of the GFP. For the C-terminal GFP fusions the reverse primers for amplification of the gene of interest contained complementary overhanging sequences in the region of the GSG or GSGMGGSG linker. This is complementary to a region in the forward primer used for the amplification of GFP from a pCR8GWu-GFP vector. This meant that the GFP and extended GFP PCR product utilised for USER cloning of the C-terminal GFP fusions was compatible with all the genes of interest to be cloned.

For the formation of the C-terminal GFP fusions: the PCR product of the gene of interest, either GFP or GFP-extended PCR product, linearised pCR8-GWu vector, and USER enzyme were incubated together in a reaction volume of 10 µl at 37 °C for 40 min, and subsequently kept on ice for 15 min. This reaction mixture was used to transform 40 µl TOP10 *E. coli* by heat shock without prior ligation. The *E. coli* was recovered and grown in 200 µl SOC media at 37 °C for 40 min, and this was then spread onto spectinomycin selective agar plates and incubated at 37 °C overnight. Three colonies from each plate were grown up in 5 ml overnight cultures. Plasmid DNA was extracted by a Qiagen Miniprep kit and tested for the insert by PCR using gene specific primers and positive plasmids were sent for sequencing.

Table 24 Primers for the generation of pCR8-GW vectors

Vector	PCR product		Primers	Sequence (5'-3')
GFP for C-terminal fusion	GFP_Cterm fusion	pCR8GW	Fwrd	ATCTGGAAU GACTAGCAAAGGAGAAGAACTTTTCACTG
			Rev	GGTTTAAU TTA TTT GTA TAG TTC ATC CAT GCC ATG TGT AAT C
GFPextended for C-terminal fusion	GFP_Cterm fusion_extended	pCR8GW	Fwrd	ATCTGGAAU GGGAGGATCTGGAACTAGCAAGGAGAAGAACTTTCACTG
			Rev	GGTTTAAU TTA TTT GTA TAG TTC ATC CAT GCC ATG TGT AAT C
CrMATE1952_full_length_native	MATE	pCR8GW	Fwrd	GGCCTTAAU ATG GGT TCC AAA CAA AAC TAT GAA ATA AAC CA
			Rev	GGTTTAAU TTA TTC ATT GGA CAA AGA TTT TGG CTT GT
CrMATE10740_full_length_native	MATE	pCR8GW	Fwrd	GGCCTTAAU ATG GAT AGT AGT AAT AAG AAT CCG TTG CTA GTG AAA
			Rev	GGTTTAAU TCA AGC TTC AGC TAT TTG TTG TTC ATT TTC TT
CrNPF2.9_full_length_native	NPF transporter	pCR8GW	Fwrd	GGCCTTAAU ATG GGA GAC ACC GAA GCA C
			Rev	GGTTTAAU TTA CAT ATA TAT TTT ATT ATC ATT TGT TTC ATC TTC ATC CTC CAT
CrMATE1952_full_length_GFPfusion	MATE	pCR8GW	Fwrd	GGCCTTAAU ATG GGT TCC AAA CAA AAC TAT GAA ATA AAC CA
			Rev	ATTCCAGAUCC TTC ATT GGA CAA AGA TTT TGG CTT GT
CrMATE10740_full_length_GFPfusion	MATE	pCR8GW	Fwrd	GGCCTTAAU ATG GAT AGT AGT AAT AAG AAT CCG TTG CTA GTG AAA
			Rev	AAGGATCU GGA ATGGATAGTA GTAATAAGAA TCCGTTGCTA GTGAAA
CrNPF2.9_full_length_GFPfusion	NPF transporter	pCR8GW	Fwrd	GGCCTTAAU ATG GGA GAC ACC GAA GCA C
			Rev	ATTCCAGAU CC CAT ATA TAT TTT ATT ATC ATT TGT TTC ATC TTC ATC CTC CAT GAC
CrNPF2.9_K>R_GFPfusion	NPF transporter	pCR8GW	Fwrd	GGCCTTAAU ATG GGA GAC ACC GAA GCA C
			Rev	ATTCCAGAU CC CAT ATA TAT TCT ATT ATC ATT TGT TTC ATC TTC ATC CTC CAT GAC
CrNPF2.9_GFPfusion_extended	NPF transporter	pCR8GW	Fwrd	GGCCTTAAU ATG GGA GAC ACC GAA GCA C
			Rev	ATTCCAGAU CC CAT ATA TAT TTT ATT ATC ATT TGT TTC ATC TTC ATC CTC CAT GAC
CrNPF2.9_K>R_GFPfusion_extended	NPF transporter	pCR8GW	Fwrd	GGCCTTAAU ATG GGA GAC ACC GAA GCA C
			Rev	ATTCCAGAU CC CAT ATA TAT TCT ATT ATC ATT TGT TTC ATC TTC ATC CTC CAT GAC
1443_CrPUP_full_length_GFPfusion	Purine permease	pCR8GW	Fwrd	GGCCTTAAU ATG GAA GAT CAA GTT AGG AGC AGG AG
			Rev	ATTCCAGAU CC CGG GAC ATT TGA AAT TAC TTC AAT TTC CTT TTT G
CrRO_full_length_native	Reticuline oxidase like protein	pCR8GW	Fwrd	GGCCTTAAU ATGATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTG
			Rev	GGTTTAAU TCAAAGTTCGACTTGTAAATGGAGAGGG
AtTPK1_full_length_GFPfusion	Arabidopsis TPK1	pCR8GW	Fwrd	GGCCTTAAU ATGTCGAGTGATGCAGCTCGTAC
			Rev	ATTCCAGAU CC CCT TTG AAT CTG AGA CGT GGT CTG A

Table 25 pCR8-GW vectors generated in this study

Gene name	Type	Vector
Green Fluorescent Protein	GFP	pCR8GW
CrMATE1952_full_length_native	MATE	pCR8GW
CrMATE10740_full_length_native	MATE	pCR8GW
CrNPF2.9_full_length_native	Nitrate/Peptide family transporter	pCR8GW
CrMATE1952_full_length_GFPfusion	MATE	pCR8GW
CrMATE10740_full_length_GFPfusion	MATE	pCR8GW
CrNPF2.9_full_length_GFPfusion	Nitrate/Peptide family transporter	pCR8GW
CrNPF2.9_K>R_GFPfusion	Nitrate/Peptide family transporter	pCR8GW
CrNPF2.9_GFPfusion_extended	Nitrate/Peptide family transporter	pCR8GW
CrNPF2.9_K>R_GFPfusion_extended	Nitrate/Peptide family transporter	pCR8GW
CrPUP_1443_full_length_GFPfusion	Purine permease	pCR8GW
CrRO_full_length	Reticuline oxidase like	pCR8GW
AtTPK1_GFPfusion	Potassium channel	pCR8GW

7.9.2 Heterologous Expression in S. cerevisiae

7.9.2.1 Cloning into pYES-DEST52

As mentioned pCR8-GW is an entry vector for Gateway® compatible technology and allows for the cloning into multiple downstream destination vectors through the use of LR clonase mediated recombination. Briefly, each entry clone that contains *att*L sites flanking the gene of interest, is incubated with the destination vector containing *att*R sites and the LR Clonase™ enzyme mix, with the resultant recombination reaction occurring resulting in the gene of interest localised to the destination vector, and this reaction is used for the transformation of TOP10 *E. coli* by heat shock and selection on the antibiotic marker specific for the destination vector.

pYES-DEST52 is a Gateway® compatible destination vector that allows for the expression of the protein of interest in *S. cerevisiae* under control of a galactose inducible promoter.

For the pCR8-GW entry vectors generated, 50-100 ng of the entry vector was incubated with 150 ng of the pYESDEST52 destination vector, 1 µl LR-Clonase™ enzyme mix, and TE buffer pH 8.0 to a final reaction volume of 5 µl at 25 ^{o}C for 1 h. To terminate the reaction 0.25 µl Proteinase K was added to each reaction and incubated at 37 ^{o}C for 10 min. 2.5 µl of each reaction was used to transform 50 µl chemically competent TOP10 *E. coli* by heat shock, and positive transformants were selected on carbenicillin agar plates. Transformants were grown in 5 ml culture overnight at 37 ^{o}C in LB + carbenicllin, plasmid extracted by Miniprep, and the plasmid tested for insertion of the gene of interest by PCR. pYES-DEST52 plasmids testing positive for the gene of interest were used for transformation of *S. cerevisiae*.

Table 26 Overview of pYES-DEST52 vectors generated in this study

Name	Gene of interest	Vector
Green Fluorescent Protein	GFP	pYesDest52
CrMATE1952_full_length_native	MATE	pYesDest52
CrMATE10740_full_length_native	MATE	pYesDest52
CrNPF2.9_full_length_native	NPF transporter	pYesDest52
CrMATE1952_full_length_GFPfusion	MATE	pYesDest52
CrMATE10740_full_length_GFPfusion	MATE	pYesDest52
CrNPF2.9_full_length_GFPfusion	NPF transporter	pYesDest52
CrNPF2.9_K>R_GFPfusion	NPF transporter	pYesDest52
CrNPF2.9_GFPfusion_extended	NPF transporter	pYesDest52
CrNPF2.9_K>R_GFPfusion_extended	NPF transporter	pYesDest52
1443_CrPUP_full_length_GFPfusion	Purine permease	pYesDest52
CrRO_full_length_native	Reticuline oxidase like protein	pYesDest52

7.9.2.2 Overexpression of proteins in S. cerevisiae

The pYES-DEST52 vector is an inducible expression system for protein expression in yeast under control of the *GAL1* promoter. For most *S. cerevisiae* laboratory strains, the presence of glucose represses transcription from the plasmid, whilst adding galactose induces transcription.

The glycerol stock of the yeast strain to test protein expression was used to inoculate a starter culture of 15 ml SC-uracil medium containing 2% glucose as a carbon source. This was grown overnight at 30 ^{o}C and used to inoculate 100 ml SC-uracil containing 2% glucose, which was then grown to an OD_{600} of between 1-2. The culture was centrifuged at 4000 x g and resuspended in induction medium; SC-uracil containing 2% galactose and 2% raffinose. The culture was then grown for between 4-10 h at 30 ^{o}C.

7.9.2.3 Confocal microscopy of S. cerevisiae expressing proteins

The induced culture was used directly for analysis of protein expression under confocal microscopy. 10 µl of culture was placed on a microscope slide under a cover slip, and visualised by a Leica SP5 confocal microscope using a HCX PL APO CS 63.0x1.20 water immersion lens. For visualisation of the GFP, the sample was excited at 488 nm and the emission detected at 500-580 nm.

7.9.2.4 Preparation of yeast membrane vesicles

For preparation of yeast membrane vesicles, the induced culture was centrifuged at 5000 x g for 5 min at room temperature, washed with 40 ml dH_2O, and pelleted at 5000 x g for 5 min at room temperature. The yeast cell pellet was resuspended in softening medium (0.1 M Tris-HCl pH 9.4, 50 mM β-mercaptoethanol and 0. 1 M glucose), and the mixture incubated for 30 min at room temperature, before centrifugation at 5000 x g. The yeast pellet was resuspended in 100 ml zymolyase medium (50 mM Tris-MES pH 7.6, 0.9 M sorbitol, 0. 1 M glucose, 5 mM DTT and 10 U/ml). This solution was incubated with shaking at 30 ^{o}C for 3 h and the resulting spheroblasts collected at 5000 x g for 10 min and washed successively with 1 M sorbitol. For preparation of yeast membranes, the spheroblasts were resuspended in 20 ml lysis buffer (50 mM Tris-MES pH 7.6, 1.1 M glycerol, 5 mM EDTA, 1 mM DTT, 0.2% (w/v) BSA, 1.5% (w/v) polyvinylpyrrolidone, and cOmplete™ protease cocktail inhibitor tablets). This suspension was homogenised with a hand held homogeniser, and centrifuged at 4000 x g, and the supernatant fraction was collected. The pellet was successively washed twice with lysis buffer and the supernatant upon centrifugation was pooled. This supernatant was centrifuged at 120,000 x g for 1 h, and the resulting membranes were resuspended in resuspension buffer (5 mM Tris-MES pH 7.6 or 5 mM MOPS pH 7.6, 0.3 M sorbitol, 1 mM DTT, 0.1 M KCl and cOmplete™ protease cocktail inhibitor tablets). The total protein concentration of the yeast microsomal fraction was assessed by a BCA assay, and adjusted to 10 µg/ml. The vesicles were run by SDS-PAGE and a western blot performed using a commercially available mouse-monoclonal anti-GFP antibody (Abcam, cat no 9F9.F9 ab1218) as the primary anti-body and rabbit mono-clonal anti-mouse IgG and IgM monoclonal antibody.

7.9.2.5 Feeding Studies with S. cerevisiae

To assess whether the enzymes expressed in *S. cerevisiae* were able to utilise compounds accumulating in VIGS tissue as substrates, the enzymes were expressed in *S. cerevisiae* as described in the section above and incubated with the substrate. For each yeast liquid culture expressing the gene of interest, 1 ml of VIGS methanol extract was dried under vacuum using a Savant SC210A speed vac concentrator coupled to a Savant RVT4104 refrigerated vapour trap. 1 ml of yeast culture was used to resuspend the lyophilised material, and the yeast culture containing the VIGS compounds was incubated for 10 h at 30 ^{o}C. After 10 h the culture was centrifuged at 10,000 x g. The supernatant was removed and mixed 1:1 with MeOH and analysed

directly by LC-MS. Additionally, the yeast pellet was extracted in 100 µl MeOH, centrifuged at 10,000 x g, and the MeOH extraction mixed 1:1 with H_2O for analysis by LC-MS.

7.9.3 Heterologous Expression in N. benthamiana

7.9.3.1 Cloning into pEAQ-HT vector

pEAQ-HT DEST1 is a Gateway® compatible destination vector that allows for the expression of the protein of interest in plants that are conducive to *Agrobacterium* infiltration (Sainsbury et al., 2009). In particular we have been using this expression system for the overproduction of proteins in *N. benthamiana.*

For the pCR8-GW entry vectors 50-100 ng of the entry vector was incubated with 150 ng of the pEAQ-HT DEST1 destination vector, 1 µl LR-Clonase™ enzyme mix, and TE buffer pH 8.0 to a final reaction volume of 5 µl at 25 ^{o}C for 1 h. To terminate the reaction 0.25 µl Proteinase K was added to each reaction and incubated at 37 ^{o}C for 10 min. 2.5 µl of each reaction was used to transform 50 µl chemically competent TOP10 *E. coli* by heat shock, and positive transformants were selected on kanamycin agar plates. Transformants were grown in 5 ml culture overnight at 37 ^{o}C in LB + kanamycin, plasmid extracted by miniprep, and the plasmid tested for insertion of the gene of interest by PCR. pEAQ-HT DEST1 plasmids testing positive for the gene of interest were used for transformation of *Agrobacterium tumefaciens* strain LBA4404.

The liquid culture of colonies of the *Agrobacterium tumefaciens* strain LBA4404 testing positive for the insert by PCR were mixed 1:1 with 25% glycerol and stored at -80 ^{o}C for future use.

Table 27 Overview of pEAQ-HT DEST1 vectors generated in this study

Name	Gene of interest	Vector
Green Fluorescent Protein	GFP	pEAQ-DEST1
CrMATE1952_full_length_native	MATE	pEAQ-DEST1
CrMATE10740_full_length_native	MATE	pEAQ-DEST1
CrNPF2.9_full_length_native	NPF transporter	pEAQ-DEST1
CrMATE1952_full_length_GFPfusion	MATE	pEAQ-DEST1
CrMATE10740_full_length_GFPfusion	MATE	pEAQ-DEST1
CrNPF2.9_full_length_GFPfusion	NPF transporter	pEAQ-DEST1
CrRO_full_length_native	Reticuline oxidase like protein	pEAQ-DEST1
***Arabidopsis* AtTPK1**	*Arabidopsis* TPK1	pEAQ-DEST1

7.9.3.2 Agroinfiltration of N. benthamiana leaf tissue

Eight week old *N. benthamiana* plants were selected for Agroinfiltration. *Agrobacterium tumefaciens* strain LBA4404 containing pEAQ-HT DEST1 expression vectors containing the gene of interest for overexpression were grown overnight in a 10 ml starter culture containing LB supplemented with rifampicin, streptomycin and kanamycin at 28 ^{o}C. The cells were centrifuged at 4000 x g, and resuspended in *Agrobacterium* infiltration buffer (10 mM MES, 10 mM $MgCl_2$, 100 µM acetosyringone, pH 5.5) to an optical density of 0.2-0.4 at 600 nm, and incubated for 2 h at room temperature.

This solution was used for the infiltration. Mid-sized *N. benthamiana* leaves were infiltrated by syringe infiltration. Pressure was applied to the leaf in order to force the agrobacterium through the stomatal pores of the underside of the leaf. ~5 ml of infiltration solution was applied to each leaf until the entire leaf was infiltrated. For each construct two plants were infiltrated utilising three leaves per plant. *Agrobacterium* containing a pEAQ-HT DEST1 construct for the expression of GFP was used as a positive control. The *N. benthamiana* plants were left for between 4 and 7 days for protein expression before harvesting.

7.9.3.3 Visualisation of protein expression in N. benthamiana by confocal microscopy

Leaves expressing the protein of interest were cut into discs 1 cm in diameter, placed on a microscope slide and visualised by a Leica SP5 confocal microscope using a HCX PL APO CS 63.0x1.20 water immersion lens. For visualisation of the GFP, the sample was excited at 488 nm and the emission detected at 500-580 nm.

7.9.3.4 Generation of protoplasts

The tissue of three leaves was used per construct for the generation of protoplasts. Agroinfiltrated *N. benthamiana* leaf tissue was harvested after 4-7 days post agroinfiltration and the mid-vein, large veins and injection site removed from the tissue with a razor blade. On the remaining tissue the underside of the leaf was gently scored with a razor blade at ~2 mm intervals and the leaf tissue incubated in protoplasting solution (0.4 M mannitol, 10 mM $CaCl_2$, 20 mM KCl, 0.1% BSA, 20 mM MES, 0.25% cellulose R10 (Sigma; C1794 from *Trichoderma viride*) and 0.25% macerozyme R10 (Sigma; P2401 pectinase from *Rhizopus spp*) pH5.7) for 4 h with gentle shaking in the dark. The resulting solution was filtered through muslin, washed gently with the protoplasting buffer, and centrifuged at 50 x g to retain the protoplasts. The protoplasts were visualised under an AF6000 stereomicroscope.

7.9.3.5 Generation of vacuoles from protoplasts

Protoplasts were lysed by osmotic lysis by incubating the protoplasts in lysis buffer (0.2 M mannitol, 10% (w/v) Ficoll 400, 15 mM EDTA, 10 mM MOPS pH 8.0, 150 µg/ml BSA and 1 mM DTT). This mixture was stained with neutral red solution in order to visualise the vacuoles. A 200 x dilution of 0.1% neutral red solution (0.1 g neutral red, 200 µl 1% acetic acid, 50 µl chloroform in a total volume of 100 ml water) was added to the lysed protoplast solution. The vacuoles were subsequently purified by ultracentrifugation at 72,000 x g for 1 h using a stepwise Ficoll gradient. The gradient consisted of a lower phase of one volume lysed protoplast solution, a middle phase of two volumes 1:1 mixture of lysis buffer and an upper phase of vacuole buffer. The vacuole buffer contained (0.4 M betaine, 20 mM HEPES-KOH, pH 7.5, 20 mM KCl, 15 mg/ml BSA and 1 mM DTT). Vacuoles were harvested at the interface between the upper and middle phases, and were visualised by an AF6000 stereomicroscope.

7.9.3.6 Subcellular Localisation in C. roseus cell suspension culture (This work was performed by Vincent Courdavault)

The pEAQ-DEST1 vectors containing CrMATE1952, CrMATE10740, CrNPF2.9, and AtTPK1 were used as a template to subclone the genes into the pSCA-cassette vector as either a YFP fusion (CrMATE1952, CrMATE10740, CrNPF2.9) or a CFP fusion AtTPK1. Transient transformation of *C. roseus* cells by particle bombardment and fluorescence imaging were performed following the procedures previously described (Guirimand et al., 2009; Guirimand et al., 2010). *C. roseus* cells were bombarded with DNA-coated gold particles (1 µm) and 1,100 psi rupture disc at a stopping-

screen-to-target distance of 6 cm, using the Bio-Rad PDS1000/He system. Cells were cultivated for 16 h to 38 h before being harvested and observed. The subcellular localisation was determined using an Olympus BX-51 epifluorescence microscope equipped with an Olympus DP-71 digital camera and a combination of YFP and CFP filters.

7.9.4 Heterologous Expression in Xenopus oocytes

7.9.4.1 Cloning of CrNPF2.9 transporter into Xenopus laevis oocyte expression vector

7.9.4.1.1 Cloning into the pT7TS vector

pT7TS is a Gateway® compatible destination vector, and was donated from the lab of Tony Miller (John Innes Centre). The pCR8-GW entry vector containing the full length CrNPF2.9 gene, 50-100 ng of the entry vector was incubated with 150 ng of the pT7TS destination vector, 1 µl LR-Clonase™ enzyme mix, and TE buffer pH 8.0 to a final reaction volume of 5 µl at 25 ^{o}C for 1 h. To terminate the reaction 0.25 µl Proteinase K was added to each reaction and incubated at 37 ^{o}C for 10 min. 2.5 µl of each reaction was used to transform 50µl chemically competent TOP10 *E. coli* by heat shock, and positive transformants were selected on carbenicillin agar plates. Transformants were grown in 5 ml culture overnight at 37 ^{o}C in LB+carbenicllin, plasmid extracted by miniprep, and the plasmid tested for insertion of the gene of interest by PCR. pT7TS plasmids testing positive for the gene of interest were used for *in vitro* transcription for the production of cRNA to be injected into the *Xenopus laevis* oocyte.

7.9.4.1.2 Cloning into pLIFE0016 vector by USER cloning

Full length CrNPF2.9 was amplified from *C. roseus* cDNA using the primers in Table 28. The pLIFE0016 vector is a USER compatible vector that can be used for the generation of cRNA compatible for microinjection into *Xenopus laevis* oocytes, and was donated by Hussam Hassan Nour-Eldin (University of Copenhagen). Linearised pLIFE0016 and the CrNPF2.9 PCR product was incubated with USER in a reaction volume of 10 µl at 37 ^{o}C for 40 min, and subsequently kept on ice for 15 min. This reaction mixture was used to transform 40 µl TOP10 *E.coli* by heat shock without prior ligation. The *E. coli* was recovered and grown in 200 µl SOC media at 37 ^{o}C for 40 min, and this was then spread onto carbenicillin selective agar plates and incubated at 37 ^{o}C overnight. Plasmid from positive colonies was used for the generation of linear cRNA.

Table 28 Primers for CrNPF2.9 cloning into pLIFE0016

Primer	Plasmid	Sequence (5'-3')
CrNPF2.9 fwrd	pLIFE0016	GGCTTAAU ATG GGA GAC ACC GAA GCA C
CrNPF2.9 rev	pLIFE0016	GGTTTAAU TTA CAT ATA TAT TTT ATT ATC ATT TGT TTC ATC TTC ATC CTC CAT

Table 29 *Xenopus* oocyte expression vectors generated in this thesis

Gene	Plasmid	Enzyme for linearisation
CrNPF2.9	pT7TS	SalI-HF
CrNPF2.9	pLIFE0016	NdeI or NotI

7.9.4.2 Generation of cRNA for microinjection

7.9.4.2.1 Linearisation of pLIFE0016 and pT7TS plasmids

Both the pT7TS or pLIFE0016 vectors were used for generation of linear RNA for microinjection into *Xenopus laevis* oocytes. For the formation of linear RNA that is compatible with over expression in *Xenopus* oocytes, the mMESSAGE mMACHINE® T7 Ultra kit (Ambion catalogue no:1345) was used. This kit allows for the *in vitro* synthesis of capped RNA. For the efficient formation of RNA for expression in *Xenopus* oocytes using this method the vector must have a T7 promoter, which is followed by a 5' UTR of *Xenopus* β-Globin gene, the gene of interest, and then the 3' UTR of the *Xenopus* β-Globin gene and a poly(A) tail.

Generation of linear RNA using the T7 system requires a linear DNA template containing a T7 promoter that does not contain 3' overhanging ends. As such the pLIFE0016 and pT7TS vectors need to be linearised prior to subjection to *in vitro* transcription.

For the pT7TS vector, 6000 µg of plasmid DNA was linearised using SalI-HF, whilst for the pLIFE0016 vector 6000 µg of plasmid DNA was linearised using NdeI or NotI, with the appropriate enzyme buffer, in a total volume of 20 µl. The restriction digest was assessed on a 1% agarose gel. These restriction enzymes only cut once on the vector, do not cut internally on the cloned gene, and do not cut between the T7 promoter and poly(A) tail, thereby generating a linear template, compatible for linear RNA generation.

The linearised plasmid DNA was treated with 0.4 µl Proteinase K (20 mg/ml) and 20 µl SDS and incubated for 30 min at 50 ^{o}C. RNase free H_2O was added to a total volume of 100 µl, to which

100 µl of phenol:chloroform (1:1) was added. This was vortexed and centrifuged at 10,000 x g for 3 min at room temperature, and the aqueous phase retained. To this aqueous phase 3 M sodium acetate (pH 5.3) at 1/10 (v/v) was added together with 2 volumes of 100% ethanol. This was incubated overnight at -20 ^{o}C. The solution was mixed well and centrifuged at 10,000 x g for 30 min at 4 ^{o}C. The pellet was retained and washed twice with 200 µl 70% ethanol, dried and resuspended in RNase free H_2O to a final concentration of ~0.5 µg/µl.

7.9.4.2.2 Generation of cRNA

1 µg of the linear plasmid was used for the *in vitro* transcription of cRNA using the mMESSAGE mMACHINE® T7 Ultra kit (Ambion catalogue no:1345) kit in accordance with the manufacturer's instructions. Briefly, the 1 µg linear plasmid, 2 µl T7 enzyme mix, 10 µl T7 nucleotides and ARCA (Anti-reverse cap analogue, for efficient RNA capping) and H_2O were combined in a reaction volume of 20 µl, and incubated for 2 hr at 37 ^{o}C. For removal of the template DNA, 1 µl of DNaseI was added and incubated at 37 ^{o}C for 15 min.

The RNA formed was purified again by phenol:chloroform extraction. 115 µl RNase free H_2O, 15 µl 5M sodium acetate and 150 µl phenol:chloroform, was added to the reaction mixture. The solution was vortexed and then centrifuged for 3 min at 10,000 x g at room temperature, and the aqueous phase retained. This was repeated with 50 µl RNase free H_2O and pooled. To the 200 µl retained aqueous phase, 200 µl chloroform was added, which was again vortexed, centrifuged for 3 min at 10,000 x g and the aqueous phase retained. 200 µl isopropanol was added to the aqueous phase and incubated at -20 ^{o}C overnight for precipitation of the RNA. The RNA was centrifuged at 10,000 x g at 4 ^{o}C for 30 min, the pellet washed with 70% ethanol, centrifuged at 10,000 x g and dried. The RNA was resuspended in RNase free H_2O to a concentration of 2 µg/µl.

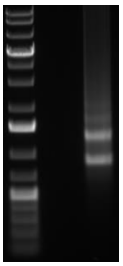


Figure 7-3 Representative cRNA generated from in vitro transcription using the linearised pT7TS-CrNPF2.9 vector as template.

7.9.4.3 Preparation of Xenopus laevis oocytes

Xenopus laevis oocytes were kindly donated from the Gurdon Institute, Cambridge, and stored in modified barths saline (MBS: 88 mM NaCl, 1 mM KCl, 0.82 mM $MgSO_4$, 10 mM Hepes, 0.33 mM $Ca(NO_3)_2$, 0.41 mM $CaCl_2$, pH 7.5). The *Xenopus laevis* oocytes were defolliculated by collagenase treatment (5 mg/ml collagenase) in calcium-free MBS, and the defolliculated oocytes were stored in MBS solution at 18 ^oC until further use.

7.9.4.3.1 Microinjection of linear RNA into Xenopus laevis oocytes

Drummond microcaps (cat no 1-000-0500) were pulled using a Narishige PE-2 glass microelectrode puller to produce glass needles for microinjection into the *Xenopus laevis* oocyte. Each needle was normalised to inject 50 nl of liquid into each oocyte. Injection into the oocyte was performed using a PLI-100 microinjector (Digitimer Ltd) utilising a micromanipulator (Prior Scientific). For each assay ~ 20 oocytes were microinjected with 50 nl 1 µg/µl cRNA and ~20 oocytes microinjected with 50 nl H_2O as a negative control. This same method was used for the microinjection of secologanin into oocytes.

7.9.4.3.2 Xenopus laevis oocyte uptake assay

Xenopus oocytes were incubated in pH 7.5 MBS buffer for three days to induce protein expression, with the buffer routinely changed twice daily. For the assay, the oocytes injected with H_2O or cRNA were incubated in 100 µM strictosidine in Kulori buffer (90 mM NaCl, 1 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 5 mM MES, pH 5) for 1h. The oocytes were washed four times in ice cold Kulori buffer to stop the assay, pooled into batches of five oocytes, flash frozen in liquid N_2 and stored at -80 ^oC.

Extraction of the metabolite content in the oocyte was performed by homogenising the oocytes in MeOH containing 40 µM caffeine as an internal standard, and extracted at 60 ^oC for 2 h. This suspension was centrifuged at 10,000 x g, the supernatant removed and evaporated to dryness. This dried material was resuspended in 80 µl 50% MeOH, centrifuged at 10,000 x g and the supernatant used directly for analysis by LC-MS.

7.9.5 Heterologous expression in E. coli and Sf9 Insect cells.

The pOPIN vector suite designed by the Oxford Protein Production Facility (OPPF) was utilised for the cloning of the reticuline oxidase CrRO gene from *C. roseus*. This suite of vectors is compatible for screening of protein expression in multiple hosts, with many of the vectors compatible with *E. coli*, Sf9 Insect cells and mammalian HEK293T cells, and therefore allows for parallel screening in multiple expression systems (Berrow et al., 2007).

7.9.5.1 Cloning into pOPIN vectors

The three pOPIN vectors used in this study were pOPINF,pOPINE and pOPING. pOPINF and pOPINE allow for the introduction of an N- or C-terminal histidine tag respectively, whilst pOPING allows for the introduction of an insect cell secretion signal at the N-terminus of the cloned gene, together with a C-terminal histidine tag. For linearisation of the pOPINF, pOPINE and pOPIN G vectors, pOPINF was NcoI and MscI, pOPINE was incubated with NcoI and PmeI, and pOPING was incubated with KpnI and PmeI, all in CutSmart buffer in accordance with manufacturer's guidelines, overnight at 37 ^{o}C. Each linearised plasmid was gel purified on a 0.7% (w/v) agarose gel using stored at -20 ^{o}C for future use.

The full length CrRO, together with designed N- and C-terminal truncated variants of this gene were PCR amplified from *C. roseus* cDNA and gel purified on a 1% agarose gel. The pOPIN suite of vectors is compatible with the In-Fusion enzyme. Each PCR product was incubated with its compatible linearised pOPIN vector together with the In-Fusion enzyme in a reaction volume of 10 µl for 15 min at 50 ^{o}C. 2 µl of this reaction mixture was used to transform 40 µl TOP10 *E. coli,* and transformants were selected on LB + carbenicillin agar plates. Three colonies were grown for plasmid extraction, with the resultant plasmid tested for insertion of the gene by PCR and positive colonies are confirmed by sequencing. The resultant plasmid was stored at -20 ^{o}C. A Table of the CrRO gene variants and the pOPIN vectors they were cloned into is presented in Table 9 and the primers used for cloning are presented in Table 30.

Table 30 Primers used for the cloning of CrRO into pOPIN vectors

Vector Name	Vector		Primer Sequence (5'-3')
CrRO_full_length_pOPINF	pOPINF	Fwrd	AAGTTCTGTTTCAGGGCCCG ATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTG
		Rev	ATGGTCTAGAAAGCTTTA AAGTTCGACTTGTAAATGGAGAGGGG
CrRO_full_length_pOPINE	pOPINE	Fwrd	AGGAGATATACCATG ATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTG
		Rev	GTGATGGTGATGTTT AAGTTCGACTTGTAAATGGAGAGGGG
CrRO_full_length_pOPING	pOPING	Fwrd	GCGTAGCTGAAACCGGC ATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTG
		Rev	GTGATGGTGATGTTT AAGTTCGACTTGTAAATGGAGAGGGG
CrRO_Ntermtruncated_pOPINF	pOPINF	Fwrd	AAG TTC TGT TTC AGG GCC CGT CAA TTC CTG AAG CTT TTC TCA ATT GTA TTT CC
		Rev	ATGGTCTAGAAAGCTTTA AAGTTCGACTTGTAAATGGAGAGGGG
CrRO_Ntermtruncated_pOPINE	pOPINE	Fwrd	AGGAGATATACCATG T CAA TTC CTG AAG CTT TTC TCA ATT GTA TTT CC
		Rev	GTGATGGTGATGTTT AAGTTCGACTTGTAAATGGAGAGGGG
CrRO_Ntermtruncated_pOPING	pOPING	Fwrd	GCGTAGCTGAAACCGGC T CAA TTC CTG AAG CTT TTC TCA ATT GTA TTT CC
		Rev	GTGATGGTGATGTTT AAGTTCGACTTGTAAATGGAGAGGGG
CrRO_Ctermtruncated_pOPINF	pOPINF	Fwrd	AAGTTCTGTTTCAGGGCCCG ATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTG
		Rev	ATGGTCTAGAAAGCTTTAATGGAGAGGGGGAATGCTTTGT
CrRO_Ctermtruncated_pOPINE	pOPINE	Fwrd	AGGAGATATACCATG ATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTG
		Rev	GTGATGGTGATGTTTATGGAGAGGGGGAATGCTTTGT
CrRO_Ctermtruncated_pOPING	pOPING	Fwrd	GCGTAGCTGAAACCGGC ATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTG
		Rev	GTGATGGTGATGTTTATGGAGAGGGGGAATGCTTTGT
CrRO_NandCtermtruncated_pOPINF	pOPINF	Fwrd	AAG TTC TGT TTC AGG GCC CGT CAA TTC CTG AAG CTT TTC TCA ATT GTA TTT CC
		Rev	ATGGTCTAGAAAGCTTTAATGGAGAGGGGGAATGCTTTGT
CrRO_NandCtermtruncated_pOPINE	pOPINE	Fwrd	AGGAGATATACCATG T CAA TTC CTG AAG CTT TTC TCA ATT GTA TTT CC
		Rev	GTGATGGTGATGTTTATGGAGAGGGGGAATGCTTTGT
CrRO_NandCtermtruncated_pOPING	pOPING	Fwrd	GCGTAGCTGAAACCGGC T CAA TTC CTG AAG CTT TTC TCA ATT GTA TTT CC
		Rev	GTGATGGTGATGTTTATGGAGAGGGGGAATGCTTTGT

7.9.5.2 Expression testing E. coli

Expression testing of the CrRO protein was performed in the *E. coli* strain SoluBL21. The CrRO_full_length_pOPINF and CrRO_Ntermtruncated_pOPINF vectors that were confirmed by sequencing were transformed into the *E. coli* expression strain SoluBL21 by heat shock and positive transformants were checked by colony PCR.

A single positive colony was used to grow an overnight 10 ml starter culture at 37 ^{o}C with shaking at 200 rpm in LB supplemented with carbenicillin. This culture was used to start a 100 ml culture grown at 37 ^{o}C in LB supplemented with carbenicillin until the optical density of the media reached an optical density at 600 nm (OD_{600}) of 0.5-0.8. After reaching this optical density the cultures were chilled on ice for 1 h. 1 ml of the non-induced culture was removed to assess protein composition in the non-induced cells. This 1 ml of culture was centrifuged at 4000 x g and lysed with BugBuster™ (Millipore). To the remaining culture, the media was supplemented with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.1 mg/ml to induce protein expression and was incubated with shaking overnight at 18 ^{o}C. 1 ml of the culture was removed to test for insoluble protein. This 1 ml of the culture was centrifuged at 4000 x g and the pellet dissolved in 50 µl 8M urea.

The remaining 100 ml of culture was centrifuged at 4000 x g and the pellet resuspended in buffer A1 (50 mM Tris-HCl, 50 mM glycine, 500 mM NaCl, 20 mM imidazole, 5% (w/v) glycerol), and subjected to cell lysis using the cell disruptor at 25 KPsi. The lysed suspension was centrifuged at 35,000 x g for 30 min and the supernatant collected. The supernatant was incubated with 200 µl Ni-NTA resin for 1 h at 4 ^{o}C. The Ni-NTA resin was washed 3 times with buffer A1, and the flow through was pooled. This represents proteins that do not bind to the Ni-NTA resin. The protein was eluted in 600 µl of elution buffer (50 mM Tris-HCl, 50 mM glycine, 500 mM NaCl, 500 mM imidazole, 5% (w/v) glycerol). Protein samples from the uninduced cell culture, insoluble cell pellet, flow through and eluted protein were run by SDS-PAGE to assess protein expression.

7.9.5.3 Expression testing in Sf9 Insect cells.

This work was performed by Joanne Nettleship at the Oxford Protein Production Facility.

7.9.5.3.1 Sf9 Insect cells

Spodoptera frugiperda (Sf9) cells were used for the purpose of recombinant protein expression using insect-specific viruses called baculoviruses. The number of cells in 10 µl of Sf9 cell stock from a shake flask was counted using 10 µl trypan blue staining. The ratio of cell stock to Sf900II media was calculated to give $4x10^5$ cells/ml, and this media and cell stock was transferred to a sterile flask and incubated at 27 ^oC with shaking at 120 rpm. Every 3-4 days the cells were passaged to a cell density of $4x10^5$ cells/ml.

7.9.5.3.2 Bacmid Preparation.

100 ml of LB supplemented with kanamycin and chloramphenicol was used to grow up an overnight culture of *E. coli* containing the bacmid, at 37 ^oC. This bacmid DNA was extracted using the BACMAX™ DNA extraction kit (Epicentre) in accordance with the manufacturer's protocol, and normalised to ~100 ng/µl. The bacmid DNA was linearised using the restriction enzyme Bsu36I utilising; 1 µl Bsu36I per 6 µg DNA for 2 h at 37 ^oC and a second digestion step again using 1 µl Bsu36I per 6 µg DNA for 2 h at 37 ^oC. The reaction was stopped at 72 ^oC for 20 min, and no further purification was needed.

7.9.5.3.3 Sf9 Transfection

Sf9 cells were cultured in 24 well culture plates. 500 µl of Sf9 cells, at a density of $5x10^5$ cells/ml in Sf900II media, were attached to the plate for 1 h at room temperature. The transfection mix, consisting of; 250 ng of the bacmid DNA, 100-500 ng of pOPIN vector DNA, 50 µl Sf900II, 1.5 µl FuGeneHD was incubated at room temperature for 30 min, before being added to the appropriate well containing Sf9 cells. This was incubated for 6-7 days at 27 ^oC, and the resultant viral supernatant was stored at 4 ^oC in the dark. This is the P0 viral stock.

The P0 viral stock was used to amplify the virus, resulting in a P1 virus which can then be used for small scale screening. 500 µl of Sf9 cells at a density of $1x10^6$ cells/ml was used to generate a monolayer of Sf9 cells in a 24 well plate, and was attached to the plate for 1 h at room temperature. 5 µl of the P0 virus stock was added to each well and incubated for 6-7 days at 27

°C. The supernatant contains the P1 virus stock, which was stored at 4 °C in the dark. For long term storage BSA is added to a final concentration of 10% and stored at -80 °C.

7.9.5.3.4 Expression Testing

24 deep well culture plates were used for Sf9 cultivation with 3 ml of cells at a density of $1x10^6$ cells/ml added to each well. 3 µl or 30 µl of the P1 virus was added to each well and the cells incubated with shaking for 3 days at 27 °C at 250 rpm.

To assess for intracellular protein expression, 1 ml of the Sf9 cells were placed in a 96-well plate, centrifuged for 15 min at 6,000 x g, the supernatant removed (to test for secreted protein) and the resultant cells frozen at -80 °C for at least 30 min. The defrosted pellet was resuspended in 210 µl NPI-10-Tween buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1% Tween-20, pH 8.0), containing cOmplete™ protease inhibitors and 400 U/ml DNaseI, and agitated for 30 min at 1000 rpm. The resulting solution was centrifuged at 6000 x g for 30 min, and was used to test for intracellular protein expression. 200 µl of the supernatant for testing for secreted proteins, and 200 µl of the supernatant for testing intracellular protein expression, was incubated with 20 µl of NiNTA magnetic beads, and each solution agitated for 30 min at 600 rpm to bind the protein to the beads, before being placed on a Qiagen Magnet Type B for 2 min, and the supernatant removed. The magnetic beads were washed twice with 200 µl NPI-20-Tween buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 0.05% Tween-20, pH 8.0), before the protein was eluted off the beads by incubation in 50 µl NPI-250-Tween buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 0.05% Tween-20, pH 8.0), with shaking for 10 min at 600 rpm for protein elution. The eluted samples were run by SDS-PAGE to assess for protein expression, and a western blot with a mouse-monoclonal anti-His antibody (Abcam, cat no 9F9.F9 ab1218) as the primary anti-body and rabbit mono-clonal anti-mouse IgG and IgM monoclonal antibody.

7.10 Methods specific for the purification of substrates

7.10.1 Purification of secologanin

Secologanin accumulates in the fruit of snowberry (*Symphoricarpos*) as well as in the leaves of honeysuckle (*Lonicera*). This accumulation is to a higher extent than in the leaves of *C. roseus,* and therefore these plants are a good natural resource for the purification of secologanin for use as a substrate in downstream assays and as a substrate for strictosidine biosynthesis.

100 g of snowberry fruit was harvested and ground under liquid N_2 using a mortar and pestle and transferred to a glass duran bottle. The chemical content of the ground material was extracted in 100 ml methanol, in an ultrasonic waterbath for 15 min. The methanol extraction was centrifuged and subsequently filtered (using Whatmann filter paper) into a round bottomed flask. This extraction procedure was repeated three times and the resultant pooled methanolic fraction concentrated to between 50-100 ml on a rotary evaporator. The methanolic fraction was washed twice 1:1 (v/v) with the hexane fraction from petroleum for the removal of lipid soluble compounds, and collected using a separating funnel. The methanolic fraction was concentrated to 60 ml for solid phase extraction (SPE, Discovery® DSC18) and removal of hydrophilic sugars. The SPE columns were primed with 1 column volume of methanol and 3 column volumes of water. The 60 ml methanolic fraction was layered on top of the C18 column and allowed to run through the column by gravity. The column was eluted with 1 column volume of water, this fraction was discarded, and 1 column volume of methanol, which was retained. The methanol fraction from the column was concentrated by rotary evaporation, and water removed by lyophilisation overnight.

The lyophilised extract was redissolved in a minimum amount of methanol and eluted on a silica column using MeOH:$CHCl_3$ (2:8 v/v) as the mobile phase. The fractions collected were monitored for secologanin content using thin layer chromatography on glass backed silica plates, again using MeOH:$CHCl_3$ (2:8 v/v) as the mobile phase, and the TLC plates were visualised under UV light at 254 nm wavelength. An authentic standard of secologanin has an Rf-0.5, and visualised at 254 nm.

The fractions containing secologanin were pooled, concentrated and lyophilised. The powder was resuspended in MeOH for purification by preparative high performance liquid chromatography (prepHPLC). The HPLC used was a Dionex Ultimate 3000 pump, with a variable wavelength UV detector. The column used for chromatographic separation was a Phenomenex Luna 5µ C18(2) 100 A (250 x 30 mm), with the binary solvent system consisting of 0.1% TFA and acetonitrile. The elution programme was the following: 1 min isocratic at 1% ACN, 11 min gradient up to 40% ACN, 2 min gradient up to 100% ACN, 2 min isocratic at 100% ACN, 1min gradient down to 1% ACN, and 6 min isocratic at 1% ACN. The compounds were monitored by UV 241nm, and the fractions collected at the retention time 12.9 min. The resulting fractions were lyophilised and stored at -20 ^{o}C until further use.

7.10.2 Generation of secologanol

Secologanol is the alcohol produced by the reduction of the aldehyde of secologanin. It was generated by incubating 1 ml 1 mM secologanin suspended in MeOH with 1 mg $NaBH_4$. 25 µl of this sample was mixed with 25 µl MeOH and 50 µl H_2O and analysed by LC-MS.

7.10.3 Expression and purification of strictosidine synthase

E. coli expression vectors for expression of strictosidine synthase were generated by Nancy Yerkes, and cloned into the *E. coli* expression strain SoluBL21.

A single positive colony for each construct was used to grow an overnight 10 ml starter culture at 37 ^oC with shaking at 200 rpm in LB supplemented with kanamycin. This culture was used to start a 500 ml culture grown at 37 ^oC in LB supplemented with kanamycin until the media reached an optical density at 600 nm (OD_{600}) of 0.5-0.8. After reaching this optical density the cultures were chilled on ice for 1 h. 1 ml of the non-induced culture was removed to assess protein composition in the non-induced cells. This 1 ml of culture was centrifuged at 4000 x g and lysed with BugBuster. To the remaining culture, the media was supplemented with IPTG at a final concentration of 0.1 mg/ml to induce protein expression and was incubated with shaking overnight at 18 ^oC. 1 ml of the culture was removed to test for insoluble protein. This 1 ml of the culture was centrifuged at 4000 x g and the pellet dissolved in 50 µl 8 M urea.

The remaining 500 ml of culture was centrifuged at 4000 x g and the pellet resuspended in binding buffer (50 mM phosphate buffer pH 7.4, 30 mM imidazole, 300 mM NaCl, cOmplete™ protease inhibitor tablets), and subjected to cell lysis using the cell disruptor in single shot mode at 25 KPsi. The lysed suspension was centrifuged at 35,000 x g for 30 min and the supernatant collected. The supernatant was incubated with 600 µl Ni-NTA resin for 1 h at 4^oC. The Ni-NTA resin and supernatant was stacked into a column, and was washed with binding buffer 3 times resulting in a total wash volume of 30 ml. This flow through was pooled and represents proteins that do not bind to the Ni-NTA resin. The protein was eluted in a total 5 ml volume of elution buffer (50 mM phosphate buffer pH 7.4, 500 mM imidazole, 300 mM NaCl) in 1 ml fractions. Protein samples from the insoluble cell pellet, flow through and eluted protein were run by SDS-PAGE to assess protein expression (Figure 7-4). Fractions from the elution containing a protein of the right size were concentrated using Millipore 10 kDa spin filters, and the protein was buffer exchanged into phosphate buffer pH 7. A BCA assay was used to measure protein concentration, and the final concentration of the strictosidine synthase and strictosidine glucosidase was

adjusted to ~0.3 mg/ml. Glycerol was added to the protein in phosphate buffer to 10% (v/v), and the protein samples aliquoted and stored at -20 °C.

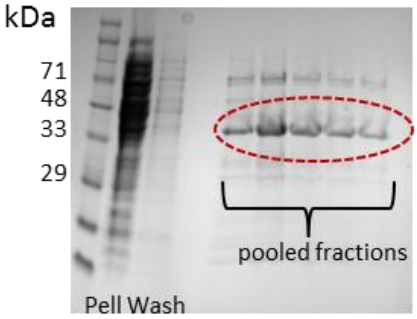


Figure 7-4 Purification of strictosidine synthase
Pooled fractions were used for the enzymatic generation of strictosidine.

7.10.4 Enzymatic generation of strictosidine

4 mM tryptamine, 2 mM purified secologanin and 2 µM strictosidine synthase was incubated in 50 mM Phosphate buffer pH 7.4 buffer in ~50 ml overnight at 37 °C. This assay was checked for strictosidine production and consumption of secologanin by LC-MS. The solution was purified by SPE, eluted in 100% MeOH, and dried under vacuum. The dried product was resuspended in H_2O, filtered with 2 µm filters and purified by HPLC in 2 ml aliquots.

7.10.5 Purification of strictosidine

The strictosidine generated was purified by preparative HPLC. Again the HPLC used was a Dionex Ultimate 3000 pump, with a variable wavelength UV detector. The column used for chromatographic separation was a Phenomenex Luna 5µ C18(2) 100 A (250 x 30 mm), with the binary solvent system consisting of 0.1% TFA and acetonitrile. The elution programme was the following: 1 min isocratic at 10% ACN, 8 min gradient up to 30% ACN, 4 min gradient up to 100% ACN, 6 min isocratic at 100% ACN, 1 min gradient down to 10% ACN, and 6 min isocratic at 10% ACN. The compounds were monitored by UV 241 nm, and the fractions collected at the retention time 13.5 min. The resulting fractions were lyophilised and stored at -20 °C until further use.

7.10.6 Nuclear Magenetic Resonance (NMR) Spectroscopy

NMR spectra (^{1}H NMR) were acquired using a Bruker Avance III 400 NMR spectrometer, operating at 400 MHz for ^{1}H. 2.8 mg strictosidine were disolved in 500 µl MeOH-d_4, and 500 µm Trimethylsilyl propanoic acid (TMSP) has been used as internal standard for chemical shift reference (δ 0 ppm) and quantification. The purity of strictosidine was calculated according to the formula $(I_{analyte}/I_{standard})=(n_{hanalyte}/n_{hstandard})*(C_{analyte}/C_{standard})$, where I is the intensity of the peak, n_h is the number of protons, and C is the concentration.

8 References

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouché, N., Gasciolli, V. & Vaucheret, H. 2006. DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Current Biology,* 16**,** 927-932.

Ageorges, A., Fernandez, L., Vialet, S., Merdinoglu, D., Terrier, N. & Romieu, C. 2006. Four specific isogenes of the anthocyanin metabolic pathway are systematically co-expressed with the red colour of grape berries. *Plant Science,* 170**,** 372-383.

Aguilar-Barragán, A. & Ochoa-Alejo, N. 2014. Virus-induced silencing of MYB and WD40 transcription factor genes affects the accumulation of anthocyanins in chilli pepper fruit. *Biologia Plantarum,* 58**,** 567-574.

Ahmad, V., Ahmad, V. U., Atta-Ur-Rahman, E., Zaman, K., Atta-Ur-Rahman, E., Choudhary, M., Shekhani, M., Choudhary, E. M., Atta-Ur-Rahman, M. & Basha, A. 1983. Biosynthesis of indole alkaloids.

Alfenito, M. R., Souer, E., Goodman, C. D., Buell, R., Mol, J., Koes, R. & Walbot, V. 1998. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *The Plant Cell,* 10**,** 1135-1149.

Alhagdow, M., Mounet, F., Gilbert, L., Nunes-Nesi, A., Garcia, V., Just, D., Petit, J., Beauvoit, B., Fernie, A. R., Rothan, C. & Baldet, P. 2007. Silencing of the mitochondrial ascorbate synthesizing enzyme L-galactono-1, 4-lactone dehydrogenase affects plant and fruit development in tomato. *Plant Physiology,* 145**,** 1408-1422.

Almagro, A., Lin, S. H. & Tsay, Y. F. 2008. Characterization of the *Arabidopsis* nitrate transporter NRT1. 6 reveals a role of nitrate in early embryo development. *The Plant Cell,* 20**,** 3289-3299.

Amann, M., Nagakura, N. & Zenk, M. 1984. (S)-tetrahydroprotoberberine oxidase the final enzyme in protoberberine biosynthesis. *Tetrahedron letters,* 25**,** 953-954.

Appelhagen, I., Thiedig, K., Nordholt, N., Schmidt, N., Huep, G., Sagasser, M. & Weisshaar, B. 2014. Update on transparent testa mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. *Planta,* 240**,** 955-970.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A. & Gasteiger, E. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic acids research,* 40**,** W597-W603.

Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H. & De Luca, V. 2013. A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *The Plant Cell,* 25**,** 4123-4134.

Axarli, I., Dhavala, P., Papageorgiou, A. C. & Labrou, N. E. 2009. Crystallographic and functional characterization of the fluorodifen-inducible glutathione transferase from *Glycine max* reveals an active site topography suited for diphenylether herbicides and a novel L-site. *Journal of molecular biology,* 385**,** 984-1002.

Bauer, P., Munkert, J., Brydziun, M., Burda, E., Müller-Uri, F., Gröger, H., Muller, Y. A. & Kreis, W. 2010. Highly conserved progesterone 5β-reductase genes (P5βR) from 5β-cardenolide-free and 5β-cardenolide-producing angiosperms. *Phytochemistry,* 71**,** 1495-1505.

Baulcombe, D. C. 1999. Fast forward genetics based on virus-induced gene silencing. *Current opinion in plant biology,* 2**,** 109-113.

Baumberger, N. & Baulcombe, D. 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences* 102**,** 11928-11933.

Beaudoin, G. A. & Facchini, P. J. 2014. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta,* 240**,** 19-32.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic acids research,* 35, e45.

Bertl, A., Rapp, U. & Hedrich, R. 2008. TPK1, a Ca^{2+}-regulated *Arabidopsis* vacuole two-pore K channel is activated by 14-3-3 proteins. *The Plant Journal,* 54, 963.

Besseau, S., Kellner, F., Lanoue, A., Thamm, A. M., Salim, V., Schneider, B., Geu-Flores, F., Höfer, R., Guirimand, G. & Guihur, A. 2013. A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. *Plant physiology,* 163, 1792-1803.

Bhattacharjee, S., Zamora, A., Azhar, M. T., Sacco, M. A., Lambert, L. H. & Moffett, P. 2009. Virus resistance induced by NB–LRR proteins involves Argonaute 4-dependent translational control. *The Plant Journal,* 58, 940-951.

Bird, D. A. & Facchini, P. J. 2001. Berberine bridge enzyme, a key branch-point enzyme in benzylisoquinoline alkaloid biosynthesis, contains a vacuolar sorting determinant. *Planta,* 213, 888-897.

Bones, A. M. & Rossiter, J. T. 1996. The myrosinase-glucosinolate system, its organisation and biochemistry. *Physiologia Plantarum,* 97, 194-208.

Bouvier, F., Linka, N., Isner, J.-C., Mutterer, J., Weber, A. P. & Camara, B. 2006. *Arabidopsis* SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. *The Plant Cell,* 18, 3088-3105.

Bowman, L. C., Houghton, J. A. & Houghton, P. J. 1988. Formation and stability of vincristine-tubulin complex in kidney cytosols: Role of GTP and GTP hydrolysis. *Biochemical pharmacology,* 37, 1251-1257.

Bradbury, L. M., Niehaus, T. D. & Hanson, A. D. 2013. Comparative genomics approaches to understanding and manipulating plant metabolism. *Current opinion in biotechnology,* 24, 278-284.

Brown, M. H., Paulsen, I. T. & Skurray, R. A. 1999. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Molecular microbiology,* 31, 394-395.

Brown, S., Clastre, M., Courdavault, V. & O' Connor, S. E. 2015. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proceedings of the National Academy of Sciences,* 112, 3205-3210.

Brown, V. L. & He, L. 2015. *Current Status and Future Prospects of Mass Spectrometry Imaging of Small Molecules*, Springer.

Burch-Smith, T. M., Anderson, J. C., Martin, G. B. & Dinesh-Kumar, S. P. 2004. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *The Plant Journal,* 39, 734-746.

Burhenne, K., Kristensen, B. K. & Rasmussen, S. K. 2003. A New Class of N-Hydroxycinnamoyltransferases; purification, cloning, and expression of a barley agmatine coumaroyltransferase *Journal of Biological Chemistry,* 278, 13919-13927.

Bürkle, L., Cedzich, A., Döpke, C., Stransky, H., Okumoto, S., Gillissen, B., Kühn, C. & Frommer, W. B. 2003. Transport of cytokinins mediated by purine transporters of the PUP family expressed in phloem, hydathodes, and pollen of *Arabidopsis*. *The Plant Journal,* 34, 13-26.

Burko, Y., Geva, Y., Refael-Cohen, A., Shleizer-Burko, S., Shani, E., Berger, Y., Halon, E., Chuck, G., Moshelion, M. & Ori, N. 2011. From organelle to organ: ZRIZI MATE-Type transporter is an organelle transporter that enhances organ initiation. *Plant and Cell Physiology,* 52, 518-527.

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.

Canel, C., Lopes-Cardoso, M. I., Whitmer, S., Van Der Fits, L., Pasquali, G., Van Der Heijden, R., Hoge, J. H. C. & Verpoorte, R. 1998. Effects of over-expression of strictosidine synthase and tryptophan decarboxylase on alkaloid production by cell cultures of *Catharanthus roseus*. *Planta,* 205**,** 414-419.

Carqueijeiro, I., Masini, E., Foureau, E., Sepúlveda, L. J., Marais, E., Lanoue, A., Besseau, S., Papon, N., Clastre, M. & Dugé De Bernonville, T. 2015. Virus-induced gene silencing in *Catharanthus roseus* by biolistic inoculation of tobacco rattle virus vectors. *Plant Biology*.

Carqueijeiro, I., Noronha, H., Duarte, P., Gerós, H. & Sottomayor, M. 2013. Vacuolar transport of the medicinal alkaloids from *Catharanthus roseus* is mediated by a proton-driven antiport. *Plant Physiology,* 162**,** 1486-1496.

Carter, C., Pan, S., Zouhar, J., Avila, E. L., Girke, T. & Raikhel, N. V. 2004. The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *The Plant Cell,* 16**,** 3285-3303.

Carter, C. J. & Thornburg, R. W. 2004. Tobacco nectarin V is a flavin-containing berberine bridge enzyme-like protein with glucose oxidase activity. *Plant Physiology,* 134**,** 460-469.

Chai, Y.-R., Lei, B., Huang, H.-L., Li, J.-N., Yin, J.-M., Tang, Z.-L., Wang, R. & Chen, L. 2009. Transparent testa 12 genes from *Brassica napus* and parental species: cloning, evolution, and differential involvement in yellow seed trait. *Molecular Genetics and Genomics,* 281**,** 109-123.

Chen, B., Ling, H. & Chang, M. W. 2013. Transporter engineering for improved tolerance against alkane biofuels in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels,* 6**,** 21.

Chen, C.-Z., Lv, X.-F., Li, J.-Y., Yi, H.-Y. & Gong, J.-M. 2012. *Arabidopsis* NRT1. 5 is another essential component in the regulation of nitrate reallocation and stress tolerance. *Plant Physiology,* 159**,** 1582-1590.

Chen, L., Liu, Y., Liu, H., Kang, L., Geng, J., Gai, Y., Ding, Y., Sun, H. & Li, Y. 2015. Identification and expression analysis of MATE genes involved in flavonoid transport in blueberry plants. *PloS One,* 10.

Chen, Z.-S., Lee, K., Walther, S., Raftogianis, R. B., Kuwano, M., Zeng, H. & Kruh, G. D. 2002. Analysis of Methotrexate and Folate Transport by Multidrug Resistance Protein 4 (ABCC4) MRP4 Is a Component of the Methotrexate Efflux System. *Cancer Research,* 62**,** 3144-3150.

Chen, Z., Chen, Y.-Z., Wang, X.-F., Wang, C., Yan, R.-X. & Zhang, Z. 2011. Prediction of ubiquitination sites by using the composition of k-spaced amino acid pairs. *PLoS One,* 6**,** e22930.

Chiang, C.-S., Stacey, G. & Tsay, Y.-F. 2004. Mechanisms and functional properties of two peptide transporters, AtPTR2 and fPTR2. *Journal of Biological Chemistry,* 279**,** 30150-30157.

Chiu, C.-C., Lin, C.-S., Hsia, A.-P., Su, R.-C., Lin, H.-L. & Tsay, Y.-F. 2004. Mutation of a nitrate transporter, AtNRT1: 4, results in a reduced petiole nitrate content and altered leaf development. *Plant and Cell Physiology,* 45**,** 1139-1148.

Cho, M. & Cho, H. 2013. The function of ABCB transporters in auxin transport. *Plant signaling & behavior,* 8**,** 642-54.

Contin, A., Van Der Heijden, R., Lefeber, A. W. & Verpoorte, R. 1998. The iridoid glucoside secologanin is derived from the novel triose phosphate/pyruvate pathway in a *Catharanthus roseus* cell culture. *FEBS letters,* 434**,** 413-416.

Contin, A., Van Der Heijden, R. & Verpoorte, R. 1999a. Accumulation of loganin and secologanin in vacuoles from suspension cultured *Catharanthus roseus* cells. *Plant science,* 147**,** 177-183.

Contin, A., Van Der Heijden, R. & Verpoorte, R. 1999b. Effects of alkaloid precursor feeding and elicitation on the accumulation of secologanin in a *Catharanthus roseus* cell suspension culture. *Plant cell, tissue and organ culture,* 56**,** 111-119.

Courdavault, V., Papon, N., Clastre, M., Giglioli-Guivarc'h, N., St-Pierre, B. & Burlat, V. 2014. A look inside an alkaloid multisite plant: the *Catharanthus* logistics. *Current opinion in plant biology,* 19**,** 43-50.

Crouzet, J., Roland, J., Peeters, E., Trombik, T., Ducos, E., Nader, J. & Boutry, M. 2013. NtPDR1, a plasma membrane ABC transporter from *Nicotiana tabacum*, is involved in diterpene transport. *Plant molecular biology,* 82**,** 181-192.

Cummins, I., O'hagan, D., Jablonkai, I., Cole, D. J., Hehn, A., Werck-Reichhart, D. & Edwards, R. 2003. Cloning, characterization and regulation of a family of phi class glutathione transferases from wheat. *Plant molecular biology,* 52**,** 591-603.

Custers, J. H., Harrison, S. J., Sela-Buurlage, M. B., Van Deventer, E., Lageweg, W., Howe, P. W., Van Der Meijs, P. J., Ponstein, A. S., Simons, B. H. & Melchers, L. S. 2004. Isolation and characterisation of a class of carbohydrate oxidases from higher plants, with a role in active defence. *The Plant Journal,* 39**,** 147-160.

Daniel, B., Pavkov-Keller, T., Steiner, B., Dordic, A., Gutmann, A., Nidetzky, B., Sensen, C. W., Van Der Graaff, E., Wallner, S. & Gruber, K. 2015. Oxidation of Monolignols by Members of the Berberine Bridge Enzyme Family Suggests a Role in Plant Cell Wall Metabolism. *Journal of Biological Chemistry,* 290**,** 18770-18781.

Daniel, H. & Kottra, G. 2004. The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Pflügers Archiv,* 447**,** 610-618.

David, P. D., Jonathan, D. S. & Robert, E. 2011. The *Arabidopsis* phi class glutathione transferase AtGSTF2: binding and regulation by biologically active heterocyclic ligands. *Biochemical Journal,* 438**,** 63-70.

Davoine, C., Falletti, O., Douki, T., Iacazio, G., Ennar, N., Montillet, J.-L. & Triantaphylidès, C. 2006. Adducts of oxylipin electrophiles to glutathione reflect a 13 specificity of the downstream lipoxygenase pathway in the tobacco hypersensitive response. *Plant Physiology,* 140**,** 1484-1493.

De Bernonville, T. D., Foureau, E., Parage, C., Lanoue, A., Clastre, M., Londono, M. A., Oudin, A., Houillé, B., Papon, N. & Besseau, S. 2015. Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced de novo assembly of a *Catharanthus roseus* transcriptome. *BMC genomics,* 16**,** 619.

De Colibus, L. & Mattevi, A. 2006. New frontiers in structural flavoenzymology. *Current opinion in structural biology,* 16**,** 722-728.

De Crécy-Lagard, V. & Hanson, A. D. 2007. Finding novel metabolic genes through plant-prokaryote phylogenomics. *Trends in microbiology,* 15**,** 563-570.

De Geyter, N., Gholami, A., Goormachtig, S. & Goossens, A. 2012. Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends in Plant Science,* 17**,** 349-359.

De Luca, V. & Cutler, A. J. 1987. Subcellular localization of enzymes involved in indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiology,* 85**,** 1099-1102.

De Luca, V., Salim, V., Levac, D., Atsumi, S. M. & Yu, F. 2012. Discovery and functional analysis of monoterpenoid indole alkaloid pathways in plants. *Methods Enzymology,* 515**,** 207-229.

Debeaujon, I., Peeters, A. J., Léon-Kloosterziel, K. M. & Koornneef, M. 2001. The Transparent testa 12 gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *The Plant Cell,* 13**,** 853-871.

Deloache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J. & Dueber, J. E. 2015. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nature chemical biology*.

Deus-Neumann, B. & Zenk, M. 1984. A highly selective alkaloid uptake system in vacuoles of higher plants. *Planta,* 162**,** 250-260.

Deus-Neumann, B. & Zenk, M. 1986. Accumulation of alkaloids in plant vacuoles does not involve an ion-trap mechanism. *Planta,* 167**,** 44-53.

Dewick, P. M. 2002. *Medicinal natural products: a biosynthetic approach*, John Wiley & Sons.

Dewitt, Å. M., Andersson, K., Peltre, G. & Lidholm, J. 2006. Cloning, expression and immunological characterization of full-length timothy grass pollen allergen Phl p 4, a berberine bridge enzyme-like protein with homology to celery allergen Api g 5. *Clinical & Experimental Allergy,* 36**,** 77-86.

Di Matteo, A., Ruggieri, V., Sacco, A., Rigano, M. M., Carriero, F., Bolger, A., Fernie, A. R., Frusciante, L. & Barone, A. 2013. Identification of candidate genes for phenolics accumulation in tomato fruit. *Plant science,* 205**,** 87-96.

Diener, A. C., Gaxiola, R. A. & Fink, G. R. 2001. *Arabidopsis* ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. *The Plant Cell,* 13**,** 1625-1638.

Dietrich, D., Hammes, U., Thor, K., Suter-Grotemeyer, M., Flückiger, R., Slusarenko, A. J., Ward, J. M. & Rentsch, D. 2004. AtPTR1, a plasma membrane peptide transporter expressed during seed germination and in vascular tissue of *Arabidopsis*. *The Plant Journal,* 40**,** 488-499.

Dittrich, H. & Kutchan, T. M. 1991. Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. *Proceedings of the National Academy of Sciences,* 88**,** 9969-9973.

Dixon, D. P., Davis, B. G. & Edwards, R. 2002. Functional Divergence in the Glutathione Transferase Superfamily in Plants identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. *Journal of Biological Chemistry,* 277**,** 30859-30869.

Dixon, D. P. & Edwards, R. 2009. Selective binding of glutathione conjugates of fatty acid derivatives by plant glutathione transferases. *Journal of Biological Chemistry,* 284**,** 21249-21256.

Dixon, D. P. & Edwards, R. 2010. Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. *Journal of Biological Chemistry,* 285**,** 36322-36329.

Dixon, D. P., Lapthorn, A., Madesis, P., Mudd, E. A., Day, A. & Edwards, R. 2008. Binding and glutathione conjugation of porphyrinogens by plant glutathione transferases. *Journal of Biological Chemistry,* 283**,** 20268-20276.

Dixon, D. P., Skipsey, M. & Edwards, R. 2010. Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry,* 71**,** 338-350.

Doyle, A., Mcqueen, E. & Smirk, F. 1955. Treatment of hypertension with reserpine, with reserpine in combination with pentapyrrolidinium, and with reserpine in combination with veratrum alkaloids. *Circulation,* 11**,** 170-181.

Drew, D., Newstead, S., Sonoda, Y., Kim, H., Von Heijne, G. & Iwata, S. 2008. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nature protocols,* 3**,** 784-798.

Durrett, T. P., Gassmann, W. & Rogers, E. E. 2007. The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. *Plant Physiology,* 144**,** 197-205.

Dyke, R. W. & Nelson, R. L. 1977. Phase I anti-cancer agents vindesine (desacetyl vinblastine amide sulfate). *Cancer treatment reviews,* 4**,** 135-142.

Dym, O. & Eisenberg, D. 2001. Sequence-structure analysis of FAD-containing proteins. *Protein Science,* 10**,** 1712-1728.

Egner, R. & Kuchler, K. 1996. The yeast multidrug transporter Pdr5 of the plasma membrane is ubiquitinated prior to endocytosis and degradation in the vacuole. *FEBS letters,* 378**,** 177-181.

Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences,* 95**,** 14863-14868.

Eybishtz, A., Peretz, Y., Sade, D., Gorovits, R. & Czosnek, H. 2010. Tomato yellow leaf curl virus infection of a resistant tomato line with a silenced sucrose transporter gene LeHT1 results in inhibition of growth, enhanced virus spread, and necrosis. *Planta,* 231**,** 537-548.

Facchini, P. J., Penzes, C., Johnson, A. G. & Bull, D. 1996. Molecular characterization of berberine bridge enzyme genes from opium poppy. *Plant Physiology,* 112**,** 1669-1677.

Fan, S.-C., Lin, C.-S., Hsu, P.-K., Lin, S.-H. & Tsay, Y.-F. 2009. The *Arabidopsis* nitrate transporter NRT1. 7, expressed in phloem, is responsible for source-to-sink remobilization of nitrate. *The Plant Cell,* 21**,** 2750-2761.

Fernández-Cañón, J. M. & Peñalva, M. A. 1998. Characterization of a fungal maleylacetoacetate isomerase gene and identification of its human homologue. *Journal of Biological Chemistry,* 273**,** 329-337.

Ferreyra, M. L. F., Rius, S. P. & Casati, P. 2012. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Frontiers in plant science,* 3.

Fossati, E., Ekins, A., Narcross, L., Zhu, Y., Falgueyret, J.-P., Beaudoin, G. A., Facchini, P. J. & Martin, V. J. 2014. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. *Nature communications,* 5.

Fossati, E., Narcross, L., Ekins, A., Falgueyret, J. P. & Martin, V. J. J. 2015. Synthesis of Morphinan Alkaloids in *Saccharomyces cerevisiae*. *PLoS One,* 10**,** 15.

Francisco, R. M., Regalado, A., Ageorges, A., Burla, B. J., Bassin, B., Eisenach, C., Zarrouk, O., Vialet, S., Marlin, T. & Chaves, M. M. 2013. ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-glucosides. *The Plant Cell,* 25**,** 1840-1854.

Frébortová, J., Fraaije, M., Galuszka, P., Sebela, M., Pec, P., Hrbac, J., Novak, O., Bilyeu, K., English, J. & Frébort, I. 2004. Catalytic reaction of cytokinin dehydrogenase: preference for quinones as electron acceptors. *Biochem. J,* 380**,** 121-130.

Frébortová, J., Novák, O., Frébort, I. & Jorda, R. 2010. Degradation of cytokinins by maize cytokinin dehydrogenase is mediated by free radicals generated by enzymatic oxidation of natural benzoxazinones. *The Plant Journal,* 61**,** 467-481.

Frick, S., Chitty, J. A., Kramell, R., Schmidt, J., Allen, R. S., Larkin, P. J. & Kutchan, T. M. 2004. Transformation of opium poppy (*Papaver somniferum* L.) with antisense berberine bridge enzyme gene (anti-bbe) via somatic embryogenesis results in an altered ratio of alkaloids in latex but not in roots. *Transgenic research,* 13**,** 607-613.

Frommer, W. B., Hummel, S. & Rentsch, D. 1994. Cloning of an Arabidopsis histidine transporting protein related to nitrate and peptide transporters. *FEBS letters,* 347**,** 185-189.

Furukawa, J., Yamaji, N., Wang, H., Mitani, N., Murata, Y., Sato, K., Katsuhara, M., Takeda, K. & Ma, J. F. 2007. An aluminum-activated citrate transporter in barley. *Plant and Cell Physiology,* 48**,** 1081-1091.

Galan, J. M., Moreau, V., Andre, B., Volland, C. & Haguenauer-Tsapis, R. 1996. Ubiquitination mediated by the Npi1p/Rsp5p ubiquitin-protein ligase is required for endocytosis of the yeast uracil permease. *Journal of Biological Chemistry,* 271**,** 10946-10952.

Gamas, P., De Carvalho Niebel, F., Lescure, N. & Cullimore, J. V. 1996. Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. *MPMI-Molecular Plant Microbe Interactions,* 9**,** 233-242.

George, G. M., Bauer, R., Blennow, A., Kossmann, J. & Lloyd, J. R. 2012. Virus-induced multiple gene silencing to study redundant metabolic pathways in plants: Silencing the starch degradation pathway in *Nicotiana benthamiana*. *Biotechnology journal,* 7**,** 884-890.

Gesell, A., Chávez, M. L. D., Kramell, R., Piotrowski, M., Macheroux, P. & Kutchan, T. M. 2011. Heterologous expression of two FAD-dependent oxidases with (S)-tetrahydroprotoberberine oxidase activity from *Argemone mexicana* and *Berberis wilsoniae* in insect cells. *Planta,* 233**,** 1185-1197.

Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T. & Halkier, B. A. 2007. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic acids research,* 35**,** e55.

Geu-Flores, F., Sherden, N. H., Courdavault, V., Burlat, V., Glenn, W. S., Wu, C., Nims, E., Cui, Y. & O' Connor, S. E. 2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature,* 492**,** 138-142.

Gillissen, B., Bürkle, L., André, B., Kühn, C., Rentsch, D., Brandl, B. & Frommer, W. B. 2000. A new family of high-affinity transporters for adenine, cytosine, and purine derivatives in *Arabidopsis*. *The Plant Cell,* 12**,** 291-300.

Ginglinger, J.-F., Boachon, B., Höfer, R., Paetz, C., Köllner, T. G., Miesch, L., Lugan, R., Baltenweck, R., Mutterer, J. & Ullmann, P. 2013. Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in *Arabidopsis* flowers. *The Plant Cell Online,* 25**,** 4640-4657.

Glenn, W. S., Nims, E. & O' Connor, S. E. 2011. Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *Journal of the American Chemical Society,* 133**,** 19346-19349.

Gomes, D., Agasse, A., Thiébaud, P., Delrot, S., Gerós, H. & Chaumont, F. 2009. Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1788**,** 1213-1228.

Gomez, C., Terrier, N., Torregrosa, L., Vialet, S., Fournier-Level, A., Verries, C., Souquet, J.-M., Mazauric, J.-P., Klein, M., Cheynier, V. & Ageorges, A. 2009. Grapevine MATE-Type Proteins Act as Vacuolar H+-Dependent Acylated Anthocyanin Transporters. *Plant Physiology,* 150**,** 402-415.

Góngora-Castillo, E., Childs, K. L., Fedewa, G., Hamilton, J. P., Liscombe, D. K., Magallanes-Lundback, M., Mandadi, K. K., Nims, E., Runguphan, W. & Vaillancourt, B. 2012. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS One,* 7**,** e52506.

Goodman, C. D., Casati, P. & Walbot, V. 2004. A multidrug resistance–associated protein involved in anthocyanin transport in *Zea mays*. *The Plant Cell,* 16**,** 1812-1826.

Green, L. S. & Rogers, E. E. 2004. FRD3 controls iron localization in *Arabidopsis*. *Plant Physiology,* 136**,** 2523-2531.

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

Guirimand, G., Courdavault, V., Lanoue, A., Mahroug, S., Guihur, A., Blanc, N., Giglioli-Guivarc'h, N., St-Pierre, B. & Burlat, V. 2010. Strictosidine activation in Apocynaceae: towards a nuclear time bomb? *BMC Plant Biology,* 10**,** 182.

Guirimand, G., Guihur, A., Ginis, O., Poutrain, P., Héricourt, F., Oudin, A., Lanoue, A., St-Pierre, B., Burlat, V. & Courdavault, V. 2011a. The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. *FEBS Journal,* 278**,** 749-763.

Guirimand, G., Guihur, A., Poutrain, P., Héricourt, F., Mahroug, S., St-Pierre, B., Burlat, V. & Courdavault, V. 2011b. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *Journal of plant physiology,* 168**,** 549-557.

Hagel, J. M., Beaudoin, G. A., Fossati, E., Ekins, A., Martin, V. J. & Facchini, P. J. 2012. Characterization of a flavoprotein oxidase from opium poppy catalyzing the final steps in sanguinarine and papaverine biosynthesis. *Journal of Biological Chemistry,* 287**,** 42972-42983.

Hawkins, K. M. & Smolke, C. D. 2008. Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. *Nature chemical biology,* 4**,** 564-573.

Heinig, U., Gutensohn, M., Dudareva, N. & Aharoni, A. 2013. The challenges of cellular compartmentalization in plant metabolic engineering. *Current opinion in biotechnology,* 24**,** 239-246.

Hemalatha, R. & Pradeep, T. 2013. Understanding the molecular signatures in leaves and flowers by desorption electrospray ionization mass spectrometry (DESI MS) imaging. *Journal of agricultural and food chemistry,* 61**,** 7477-7487.

Hetherington, A. M. & Woodward, F. I. 2003. The role of stomata in sensing and driving environmental change. *Nature,* 424**,** 901-908.

Heuts, D., Winter, R., Damsma, G., Janssen, D. & Fraaije, M. 2008. The role of double covalent flavin binding in chito-oligosaccharide oxidase from *Fusarium graminearum*. *Biochem. J,* 413**,** 175-183.

Heuts, D. P., Scrutton, N. S., Mcintire, W. S. & Fraaije, M. W. 2009. What's in a covalent bond? *FEBS Journal,* 276**,** 3405-3427.

Hibi, N., Higashiguchi, S., Hashimoto, T. & Yamada, Y. 1994. Gene expression in tobacco low-nicotine mutants. *The Plant Cell,* 6**,** 723-735.

Hicke, L. & Riezman, H. 1996. Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. *Cell,* 84**,** 277-287.

Hildreth, S. B., Gehman, E. A., Yang, H., Lu, R.-H., Ritesh, K., Harich, K. C., Yu, S., Lin, J., Sandoe, J. L., Okumoto, S., Murphy, A. S. & Jelesko, J. G. 2011. Tobacco nicotine uptake permease (NUP1) affects alkaloid metabolism. *Proceedings of the National Academy of Sciences,* 108**,** 18179-18184.

Hileman, L. C., Drea, S., Martino, G., Litt, A. & Irish, V. F. 2005. Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species *Papaver somniferum* (opium poppy). *The Plant Journal,* 44**,** 334-341.

Hong, G.-J., Xue, X.-Y., Mao, Y.-B., Wang, L.-J. & Chen, X.-Y. 2012. *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. *The Plant Cell Online,* 24**,** 2635-2648.

Horák, J. 2003. The role of ubiquitin in down-regulation and intracellular sorting of membrane proteins: insights from yeast. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1614**,** 139-155.

Huang, C.-H., Lai, W.-L., Lee, M.-H., Chen, C.-J., Vasella, A., Tsai, Y.-C. & Liaw, S.-H. 2005. Crystal structure of glucooligosaccharide oxidase from *Acremonium strictum*: a novel flavinylation of 6-S-cysteinyl, 8α-N1-histidyl FAD. *Journal of Biological Chemistry,* 280**,** 38831-38838.

Huang, C.-H., Winkler, A., Chen, C.-L., Lai, W.-L., Tsai, Y.-C., Macheroux, P. & Liaw, S.-H. 2008. Functional roles of the 6-S-cysteinyl, 8α-N1-histidyl FAD in glucooligosaccharide oxidase from *Acremonium strictum*. *Journal of Biological Chemistry,* 283**,** 30990-30996.

Huang, N.-C., Chiang, C.-S., Crawford, N. M. & Tsay, Y.-F. 1996. CHL1 encodes a component of the low-affinity nitrate uptake system in *Arabidopsis* and shows cell type-specific expression in roots. *The Plant Cell,* 8**,** 2183-2191.

Huang, N.-C., Liu, K.-H., Lo, H.-J. & Tsay, Y.-F. 1999. Cloning and functional characterization of an *Arabidopsis* nitrate transporter gene that encodes a constitutive component of low-affinity uptake. *The Plant Cell,* 11**,** 1381-1392.

Hunter, W. N. 2007. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *Journal of Biological Chemistry,* 282**,** 21573-21577.

Ifa, D. R., Srimany, A., Eberlin, L. S., Naik, H. R., Bhat, V., Cooks, R. G. & Pradeep, T. 2011. Tissue imprint imaging by desorption electrospray ionization mass spectrometry. *Analytical Methods,* 3**,** 1910-1912.

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.

Jack, D. L., Yang, N. M. & H Saier, M. 2001. The drug/metabolite transporter superfamily. *European Journal of Biochemistry,* 268**,** 3620-3639.

Jelesko, J. G. 2012. An expanding role for purine uptake permease-like transporters in plant secondary metabolism. *Frontiers in plant science,* 3.

Jeong, J., Suh, S., Guan, C., Tsay, Y.-F., Moran, N., Oh, C. J., An, C. S., Demchenko, K. N., Pawlowski, K. & Lee, Y. 2004. A nodule-specific dicarboxylate transporter from alder is a member of the peptide transporter family. *Plant Physiology,* 134**,** 969-978.

Ji, Y., Xiao, J., Shen, Y., Ma, D., Li, Z., Pu, G., Li, X., Huang, L., Liu, B., Ye, H. & Wang, H. 2014. Cloning and Characterization of AabHLH1, a bHLH Transcription Factor that Positively Regulates Artemisinin Biosynthesis in *Artemisia annua*. *Plant and Cell Physiology,* 55**,** 1592-1604.

Jones, D. T., Taylor, W. R. & Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS,* 8**,** 275-282.

Kajikawa, M., Shoji, T., Kato, A. & Hashimoto, T. 2011. Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. *Plant Physiology,* 155**,** 2010-2022.

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y. & Martinoia, E. 2011. Plant ABC transporters. *The Arabidopsis book/American Society of Plant Biologists,* 9.

Kanno, Y., Hanada, A., Chiba, Y., Ichikawa, T., Nakazawa, M., Matsui, M., Koshiba, T., Kamiya, Y. & Seo, M. 2012. Identification of an abscisic acid transporter by functional screening using the receptor complex as a sensor. *Proceedings of the National Academy of Sciences,* 109**,** 9653-9658.

Kanno, Y., Kamiya, Y. & Seo, M. 2013. Nitrate does not compete with abscisic acid as a substrate of AtNPF4. 6/NRT1. 2/AIT1 in *Arabidopsis*. *Plant signaling & behavior,* 8**,** e26624.

Karim, S., Holmström, K.-O., Mandal, A., Dahl, P., Hohmann, S., Brader, G., Palva, E. T. & Pirhonen, M. 2007. AtPTR3, a wound-induced peptide transporter needed for defence against virulent bacterial pathogens in *Arabidopsis*. *Planta,* 225**,** 1431-1445.

Kato, K., Shitan, N., Shoji, T. & Hashimoto, T. 2015. Tobacco NUP1 transports both tobacco alkaloids and vitamin B_6. *Phytochemistry,* 113**,** 33-40.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

Kellner, F., Geu-Flores, F., Sherden, N. H., Brown, S., Foureau, E., Courdavault, V. & O' Connor, S. E. 2015a. Discovery of a P450-catalyzed step in vindoline biosynthesis: A link between the aspidosperma and eburnamine alkaloids. *Chemical Communications,* 51**,** 7626-7628.

Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., Maclay, K., Buell, R. & O' Connor, S. 2015b.

Genome-guided investigation of plant natural product biosynthesis. *The Plant Journal,* 82**,** 680-692.

Kitamura, S., Shikazono, N. & Tanaka, A. 2004. Transparent testa 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *The Plant Journal,* 37**,** 104-114.

Klement, Z., Rudolph, K. & Sands, D. 1990. *Methods in phytobacteriology*, Akademiai Kiado.

Kohonen, T. 1990. The self-organizing map. *Proceedings of the IEEE,* 78**,** 1464-1480.

Kölling, R. & Hollenberg, C. P. 1994. The ABC-transporter Ste6 accumulates in the plasma membrane in a ubiquitinated form in endocytosis mutants. *The EMBO journal,* 13**,** 3261.

Komarova, N. Y., Meier, S., Meier, A., Grotemeyer, M. S. & Rentsch, D. 2012. Determinants for *Arabidopsis* peptide transporter targeting to the tonoplast or plasma membrane. *Traffic,* 13**,** 1090-1105.

Komarova, N. Y., Thor, K., Gubler, A., Meier, S., Dietrich, D., Weichert, A., Grotemeyer, M. S., Tegeder, M. & Rentsch, D. 2008. AtPTR1 and AtPTR5 transport dipeptides *in planta*. *Plant Physiology,* 148**,** 856-869.

Krouk, G., Lacombe, B., Bielach, A., Perrine-Walker, F., Malinska, K., Mounier, E., Hoyerova, K., Tillard, P., Leon, S. & Ljung, K. 2010. Nitrate-regulated auxin transport by NRT1. 1 defines a mechanism for nutrient sensing in plants. *Developmental cell,* 18**,** 927-937.

Kunjapur, A. M., Tarasova, Y. & Prather, K. L. 2014. Synthesis and accumulation of aromatic aldehydes in an engineered strain of *Escherichia coli*. *Journal of the American Chemical Society,* 136**,** 11644-11654.

Kutchan, T. M. & Dittrich, H. 1995. Characterization and mechanism of the berberine bridge enzyme, a covalently flavinylated oxidase of benzophenanthridine alkaloid biosynthesis in plants. *Journal of Biological Chemistry,* 270**,** 24475-24481.

Lane, N. & Martin, W. 2010. The energetics of genome complexity. *Nature,* 467**,** 929-934.

Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & Lopez, R. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

Leferink, N. G., Fraaije, M. W., Joosten, H.-J., Schaap, P. J., Mattevi, A. & Van Berkel, W. J. 2009. Identification of a gatekeeper residue that prevents dehydrogenases from acting as oxidases. *Journal of Biological Chemistry,* 284**,** 4392-4397.

Leferink, N. G., Heuts, D. P., Fraaije, M. W. & Van Berkel, W. J. 2008a. The growing VAO flavoprotein family. *Archives of biochemistry and biophysics,* 474**,** 292-301.

Leferink, N. G., Van Den Berg, W. A. & Van Berkel, W. J. 2008b. l-Galactono-γ-lactone dehydrogenase from *Arabidopsis thaliana*, a flavoprotein involved in vitamin C biosynthesis. *FEBS Journal,* 275**,** 713-726.

Leonard, E., Runguphan, W., O' Connor, S. & Prather, K. J. 2009. Opportunities in metabolic engineering to facilitate scalable alkaloid production. *Nature chemical biology,* 5**,** 292-300.

Léran, S., Garg, B., Boursiac, Y., Corratgé-Faillie, C., Brachet, C., Tillard, P., Gojon, A. & Lacombe, B. 2015. AtNPF5. 5, a nitrate transporter affecting nitrogen accumulation in *Arabidopsis* embryo. *Scientific reports,* 5.

Leran, S., Varala, K., Boyer, J.-C., Chiurazzi, M., Crawford, N., Daniel-Vedele, F., David, L., Dickstein, R., Fernandez, E., Forde, B., Gassmann, W., Geiger, D., Gojon, A., Gong, J.-M., Halkier, B. A., Harris, J. M., Hedrich, R., Limami, A. M., Rentsch, D., Seo, M., Tsay, Y.-F., Zhang, M., Coruzzi, G. & Lacombe, B. 2014. A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends in Plant Science,* 19**,** 5-9.

Levac, D., Murata, J., Kim, W. S. & De Luca, V. 2008. Application of carborundum abrasion for investigating the leaf epidermis: molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-O-methyltransferase. *The Plant Journal,* 53**,** 225-236.

Lewis, R. S., Lopez, H. O., Bowen, S. W., Andres, K. R., Steede, W. T. & Dewey, R. E. 2015. Transgenic and Mutation-Based Suppression of a Berberine Bridge Enzyme-Like (BBL) Gene Family Reduces Alkaloid Content in Field-Grown Tobacco. *PloS One,* 10.

Li, B., Bjarnholt, N., Hansen, S. H. & Janfelt, C. 2011. Characterization of barley leaf tissue using direct and indirect desorption electrospray ionization imaging mass spectrometry. *Journal of Mass Spectrometry,* 46**,** 1241-1246.

Li, B., Knudsen, C., Hansen, N. K., Jørgensen, K., Kannangara, R., Bak, S., Takos, A., Rook, F., Hansen, S. H. & Møller, B. L. 2013. Visualizing metabolite distribution and enzymatic conversion in plant tissues by desorption electrospray ionization mass spectrometry imaging. *The Plant Journal,* 74**,** 1059-1071.

Li, J.-Y., Fu, Y.-L., Pike, S. M., Bao, J., Tian, W., Zhang, Y., Chen, C.-Z., Zhang, Y., Li, H.-M. & Huang, J. 2010. The *Arabidopsis* nitrate transporter NRT1. 8 functions in nitrate removal from the xylem sap and mediates cadmium tolerance. *The Plant Cell,* 22**,** 1633-1646.

Li, L., He, Z., Pandey, G. K., Tsuchiya, T. & Luan, S. 2002. Functional cloning and characterization of a plant efflux carrier for multidrug and heavy metal detoxification. *Journal of Biological Chemistry,* 277**,** 5360-5368.

Li, R., Li, J., Li, S., Qin, G., Novák, O., Pěnčík, A., Ljung, K., Aoyama, T., Liu, J. & Murphy, A. 2014. ADP1 affects plant architecture by regulating local auxin biosynthesis. *PLoS genetics,* 10**,** e1003954.

Li, R., Reed, D. W., Liu, E., Nowak, J., Pelcher, L. E., Page, J. E. & Covello, P. S. 2006. Functional genomic analysis of alkaloid biosynthesis in *Hyoscyamus niger* reveals a cytochrome P450 involved in littorine rearrangement. *Chemistry & Biology,* 13**,** 513-520.

Li, Z.-S., Zhao, Y. & Rea, P. A. 1995. Magnesium adenosine 5 [prime]-triphosphate-energized transport of glutathione-S-conjugates by plant vacuolar membrane vesicles. *Plant Physiology,* 107**,** 1257-1268.

Lin, S.-H., Kuo, H.-F., Canivenc, G., Lin, C.-S., Lepetit, M., Hsu, P.-K., Tillard, P., Lin, H.-L., Wang, Y.-Y. & Tsai, C.-B. 2008. Mutation of the *Arabidopsis* NRT1. 5 nitrate transporter causes defective root-to-shoot nitrate transport. *The Plant Cell,* 20**,** 2514-2528.

Linster, C. L., Van Schaftingen, E. & Hanson, A. D. 2013. Metabolite damage and its repair or pre-emption. *Nature chemical biology,* 9**,** 72-80.

Lipp, J. 1991. Possible mechanisms of morphine analgesia. *Clinical neuropharmacology,* 14**,** 131-147.

Liscombe, D. K. & O'connor, S. E. 2011. A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. *Phytochemistry,* 72**,** 1969-1977.

Liscombe, D. K., Usera, A. R. & O'connor, S. E. 2010. Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis. *Proceedings of the National Academy of Sciences,* 107**,** 18793-18798.

Liu, E. & Page, J. E. 2008. Optimized cDNA libraries for virus-induced gene silencing (VIGS) using tobacco rattle virus. *Plant methods,* 4**,** 1-13.

Liu, J., Magalhaes, J. V., Shaff, J. & Kochian, L. V. 2009. Aluminum-activated citrate and malate transporters from the MATE and ALMT families function independently to confer *Arabidopsis* aluminum tolerance. *The Plant Journal,* 57**,** 389-399.

Liu, K.-H., Huang, C.-Y. & Tsay, Y.-F. 1999. CHL1 is a dual-affinity nitrate transporter of *Arabidopsis* involved in multiple phases of nitrate uptake. *The Plant Cell,* 11**,** 865-874.

Liu, Y., Schiff, M., Marathe, R. & Dinesh-Kumar, S. 2002. Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. *The Plant Journal,* 30**,** 415-429.

Loe, D. W., Deeley, R. G. & Cole, S. P. 1998. Characterization of vincristine transport by the Mr 190,000 multidrug resistance protein (MRP): evidence for cotransport with reduced glutathione. *Cancer Research,* 58**,** 5130-5136.

Lopachin, R. M. & Gavin, T. 2014. Molecular mechanisms of aldehyde toxicity: a chemical perspective. *Chemical research in toxicology,* 27**,** 1081-1091.

Lu, Y.-P., Li, Z.-S., Drozdowicz, Y. M., Hörtensteiner, S., Martinoia, E. & Rea, P. A. 1998. AtMRP2, an *Arabidopsis* ATP binding cassette transporter able to transport glutathione S-conjugates and chlorophyll catabolites: functional comparisons with AtMRP1. *The Plant Cell,* 10**,** 267-282.

Macgurn, J. A., Hsu, P.-C. & Emr, S. D. 2012. Ubiquitin and membrane protein turnover: from cradle to grave. *Annual review of biochemistry,* 81**,** 231-259.

Magalhaes, J. V., Liu, J., Guimaraes, C. T., Lana, U. G., Alves, V. M., Wang, Y.-H., Schaffert, R. E., Hoekenga, O. A., Pineros, M. A. & Shaff, J. E. 2007. A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nature genetics,* 39**,** 1156-1161.

Maitai, C. 1975. D-norpseudoephedrine as an anti-asthmatic: comparison with l-ephedrine, d-pseudoephedrine. *East African medical journal,* 52**,** 330-332.

Malito, E., Coda, A., Bilyeu, K. D., Fraaije, M. W. & Mattevi, A. 2004. Structures of Michaelis and product complexes of plant cytokinin dehydrogenase: implications for flavoenzyme catalysis. *Journal of molecular biology,* 341**,** 1237-1249.

Maresh, J. J., Giddings, L.-A., Friedrich, A., Loris, E. A., Panjikar, S., Trout, B. L., Stöckigt, J., Peters, B. & O'connor, S. E. 2008. Strictosidine synthase: mechanism of a Pictet-Spengler catalyzing enzyme. *Journal of the American Chemical Society,* 130**,** 710-723.

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J.-M., Debeaujon, I. & Klein, M. 2007. The *Arabidopsis* MATE transporter TT12 acts as a vacuolar flavonoid/H^+-antiporter active in proanthocyanidin-accumulating cells of the seed coat. *The Plant Cell,* 19**,** 2023-2038.

Markel, M. L., Miles, W. M., Luck, J. C., Klein, L. S. & Prystowsky, E. N. 1993. Differential effects of isoproterenol on sustained ventricular tachycardia before and during procainamide and quinidine antiarrhythmic drug therapy. *Circulation,* 87**,** 783-792.

Maron, L. G., Piñeros, M. A., Guimarães, C. T., Magalhaes, J. V., Pleiman, J. K., Mao, C., Shaff, J., Belicuas, S. N. & Kochian, L. V. 2010. Two functionally distinct members of the MATE (multi-drug and toxic compound extrusion) family of transporters potentially underlie two major aluminum tolerance QTLs in maize. *The Plant Journal,* 61**,** 728-740.

Marrs, K. A., Alfenito, M. R., Lloyd, A. M. & Walbot, V. 1995. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature,* 375**,** 397-400.

Martinoia, E., Grill, E., Tommasini, R., Kreuz, K. & Amrhein, N. 1993. ATP-dependent glutathione S-conjugate 'export' pump in the vacuolar membrane of plants. *Nature,* 364**,** 247-249.

Mathews, H., Clendennen, S. K., Caldwell, C. G., Liu, X. L., Connors, K., Matheis, N., Schuster, D. K., Menasco, D., Wagoner, W. & Lightner, J. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *The Plant Cell,* 15**,** 1689-1703.

Matros, A. & Mock, H.-P. 2013. Mass spectrometry based imaging techniques for spatially resolved analysis of molecules. *Frontiers in plant science,* 4.

Mattevi, A. 2006. To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes. *Trends in biochemical sciences,* 31**,** 276-283.

Mcfarlane, H. E., Shin, J. J., Bird, D. A. & Samuels, A. L. 2010. *Arabidopsis* ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. *The Plant Cell,* 22**,** 3066-3075.

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., Van Der Krol, S., Lugan, R., Ilc, T., Verpoorte, R., Oksman-Caldentey, K.-M., Martinoia, E., Bouwmeester, H., Goossens, A., Memelink, J. & Werck-Reichhart, D. 2014. The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications,* 5.

Minami, H., Kim, J.-S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H. & Sato, F. 2008. Microbial production of plant benzylisoquinoline alkaloids. *Proceedings of the National Academy of Sciences,* 105**,** 7393-7398.

Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A. & Shlomi, T. 2012. Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proceedings of the National Academy of Sciences,* 109**,** 339-344.

Morgan, J. A. & Shanks, J. V. 2000. Determination of metabolic rate-limitations by precursor feeding in *Catharanthus roseus* hairy root cultures. *Journal of Biotechnology,* 79**,** 137-145.

Morita, M., Shitan, N., Sawada, K., Van Montagu, M. C., Inzé, D., Rischer, H., Goossens, A., Oksman-Caldentey, K.-M., Moriyama, Y. & Yazaki, K. 2009. Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. *Proceedings of the National Academy of Sciences,* 106**,** 2447-2452.

Morita, Y., Kodama, K., Shiota, S., Mine, T., Kataoka, A., Mizushima, T. & Tsuchiya, T. 1998. NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrobial agents and chemotherapy,* 42**,** 1778-1782.

Motohashi, H. & Inui, K.-I. 2013. Organic cation transporter OCTs (SLC22) and MATEs (SLC47) in the human kidney. *The AAPS journal,* 15**,** 581-588.

MüLler, T., Oradu, S., Ifa, D. R., Cooks, R. G. & KräUtler, B. 2011. Direct plant tissue analysis and imprint imaging by desorption electrospray ionization mass spectrometry. *Analytical chemistry,* 83**,** 5754-5761.

Munkert, J., Pollier, J., Miettinen, K., Van Moerkercke, A., Payne, R., Müller-Uri, F., Burlat, V., O'connor, S. E., Memelink, J. & Kreis, W. 2015. Iridoid synthase activity is common among the plant progesterone 5β-reductase family. *Molecular Plant,* 8**,** 136-152.

Murata, J. & De Luca, V. 2005. Localization of tabersonine 16-hydroxylase and 16-OH tabersonine-16-O-methyltransferase to leaf epidermal cells defines them as a major site of precursor biosynthesis in the vindoline pathway in *Catharanthus roseus*. *The Plant Journal,* 44**,** 581-594.

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.

Nawrath, C., Heck, S., Parinthawong, N. & Métraux, J.-P. 2002. EDS5, an essential component of salicylic acid–dependent signaling for disease resistance in *Arabidopsis*, is a member of the MATE transporter family. *The Plant Cell,* 14**,** 275-286.

Newstead, S., Drew, D., Cameron, A. D., Postis, V. L., Xia, X., Fowler, P. W., Ingram, J. C., Carpenter, E. P., Sansom, M. S. & Mcpherson, M. J. 2011. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide–proton symporters, PepT1 and PepT2. *The EMBO journal,* 30**,** 417-426.

Newstead, S., Kim, H., Von Heijne, G., Iwata, S. & Drew, D. 2007. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences,* 104**,** 13936-13941.

Niehaus, T. D., Richardson, L. G., Gidda, S. K., Elbadawi-Sidhu, M., Meissen, J. K., Mullen, R. T., Fiehn, O. & Hanson, A. D. 2014. Plants utilize a highly conserved system for repair of NADH and NADPH hydrates. *Plant Physiology,* 165**,** 52-61.

Noinaj, N., Bosserman, M. A., Schickli, M. A., Piszczek, G., Kharel, M. K., Pahari, P., Buchanan, S. K. & Rohr, J. 2011. The crystal structure and mechanism of an unusual oxidoreductase, GilR, involved in gilvocarcin V biosynthesis. *Journal of Biological Chemistry,* 286**,** 23533-23543.

Nour-Eldin, H. H., Andersen, T. G., Burow, M., Madsen, S. R., Jørgensen, M. E., Olsen, C. E., Dreyer, I., Hedrich, R., Geiger, D. & Halkier, B. A. 2012. NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature,* 488**,** 531-534.

Nour-Eldin, H. H., Geu-Flores, F. & Halkier, B. A. 2010. USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. *Plant Secondary Metabolism Engineering*. Springer.

Nour-Eldin, H. H. & Halkier, B. A. 2013. The emerging field of transport engineering of plant specialized metabolites. *Current opinion in biotechnology,* 24**,** 263-270.

Nour-Eldin, H. H., Nørholm, M. H. & Halkier, B. A. 2006. Screening for plant transporter function by expressing a normalized *Arabidopsis* full-length cDNA library in *Xenopus* oocytes. *Plant methods,* 2**,** 17.

Nützmann, H.-W. & Osbourn, A. 2014. Gene clustering in plant specialized metabolism. *Current opinion in biotechnology,* 26**,** 91-99.

O'connor, S. E. & Maresh, J. J. 2006. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Natural product reports,* 23**,** 532-547.

Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M. & Moriyama, Y. 2006. The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends in Pharmacological Sciences,* 27**,** 587-593.

Onofri, C., De Meijer, E. P. & Mandolino, G. 2015. Sequence heterogeneity of cannabidiolic-and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry*.

Osbourn, A. 2010. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends in genetics,* 26**,** 449-457.

Pao, S. S., Paulsen, I. T. & Saier, M. H. 1998. Major facilitator superfamily. *Microbiology and Molecular Biology Reviews,* 62**,** 1-34.

Papon, N., Vansiri, A., Gantet, P., Chénieux, J.-C., Rideau, M. & Crèche, J. 2004. Histidine-containing phosphotransfer domain extinction by RNA interference turns off a cytokinin signalling circuitry in *Catharanthus roseus* suspension cells. *FEBS letters,* 558**,** 85-88.

Parinthawong, N., Cottier, S., Buchala, A., Nawrath, C. & Métraux, J.-P. 2015. Localization and expression of EDS5H a homologue of the SA transporter EDS5. *BMC Plant Biology,* 15**,** 135.

Park, S.-U., Yu, M. & Facchini, P. J. 2002. Antisense RNA-mediated suppression of benzophenanthridine alkaloid biosynthesis in transgenic cell cultures of California poppy. *Plant Physiology,* 128**,** 696-706.

Park, S.-U., Yu, M. & Facchini, P. J. 2003. Modulation of berberine bridge enzyme levels in transgenic root cultures of California poppy alters the accumulation of benzophenanthridine alkaloids. *Plant molecular biology,* 51**,** 153-164.

Parker, J. L. & Newstead, S. 2014. Molecular basis of nitrate uptake by the plant nitrate transporter NRT1. 1. *Nature,* 507**,** 68-72.

Peebles, C. A., Hong, S. B., Gibson, S. I., Shanks, J. V. & San, K. Y. 2006. Effects of terpenoid precursor feeding on *Catharanthus roseus* hairy roots over-expressing the alpha or the alpha and beta subunits of anthranilate synthase. *Biotechnology and bioengineering,* 93**,** 534-540.

Pellizzaro, A., Clochard, T., Cukier, C., Bourdin, C., Juchaux, M., Montrichard, F., Thany, S., Raymond, V., Planchet, E. & Limami, A. M. 2014. The Nitrate Transporter MtNPF6. 8 (MtNRT1. 3) Transports Abscisic Acid and Mediates Nitrate Regulation of Primary Root Growth in *Medicago truncatula*. *Plant physiology,* 166**,** 2152-2165.

Pennings, E. J., Groen, B. W., Duine, J. A. & Verpoorte, R. 1989. Tryptophan decarboxylase from *Catharanthus roseus* is a pyridoxoquinoprotein. *FEBS letters,* 255**,** 97-100.

Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods,* 8**,** 785-786.

Pil, J. S., Jungmin, P., Mi-Jeong, P., Youn-Sung, K., Sang-Gyu, K., Jae-Hoon, J. & Chung-Mo, P. 2012. A Golgi-localized MATE transporter mediates iron homoeostasis under osmotic stress in *Arabidopsis*. *Biochemical Journal,* 442**,** 551-561.

Pollier, J., Bossche, R. V., Rischer, H. & Goossens, A. 2014. Selection and validation of reference genes for transcript normalization in gene expression studies in *Catharanthus roseus*. *Plant Physiology and Biochemistry,* 83**,** 20-25.

Qu, F., Ye, X. & Morris, T. J. 2008. *Arabidopsis* DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. *Proceedings of the National Academy of Sciences,* 105**,** 14732-14737.

Qu, Y., Easson, M. L., Froese, J., Simionescu, R., Hudlicky, T. & De Luca, V. 2015. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences,* 112**,** 6224-6229.

Ranocha, P., Denancé, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., Köhler, L., Pouzet, C., Renou, J. P. & Sundberg, B. 2010. Walls are thin 1 (WAT1), an *Arabidopsis* homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *The Plant Journal,* 63**,** 469-483.

Ratcliff, F., Martin-Hernandez, A. M. & Baulcombe, D. C. 2001. Technical advance: tobacco rattle virus as a vector for analysis of gene function by silencing. *The Plant Journal,* 25**,** 237-245.

Rea, P. A. 2007. Plant ATP-binding cassette transporters. *Annu. Rev. Plant Biol.,* 58**,** 347-375.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I. & Saier, M. H. 2012. The major facilitator superfamily (MFS) revisited. *FEBS Journal,* 279**,** 2022-2035.

Remy, E. & Duque, P. 2014. Beyond cellular detoxification: a plethora of physiological roles for MDR transporter homologs in plants. *Frontiers in physiology,* 5.

Rentsch, D., Laloi, M., Rouhara, I., Schmelzer, E., Delrot, S. & Frommer, W. B. 1995. NTR1 encodes a high affinity oligopeptide transporter in *Arabidopsis*. *FEBS letters,* 370**,** 264-268.

Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S. & Kirby, J. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature,* 440**,** 940-943.

Roytrakul, S. & Verpoorte, R. 2007. Role of vacuolar transporter proteins in plant secondary metabolism: *Catharanthus roseus* cell culture. *Phytochemistry Reviews,* 6**,** 383-396.

Ruetsch, Y. A., Boni, T. & Borgeat, A. 2001. From cocaine to ropivacaine: the history of local anesthetic drugs. *Current topics in medicinal chemistry,* 1**,** 175-182.

Runguphan, W., Maresh, J. J. & O'connor, S. E. 2009. Silencing of tryptamine biosynthesis for production of nonnatural alkaloids in plant culture. *Proceedings of the National Academy of Sciences,* 106**,** 13673-13678.

Saeed, A., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T. & Thiagarajan, M. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques,* 34**,** 374.

Saier, M. H., Yen, M. R., Noto, K., Tamang, D. G. & Elkan, C. 2009. The transporter classification database: recent advances. *Nucleic acids research,* 37**,** D274-D278.

Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant biotechnology journal,* 7**,** 682-693.

Salim, V., Wiens, B., Masada-Atsumi, S., Yu, F. & De Luca, V. 2014. 7-Deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in *Catharanthus roseus* iridoid biosynthesis. *Phytochemistry,* 101**,** 23-31.

Salim, V., Yu, F., Altarejos, J. & Luca, V. 2013. Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *The Plant Journal,* 76**,** 754-765.

Sappl, P. G., Carroll, A. J., Clifton, R., Lister, R., Whelan, J., Harvey Millar, A. & Singh, K. B. 2009. The *Arabidopsis* glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress. *The Plant Journal,* 58**,** 53-68.

Sawaki, Y., Kihara-Doi, T., Kobayashi, Y., Nishikubo, N., Kawazu, T., Kobayashi, Y., Koyama, H. & Sato, S. 2013. Characterization of Al-responsive citrate excretion and citrate-transporting MATEs in *Eucalyptus camaldulensis*. *Planta,* 237**,** 979-989.

Schaedler, T. A., Thornton, J. D., Kruse, I., Schwarzländer, M., Meyer, A. J., Van Veen, H. W. & Balk, J. 2014. A conserved mitochondrial ATP-binding cassette transporter exports glutathione polysulfide for cytosolic metal cofactor assembly. *Journal of Biological Chemistry,* 289**,** 23264-23274.

Schmidt, G. W., Jirschitzka, J., Porta, T., Reichelt, M., Luck, K., Torre, J. C. P., Dolke, F., Varesio, E., Hopfgartner, G., Gershenzon, J. & D'auria, J. C. 2015. The Last Step in Cocaine Biosynthesis Is Catalyzed by a BAHD Acyltransferase. *Plant Physiology,* 167**,** 89-101.

Schmülling, T., Werner, T., Riefler, M., Krupková, E. & Y Manns, I. B. 2003. Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, *Arabidopsis* and other species. *Journal of plant research,* 116**,** 241-252.

Segonzac, C., Boyer, J.-C., Ipotesi, E., Szponarski, W., Tillard, P., Touraine, B., Sommerer, N., Rossignol, M. & Gibrat, R. 2007. Nitrate efflux at the root plasma membrane: identification of an *Arabidopsis* excretion transporter. *The Plant Cell,* 19**,** 3760-3777.

Serrano, M., Wang, B., Aryal, B., Garcion, C., Abou-Mansour, E., Heck, S., Geisler, M., Mauch, F., Nawrath, C. & Métraux, J.-P. 2013. Export of salicylic acid from the chloroplast requires the multidrug and toxin extrusion-like transporter EDS5. *Plant Physiology,* 162**,** 1815-1821.

Shi, J., Wang, H., Schellin, K., Li, B., Faller, M., Stoop, J. M., Meeley, R. B., Ertl, D. S., Ranch, J. P. & Glassman, K. 2007. Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nature biotechnology,* 25**,** 930-937.

Shitan, N., Bazin, I., Dan, K., Obata, K., Kigawa, K., Ueda, K., Sato, F., Forestier, C. & Yazaki, K. 2003. Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proceedings of the National Academy of Sciences,* 100**,** 751-756.

Shitan, N., Minami, S., Morita, M., Hayashida, M., Ito, S., Takanashi, K., Omote, H., Moriyama, Y., Sugiyama, A., Goossens, A., Moriyasu, M. & Yazaki, K. 2014. Involvement of the Leaf-Specific Multidrug and Toxic Compound Extrusion (MATE) Transporter Nt-JAT2 in Vacuolar Sequestration of Nicotine in *Nicotiana tabacum*. *Plos One,* 9.

Shoji, T., Inai, K., Yazaki, Y., Sato, Y., Takase, H., Shitan, N., Yazaki, K., Goto, Y., Toyooka, K., Matsuoka, K. & Hashimoto, T. 2009. Multidrug and Toxic Compound Extrusion-Type Transporters Implicated in Vacuolar Sequestration of Nicotine in Tobacco Roots. *Plant Physiology,* 149**,** 708-718.

Shoyama, Y., Tamada, T., Kurihara, K., Takeuchi, A., Taura, F., Arai, S., Blaber, M., Shoyama, Y., Morimoto, S. & Kuroki, R. 2012. Structure and Function ofΔ 1-Tetrahydrocannabinolic Acid (THCA) Synthase, the Enzyme Controlling the Psychoactivity of *Cannabis sativa*. *Journal of molecular biology,* 423**,** 96-105.

Sikorskaite, S., Rajamäki, M.-L., Baniulis, D., Stanys, V. & Valkonen, J. P. 2013. Protocol: Optimised methodology for isolation of nuclei from leaves of species in the Solanaceae and Rosaceae families. *Plant methods,* 9**,** 31.

Simkin, A. J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S. & St-Pierre, B. 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.

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y. & Taura, F. 2004. The Gene Controlling Marijuana Psychoactivity: Molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from *Cannabis sativa* (L.). *Journal of Biological Chemistry,* 279**,** 39767-39774.

Sirikantaramas, S., Taura, F., Tanaka, Y., Ishikawa, Y., Morimoto, S. & Shoyama, Y. 2005. Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. *Plant and Cell Physiology,* 46**,** 1578-1582.

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

Smith, D. E., Clémençon, B. & Hediger, M. A. 2013. Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Molecular aspects of medicine,* 34**,** 323-336.

Sottomayor, M., Lopez-Serrano, M., Dicosmo, F. & Barceló, A. R. 1998. Purification and characterization of α-3′, 4′-anhydrovinblastine synthase (peroxidase-like) from *Catharanthus roseus* (L.) G. Don. *FEBS letters,* 428**,** 299-303.

Souza, P. S., Vasconcelos, F. C., De Souza Reis, F. R., De Moraes, G. N. & Maia, R. C. 2011. P-glycoprotein and survivin simultaneously regulate vincristine-induced apoptosis in chronic myeloid leukemia cells. *International journal of oncology,* 39**,** 925-933.

Springael, J.-Y., Galan, J.-M., Haguenauer-Tsapis, R. & André, B. 1999. NH_4^+-induced down-regulation of the *Saccharomyces cerevisiae* Gap1p permease involves its ubiquitination with lysine-63-linked chains. *Journal of cell science,* 112**,** 1375-1383.

Spyropoulou, E. A., Haring, M. A. & Schuurink, R. C. 2014. RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. *BMC Genomics,* 15**,** 402.

St-Pierre, B. & De Luca, V. 1995. A cytochrome P-450 monooxygenase catalyzes the first step in the conversion of tabersonine to vindoline in *Catharanthus roseus*. *Plant Physiology,* 109**,** 131-139.

St-Pierre, B., Vazquez-Flota, F. A. & De Luca, V. 1999. Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. *The Plant Cell,* 11**,** 887-900.

St-Pierre, B., Laflamme, P., Alarco, A. M. & Luca, E. 1998. The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. *The Plant Journal,* 14**,** 703-713.

Staud, F., Cerveny, L., Ahmadimoghaddam, D. & Ceckova, M. 2013. Multidrug and toxin extrusion proteins (MATE/SLC47); role in pharmacokinetics. *The international journal of biochemistry & cell biology,* 45**,** 2007-2011.

Stavrinides, A., Tatsis, E. C., Foureau, E., Caputi, L., Kellner, F., Courdavault, V. & O'connor, S. E. 2015. Unlocking the diversity of alkaloids in *Catharanthus roseus*: nuclear localization suggests metabolic channeling in secondary metabolism. *Chemistry & Biology,* 22**,** 336-341.

Steffens, P., Nagakura, N. & Zenk, M. H. 1985. Purification and characterization of the berberine bridge enzyme from *Berberis beaniana* cell cultures. *Phytochemistry,* 24**,** 2577-2583.

Stevens, L. H., Blom, T. J. & Verpoorte, R. 1993. Subcellular localization of tryptophan decarboxylase, strictosidine synthase and strictosidine glucosidase in suspension cultured cells of *Catharanthus roseus* and *Tabernaemontana divaricata*. *Plant cell reports,* 12**,** 573-576.

Sun, J., Bankston, J. R., Payandeh, J., Hinds, T. R., Zagotta, W. N. & Zheng, N. 2014. Crystal structure of the plant dual-affinity nitrate transporter NRT1. 1. *Nature,* 507**,** 73-77.

Sun, X., Gilroy, E. M., Chini, A., Nurmberg, P. L., Hein, I., Lacomme, C., Birch, P. R., Hussain, A., Yun, B. W. & Loake, G. J. 2011. ADS1 encodes a MATE-transporter that negatively regulates plant disease resistance. *New Phytologist,* 192**,** 471-482.

Sweetlove, L. J. & Fernie, A. R. 2013. The spatial organization of metabolism within the plant cell. *Annual Review of Plant Biology,* 64**,** 723-746.

Szydlowski, N., Bürkle, L., Pourcel, L., Moulin, M., Stolz, J. & Fitzpatrick, T. B. 2013. Recycling of pyridoxine (vitamin B_6) by PUP1 in *Arabidopsis*. *The Plant Journal,* 75**,** 40-52.

Tahallah, N., Van Den Heuvel, R. H., Van Den Berg, W. A., Maier, C. S., Van Berkel, W. J. & Heck, A. J. 2002. Cofactor-dependent assembly of the flavoenzyme vanillyl-alcohol oxidase. *Journal of Biological Chemistry,* 277**,** 36425-36432.

Takahashi, H., Kopriva, S., Giordano, M., Saito, K. & Hell, R. 2011. Sulfur assimilation in photosynthetic organisms: molecular functions and regulations of transporters and assimilatory enzymes. *Annual review of plant biology,* 62**,** 157-184.

Takanashi, K., Yokosho, K., Saeki, K., Sugiyama, A., Sato, S., Tabata, S., Ma, J. F. & Yazaki, K. 2013. LjMATE1: a citrate transporter responsible for iron supply to the nodule infection zone of *Lotus japonicus*. *Plant and Cell Physiology,* 54**,** 585-594.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution,* 28**,** 2731-2739.

Taura, F., Morimoto, S. & Shoyama, Y. 1996. Purification and characterization of cannabidiolic-acid synthase from *Cannabis sativa* L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. *Journal of Biological Chemistry,* 271**,** 17411-17416.

Taura, F., Morimoto, S., Shoyama, Y. & Mechoulam, R. 1995. First direct evidence for the mechanism of. DELTA. 1-tetrahydrocannabinolic acid biosynthesis. *Journal of the American Chemical Society,* 117**,** 9766-9767.

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y. & Morimoto, S. 2007. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. *FEBS letters,* 581**,** 2929-2934.

Team, R. C. 2015. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. *Document freely available on the internet at: http://www. r-project. org*.

Thunig, J., Hansen, S. H. & Janfelt, C. 2011. Analysis of secondary plant metabolites by indirect desorption electrospray ionization imaging mass spectrometry. *Analytical chemistry,* 83**,** 3256-3259.

Tiwari, M., Sharma, D., Singh, M., Tripathi, R. D. & Trivedi, P. K. 2014. Expression of OsMATE1 and OsMATE2 alters development, stress responses and pathogen susceptibility in *Arabidopsis*. *Scientific reports,* 4.

Todd, A. T., Liu, E., Polvi, S. L., Pammett, R. T. & Page, J. E. 2010. A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. *The Plant Journal,* 62**,** 589-600.

Treimer, J. F. & Zenk, M. H. 1979. Purification and properties of strictosidine synthase, the key enzyme in indole alkaloid formation. *European Journal of Biochemistry,* 101**,** 225-233.

Truernit, E. & Haseloff, J. 2008. A simple way to identify non-viable cells within living plant tissue using confocal microscopy. *Plant methods,* 4**,** 15.

Tsay, Y.-F., Schroeder, J. I., Feldmann, K. A. & Crawford, N. M. 1993. The herbicide sensitivity gene CHL1 of *Arabidopsis* encodes a nitrate-inducible nitrate transporter. *Cell,* 72**,** 705-713.

Tuteja, N. 2007. Abscisic acid and abiotic stress signaling. *Plant signaling & behavior,* 2**,** 135-138.

Van Der Fits, L. & Memelink, J. 2000. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science,* 289**,** 295-297.

Van Moerkercke, A., Fabris, M., Pollier, J., Baart, G. J., Rombauts, S., Hasnain, G., Rischer, H., Memelink, J., Oksman-Caldentey, K.-M. & Goossens, A. 2013. CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data. *Plant and Cell Physiology,* 54**,** 673-685.

Van Moerkercke, A., Steensma, P., Schweizer, F., Pollier, J., Gariboldi, I., Payne, R., Bossche, R. V., Miettinen, K., Espoz, J., Purnama, P. C., Kellner, F., Seppanen-Laakso, T., O'connor, S. E., Rischer, H., Memelink, J. & Goossens, A. 2015. The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences* 112**,** 8130-8135.

Vazquez-Flota, F. A. & De Luca, V. 1998. Developmental and light regulation of desacetoxyvindoline 4-hydroxylase in *Catharanthus roseus* (L.) G. Don. Evidence of a multilevel regulatory mechanism. *Plant Physiology,* 117**,** 1351-1361.

Vesanto, J. & Alhoniemi, E. 2000. Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on,* 11**,** 586-600.

Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., Van Den Heuvel, S. & Vidal, M. 1999. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in enzymology,* 328**,** 575-592.

Walsh, C. T. & Wencewicz, T. A. 2013. Flavoenzymes: versatile catalysts in biosynthetic pathways. *Natural product reports,* 30**,** 175-200.

Walter, S., Kahla, A., Arunachalam, C., Perochon, A., Khan, M. R., Scofield, S. R. & Doohan, F. M. 2015. A wheat ABC transporter contributes to both grain formation and mycotoxin tolerance. *Journal of experimental botany***,** erv048.

Wang, R., Liu, D. & Crawford, N. M. 1998. The *Arabidopsis* CHL1 protein plays a major role in high-affinity nitrate uptake. *Proceedings of the National Academy of Sciences,* 95**,** 15134-15139.

Wang, R., Liu, X., Liang, S., Ge, Q., Li, Y., Shao, J., Qi, Y., An, L. & Yu, F. 2015. A subgroup of MATE transporter genes regulates hypocotyl cell elongation in *Arabidopsis*. *Journal of experimental botany***,** erv344.

Wang, Y.-Y. & Tsay, Y.-F. 2011. *Arabidopsis* nitrate transporter NRT1. 9 is important in phloem nitrate transport. *The Plant Cell,* 23**,** 1945-1957.

Watson, J. M. & Wang, M.-B. 2012. *Antiviral resistance in plants: methods and protocols*, Humana Press.

Wege, S., Scholz, A., Gleissberg, S. & Becker, A. 2007. Highly efficient virus-induced gene silencing (VIGS) in California poppy (*Eschscholzia californica*): an evaluation of VIGS as a strategy to obtain functional data from non-model plants. *Annals of Botany,* 100**,** 641-649.

Wehrens, R. & Buydens, L. M. 2007. Self-and super-organizing maps in R: the Kohonen package. *Journal of Statistical Software,* 21**,** 1-19.

Wei, H., Persson, S., Mehta, T., Srinivasasainagendra, V., Chen, L., Page, G. P., Somerville, C. & Loraine, A. 2006. Transcriptional coordination of the metabolic network in *Arabidopsis*. *Plant Physiology,* 142**,** 762-774.

Weichert, A., Brinkmann, C., Komarova, N. Y., Dietrich, D., Thor, K., Meier, S., Grotemeyer, M. S. & Rentsch, D. 2012. AtPTR4 and AtPTR6 are differentially expressed, tonoplast-localized members of the peptide transporter/nitrate transporter 1 (PTR/NRT1) family. *Planta,* 235**,** 311-323.

Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H. & Schmülling, T. 2003. Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *The Plant Cell,* 15**,** 2532-2550.

Whitmer, S., Canel, C., Hallard, D., Gonçalves, C. & Verpoorte, R. 1998. Influence of precursor availability on alkaloid accumulation by transgenic cell line of *Catharanthus roseus*. *Plant Physiology,* 116**,** 853-857.

Winde, I. & Wittstock, U. 2011. Insect herbivore counteradaptations to the plant glucosinolate–myrosinase system. *Phytochemistry,* 72**,** 1566-1575.

Winkel-Shirley, B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology,* 126**,** 485-493.

Winkler, A., Hartner, F., Kutchan, T. M., Glieder, A. & Macheroux, P. 2006. Biochemical evidence that berberine bridge enzyme belongs to a novel family of flavoproteins containing a bi-covalently attached FAD cofactor. *Journal of Biological Chemistry,* 281**,** 21276-21285.

Winkler, A., Kutchan, T. M. & Macheroux, P. 2007. 6-S-cysteinylation of bi-covalently attached FAD in berberine bridge enzyme tunes the redox potential for optimal activity. *Journal of Biological Chemistry,* 282**,** 24437-24443.

Winkler, A., Łyskowski, A., Riedl, S., Puhl, M., Kutchan, T. M., Macheroux, P. & Gruber, K. 2008. A concerted mechanism for berberine bridge enzyme. *Nature chemical biology,* 4**,** 739-741.

Winkler, A., Motz, K., Riedl, S., Puhl, M., Macheroux, P. & Gruber, K. 2009a. Structural and mechanistic studies reveal the functional role of bicovalent flavinylation in berberine bridge enzyme. *Journal of Biological Chemistry,* 284**,** 19993-20001.

Winkler, A., Puhl, M., Weber, H., Kutchan, T. M., Gruber, K. & Macheroux, P. 2009b. Berberine bridge enzyme catalyzes the six electron oxidation of (S)-reticuline to dehydroscoulerine. *Phytochemistry,* 70**,** 1092-1097.

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., Li, Y., Dowle, A. A., Cartwright, J. & Bates, R. 2015. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science,* 349**,** 309-312.

Woldemariam, M. G., Dinh, S. T., Oh, Y., Gaquerel, E., Baldwin, I. T. & Galis, I. 2013. NaMYC2 transcription factor regulates a subset of plant defense responses in *Nicotiana attenuata*. *BMC Plant Biology,* 13**,** 73.

Wuts, P. G. & Greene, T. W. 2006. *Greene's protective groups in organic synthesis*, John Wiley & Sons.

Xiao, M., Zhang, Y., Chen, X., Lee, E.-J., Barber, C. J., Chakrabarty, R., Desgagné-Penix, I., Haslam, T. M., Kim, Y.-B. & Liu, E. 2013. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of Biotechnology,* 166**,** 122-134.

Yamasaki, K., Motomura, Y., Yagi, Y., Nomura, H., Kikuchi, S., Nakai, M. & Shiina, T. 2013. Chloroplast envelope localization of EDS5, an essential factor for salicylic acid biosynthesis in *Arabidopsis thaliana*. *Plant signaling & behavior,* 8**,** e23603.

Yan, N. 2013. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends in biochemical sciences,* 38**,** 151-159.

Yan, N. 2015. Structural Biology of the Major Facilitator Superfamily (MFS) Transporters. *Annual Review of Biophysics,* 44.

Yang, X. Y., Yang, J. L., Zhou, Y., Pineros, M. A., Kochian, L. V., Li, G. X. & Zheng, S. J. 2011. A *de novo* synthesis citrate transporter, *Vigna umbellata* multidrug and toxic compound extrusion, implicates in Al-activated citrate efflux in rice bean (*Vigna umbellata*) root apex. *Plant, cell & environment,* 34**,** 2138-2148.

Yendrek, C. R., Lee, Y. C., Morris, V., Liang, Y., Pislariu, C. I., Burkart, G., Meckfessel, M. H., Salehin, M., Kessler, H. & Wessler, H. 2010. A putative transporter is essential for integrating nutrient and hormone signaling with lateral root growth and nodule development in *Medicago truncatula*. *The Plant Journal,* 62**,** 100-112.

Yokosho, K., Yamaji, N. & Ma, J. F. 2010. Isolation and characterisation of two MATE genes in rye. *Functional Plant Biology,* 37**,** 296-303.

Yokosho, K., Yamaji, N., Ueno, D., Mitani, N. & Ma, J. F. 2009. OsFRDL1 is a citrate transporter required for efficient translocation of iron in rice. *Plant Physiology,* 149**,** 297-305.

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., Watanabe-Takahashi, A., Inoue, E. & Saito, K. 2008. Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in *Arabidopsis*. *The Plant Cell,* 20**,** 2160-2176.

Yu, F. & De Luca, V. 2013. ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences,* 110**,** 15830-15835.

Zafred, D., Steiner, B., Teufelberger, A. R., Hromic, A., Karplus, P. A., Schofield, C. J., Wallner, S. & Macheroux, P. 2015. Rationally engineered flavin-dependent oxidase reveals steric control of dioxygen reduction. *FEBS Journal*.

Zhang, H., Zhu, H., Pan, Y., Yu, Y., Luan, S. & Li, L. 2014. A DTX/MATE-type transporter facilitates abscisic acid efflux and modulates ABA sensitivity and drought tolerance in *Arabidopsis*. *Molecular Plant,* 7**,** 1522-1532.

Zhao, J. 2015. Flavonoid transport mechanisms: how to go, and with whom. *Trends in Plant Science,* 20**,** 576-585.

Zhao, J. & Dixon, R. A. 2009. MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. *The Plant Cell,* 21**,** 2323-2340.

Zhao, J., Huhman, D., Shadle, G., He, X.-Z., Sumner, L. W., Tang, Y. & Dixon, R. A. 2011. MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in *Medicago truncatula*. *The Plant Cell,* 23**,** 1536-1555.

Zhou, G., Delhaize, E., Zhou, M. & Ryan, P. R. 2013. The barley MATE gene, HvAACT1, increases citrate efflux and Al^{3+} tolerance when expressed in wheat and barley. *Annals of botany,* 112**,** 603-612.

Zhou, G., Pereira, J. F., Delhaize, E., Zhou, M., Magalhaes, J. V. & Ryan, P. R. 2014. Enhancing the aluminium tolerance of barley by expressing the citrate transporter genes SbMATE and FRD3. *Journal of experimental botany***,** eru121.

9 Appendix

9.1 Gene List of 5 neighbouring nodes highlighted in the Self Organising Map

Highlighted in green are known genes in the MIA pathway, highlighted in yellow are the candidates selected in this study.

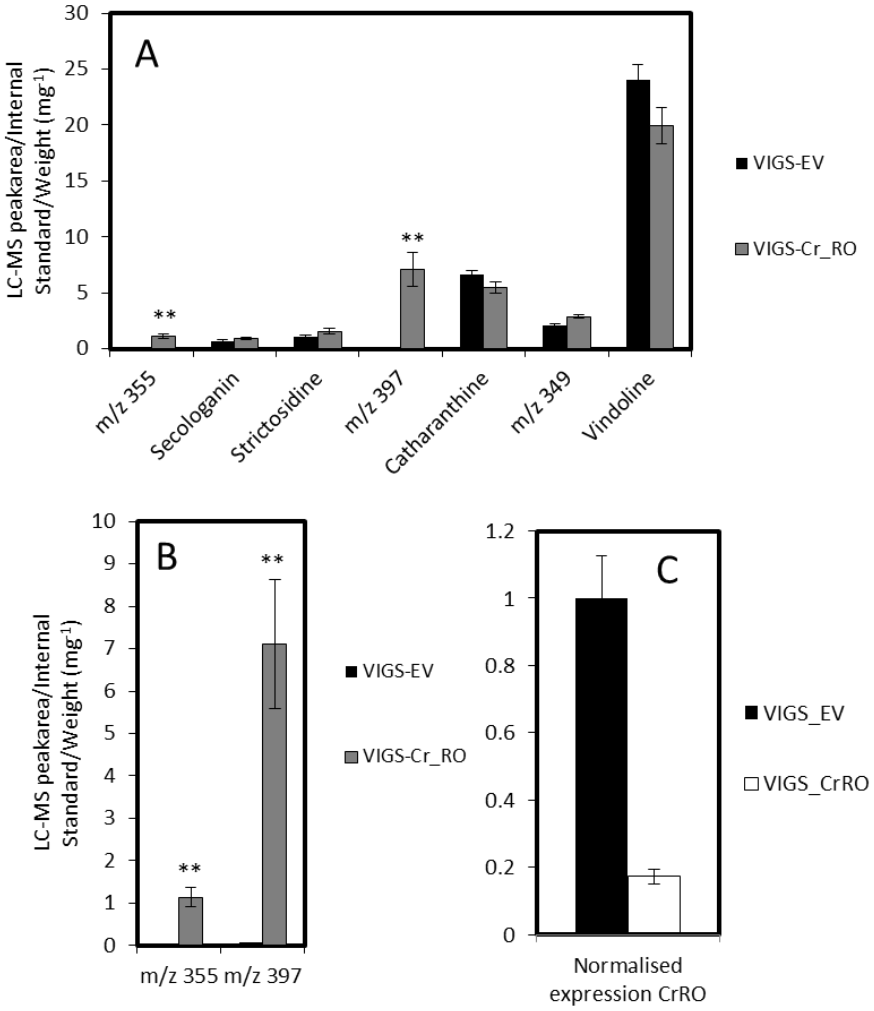
locus_ref	contig	function.	neuron_number	imp_contig	transporter	mature_leaf	immature_leaf
50822	cra_locus_50822_iso_2_len_357_ver_3	Gene of unknown function	9	n	n	2.034662966	1.814853901
17962	cra_locus_17962_iso_6_len_1181_ver_3	Kinesin	9	n	n	2.203182359	2.266705598
107691	cra_locus_107691_iso_1_len_279_ver_3	Gene of unknown function	9	n	n	2.426431158	2.521729197
21267	cra_locus_21267_iso_1_len_2314_ver_3	Pentatricopeptide repeat-containing protein	9	n	n	2.637359845	3.054322344
4824	cra_locus_4824_iso_1_len_889_ver_3	Major latex protein	9	n	n	2.718021806	3.647510233
10286	cra_locus_10286_iso_1_len_1071_ver_3	Aquaporin, MIP family, NIP subfamily	9	n	n	3.026038682	3.481195593
8587	cra_locus_8587_iso_2_len_2384_ver_3	Wall-associated kinase	9	n	n	3.20487002	2.714302675
14470	cra_locus_14470_iso_1_len_734_ver_3	Conserved gene of unknown function	9	n	n	3.252663919	2.845875387
5848	cra_locus_5848_iso_1_len_1276_ver_3	Esterase/lipase/thioesterase family protein	9	n	n	3.601506363	4.190282986
42757	cra_locus_42757_iso_2_len_1176_ver_3	Transportin	9	n	y	3.923149181	4.391747102
3419	cra_locus_3419_iso_2_len_582_ver_3	Peroxidase 2a	9	n	n	4.027065562	2.809498822
5706	cra_locus_5706_iso_3_len_1913_ver_3	Carboxypeptidase B2	9	n	n	4.078481749	4.233121118
775	cra_locus_775_iso_7_len_1694_ver_3	Nodulin MtN21	9	n	n	4.417393479	4.009329255
5555	cra_locus_5555_iso_2_len_3030_ver_3	Calcium-dependent protein kinase	9	n	n	4.60847218	4.52990838
13640	cra_locus_13640_iso_3_len_1323_ver_3	RNA-binding protein	9	n	n	4.69540927	5.220306735
1614	cra_locus_1614_iso_6_len_1191_ver_3	Ubiquitin-conjugating enzyme E2 C	9	n	n	4.766022634	4.467449046
10099	cra_locus_10099_iso_3_len_1907_ver_3	CYP72A57	9	n	n	4.816200325	4.026348593
679	cra_locus_679_iso_6_len_1771_ver_3	4-coumarate--CoA ligase	9	n	n	4.897395338	4.267011359
5186	cra_locus_5186_iso_5_len_2851_ver_3	NBS-LRR type resistance protein	9	n	n	5.187597533	5.709494672
3934	cra_locus_3934_iso_2_len_1257_ver_3	GTP binding protein	9	n	n	5.496903297	3.735338038
18109	cra_locus_18109_iso_3_len_1194_ver_3	Conserved gene of unknown function	9	n	n	5.500231617	5.910444862
21742	cra_locus_21742_iso_1_len_1132_ver_3	Annexin 11	10	n	n	2.24652362	2.116697443
13905	cra_locus_13905_iso_2_len_1077_ver_3	Protein 2	10	n	n	2.586596629	3.389621875
3676	cra_locus_3676_iso_2_len_1708_ver_3	Anthocyanidine rhamnosyl-transferase	10	n	n	2.906634735	2.862625952
49421	cra_locus_49421_iso_1_len_435_ver_3	Gene of unknown function	10	n	n	3.060841167	3.583579255
16069	cra_locus_16069_iso_2_len_1960_ver_3	Enolase	10	n	n	3.062043978	4.299808787
4194	cra_locus_4194_iso_2_len_1818_ver_3	laa-amino acid hydrolase 11	10	n	n	3.84317989	3.987693772
8319	cra_locus_8319_iso_4_len_2366_ver_3	Nitrate transporter	10	n	y	4.049238678	4.271881302
4773	cra_locus_4773_iso_3_len_2437_ver_3	Amino acid binding protein	10	n	n	4.112791862	4.883757651
26	cra_locus_26_iso_7_len_3347_ver_3	Carbonyl reductase	10	n	n	4.115299644	4.113858827
2335	cra_locus_2335_iso_6_len_2732_ver_3	Enoyl-CoA hydratase, mitochondrial	10	n	n	4.118982589	4.396426681
6504	cra_locus_6504_iso_8_len_3035_ver_3	Anthranilate synthase alpha subunit	10	y	n	4.301463271	4.929506975
4873	cra_locus_4873_iso_8_len_1642_ver_3	Myricetin O-methyltransferase	10	n	n	4.474598281	5.37834168
6672	cra_locus_6672_iso_4_len_1013_ver_3	Conserved gene of unknown function	10	n	n	4.568092925	5.044350401
9393	cra_locus_9393_iso_1_len_941_ver_3	Cytochrome P450	10	n	n	4.79469658	5.048746236
16283	cra_locus_16283_iso_1_len_809_ver_3	NADP specific isocitrate dehydrogenase	10	n	n	4.907645409	4.980057299
4121	cra_locus_4121_iso_1_len_1649_ver_3	Oxidoreductase	10	n	n	4.948077495	5.312498185
995	cra_locus_995_iso_1_len_1587_ver_3	Tobamovirus multiplication 1	10	n	n	4.948199005	5.191886337
539	cra_locus_539_iso_3_len_1184_ver_3	Epoxide hydrolase	10	n	n	5.184930868	5.152751297
6518	cra_locus_6518_iso_7_len_4746_ver_3	D-3-phosphoglycerate dehydrogenase	10	n	n	5.193298651	5.457804391
489	cra_locus_489_iso_9_len_2848_ver_3	Glucose-6-phosphate 1-dehydrogenase, chlorc	10	n	n	5.233105775	4.77119892
2046	cra_locus_2046_iso_8_len_2228_ver_3	Strictosidine-O-beta-D-glucosidase	10	y	n	5.348855601	6.190936784
3591	cra_locus_3591_iso_3_len_1774_ver_3	Indole-3-glycerol phosphate lyase IGL1	10	n	n	5.504560014	6.131328133
1726	cra_locus_1726_iso_4_len_1538_ver_3	DUF300 family protein	10	n	n	5.530354832	5.085981197
1763	cra_locus_1763_iso_10_len_5663_ver_3	Multidrug resistance-associated protein	10	n	y	5.563542554	5.634375454
10740	cra_locus_10740_iso_5_len_2322_ver_3	Mate efflux family protein	10	n	y	5.694651794	5.760449843
6036	cra_locus_6036_iso_3_len_2256_ver_3	6-phosphogluconate dehydrogenase	10	n	n	5.882158809	6.523692106
1952	cra_locus_1952_iso_7_len_1704_ver_3	Mate efflux family protein	10	n	y	5.913109615	6.625469226
2478	cra_locus_2478_iso_5_len_2072_ver_3	Auxin-induced protein 5NG4	10	n	n	5.971180248	5.796345966
23587	cra_locus_23587_iso_1_len_537_ver_3	Gene of unknown function	10	n	n	6.086340073	5.622716624
2461	cra_locus_2461_iso_2_len_435_ver_3	Gene of unknown function	10	n	n	6.223785495	6.545098846
7644	cra_locus_7644_iso_3_len_1034_ver_3	Ferredoxin-3, chloroplast	10	n	n	6.331178511	6.590615492
2176	cra_locus_2176_iso_5_len_1213_ver_3_stricto	Strictosidine synthase	10	y	n	6.344224881	6.975595797
580	cra_locus_580_iso_6_len_2366_ver_3	10-hydroxygeraniol oxidoreductase	10	y	n	6.414739349	6.88381629
518	cra_locus_518_iso_8_len_1140_ver_3	Tryptophan synthase beta chain 2, chloroplasti	10	n	n	6.486075572	7.417042353
11931	cra_locus_11931_iso_1_len_427_ver_3	Cytochrome P450	10	n	n	6.767350232	7.721666583
17400	cra_locus_17400_iso_1_len_272_ver_3	Secologanin synthase	10	y	n	6.841520319	7.959828103
5091	cra_locus_5091_iso_1_len_297_ver_3	Elicitor-inducible protein EIG-J7	10	n	n	6.968447925	7.546346595
2442	cra_locus_2442_iso_7_len_1193_ver_3	Cytochrome P450	10	n	n	7.194224978	8.004742945
2479	cra_locus_2479_iso_3_len_905_ver_3	Lemir	10	n	n	7.519879957	7.559354145
1734	cra_locus_1734_iso_3_len_2168_ver_3	NADPH--cytochrome P450 reductase	10	y	n	7.613178008	7.572306108
1389	cra_locus_1389_iso_2_len_318_ver_3	Secologanin synthase	10	y	n	7.705722229	9.412709558
13307	cra_locus_13307_iso_1_len_1915_ver_3	Nitrate transporter	11	n	y	2.031324613	3.109462531
10830	cra_locus_10830_iso_3_len_1683_ver_3	Calmodulin binding protein	11	n	n	2.081902005	3.024748736
5465	cra_locus_5465_iso_1_len_2209_ver_3	Endonuclease/exonuclease/phosphatase	11	n	n	2.282994239	3.392234981
14550	cra_locus_14550_iso_1_len_1120_ver_3	Sugar transporter	11	n	y	2.300779745	3.597936335
59453	cra_locus_59453_iso_1_len_453_ver_3	Superoxide dismutase [Mn], mitochondrial	11	n	n	2.323257647	3.313908991
88388	cra_locus_88388_iso_1_len_535_ver_3	Conserved gene of unknown function	11	n	n	2.446695589	2.743458957
10356	cra_locus_10356_iso_1_len_1833_ver_3	Leucine-rich repeat/extensin	11	n	n	2.526975221	4.579759427
5964	cra_locus_5964_iso_3_len_2202_ver_3	Conserved gene of unknown function	11	n	n	2.590197189	3.020075437
15951	cra_locus_15951_iso_1_len_2216_ver_3	Kinase	11	n	n	2.656880733	3.686085868
9952	cra_locus_9952_iso_1_len_1019_ver_3	Phosphoglycerate mutase	11	n	n	2.762859038	3.188987949
18723	cra_locus_18723_iso_3_len_1416_ver_3	Ku70-binding family protein	11	n	n	2.8254763	3.011887389
1926	cra_locus_1926_iso_3_len_1400_ver_3	Optic atrophy 3 protein	11	n	n	2.831798215	2.776815683
14334	cra_locus_14334_iso_1_len_870_ver_3	PnFL-1	11	n	n	2.871268032	2.933211919
7314	cra_locus_7314_iso_1_len_2533_ver_3	RNA binding protein	11	n	n	3.185184671	3.319417064
19017	cra_locus_19017_iso_4_len_2003_ver_3	Ankyrin repeat family protein	11	n	n	3.191734766	2.831490174

8604	cra_locus_8604_iso_2_len_1310_ver_3	Pentatricopeptide repeat-containing protein	11	n	n	3.213549478	3.446137103
10234	cra_locus_10234_iso_1_len_2183_ver_3	Fructan 1-exohydrolase I	11	n	n	3.236391588	3.730139952
10319	cra_locus_10319_iso_6_len_1803_ver_3	Galactosyltransferase family protein	11	n	n	3.270359995	3.676572076
4905	cra_locus_4905_iso_2_len_894_ver_3	CHCH domain containing protein	11	n	n	3.270373452	4.189642728
5360	cra_locus_5360_iso_1_len_1168_ver_3	Mitochondrion protein	11	n	n	3.321652514	3.443619912
7501	cra_locus_7501_iso_6_len_2278_ver_3	Integrin-linked protein kinase	11	n	n	3.528858829	4.043921887
7263	cra_locus_7263_iso_2_len_2518_ver_3	Oxysterol-binding family protein	11	n	n	3.621161028	4.301251076
6827	cra_locus_6827_iso_1_len_1237_ver_3	Phytol kinase 1, chloroplast	11	n	n	3.626976397	3.569904204
5578	cra_locus_5578_iso_3_len_1371_ver_3	Pyridoxal kinase	11	n	n	3.75411786	3.76544988
8590	cra_locus_8590_iso_5_len_1451_ver_3	Gene of unknown function	11	n	n	3.820332161	4.454208801
2710	cra_locus_2710_iso_4_len_2087_ver_3	Bem46	11	n	n	4.050153389	4.632460174
10281	cra_locus_10281_iso_2_len_1155_ver_3	DECOY	11	n	n	4.301916824	4.717468057
11582	cra_locus_11582_iso_6_len_2036_ver_3	Isoform 2 of Probable beta-1,3-galactosyltranst	11	n	n	4.310456462	4.362455862
4536	cra_locus_4536_iso_4_len_1291_ver_3	Methionyl-tRNA synthetase	11	n	n	4.435248537	4.837201526
17470	cra_locus_17470_iso_1_len_1823_ver_3	Arogenate dehydratase	11	n	n	4.449066085	4.383441646
3481	cra_locus_3481_iso_5_len_2753_ver_3	Hydroxyproline-rich glycoprotein family protein	11	n	n	4.606223083	5.614097305
8473	cra_locus_8473_iso_2_len_1124_ver_3	DH putative beta-hydroxyacyl-ACP dehydratas	11	n	n	4.737567794	6.263073854
838	cra_locus_838_iso_3_len_2096_ver_3	Nuclear transport factor	11	n	y	4.843491476	4.856622232
9664	cra_locus_9664_iso_1_len_1014_ver_3	Glutaredoxin-C4	11	n	n	4.844164672	5.249650126
216	cra_locus_216_iso_7_len_1193_ver_3	Protein pof4	11	n	n	4.857259467	5.471047607
4232	cra_locus_4232_iso_1_len_681_ver_3	Mitochondrial import inner membrane translocá	11	n	n	4.893590344	5.873576994
22285	cra_locus_22285_iso_1_len_722_ver_3	Ubiquinol-cytochrome c reductase complex 6.1	11	n	n	4.897400178	4.728268837
4768	cra_locus_4768_iso_4_len_2787_ver_3	NADH-ubiquinone oxidoreductase 39 kD subur	11	n	n	5.010650062	5.164029984
4649	cra_locus_4649_iso_2_len_2141_ver_3	NAD-malate dehydrogenase	11	n	n	5.072873261	5.253202677
10929	cra_locus_10929_iso_4_len_1221_ver_3	Gene of unknown function	11	n	n	5.144287071	4.768824792
2085	cra_locus_2085_iso_1_len_927_ver_3	NADH dehydrogenase	11	n	n	5.267404706	5.498445554
6855	cra_locus_6855_iso_4_len_1139_ver_3	Cytochrome P450	11	n	n	5.268733801	6.515376575
4159	cra_locus_4159_iso_4_len_1417_ver_3	ATP synthase 24 kDa subunit, mitochondrial	11	n	n	5.284810223	5.524016647
6335	cra_locus_6335_iso_1_len_1143_ver_3	Clathrin adaptor complex small chain family pr	11	n	n	5.319668509	5.146712238
6962	cra_locus_6962_iso_3_len_1689_ver_3	Vinorine synthase	11	n	n	5.367671596	6.210924554
14850	cra_locus_14850_iso_3_len_734_ver_3	Conserved gene of unknown function	11	n	n	5.3953062	5.752748591
15222	cra_locus_15222_iso_1_len_911_ver_3	Multidrug resistance-associated protein 2, 6 (M	11	n	y	5.397132103	5.234087398
635	cra_locus_635_iso_4_len_2478_ver_3	Monodehydroascorbate reductase	11	n	n	5.482499826	5.608809243
7753	cra_locus_7753_iso_3_len_1253_ver_3	Conserved gene of unknown function	11	n	n	5.545109661	6.001302346
8320	cra_locus_8320_iso_6_len_1679_ver_3	Negatively light-regulated protein	11	n	n	5.558013487	5.610304377
30243	cra_locus_30243_iso_1_len_701_ver_3	Conserved gene of unknown function	11	n	n	5.568953262	5.541291802
7376	cra_locus_7376_iso_2_len_889_ver_3	Mago nashi 1	11	n	n	5.639266899	6.112224716
9318	cra_locus_9318_iso_4_len_587_ver_3	Peptidyl-prolyl cis-trans isomerase Pin1	11	n	n	5.65316036	5.36447088
5549	cra_locus_5549_iso_8_len_1090_ver_3	ATP-dependent transporter YFL028C	11	n	y	5.657188378	6.029225376
4216	cra_locus_4216_iso_8_len_3698_ver_3	Acetyl-CoA synthetase	11	n	n	5.816351386	5.779425206
8318	cra_locus_8318_iso_3_len_1309_ver_3	Nucleoside diphosphate kinase 1	11	n	n	5.917756349	5.534404052
351	cra_locus_351_iso_4_len_2032_ver_3	ATP-citrate synthase	11	n	n	5.924532249	6.156627177
8420	cra_locus_8420_iso_4_len_1918_ver_3	Aromatic-L-amino-acid decarboxylase	11	y	n	5.94035941	7.463540721
1737	cra_locus_1737_iso_4_len_735_ver_3	Chalcone reductase	11	n	n	6.0579744	6.910133013
570	cra_locus_570_iso_3_len_957_ver_3	Constitutive plastid-lipid associated protein	11	n	n	6.543693838	6.58369959
9037	cra_locus_9037_iso_5_len_1732_ver_3	AAA ATPase; 26S proteasome subunit P45	11	n	n	6.587220954	6.616788477
12599	cra_locus_12599_iso_2_len_782_ver_3	CXE carboxylesterase	11	n	n	6.604585845	8.278133471
2110	cra_locus_2110_iso_3_len_915_ver_3	Superoxide dismutase [Mn], mitochondrial	11	n	n	6.641991023	6.969288659
24360	cra_locus_24360_iso_1_len_527_ver_3	ATP synthase epsilon chain, mitochondrial	11	n	n	6.727321476	7.811779266
8396	cra_locus_8396_iso_1_len_873_ver_3	ATP synthase D chain, mitochondrial	11	n	n	6.842061147	6.897905945
19251	cra_locus_19251_iso_1_len_303_ver_3	Isocitrate dehydrogenase (NADP+)	11	n	n	6.875939753	5.969528122
2533	cra_locus_2533_iso_1_len_794_ver_3	Fiber protein Fb15	11	n	n	6.965853532	7.402994975
2052	cra_locus_2052_iso_4_len_1346_ver_3	ATP synthase f1, gamma subunit	11	n	n	7.019657441	7.166906212
3455	cra_locus_3455_iso_5_len_1570_ver_3	Phosphoglycerate kinase	11	n	n	7.144984469	7.429766586
4203	cra_locus_4203_iso_1_len_1034_ver_3	Adenosine kinase isoform 2S	11	n	n	7.201976853	7.70016225
11589	cra_locus_11589_iso_1_len_737_ver_3	Mitochondrial ATP synthase 6kDa subunit	11	n	n	7.510360671	7.805859928
53790	cra_locus_53790_iso_1_len_338_ver_3	Gene of unknown function	11	n	n	7.641011277	8.820612982
8885	cra_locus_8885_iso_1_len_264_ver_3	Gene of unknown function	11	n	n	8.469894114	9.687225194
17498	cra_locus_17498_iso_1_len_289_ver_3	Gene of unknown function	11	n	n	8.795227966	8.518464324
37409	cra_locus_37409_iso_1_len_301_ver_3	Gene of unknown function	35	n	n	2.378752977	1.69723115
28574	cra_locus_28574_iso_1_len_370_ver_3	Gene of unknown function	35	n	n	2.401136196	3.293988398
8512	cra_locus_8512_iso_1_len_1271_ver_3	UDP-glucose:glucosyltransferase	35	n	n	2.647197063	3.343279893
6087	cra_locus_6087_iso_1_len_2353_ver_3	Gene of unknown function	35	n	n	2.749795967	3.330185488
35985	cra_locus_35985_iso_1_len_295_ver_3	Gene of unknown function	35	n	n	2.843891412	2.253583924
13958	cra_locus_13958_iso_1_len_444_ver_3	Gene of unknown function	35	n	n	3.001908505	3.287902203
381	cra_locus_381_iso_2_len_2638_ver_3	1-deoxy-D-xylulose 5-phosphate synthase 2	35	y	n	3.103344886	3.866739868
2215	cra_locus_2215_iso_4_len_3190_ver_3	Conserved gene of unknown function	35	n	n	3.169173001	3.649741905
7219	cra_locus_7219_iso_6_len_1914_ver_3	Glutamate receptor	35	n	n	3.643313634	3.963557354
722	cra_locus_722_iso_7_len_1164_ver_3	Strictosidine synthase	35	n	n	3.723733291	4.095603786
1974	cra_locus_1974_iso_3_len_1196_ver_3	Sinapyl alcohol dehydrogenase	35	n	n	4.114949983	4.335390355
1452	cra_locus_1452_iso_2_len_1656_ver_3	Geranylgeranyl pyrophosphate synthase, chlor	35	n	n	4.312338439	4.995235757
5739	cra_locus_5739_iso_2_len_1991_ver_3	UDP-glucosyltransferase family 1 protein	35	n	n	4.600638469	5.557364136
758	cra_locus_758_iso_5_len_2527_ver_3	Methylmalonate-semialdehyde dehydrogenase	35	n	n	4.807555855	4.533606952
891	cra_locus_891_iso_7_len_1738_ver_3	4-diphosphocytidyl-methylerythritol 2-phosphat	35	n	n	4.916094423	5.064818198
3880	cra_locus_3880_iso_5_len_1914_ver_3	TPR domain containing protein	35	n	n	5.043659469	5.143507706
113	cra_locus_113_iso_6_len_1847_ver_3	Geraniol synthase	35	n	n	5.166337853	5.275841423
618	cra_locus_618_iso_7_len_2657_ver_3	1-deoxyxylulose 5-phosphate synthase	35	y	n	5.18328006	5.629143843
3243	cra_locus_3243_iso_4_len_2217_ver_3	Cytochrome P450	35	n	n	5.494940364	5.96045379
2665	cra_locus_2665_iso_3_len_1362_ver_3	Progesterone 5beta-reductase 2	35	n	n	5.74253715	5.694637867

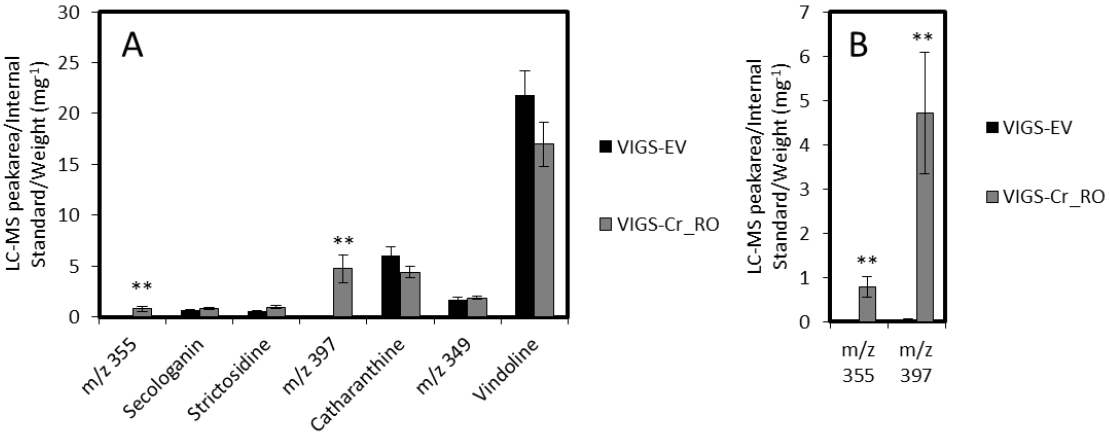
5743	cra_locus_5743_iso_7_len_900_ver_3	Alcohol dehydrogenase	35	n	n	5.881583869	5.603332584
6991	cra_locus_6991_iso_4_len_2997_ver_3	Plastidic ATP/ADP-transporter	35	n	y	5.924727008	6.308564747
5034	cra_locus_5034_iso_8_len_2450_ver_3	Beta-glucosidase	35	n	n	6.101036498	6.318400002
5068	cra_locus_5068_iso_2_len_683_ver_3	Protein S	35	n	n	6.206881518	6.013906613
4910	cra_locus_4910_iso_10_len_1394_ver_3	4-diphosphocytidyl-2-C-methyl-D-erythritol kina	35	n	n	6.334155272	6.01564921
7966	cra_locus_7966_iso_3_len_1015_ver_3	2-C-methyl-D-erythritol 2,4-cyclodiphosphate s	35	y	n	6.470305833	6.35897816
605	cra_locus_605_iso_4_len_1576_ver_3	Cytochrome P450	35	n	n	6.507582184	6.553742939
4962	cra_locus_4962_iso_5_len_2409_ver_3	1-deoxy-D-xylulose-5-phosphate reductoisome	35	y	n	6.529311854	6.024248749
2708	cra_locus_2708_iso_3_len_941_ver_3	Geraniol 10-hydroxylase	35	y	n	6.773350237	7.02103548
1424	cra_locus_1424_iso_6_len_1981_ver_3	GCPE protein	35	y	n	7.331194637	7.166564822
11857	cra_locus_11857_iso_9_len_2980_ver_3	Lecithine cholesterol acyltransferase	36	n	n	2.232009862	2.336674719
5665	cra_locus_5665_iso_4_len_1529_ver_3	Nudix hydrolase 20, chloroplastic	36	n	n	2.293979572	2.743381399
58436	cra_locus_58436_iso_1_len_586_ver_3	Gene of unknown function	36	n	n	2.308156975	2.603601422
14634	cra_locus_14634_iso_1_len_1053_ver_3	Basic helix-loop-helix (bHLH) family protein	36	n	n	2.636139678	2.519879014
45073	cra_locus_45073_iso_1_len_491_ver_3	Conserved gene of unknown function	36	n	n	2.81609995	2.348141752
40681	cra_locus_40681_iso_3_len_1541_ver_3	CBS	36	n	n	2.921895193	3.329755086
10710	cra_locus_10710_iso_5_len_2155_ver_3	Conserved gene of unknown function	36	n	n	3.015174611	3.543446391
1113	cra_locus_1113_iso_2_len_1414_ver_3	ABC-transporter	36	n	y	3.083655897	3.210119191
54804	cra_locus_54804_iso_4_len_802_ver_3	Serine hydroxymethyltransferase	36	n	n	3.134925684	1.079859356
10262	cra_locus_10262_iso_1_len_3510_ver_3	Glycosyltransferase	36	n	n	3.142209191	2.840199345
9522	cra_locus_9522_iso_6_len_3467_ver_3	Ion channel DMI1	36	n	n	3.145126355	3.327831052
4190	cra_locus_4190_iso_3_len_1505_ver_3	Protein phosphatase 2c	36	n	n	3.148192235	3.136214359
56545	cra_locus_56545_iso_1_len_469_ver_3	Gene of unknown function	36	n	n	3.245570199	3.122339751
6540	cra_locus_6540_iso_3_len_1219_ver_3	Conserved gene of unknown function	36	n	n	3.39514508	3.367440962
12417	cra_locus_12417_iso_4_len_2573_ver_3	Acyl-activating enzyme 18	36	n	n	3.412605411	3.85939315
8977	cra_locus_8977_iso_2_len_2623_ver_3	Gd2b	36	n	n	3.439955518	3.643752312
24210	cra_locus_24210_iso_1_len_1044_ver_3	Auxin-induced protein 5NG4	36	n	n	3.469912018	3.362750405
70054	cra_locus_70054_iso_1_len_292_ver_3	14-3-3 family protein	36	n	n	3.497689001	3.172965881
11693	cra_locus_11693_iso_4_len_3176_ver_3	ATP-binding cassette transporter	36	n	y	3.590613995	3.50021568
36563	cra_locus_36563_iso_1_len_945_ver_3	Isopentenyl pyrophosphate: dimethyllallyl pyro	36	y	n	3.67746314	4.563823179
6670	cra_locus_6670_iso_2_len_1585_ver_3	Diphosphonucleoside phosphohydrolase	36	n	n	3.754983679	3.784650548
5608	cra_locus_5608_iso_2_len_1934_ver_3	Reticuline oxidase	36	n	n	3.931683057	4.348919321
3978	cra_locus_3978_iso_3_len_1482_ver_3	Mitochondrial carnitine/acylcarnitine carrier pro	36	n	n	4.061534288	4.068653615
18184	cra_locus_18184_iso_1_len_922_ver_3	Chaperone protein DNAj	36	n	n	4.089964172	2.718401084
3640	cra_locus_3640_iso_3_len_1136_ver_3	ZIP transporter	36	n	y	4.097939411	4.281379264
9771	cra_locus_9771_iso_4_len_1760_ver_3	Hydrolase, alpha/beta fold family protein	36	n	n	4.141784266	3.821700012
4144	cra_locus_4144_iso_2_len_2457_ver_3	3-ketoacyl-CoA synthase	36	n	n	4.19105724	4.268569135
14176	cra_locus_14176_iso_4_len_1284_ver_3	Pentatricopeptide repeat-containing protein	36	n	n	4.209656109	5.287745619
5019	cra_locus_5019_iso_5_len_2123_ver_3	Minovincinine 19-hydroxy-O-acetyltransferase	36	y	n	4.369627007	4.747532414
7015	cra_locus_7015_iso_3_len_2535_ver_3	LEC14B protein	36	n	n	4.469664607	4.737594833
818	cra_locus_818_iso_4_len_1965_ver_3	Serine/threonine-protein kinase PBS1	36	n	n	4.49922085	4.061966241
56804	cra_locus_56804_iso_2_len_545_ver_3	Adenosylhomocysteinase	36	n	n	4.516765305	3.284311277
1810	cra_locus_1810_iso_6_len_5044_ver_3	Multidrug resistance protein ABC transporter fa	36	n	y	4.612140149	4.731188673
32432	cra_locus_32432_iso_1_len_396_ver_3	Gene of unknown function	36	n	n	4.658074349	4.235359447
55853	cra_locus_55853_iso_1_len_338_ver_3	S-adenosyl-L-homocystein hydrolase	36	n	n	4.667239303	2.79540332
2310	cra_locus_2310_iso_1_len_2099_ver_3	Tubby; Cyclin-like F-box	36	n	n	4.683460664	4.495874209
8230	cra_locus_8230_iso_5_len_1862_ver_3	Phosphoserine phosphatase	36	n	n	4.841515287	4.991997942
6747	cra_locus_6747_iso_5_len_1697_ver_3	UDP-glucosyltransferase family 1 protein	36	n	n	4.860342097	5.456030722
9086	cra_locus_9086_iso_5_len_1891_ver_3	Peptide transporter	36	n	y	4.907356992	5.09761501
8928	cra_locus_8928_iso_4_len_3066_ver_3	Potassium transporter family protein	36	n	y	4.931205546	4.850979399
7823	cra_locus_7823_iso_2_len_1499_ver_3	Catalytic/ hydrolase	36	n	n	4.978040001	4.567673211
9450	cra_locus_9450_iso_4_len_2169_ver_3	External rotenone-insensitive NADPH dehydrog	36	n	n	4.998263213	4.809043947
1227	cra_locus_1227_iso_6_len_1706_ver_3	Cytochrome P450 71D7	36	n	n	5.000572456	5.252321037
8349	cra_locus_8349_iso_3_len_1344_ver_3	Cytochrome b5	36	n	n	5.040826693	5.289797586
4729	cra_locus_4729_iso_5_len_1706_ver_3	Mitogen-activated protein kinase homolog MMI	36	n	n	5.103275218	5.032136117
126	cra_locus_126_iso_1_len_1595_ver_3	Senescence-associated protein	36	n	n	5.151639669	5.679105654
205	cra_locus_205_iso_8_len_1925_ver_3	Aldose 1-epimerase	36	n	n	5.203909798	4.937739855
6405	cra_locus_6405_iso_6_len_4766_ver_3	ATP binding protein	36	n	n	5.229149511	5.114321211
1468	cra_locus_1468_iso_7_len_2400_ver_3	Jasmonic acid-amino acid-conjugating enzyme	36	n	n	5.29495152	4.888986721
8301	cra_locus_8301_iso_3_len_2006_ver_3	Thylakoidal processing peptidase 2, chloroplas	36	n	n	5.320603815	5.476041389
1259	cra_locus_1259_iso_3_len_1357_ver_3	Pom30 protein	36	n	n	5.438133135	5.454419393
879	cra_locus_879_iso_2_len_1928_ver_3	Dehydroquinate synthase	36	n	n	5.466366093	5.387393562
5275	cra_locus_5275_iso_4_len_2521_ver_3	Receptor protein kinase	36	n	n	5.480959801	4.99933711
2720	cra_locus_2720_iso_8_len_1822_ver_3	Multidrug resistance pump	36	n	y	5.644467763	5.155923246
4226	cra_locus_4226_iso_5_len_2167_ver_3	Fructokinase 2	36	n	n	6.077300689	5.48634754
4333	cra_locus_4333_iso_6_len_2424_ver_3	Methylenetetrahydrofolate reductase	36	n	n	6.125186135	6.393503723
10652	cra_locus_10652_iso_3_len_2341_ver_3	Glucose-6-phosphate isomerase	36	n	n	6.325812727	6.127707559
399	cra_locus_399_iso_5_len_1960_ver_3	Cytosolic 6-phosphogluconate dehydrogenase	36	n	n	6.35133504	6.73794359
59882	cra_locus_59882_iso_1_len_336_ver_3	5-methyltetrahydropteroyltriglutamate–homocy	36	n	n	6.596437267	6.664184144
2522	cra_locus_2522_iso_2_len_1526_ver_3	Loganic acid methyltransferase	36	y	n	6.685155252	7.256161502
7012	cra_locus_7012_iso_4_len_835_ver_3	Cytochrome b5	36	n	n	6.972462907	6.882104995
1852	cra_locus_1852_iso_8_len_3527_ver_3	Cystathionine gamma-synthase isoform 2	36	n	n	6.994251596	6.735765855
1815	cra_locus_1815_iso_9_len_1294_ver_3	Plastid isopentenyl pyrophosphate:dimethylally	36	y	n	7.876117848	7.77363376
5563	cra_locus_5563_iso_3_len_426_ver_3	Gene of unknown function	36	n	n	7.957340052	7.673627087
3269	cra_locus_3269_iso_1_len_329_ver_3	Methionine synthase	36	n	n	7.974810345	8.179068102

9802	cra_locus_9802_iso_3_len_906_ver_3	5-methyltetrahydropteroyltriglutamate--homocy	36	n	n	7.995846278	8.006202634
16801	cra_locus_16801_iso_1_len_296_ver_3	5-methyltetrahydropteroyltriglutamate--homocy	36	n	n	8.433426872	8.720271624
7197	cra_locus_7197_iso_1_len_671_ver_3	Glutathione s-transferase	36	n	n	8.436011884	8.239631941
596	cra_locus_596_iso_2_len_582_ver_3	2-hydroxyisoflavanone dehydratase	36	n	n	8.443072006	8.350329277
5891	cra_locus_5891_iso_1_len_642_ver_3	Adenosylhomocysteinase	36	n	n	8.589310327	9.732041895
4840	cra_locus_4840_iso_6_len_2136_ver_3	Conserved gene of unknown function	435	n	n	2.125680855	2.25954066
11051	cra_locus_11051_iso_2_len_928_ver_3	Type-A response regulator	435	n	n	2.541564399	1.998191892
40962	cra_locus_40962_iso_1_len_384_ver_3	ATP-binding cassette transporter	435	n	y	2.6726109	2.394179358
24716	cra_locus_24716_iso_1_len_1549_ver_3	Cytochrome P450	435	n	n	2.822358976	2.676642027
5924	cra_locus_5924_iso_1_len_1312_ver_3	Metal transporter	435	n	y	3.140909507	1.723659033
13084	cra_locus_13084_iso_2_len_1720_ver_3	Conserved gene of unknown function	435	n	n	3.155336374	3.120688685
11045	cra_locus_11045_iso_1_len_2237_ver_3	Pathogen-inducible alpha-dioxygenase	435	n	n	4.015898936	3.191834236
2087	cra_locus_2087_iso_5_len_1776_ver_3	DNA-damage inducible protein DDI1	435	n	n	4.409798228	4.700484108
13992	cra_locus_13992_iso_2_len_843_ver_3	ATP-binding cassette transporter	435	n	y	4.495560863	3.198079201
6707	cra_locus_6707_iso_4_len_1366_ver_3	ATP-binding cassette transporter	435	n	y	4.522300615	3.61801542
8112	cra_locus_8112_iso_3_len_2201_ver_3	ERD6-like transporter	435	n	y	4.592809819	4.491827448
32767	cra_locus_32767_iso_1_len_742_ver_3	NADH dehydrogenase	435	n	n	5.397847447	5.780170968
4746	cra_locus_4746_iso_4_len_1655_ver_3	Cysteine protease	435	n	n	5.697554219	6.16538138
16158	cra_locus_16158_iso_3_len_2097_ver_3	Peptide transporter	436	n	y	2.025085509	2.148280108
57474	cra_locus_57474_iso_1_len_478_ver_3	Calmodulin	436	n	n	2.332324119	2.592557499
14769	cra_locus_14769_iso_3_len_2466_ver_3	Exostosin family protein	436	n	n	2.515447542	2.818339311
3942	cra_locus_3942_iso_9_len_1792_ver_3	UDP-glucuronic acid/UDP-N-acetylgalactosam	436	n	y	2.530395412	2.87302767
13551	cra_locus_13551_iso_1_len_437_ver_3	LysM receptor-like kinase	436	n	n	2.829413236	1.747146751
32385	cra_locus_32385_iso_2_len_1608_ver_3	Conserved gene of unknown function	436	n	n	2.903196418	3.080222743
8519	cra_locus_8519_iso_2_len_1521_ver_3	Histone deacetylase	436	n	n	2.932172696	2.719586506
7790	cra_locus_7790_iso_3_len_1644_ver_3	F-box family protein	436	n	n	3.033863452	2.667326741
16200	cra_locus_16200_iso_2_len_2362_ver_3	Transporter	436	n	y	3.059604474	2.956757379
5874	cra_locus_5874_iso_1_len_1475_ver_3	Ribonuclease III	436	n	n	3.39462389	2.992994967
4346	cra_locus_4346_iso_1_len_2082_ver_3	Mechanosensitive ion channel domain-containi	436	n	n	3.515813357	3.646669586
1697	cra_locus_1697_iso_1_len_2819_ver_3	Receptor kinase	436	n	n	3.593054922	4.123434446
131	cra_locus_131_iso_4_len_1380_ver_3	16-methoxy-2,3-dihydrotabersonine N-methyltr	436	y	n	3.60245688	3.000090166
17006	cra_locus_17006_iso_1_len_2395_ver_3	GATA transcription factor	436	n	n	3.639961431	3.192991058
7747	cra_locus_7747_iso_5_len_1534_ver_3	P-coumaroyl shikimate 3'-hydroxylase isoform	436	n	n	3.644017763	3.91798543
811	cra_locus_811_iso_6_len_1980_ver_3	Sugar binding protein	436	n	n	3.690025917	3.83163407
7567	cra_locus_7567_iso_3_len_2520_ver_3	Peptidase, trypsin-like serine and cysteine pro	436	n	n	3.700983401	3.788925792
13761	cra_locus_13761_iso_1_len_2494_ver_3	GRAS family transcription factor	436	n	n	3.768703072	4.133727907
9821	cra_locus_9821_iso_2_len_1869_ver_3	3-hydroxyisobutyrate dehydrogenase	436	n	n	3.960576578	3.700994494
6674	cra_locus_6674_iso_3_len_1498_ver_3	Zinc finger family protein	436	n	n	3.963520364	3.527633387
7640	cra_locus_7640_iso_9_len_2908_ver_3	Potassium transporter	436	n	y	4.003872042	3.838941685
1694	cra_locus_1694_iso_8_len_3396_ver_3	NRAMP family metal transporter	436	n	y	4.019497325	4.052650643
10104	cra_locus_10104_iso_2_len_1571_ver_3	Hydroxycinnamoyl transferase	436	n	n	4.033484673	4.625299717
4478	cra_locus_4478_iso_1_len_1840_ver_3	SFR6 (SENSITIVE TO FREEZING 6)	436	n	n	4.283143929	4.407950917
10132	cra_locus_10132_iso_8_len_2383_ver_3	Acyl-CoA synthetase	436	n	n	4.402817662	3.908822513
2788	cra_locus_2788_iso_4_len_3582_ver_3	Peptidyl-prolyl cis-trans isomerase CYP40	436	n	n	4.434274599	4.771690178
1067	cra_locus_1067_iso_4_len_1465_ver_3	CAPIP1	436	n	n	4.509714841	4.039954286
1665	cra_locus_1665_iso_8_len_3747_ver_3	Prf	436	n	n	4.778329497	4.437261036
9431	cra_locus_9431_iso_6_len_1440_ver_3	Trafficking protein particle complex subunit	436	n	n	4.996935529	4.737259513
23138	cra_locus_23138_iso_1_len_732_ver_3	USP family protein	436	n	n	5.222182442	6.092205324
1189	cra_locus_1189_iso_3_len_2062_ver_3	Harpin-induced protein	436	n	n	5.250290837	5.150380953
859	cra_locus_859_iso_5_len_1511_ver_3	Pyroglutamyl-peptidase I	436	n	n	5.493599303	5.101381142
6502	cra_locus_6502_iso_4_len_1395_ver_3	Glutathione S-transferase	436	n	n	5.795979888	6.436928034
1253	cra_locus_1253_iso_4_len_2391_ver_3	UDP-glucosyltransferase	436	n	n	6.714108112	5.693138438
627	cra_locus_627_iso_9_len_1333_ver_3	Lactoylglutathione lyase	436	n	n	7.038436182	6.830483517

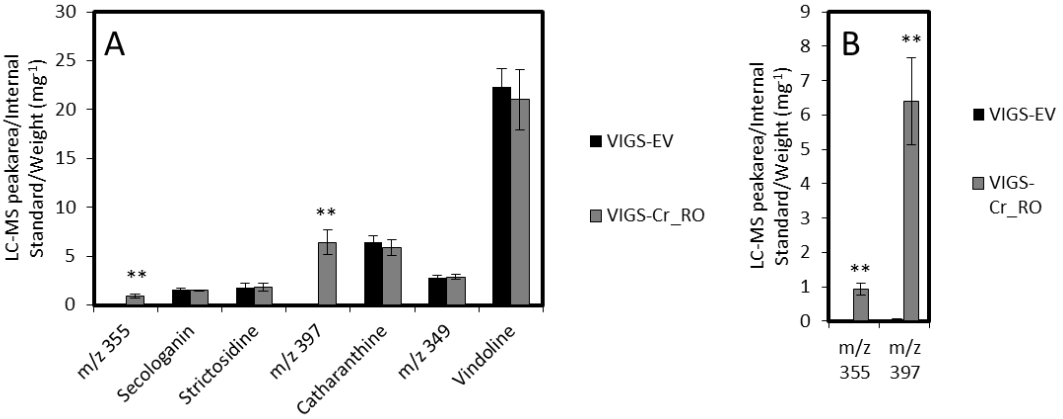
9.2 Replicates of VIGS-CrRO



Replicate1 for CrRO Metabolite and qPCR profile upon VIGS of CrRO A.) Alkaloid profile VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes VIGS-CrRO (n=18) VIGS-EV (n=8) ** p<0.01 **B.)** *m/z* 355 and *m/z* 397 profile for VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes VIG-CrRO (n=18) VIGS-EV (n=8) ** p<0.01 **C.)** qPCR normalised expression of the CrRO gene in empty vector and CrRO silenced plant tissue VIGS-CrRO (n=8) VIGS-EV (n=8). All data presented is the mean $\pm$ SEM.

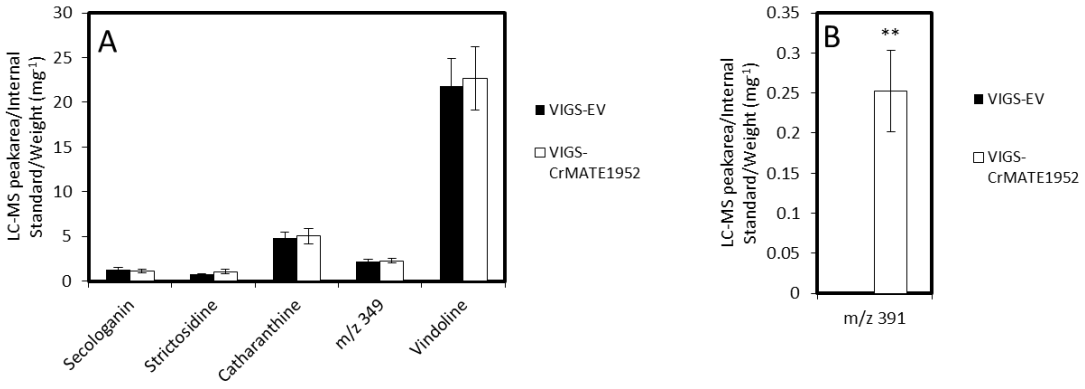


Replicate 2 for CrRO Metabolite profile upon VIGS of CrRO A.) Alkaloid profile VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes VIGS-CrRO (n=8) VIGS-EV (n=8) ** p<0.01 **B.)** *m/z* 355 and *m/z* 397 profile for VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes VIG- CrRO(n=8) VIGS-EV (n=8) ** p<0.01. All data shown are mean ± SEM.

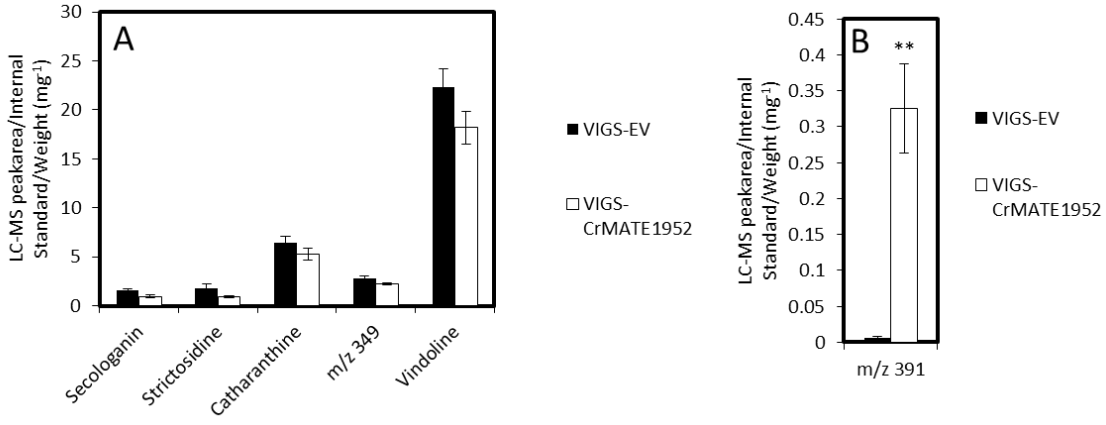


Replicate 2 for CrRO Metabolite profile upon VIGS of CrRO A.) Alkaloid profile VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes VIGS-CrRO (n=10) VIGS-EV (n=8) ** p<0.01 **B.)** *m/z* 355 and *m/z* 397 profile for VIGS of CrRO relative to empty vector control tissue in Little Bright Eyes VIG-CrRO(n=10) VIGS-EV (n=8) ** p<0.01. All data shown are mean ± SEM.

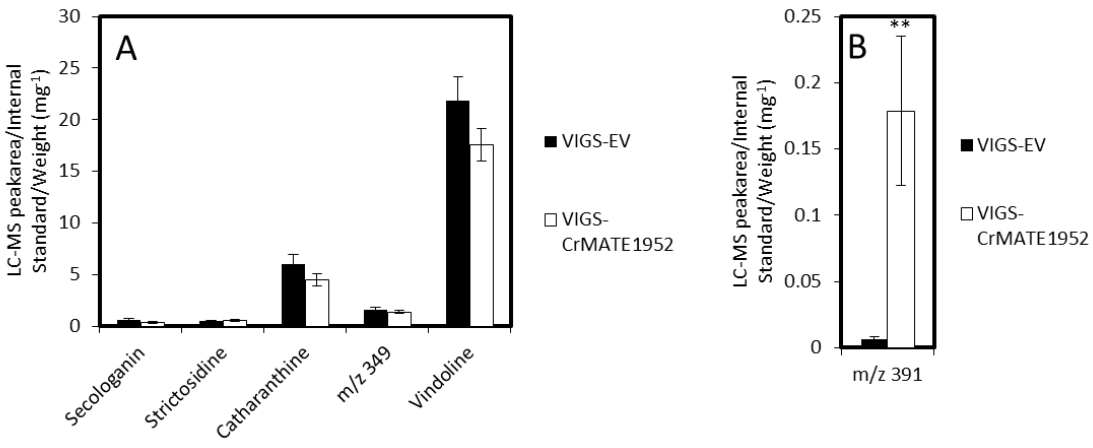
9.3 Replicates of VIGS-CrMATE1952



Replicate 1 Metabolite profile upon VIGS of CrMATE1952_A.) Alkaloid profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes VIGS-CrMATE1952 (n=12) VIGS-EV (n=12) ** p<0.01 **B.)** *m/z* 391 profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes VIG-CrMATE1952 (n=12) VIGS-EV (n=12) ** p<0.01**.** . All data shown are mean ± SEM.

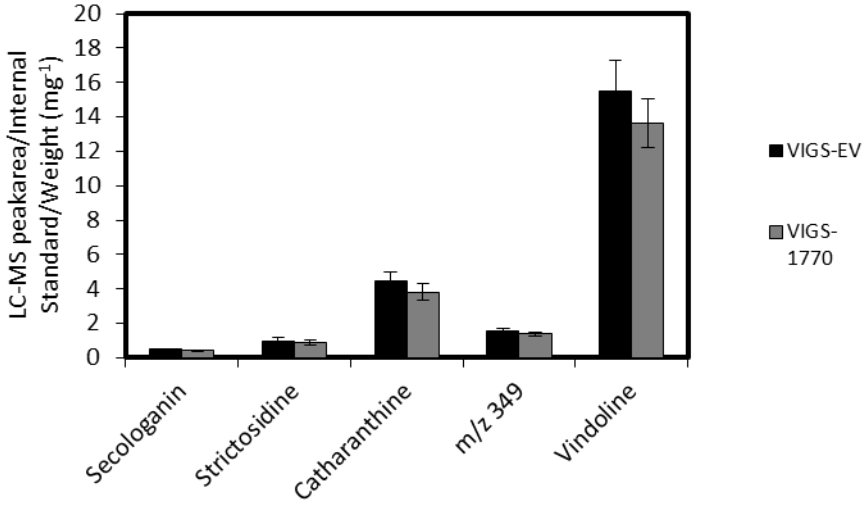


Replicate 2 Metabolite profile upon VIGS of CrMATE1952 A.) Alkaloid profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes VIGS-CrMATE1952 (n=9) VIGS-EV (n=12) ** p<0.01 **B.)** *m/z* 391 profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes VIG- CrMATE1952 (n=9) VIGS-EV (n=12) ** p<0.01. All data shown are mean ± SEM.



Replicate 3 Metabolite profile upon VIGS of CrMATE1952 A.) Alkaloid profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes VIGS-CrMATE1952 (n=6) VIGS-EV (n=8) ** p<0.01 **B.)** *m/z* 391 profile for VIGS of CrMATE1952 relative to empty vector control tissue in Little Bright Eyes VIG- CrMATE1952 (n=6) VIGS-EV (n=8) ** p<0.01. All data shown are mean ± SEM.

9.4 VIGS Amino Acid Permease



Metabolite profile upon VIGS of amino acid permease Alkaloid profile for VIGS of VIGS-1770 relative to empty vector control tissue in Little Bright Eyes VIGS-1770 (n=9) VIGS-EV (n=8) ** p<0.01. All data shown are mean ± SEM.

9.5 Full length sequences for genes in this study

Multidrug and Toxic Compound Extrusion Proteins (MATES)

CrMATE10740

ATGGATAGTAGTAATAAGAATCCGTTGCTAGTGAAAGAATACTCCGATTCCGATGGTAAAGGTAATGG
AGTAAAAACAATGATTAGTAGTTGTAGGTCTTCATCTAAAGGCTGGTGGAATAAGATATTGGATGTGG
AAGAAGCTAAAGGTCAAATTCTGTTTTCGTTGCCGATGATTCTTACTAATCTTTCTTATCACTTTATACTT
CTTACTTCCGTCATGTTCGCCGGTCATATTGGAGAACTTGAACTTGCCGCTGCAAATCTTGCTAATTCAT
GGGCGATTGTTACTGGTATTGCTCTCATGGTTGGACTAAGTGGTGCACTGGAAAGTTTGTGTGGACAA
GGATATGGAGCAAAGTTGTATAGTATGTTGGGGATATATCTACAGGCATCCTGCATAATATCATTCTTC
TTCTGTATTCTTCTATCCATCTTGTGGTGGTTCTCTGGTTCTGTCCTCACTTTATTACATCAAGATCCTCAA
ATAGCAAAAGCTGCCGGTCTCTATTTAAGGTTTCTCATACCAGGATTATTTGCATATGGAATGTTACAAA
GCCTCTTGAGATTTCTTCAGACACAATCCGTTGTTACTCCATTGGTTATCTTCTCAGTTCTACCTTTGATCC
TCCATATTGGAATTGCCTATTGTTTCACTCACTGGACATCATTTCCTTTCACCGGAGCAGCATTAGCTGCT
TCGTTCTCCTTCTGGTTATCAGTCATAATGTTAGCCTTATATGTTCTACGTGCAAGTAGATTTAAGCATAC
TTGGAAAGGTTTTTCAATGGAGTGTTTCAATCATATTGGACAGACCTTGAAGCTGGCTATACCGTGCGC
TGCAATGGTTTGTTTGGAGTACTGGGCTTTCGAGCTTCTCGTGTTATTAGCTGGCGTAATGCCAGATCC
AGAGATAACTACTTCCTTGATTGCAATATGTGTAAATACAGAATCCATTTCCTACATGTTTGCTTATGGT
CTCAGTGCAGCAACAAGCACAAGGGTTTCAAATGAGCTAGGAGCAGGGAATCCTGAGCGAGCCAAGA
AAGCTATGACGGTCACTTTGAAGCTTTGTGTGATTACGGCACTTGGTGTTGTTTCTGCTTTAGCATTTGG
CCATGATATCTGGGCTGGATTATTCAGTAATAGCCCAGTTATCATACAAAAGTATGCTTCAATGACACCA
TTTCTTGTGATCTCCATTTCCATTGATTTTTTCCAAGCAATCTTATCAGGTGTTGCCAGGGGATGTGGTTG
GCCACATTTAGCTGTGGGAATCAATTTGGGTGCATATTATTTTGTGGGCATGACAATTGCAGCTGTACT
TGGTTTCAAGGCGAAGTTGTACGCCAAGGGCTTGTGGACAGGCTTAATATGTGGTCTAAGCTGCCAGG
CTTGTGGGCTCTTGTTACTTTCATTGTTCACAAAATGGAAAAGAGTGGACTTGTCCTCCGATTGCAAAGA
AAATGAACAACAAATAGCTGAAGCTTGA

CrMATE1952

ATGGGTTCCAAACAAAACTATGAAATAAACCAACCACTGTTACTAAGTAACAACGGCGGTGGCAGCAG
TGCTATTACAGCTGCAGAGTCGCCAAAGGCGGCTAAGACGGTGGAGGAAAGGCAGTATGAGCTTAGT
GTTGAGTTAGAAAGGGTACTTTCCGATACATCTGTACCATTAGTGCCAAGACTCACCGCCGCCACTTGG
ATTGAGTTTAAGCTTCTCTTTCGGCTGGCGGCTCCAGCAGCCGCCGTTTATTTGATAAACTATGTAATGT
CTATGTTCACTCAAATATTTTCCGGGCATCTTGGGAATCTTGAGCTTGCTGCTGCTTCTCTTGGTAATAAT
GGTATCCAGACCTTCGCCTATGGTATCATGCTTGGAATGGGAAGTGCAGTGGAAACACTATGTGGACA
AGCATATGGAGCACAAAAACTAGACATGCTTGGAATATATCTCCAAAGATCAACAATTCTTCTAACCCT
AACAGGCATTGTTATTATGTTTGTTTACATATTCTCAAAACCAATCCTCCTTTTACTTGGCCAATCAGAAG
CCATAGCTGCAGCAGCTGCCTTATTTACTTATGGTCTTATCCCACAAATCTTTGCCTACGCAGCCAATTTC
CCAATCCAAAAGTTCTTACAAGCTCAAAGTATAGTGGCACCTAGTACTTATATATCAGCAGGAGCTATT
GTTTTCCATGTTTTGTTTAGTTGGTTAGCAATTTATAAGGTTGGACTTGGATTGTTCGGGGCATCTTTAG
TATTGAGTTTGTCTTGGTGGGTTGTTGTGGTTGGTCAGTTTATTTATATTCTATATAGTGATAGGACTAA
GGACACTTGGCGTGGATTTAGTGTTGAAGCTTTCCATGGACTTTGGAGCTTTTTTAAGTTGTCTGCTGCT
TCTGCTGTTATGCTTTGTTTGGAAGCTTGGTATTTTCAGATTCTTGTTCTTTTGGCTGGAATGCTTCCTGA
TCCTAAAATCGCTTTGGATTCCCTCTCCATTTGCATTACAATCTTGGGTTGGGTATTCATGATAGCCGTTG
GATTCAATGCTGCTGCCAGTGTGAGAGTAGGGAATGAACTAGGGGCAGGACATCCAAGGGCAGCTGC

ATTTTCAGTAGTAATAGTGACAACAATGTCATTCATAATAGCAGTGATAATATCATTAGTGGTACTTGCT
TTGCGCTACAAAATTAGCTATATCTTTACCGAAGGTGAAGTTGTAAGCAATGCTGTTGCCGATATGTGT
CCCTTGCTCGCCATCACTCTTGTTCTTAATGGAATTCAACCTGTTTTATCCGGTGTTGCTGTTGGATGTGG
ATGGCAAGCTTTTGTTGCATATGTGAACGTTGGCTGTTATTACATTGTTGGTATCCCAACCGGTGCACTT
CTTGGGTTCTACTTCAAACTTGGAGCCAAGGGTATCTGGTCCGGTATGATTGGTGGGACATTGATGCAA
ACTATTATTCTCATCTGGTTTACTTATCGAACGGACTGGAAAAAAGAGGTGGATATAGCTCAAAGTAGG
TTGGATACATGGGAGGACAAGCCAAAATCTTTGTCCAATGAATAA

locus_3327

ATGGATCAAAGTTGTAGTAATGGAATGAATGAAGGTTTGCTGCAAAGTAAAAAAGAATCATCATCAGT
AGTTGGGGTTGGGGAGGAAATAATTAGTTCTGAGCTTGAAGAAATATTATCAGATACAAATATATCAC
AGTTACAGCGTCTCAAACGAGCAGTAGTCGTGGAAATGCGTACGCTATTCCGTCTTGCTGCACCTGCAT
GTGTCACCGCCTTGCTCGGTAATGTTGTTTCCATGTCCACTCAAATCTTCTGCGGACATGTTAGTAATCTT
GATCTTGCTGCCTCTTCCCTCGGCAATAATGGTATCCAACTCTTCGCCTACGGCCTCATGCTGGGAATGG
CAAGTGCAGCTGAGACTCTATGCGGGCAGGCATATGGAGCACACAAGTACGAGATGCTAGGCATCTAT
TTACAAAGATCAATAATCCTCCTTATGGGAACTGGATTACTCGTCATGTTAATTTACATTTTCTGCAAGCC
AATTTTGCTACTTCTAGGAGAATCAGAACAAGTGGCATCTGCAGCGGCACTATTTGTATACGGTCTTATC
CCACAAATATTTGCATATGCAGCCAATTTTCCAATACAAAAATTTCTGCAGGCACAGAGCATTATTAATC
CAAGTGCATATATATCAGTAGCTACACTTGTGGTGCACATCCCTTTGACATGGTTGGTTTTGTTTGTATT
GGATTGGGGTTTATTAGGCGCTTCGGCAGTATTGAGTTTTTCTTGGTGGATTATTGTTTTGGCACAATTT
ATTTACATTGTGAATAGTAAAAAGTGCAAAAAGACATGGACAGGTTTCAGTTTTCAGGCATTTACTGGA
TTATGGGATTTCTTGAAATTGTCTGCATCTTCGGCTGTAATGCTTTGTCTTGAGACTTGGTACTTTGACTT
TTTGATTCTCATTGCTGGTTTGCTTCCTAATCCTCAAATAACTTTGGATGCTCTTGCTATTTGCACCACAA
TCCTTGGATGGGTTTTCATGATTTCAGTGGGCTTCAATGCGGCAGCCAGTGTGAGGGTTAGCAATGAAT
TGGGGGCTGGGCATCCAAAATCAGCAGCATTTTCAGTGAAGGTTGTGACATTGAGTTCATTGGTGATTT
CAATAGTATGTGCAATTGTGGTACTGGTGTTGCGGCGGGTGATGAGTTATGCCTTCACCAGTGGTGAA
ACTGTTGCTAACGCCGTCTCCGATCTTGCTCCTCTCTTGGCCATCTCCATTGTTCTCAACGGCGTGCAGCC
TGTTTTGTCCGGGGTTGCGGTAGGATGCGGTTGGCAAGCATTTGTGGCGTATGTTAACGTTGGTTGTTA
TTACATAGTTGGGATTCCATTGGGTGCATTGCTCGGATTCAAATTCAATCTTGGTGTTAAGGGAATATG
GACAGGAATGGTGGGAGGAGTGGCCATGCAGACTGCTATCTTACTTTGGGTCACTTTGCGAACCAACT
GGGATCGAGAAGTGGAGATCACAAGGAATCGGTTAGATAAGTGGAATAAGACAAAGGTAGAGGCTG
CTGCTTTACTGAATAATTGA

locus_8527

ATGGAGTACTCCGAAGATGAAAGCCGAGGAACAGCGGCCTTACTGCAGCCGCCATCCCCAGCTGCTGT
GGACTCCCAGCAGTCGGAGGTCAGCTCAGAGCTGGAAAAGGTGCTGAGCGACACTGAATTAAGCTACT
GGAAACGCATCCGGGTGGCCACATGGATCGAACTCAAACTTTTGTTTCGGCTGGCTGCTCCAGCAATCG
GTGTTTACTTGATAAACAACTTCATGTCCCTTTCTACTCGCATCTTTGCTGGCCATCTCGGCAATCTTCAG
CTTGCTGCTGCCTCCTTAGGCAACCAGGGTATCCAATTATTCGCATACGGCCTCATGCTGGGGATGGGA
AGCGCAGTGGAAACATTATGCGGCCAAGCATACGGCGCTGGGAGACATGAAATGCTTGGAATTTACCT
CCAAAGAGCAACAGTAGTTCTTACACTAACCGGTCTACCAATGGTCTTCATCTACATTTTCTCAAAACCA
ATTCTAATTTTGCTAGGCGAATCCAAAACTTTGGCATCAGAAGCATCAATTTTTGTTTATGGCCTAATTCC
CCAAATCTTTGCATATGCAATAAATTTCCCAGTACAAAAATATCTACAAGCTCAAAGTATAGTGAAGCCA
AGTGCCTACATTTCATTGGCTACATTATGTCTTCATTTATTATTGAGTTGGGTTGTGGTTTATAAAATTGG

GTTAGGGTTAATTGGTGCTTCATTGGTTTTGAGCTTATCATGGTGGATAATTGTGGTGGCTCAATTTGTG
TATATATTGATGAGTGAGAAATGCAAGGCTACTTGGATCGGATTCAGTTTTGAAGCTTTTAAAGGATTA
TGGGAATTTGTGAAATTATCAGCAGCATCAGCAGTTATGTTGTGTTTGGAGACATGGTATTTTCAGATT
TTGATTTTGATTACTGGATTGCTTAAGGATCCTCAGCTTGCCTTGGATGCTTTATCTGCCTGCTTGGGAG
TAAACGCTATTGTGTTCATGATTTCTGTTGGATTCAATGCAGCTGCTAGTGTGAGGGTTGGCAATGAAC
TTGGAGCAGGGCACCCCAAATCAGCTTCCTTTGCAGTTACAGTAGTGAATTTAGTTTCTTTCATCATAGC
CGTTATTGCCGCCGGAATAGTGCTTTCGTTACGCCACGTCATTAGCTACGCTTTTACCGACGGTGAAACT
GTGGCGGAGGCCGTCTCTGACTTATGCCCGCTCTTGGCTGTCACACTTGTTCTCAACGGTGTCCAACCA
GTTTTGTCCGGTGTAGCGGTGGGGTGTGGATGGCAAGCATTTGTTGCTTATGTAAACGTTGGGTGTTAC
TATGTGGTTGGCATTCCACTCGGTTGCATTCTCGGCTTCAAGTTTGATTTTGGTGTTAAGGGCATATGGT
TAGGTATGCTTGGAGGAACAATGATGCAGACCATAATCTTAGTATGGGTAACATTTCGTACCGACTGGA
ATAAGGAGGTGGAGAAAGCAAGAAACCGTCTGGACAAATGGGACGACATCAAAAAGCCTCTTTTGGT
GGATTAA

locus_2720 (partial sequence)

ATGGCGTTATGTTTGGAGCTATGGTACGTCACAGCAGTTATTCTAATGGTAGGAGGCTTGAAAGATGCT
GATGTTGCAGTTGCAGTTGATGCGGCTTCAATATGTATAAATTTGCACTTCTGGACGTTGATGCTTGCCC
TTGGTTTTAATGCATCAGTAAGTGTAAGAGTATCAAACGAATTGGGAGCTGGAAAACCAAAAGCTGCA
AAACTCTCAATTGTAGTACATATGTTAACAACTGGGGTTCTTGGAATTATATTTGCATCAACTATTTTGG
CCACAAGAAACCAATTTCCCAGAGTTTTCACAGATAAACCACAAGTAATTAAAGAAACCTCCAAGTTGG
GTTATCTTCTAGCAGCCTCTGTCTTTCTCAATGGCATTCAACCAGTTCTGCTCGGGGTGGCTGTTGGTGC
AGGATGGCAGTTCCCGGTTGCCATTCTAAATGTGATATGTTATTATGGGATTGGGCTTCCAATAGGAGC
CTTACTTGGTTTTAAGTTTAAGCTGGGCGTTCATGGCATATGGTGGGCAATGCTTGGTGGTAGCTTGCT
TCAGACAATTGCTCTGCTGATCATTGTTATTCGAACAAATTGGCACAAGCAGTCTCTTCAAGCTGAGGA
GCGGGTAAAAGAATGGGGAGGCCAAACAGAACCTCAACAAAGCTCATGA

NPF transporters

CrNPF2.9

ATGGGAGACACCGAAGCACAGCTTCTCCAGCCAGGTCACAAGCAGAAAGGCGGTTGGATCACTTTCCC
TTTCATTTTAGCAACGAGGACTTTATTGACGCTTGCGGTAGCTGGATTTAGCTCTAATTTAATCGTTTACT
TGATCAACGAATTCAATGTGAATAGAATCGATTCTGCTCAAATATATAATGTTGTGAATGGTTGCATGG
CCCTATTCCCTCTTCTCCTTGCTATAATTGCCGATACTTTTCTGGGCTGTTTCAATGTCATCTGGATTTCTA
CCCTCATCTCTTTGATGGGGATGGCTCTGTTGACATTAACTTCCTCAATCACTTCACTAAGACCTCAACCA
TGTGCCGAAGGGTCAACCTTTTGCCAACAACCATCAGCATATCAATCCAGCATTCTGTTTTTGGCTCTGG
CCTTGCCATCTATAGGCTTTGCAGGAACCAGTTTTACAGTGGGAACAATGGGTGCACATCAATTAGACG
ATCCAAAGCATCAAGAAAATTTCTTCAACTGGTTTCTCTTTATCTGGAATGCTGCTTCAGTAATCAGTGG
CATTGGAATTGTCTATGTTCAAGATAATGTGAGTTGGTCATGGGGATTTGGTATTTGTGTTGCTTCAAAT
CTTCTGGGTTTGATCATTTTCTTGGCCGGAAAGCGTCTCTATCGTGATGTACAGCCACAGAAAAGTAGC
CCATTCAAGGATTTAGCTTGTGTTGTTGTTGCTGCTCTCTCTAAGAAGAAGCTCTCCTTGTCAAACAAAG
AAGAAGATTATTACAGCGAATTACCCGATAATGCTGAGGAACAACAACAACAAGAGGGCGTTACCCTT
TTACCTGCTACTGTACCTGATGAAAGTTTCAAGTTCTTGAATCATGCGGCTCTGGTAACTTCAGCAGATA
TCCAACCAGATGGTTCAATCAAGAAATCATGGAAACTATGCACAGTTAAACAGATTGAAGATCTCAAAA

CCTTAATCAGATTATTCCCACTTTGGACAACTGGTTTCTTATTAACTATCCCAATGGGAGTACTTTCCAGT
TTAACAACCCTTCAAGCTCTAACAATGGATTGCTCCACTTTTTGGGGTCTCAAATATCCAGTAGGATCCA
TGTCTGTTTTCACTCTGTTAGCAGGTGCAATCTCTCTAACCTTTATAGACAGATTGATTTTCCCCATTTGT
CGAAAAATGGCTAAACCAATTAGACCCCTTCAAAGAATAGCCATTGGTCACATCATCAACGTGATCTCT
GTTGTCATTGCTGCCATTGTTGAACACAAGCGTCTACAATTAGCTCGAGCCCAGAAATTCCAAGGCAAA
ACGGACTCCGTCGTCGTGCCCATGTCAGTCTTCTGGTTAATTCCGCAATTGGCACTCTCCGGTACCGGA
GAAGCATTTCATTTTCCAGGACAAGCATTGTTGTATTACAAAGAATTTCCTGCATCACTCAAGAGTACTT
CGACAGCTATGTTGGCAATTCTTATAGCCATTGGGTATTATATGGGAACTTTCGTTATTGATGTTGTTAG
GAAAGTTACAGATTGGTTGCCGGAAGATATTAATCACGGAAGGCTGGATAATCTCTATTGGTTGGTTGC
TGTTCTTGGAGTACTGAATTTTGTGTATTATCTTGCTTGTGCTGGGGCGTACGAGTATAGTTCCGTCATG
GAGGATGAAGATGAAACAAATGATAATAAAATATATATGTAA

CrNPF2.4

ATGGCTGAAAAATCTTCTTCCATTGATGCTACTGAAAATGGTTTCGTTAAGATTAAAGAACCAAACCAC
AGAGGTGTCAAAGCCATGCCTTTTATCGTTGGAAACGAGACGTTTGAGAAGCTGGGAACACTTGGAAC
CTCAGCAAACCTTTTGATTTATCTGACCACTGTTTTTCATATGAAGTCAATTACTGCAACAAACCTCATTA
ACATCTTCAATGGCACATGCAATTTTGGAACCTTGCTTGGAGCTTTTTTCTCGGATACTTATTTTGGACGT
TACAAAACTCTCGGCTTTGCCTCTGTTTCCTCTACCTTGGGGATGCTTATACTAACCCTAACAGCTGCAAT
TTCAAAGTTGCATCCTCCTTCCTGTGAAAAAAACAGCAAAGCTTGTAGCGGACCGACGTCGGGGCAATT
CGCTTTCCTCTTCACTGCATTTTTGTTCTTAATAATTGGAGCAAGTGGAATAAGGCCATGTAATTTAGCC
TTTGGAGCTGATCAATTTAATCCCAATACTGAATCAGGAAGAAAAGGAACAAACAGTTTCTTCAATTGG
TATTATTTTACCTTTACATTTGCAATGATGGTTTCATTAACAGTCATTGTTTATGTTCAATCAAGTATCAG
TTGGGCACTTGGATTGGCTATTCCAACGTTTTTAATGTTGTTATCTTGCGTTTTCTTCTTCGTGGGTTCAA
GAATTTATGTTAAAGTCTTGCCGGAGGGTAGCCCTTTGACTAGTATGATGCAGGTTATTGTGGCTGCAT
TTAGAAAAAGAAAACTGAATTTACCAGATCATCAGCCTTGGGATTCTCTTTTTAATTACATTTCCCCAAG
TTCAATCAACTCAAAGCTTTCATACACAGATCAGTTCAGGTGGCTAAACAAAGCAGCAATTAAGACCTC
AAAAGACCATCTATACGCAGAAGGATCAACGGCAAATCCATGGAGACTATGCAGTATGCAACAAGTAG
AAGAAGTAAAATGTGTGATAAGATCAATTCCAATTTGGACATCAGGTGTTTTATACTATGTTGTACTAA
ACCAAATGAGCACTTATTTAGTGTTTCAAGCCCTTCAATCAGACAGAAGAATTGGCAAAACAAACTTCA
AAATCCCAGCAGCTTCTTACACAGTCTTCTCAATGCTAAGCCTTAGCATTTGGATACCAATTTATGATAG
AATCATCGTTCCATTTCTTCGAAAAATCACAAAAACAGAACAAGGAATAACAATTCTACAAAAAATGGG
ATTTGGATTGGTTATTGCAGTAATAACAATGGTTGTTTCAGCAATAATTGAAGATAAAAGAAGAGTAAT
TGCACTTACTAGGCCAATTGTTGGAATTGAAGAAAGAAAAGGGGGAATTTCTTCATTATCTGGATTTTG
GTTAGTACCTCAATTGGCTTTGGCTGGGATTTCAGAGGCATTTACTGTGATTGCTCAAGTTGAATTTTTC
TATAAACAATTTCCAGAGAATATGAGGAGCTTTGGTGGGTCATTTTTATTTTGCGGATTTGCAATTTCAA
GTTATTTGAGTAGTTTCTTGATATCAATTGTTCATAAAGTGACAAGAAATTCAAGTGGTGGAAATTGGTT
GGATGAAGATCTTAATAAGGCTAGATTGGATTATTTTTACTATTTGGTGGCTGGTTTGGAGGTTTTGAA
TTTTGGGTATTTTCTTGTTTGTGCTCATTGGTATAGATACAAAGAAACACAAAGCGGTAATGAAATTGCC
ATGGAACAAATATTGGATGAGCCCAAGAAGAACTTAGATGTGGTTTGA

CrNPF2.6

ATGGAGAAAAACGAGAAATTAGTCCCAAATGGCACAAAGGAAGATGGGAAAGTTGCTGCAAAAAATG
AGCCTAACTACAGAGGAATCAAGGCCATGCCATTTGTTCTTGGAAATGAAACATGTGAGAAATTGGGA
ACAATTGGGACCTCATCGAACTTGTCGGTTTATCTATCTACAGTGTTTCATATGAACACTATGTCTGCAA
ATAATGTAATAAACATCTTCAATGGAACTTGCTATCTTGGTTCCTTAGTTGGAGCTTTCTTGTGTGATACT
TATTTTGGTCGGTACAAAACTCTTGGGCTTGGCTCTGTAGCTTCATTCTTAGGAATGCTAATGCTGACAC

TAACTGCTGCAATATCGACATTACATCCTACAAATTGTGACAACTCAGGTGATTGTAAAGGGCCATCAG
CAGGGCAAATGGCTTTCCTGGTAGGTGGATTTGTACTAATGGTAATTGGGGCTAGTGGCATTAGGCCA
TGTAACTTGGCCTTTGGAGCTGATCAGTTTAATCCCAATACTGAATCTGGAAGAAGAGGAATCAATAGT
TTCTTCAATTGGTATTATTTCACATACACTTTTGCTGTCATGGTGTCTGCTACCATCATTGTATACGTGCA
GTCGGATATAAGCTGGTCTATTGGTTTTGCAATTCCAACATCTCTCATGTTCTTATCTTGTGCGCTATTCT
TTGCCGGTACCTGGCTATACGTCAAAGTGTTACCACAAGGCAGTCCTATGACTAGTCTTGTGCAGGTTA
TTGTGGTTTCGGTCAAGAAGAGACGATTGAAGTTACCAGGGGAACCATCTTTGCTTTTTAACCACATTC
CTTCTAATGCTATCAACTCAAAACTCCCTCGAACAAATAACTTCAGGTTTCTTGATAAAGCAGCAATTTT
GACCCCTGAGGACAAAGTTAATCCTGATGGTTCAGCAGCCAACAAATGGAAACTCTGCAGCATTCAGC
AAGTTGAAGAAGTGAAATGTGTGTTGAGAGTTCTTCCTATATGGATTGCGGGCGTAATATTTCAAATTT
CAGTAACACAGCAATCAAATTATGTACCATTTCAAGCATTCCAAAGTGATAGGCACCTAATCAAAGGCA
ACAAATTCCAGATACCAGCAGGAACTTACAATGTCTTCTCCATGTTAGCTCTAACCATTTGGCTACCTAT
ATATGATAGGATCATCATCCCTCTGCTCCGTAGATGGACAAAAAAGGAAGACGGTATCACCCTACTACA
GAGGATGGGAGTTGGCATAGTTCTCTCCATTCTTTCAATGTTAGTATCAGCATTAGTAGAATCAAGAAG
GAGAAACTTGGCATTGACTCGTCCAACGCTAGGAATTCAAGAAGGTCAGGGCTCAATTTCATCCATGTC
TGCTTTGTGGTTTGTACCTCAGTTGGCACTAATGGGGCTTGCTGAAGGATTTACTTTCATTGGTGAAAAC
GAACTTTTCTACAAAGAATTCCCAGAAAACATGAGAAGCATTGCTGCATCTTTGGTGTTATGTGGGGTG
GCAGTGGCAAGCTACTTGACTTCTTTCATTGCTGAGATTGTTCATAAGACTACAAGGTCCGGGACAAGG
GATTGGTTGGCTCAAGATCTTAACAAGGCAAGATTAGATTACTTCTTTTACATGGTTGCTGTATTAGAG
GCACTGAATTTGGTTTATTTTCTGGTATGTGCCAAGTGGTATAGATACAAAGGAACGGGTGATCTTGGA
AAACTTGAAGTGGCAATGGAAAAATTAGAAACTGAAAAGCATCTAGTTTAA

Reticuline Oxidase like enzymes

CrRO_locus_5608

ATGATAAAAAAAGTCCCAATAGTTCTTTCAATTTTCTGCTTTCTTCTTCTACTCTCATCATCCCATGGCTCA
ATTCCTGAAGCTTTTCTCAATTGTATTTCCAATAAATTTTCATTAGATGTATCCATTTTAAACATTCTTCAT
GTTCCCAGCAATTCTTCCTATGATTCTGTTCTCAAATCTACTATCCAAAATCCAAGATTCCTCAAATCACC
CAAGCCCTTAGCTATAATCACCCCAGTACTTCACTCCCATGTCCAATCTGCTGTTATCTGTACCAAACAA
GCCGGTTTACAAATTAGAATCCGAAGCGGAGGAGCTGATTACGAGGGCTTATCCTATCGTTCTGAGGTT
CCCTTTATTCTGCTAGATCTCCAGAATCTTCGATCAATTTCCGTTGATATTGAAGACAACAGCGCTTGGG
TCGAATCAGGAGCAACAATTGGTGAATTCTATCATGAGATAGCTCAGAACAGCCCTGTTCATGCGTTTC
CAGCTGGGGTCTCTTCCTCTGTTGGAATTGGCGGCCATTTGAGTAGCGGCGGTTTTGGTACATTGCTTC
GGAAATATGGATTAGCAGCCGATAATATAATCGATGCAAAAATTGTTGATGCCAGAGGCAGAATTCTT
GATAGGGAATCAATGGGAGAAGATCTATTTTGGGCTATTAGAGGAGGAGGAGGAGCTAGTTTTGGTG
TTATAGTTTCTTGGAAGGTTAAACTTGTAAAAGTCCCTCCGATGGTAACTGTTTTCATCTTGTCCAAGAC
TTATGAAGAAGGAGGTTTAGATCTTCTACACAAATGGCAATATATAGAACACAAACTCCCTGAAGATTT
ATTCCTTGCTGTAAGCATCATGGATGATTCATCTAGTGGAAATAAAACACTTATGGCAGGTTTTATGTCT
CTGTTTCTTGGAAAAACAGAGGACCTTCTGAAAGTAATGGCGGAAAATTTCCCACAACTTGGATTGAAA
AAGGAAGATTGCTTAGAAATGAATTGGATTGATGCAGCAATGTATTTTTCAGGACACCCAATTGGAGA
ATCCCGATCTGTGCTTAAAAACCGAGAATCTCATCTTCCAAAGACATGCGTTTCGATCAAATCAGACTTT
ATTCAAGAACCACAATCCATGGATGCATTGGAAAAGTTATGGAAGTTTTGTAGGGAAGAAGAAAATAG
TCCCATAATACTGATGCTTCCACTGGGGGGAATGATGAGTAAAATATCAGAATCAGAAATCCCATTTCC
TTACAGAAAAGATGTGATTTACAGTATGATATACGAAATAGTTTGGAATTGTGAAGACGATGAATCATC
GGAAGAATATATCGATGGATTGGGAAGGCTTGAGGAATTAATGACTCCATATGTGAAACAACCAAGAG
GTTCTTGGTTCAGCACCAGAAACCTTTATACCGGTAAAAATAAAGGTCCAGGAACAACTTATTCCAAAG

CTAAAGAATGGGGATTTCGGTATTTTAATAATAATTTCAAAAAGTTGGCCCTTATCAAAGGACAAGTTG
ATCCAGAAAACTTCTTCTACTATGAACAAAGCATTCCCCCTCTCCATTTACAAGTCGAACTTTGA

locus_2174

ATGTCTCTTTTTCTTCCCCTTATATCCCTCCTCTGTTTTTGCCTTCTAAGTTCTTCACACTTTGTAGCCTCAG
ACTCCAATTATGACAAATTTTCTCAATGCCTAGAAACTAATTCAATCCCAGATGACCAAATCTCTAAAATT
CTGTATTCCAAAGGCAATTCTTCTTTTAACTCTGACCAAATCTCTAAAATTCTGTATTCCAAAGGCAATTC
TTCTTTTAACTCTGTTCTTGATGCCTATGTTAGAAACCGTCGTTTCAACACTTCTTCTACATCAAAACCAG
TTATAATTGTGACCCCTTTGTCAATAACCCATGTCCAAGCAACTATTTTATGTACCAAAAATTTAGGCCTA
CAACTCAAAATCCGAAGTGGTGGACATGACTATGAGGGGATCTCGTACGTATCTGAATCTCCTTTTGTG
ATTCTTGATATGTTCAATTTCAGAAACATTAGTATTGATGTTGAAAATGAAACCGCTTGGGTCCAATCTG
GTGCTACCCTTGGTGAACTTTATTATAGAATTTGGGAACAGAGCAAAGTTCTTGGATTCCCTGCTGGTG
TTTGCACAACAGTAGGAATTGGAGGCCATATAAGTGGCGGTGGGTATGGCGCAATGCTTAGAAAATTT
GGACTCGCTGTAGATAATTTGCTCGATGCCCAAGTCGTTGATGTGAAAGGCCGGGTTTTAGATCGAAGT
AGTATGGGTGAAGATCTCTTTTGGGCCATAAGTGGTGGTGGCGGTGCTAGTTTTGGTGTTATTTTGGGA
TATAAAATAAAGCTAGTTCAAGTGCCTAAAATTGTAACCGTTTTTCGAGTTGAAAAAAGGGCAGAAGA
GAATGCTTTAGAAATTGTTCATCGATGGCAAGATATTGCGGACAAGATCGATAATGATTTGTTCATTCG
GGTTCTTTTGCAACCTGTTAGTGGAAAAGGAAAAGCAAAAGGGCAGAAAACAATTAGAGCAACATTTA
TCGCATTGTTTCTCGGTGATTCCAATAGGCTTATGTCTGTGATGAATGGTCAATTTCCCGAATTGGGACT
CAAGAAAACAGATTGTTTAGAGATGAGTTGGATTCGATCTGTACTCTGGTGGGCGAATTTTGACAACAC
CACTAGTGAAAGTGTTCTTCTGAATCGAAACCCAGACTCAGTTAATTTCTTGAAGAGAAAATCAGATTA
TGTCCAGACGCCAATTCCTAAAGATGGGTTGGAATCAATTTTCAAGAAAATGGTTGAGTTGGGAAAAA
CTGGATTTGTTTTCAATCCTTATGGAGGAAAAATGGGTGAAATTCCTGAAAATGAAAAGCCATTTCCTC
ATCGTGCTGGAATTATTTACAAGATTCAGTACTCTGTGAATTGGGAAGACGAAGATCCTGCTTTAGACA
AGAAATACATTGCTGAAACAAGAGAATTATACAAATTCATGACACCTTTTGTGTCAAAAAATCCTAGAC
AAGCATTTCTGAATTATAGAGATCTTGATATTGGAACTACGAGCAATGGTAAAAATAGCTACAACGAAG
GAGAAGTATTCGGGTCAAAATATTTTAAGGGTAATTTTGATAGATTAGTGAAAGTGAAGACGATGATT
GATCCAGAAAATTTCTTTAGGAATGAACAAAGTATTCCTCCTCTACCGGTTAGGTCGTTGGGTGGGAGG
AAAGGTGGAAATTGA

locus_5372

ATGATATCACGATCTTTTTCTATTTCAGACACAATTCTTCAAGACAAATTTTACCAATGTATCAATCTTAA
TTCTGAGCTCTATATTCCTTTCTCCACAGCGTTTGTTACCCCAAACAATGCCACATTCAACTCTGTCTTAC
AATCTACTGCACAAAATCTTAGGTGTTTAATCCCTTCTGTACCAAAACCTCAACTGATTTTCACACCTTTA
GTTGAATCCCATGTCCAAGCTGCAGTTATTTGTGCCAAAGAACTCGGTATTCATCTCAGAGTACGAAGT
GGTGGTCATGACTATGAAGGACTTTCTTATATATCTGATGATACAGAATCCCCGTTCATTATTGTTGATC
TTGTTAAACTCCGATCTGTTAATGTAAATATTGGAGATGATAGTGCCTGGGTTGAGGCCGGTGCCACTG
TTGGTGAACTTTATTATAGAATTTCGGAGAAAAGCAGAACTCATGGATTTCCTGCTGGCCTTTGTTCAAG
TTTAGGAATTGGGGGACATATAACTGGAGGAGCATACGGATCTATGATGAGAAAATATGGTCTAGGA
GCTGATAATGTGCTTGATGCTAGAATTGTGGATTCAATTGGCAGAGTTCTTGATAGGCAATCAATGGGA
GAGGATCTTTTTTGGGCAATCAGGGGAGGTGGAGGAGCTAGTTTTGGGATTATTCTTGCTTGGAAGAT
CAAATTGGTACCTGTCCCTGCAATTGTGACAGTTTTTACTGTTCCAAAAACACTAGAACAAGGTGCTACA
AAAGTTCTATACAAATGGCAAGAAGTTGCCCCTAAAATTGATGAAGATCTTTTCATTAGAGTCATTATAC
AAAAGGGGGATAAAACAATTGTAACAGCCTATAACGCGTTGTTTCTTGGGGGAACAGATAGACTTGTT

CAAGTAATGAATGAAAGTTTTCCCGAACTGGGATTGACTCGAAAAGATTGTATAGAGATGAGTTGGAT
TGAATCTGTACTATACATCGCTGGATATCCTCGGAACACGCCACCTGAAGTTCTTCTTCAGGGGAAGTC
ATTATTCAAGAACTATTTCAAAGCGAAATCAGATTTCATCAAGGAGCCAATTCCAGAAGAAGGACTTGA
AGGACTATGGAAAAGGGTTCTTGAAGAGGATTCACCATTGATGATTTGGAATCCATATGGTGGAATGA
TGAGTAAAATATCAGAATCTGAAATACCTTTTCCTCATAGAAAAGGGGTACTTTACAAAATCCAATACCT
TACAACATGGAATGATGAAAACAAAGAATCAGCAACAAAACATGTAGATTGGATAAGGAGACTATACA
ATTACATGGCTTCATATGCTTCAATGTTTCCAAGACAAGCTTATGTGAACTATAGAGATCTTGATTTGGG
AATGAACAAGAATGAGAACACAAGTTTTATTCAAGCAAGTGTTTGGGGTACTAAATATTTCAAGGACAA
TTACAATAGGCTAGTAAGAATCAAGACTAAAGTTGATCCTGATAACTTCTTCAGACATGAGCAGAGTAT
CCCAACACTTCAATTAACAAGTAATAAAAGAGGAAAGAGAAGAATCCATTGA

Other Transporters

CrPUP_1443_purine permease

ATGGAAGATCAAGTTAGGAGCAGGAGCAAGGAGGCACTAAGAAGAACAATACTATTGCTAAGTTTCAT
GTCAATGTGCATAGGTTTTTGCGGCGGTCCTTTAATCACACGCCTTTATTTCCTTCATGGAGGCAGTAGA
ATTTGGCTAGCGGGTTTTTTACAAACTTGTGCATGCCCAATTATTCTAATCCCGTTATTAATTTCTTATTTT
TCTCGTCGAAAATCGGAAGGATCCAGCACGAAGATTATTTACATGAAGATTTCAACTAGTTTAGCCTGT
TTAATTATTGGTCTTTTTGTTGGTTTAATTAATTACTTATTTTCCTATGGTTTAGGACGTTTGCCAGTTTCA
ACCGCTACACTAATTACCACAACTCAGCTTGTTTTTACTGCAGTCTTTGCTTTTCTTCTCGTCAAGCAAAA
ATTTACTCCATTTTCAATAAATGCAATTATATTGTTGACAATTGGATCAGTTGTTTTGGGCTTGCAAGGA
AGTAAAGATAGGCCAGAAGGTGAAAGTAAGAAAATGTACAATTTAGGATTTGCTTTGACGTTTTTAGG
AGCAGCCGTGAGTGGGTTCGTGGCACCTTCTATTGAATTTGTTTACAAGAAAACGGGGATGGGAGATG
GATATACAAAGGTGTTGGAGACTCAGATTGTGATTTCTTTTTCTGCTTCCATATTCTCCTTGATTGGAAT
GCTAGCAAACCATGATTTCAAGGTAATTCCACGGGAGGCAAGAGAATTTCAACTTGGCGAAGCAAAAT
ACTATGGAATTCTGGTGTCGAGTGCCTTTGTTTGGCAATTCTACCTTGTGGGAATCGTTGGAGTAATCTT
CAACGGATCGTCTCTGCTGTCCGGAATTATTGGGGCCGTTCTAGTTCCGGTTTCAGAAGTTTTAGCGGT
TATTTTTTACCATGAAAAATTTACGGCAGAAAAGGGAGTTTCTCTGGGTCTCTGTATCTGGGGTTTGGTT
TCCCACTTCTACGGTGAGACAAAACATAGCCAAAACACAAGTCAAAACCAAGCATTCGAGACAAGCTCA
GGTGTTTCTCCACAGAGCAAAAAGGAAATTGAAGTAATTTCAAATGTCCCGTGA

locus_1763_ABC transporter

ATGGCTTTCAAGCCATTTGAATGGTACTGCCAGCCAGTGAAGAATGGTGTGTGGTCAAAGGCTGTGGA
CAATGCATTTGGAGTTTATACTCCATGTGCAACAGATTCCCTAGTTATCGGATTATCATACTTGGTTGTG
TTTGTGTTGTGCTTATATCGAATTTGGCTGACTGAGAAAGATTTTAAGGTCCAGAGATTCCGTTTGAGG
TCAAAGTATTTCAATTATGTGCTGGGACTGTTGGCCTTATATTGCACTGCAGAACCACTGTTCAGACTGG
TAATGGGTATTTCAACATTTAATGTAGATGGACAGAATGGTCTGGCTCCATACGAGGTTGTATCTCTGA
TCATTGAGGCTTTGGCTTGGTGTGCTGTGCTTGTGATGATTGGTCTGGAAACAAAAATCTATATCCGTG
AGGCTCGTTGGTTTCTGAGGTTTGCAATCATTTATGCTTTGATTGGGGATGCTGTGATGCTCAATCTTGT
GCTTGGTGTGAGGGAATTTTACGAACGATCTGTACTCTATTTGTATATCAGCGAGATTGTTGTCCAGGT
TTTATTTGGAGTGCTCCTGCTCTTCTATGTTCCAAATTTGGATCCTTATCCAGGATACACCCCTGTGGAAA

CTGAATTTGTGGAGGATACTGCTTATGAAGAACTTCCTGGAGATGAGCAGATTTGTCCTGAGAGGCAT
GTCAATCTATTTTCCAAAATTACTTTTGCATGGATGCAACCCCTTATGAAACTGGGCTATAGACGACCTC
TAACAGATAAGGATGTGTGGAAACTGGACACTTGGGACAGGACGGAAACACTTATAAATGTGTTTCAG
AAATGTTGGGTTGAAGAGTCACAGAGGACTAAACCATGGCTTTTAAGAGCACTACATCGTGCACTTGG
CGGAAGATTTTGGTGGGGTGGCTTTTGGAAGATTGGTAATGATGTTTCCCAATTTATTGGGCCTCTTAT
ACTGAATCAGCTCTTGAAGTCCATGCAACAAGGAGAGCCAGCTTGGATTGGTTATATCTACGCATTTTT
GATTTTTGTTGGAGTGGTATTTGGGGTGTTATGTGAGGCACAATATTTTCAAAATGTGATGCGTGTTGG
TTATCGTTTGAGGTCAACCCTGATAGCTGCTGTTTTCAGGAAATCCCTAAGGCTAACCCATGAGAGTCG
TAAACAATTTGCAACTGGAAAGATAACCAATTTAATGACCACTGATTCTGAGGCACTACAGCAAATATG
CCAATCGCTCCATACTTTATGGTCTGCTCCTTTTCGTATCATTGTTGCCTTGGTCCTTCTTTACGATCAATT
GGGTGTTTCTTCCCTTGTTGGTGCACTAGGGCTAGTCCTCATGTTCCCAATTCAGACATTTGTGATAAGC
AAGATGCAGAAGTTATCAAAAGAAGGATTGCAGCGAACAGACAAGAGAATTGGACTCATGAATGAAA
TTTTGGCTGCCATGGACACTGTCAAATGTTATGCATGGGAGAATAGTTTCCAGAAGAAGGTTCAGGAT
GTACGTGATGATGAGCTGTCTTGGTTCCGGAAAGCCCAAATGTTAGGAGCGTTGAACAGTTTTATGCTG
AACAGTATTCCAGTTGTGGTGATTGTGATCTCATTTGGGATGTTCACTTTACTTGGAGGGGATCTTACAC
CTGCAAGGGCATTCACATCTCTTTCTTTGTTTGCTGTGTTGCGTTTTCCACTTTTCATGCTCCCAAACATA
ATAACGCAGGTGGTCAATGCAAATGTTTCTTTGAAACGTTTGGAGGAACTTCTATTAGCAGAAGAGAG
AATTCTTCTACCCAACCCCCCACTTGAACCAGGTCTTCCAGCCATCTCAATAAGGAATGGATGCTTCTCG
TGGGAATCAAAGGCAGAGAAACCTACATTATCAAATGTTAATTTGGATGTACCAGTTGGTAGCTTAGTG
GCAATTGTTGGCAGTACTGGAGAGGGAAAGACATCACTTGTATCAGCAATGCTTGGAGAGCTTCCAGC
AGTTACAGATGCAACGTCAGTTGTCCTCAGAGGATCAGTTGCTTTTGTCCCACAAGTTTCATGGATATTC
AACGCAACGGTAAGGGACAACATTCTGTTTGGATCTCCCTTTGAACCAGCACGATATGATAGGGCGAT
AGAGGTCACGTCGCTTCAGCATGACCTTGAATTGCTCCCTGGAGGTGATCTTACTGAAATTGGTGAAAG
AGGTGTAAACATTAGTGGAGGGCAAAAGCAAAGAGTGTCCATGGCTAGGGCTGTATATTCAAATTCAG
ATGTTTGCATATTCGATGACCCCCTAAGTGCTCTTGATGCTCATGTGGCTCGTCAGGTTTTCGAAAAATG
CATCAAGGGAGAACTCAGAGGAAAAACCAGAGTTCTTGTTACCAATCAGCTACACTTTCTTTCACAAGT
AGATAAAATTATCCTTGTGCACGATGGTATGGTGAAAGAGGAAGGAACCTTTGAAGAACTCTCAAATA
ATGGTGTGCTCTTTAAAAGACTGATGGAAAATGCGGGGAAAATGGAAGAATATGTGGAAGAAAAGGA
AAATGTTGAAAATACTGAAAATAAAACTTCTAATGCCATTGTTAATGGTGAGACCAATGGCTTGTTGAA
GGATCAAGACCCTGCAAACAAAAAGAAAGGAAAATCTATTCTTATCAAGCAGGAGGAACGTGAAACTG
GCGTTGTCAGTTGGAAAGTCTTAAACAGATACAAGAATGCACTGGGCGGCACTTGGGTAGTTATGATA
TTGTTTTTGTGTTATGTCCTAACGGAAGCTTTACGGATTTCAAGTAGTACATGGTTGAGTTATTGGACAG
ATAAAGGCGGTTCGAAAAATTATTCCCCCACTTTCTACAATTTGATCTACGCTCTCCTTTCATTCGGTCAA
GTAATGGTCACGCTGGCAAACTCATTTTGGTTGATCATCTCAAGTCTTTATGCTGCTCGAAAATTGCATG
ATGCCATGCTTGGCTCCATATTGAGAGCTCCAATGGTCTTCTTTCAGACGAATCCACTAGGGCGGATAA
TTAATAGATTTGCTAAGGATCTGGGTGACATAGACCGGATGGTTGCTCCATTTGTTAATATGTTTCTGG
GGCAAGTTTCACAGCTTTTTTCAACATTTGTGTTGATTGGAATAGTGAGCACCATGTCACTTTGGGCCAT
AATGCCCCTTCTTGTGCTGTTCTATGCAGCCTATCTGTACTATCAGAGTACAGCCCGTGAAGTAAAGCGC
TTAGATTCCATTAGCAGATCTCCTGTGTATGCACAGTTCGGAGAAGCACTCAATGGTCTAACAACTATTC
GTGCCTATAAAGCTTATGATAGAATGGCAAATATAAATGGGAAGTCAATGGACAACAATGTCAGGTTC
ACTCTTGTAACCATGAGTGGGAATCGCTGGCTTGCAATTCGTCTGGAGACTTTGGGGGGGCTTATGATT
TGGCTTACAGCAACTTTTGCTGTGATGCAGAATGGTAGGGCAGAAAACCAGCAAGAGTTTGCCTCCACT
ATGGGTTTACTTCTCAGTTATGCATTAAATATTACTAGTTTGTTGACTGCTGTACTTAGACTAGCTAGTAT
GGCCGAGAATAGTCTAAATTCCGTTGAGCGGGTTGGTACTTATATAGAGTTGCCTTCTGAAGGTCCAGC
TATCGTTAAGGACAACCGTCCCCCTCCTGGGTGGCCTTCTGCAGGATCTATTACTTTTGAGGATGTTGTC

CTGCGTTATAGGCCTGAACTTCCTCCAGTCCTCCATGGTATATCATTCAAAATTCCTCCAAGTGACAAGG
TTGGGATAGTTGGTAGGACTGGAGCCGGGAAATCTAGCATGCTTAATGCTTTATTCCGACTTGTAGAAC
TGGAAAAAGGAAGAATCCTGATTGATGATTGTGATATTGGAAGATTTGGATTGATGGACCTCCGTAAA
GTTCTTGGCATTATACCACAATCTCCAGTTCTTTTCTCAGGAACCGTGAGATTCAATCTTGACCCCTTCAA
CGAACACAGTGAATTTGATCTTTGGGAGGCTCTGGAGAGGGCACATTTGAAAGATGTTATAATGAGAA
ATTCTTCGGGTCTGGATACAGAGGTCTCGGAGGCAGGAGAAAATTTTAGTGTTGGACAGAGGCAACTC
TTGAGTCTTGCCCGGGCGTTGCTTCGGAGATCAAAGATTCTTGTACTTGATGAAGCAACAGCAGCTGTT
GATGTTCGAACTGATGCTCTTATCCAGAAGACCATTAGGGAGGAATTTAAATCTTGCACAATGCTTATT
ATTGCTCATCGTTTAAATACCATAATCGACTGTGACAGGATTCTTCTGCTTGATTCTGGCCGGGTTGTTG
AATATGATACTCCAGAAAGACTGCTGCATAACGAAGAGAGTTCATTCTCTAAGATGGTGCAAAGTACA
GGTGCTGCTAATGCACAATACCTGCGCAGCTTAGTACTTGATGCAGGTGACAAAGTGGAGAGCGAAAA
GCCCCATCTAGATGGTCAAAGGAGATGGCTTGCCTCCTCACGCTGGGCAGCTGCCGCCCAGTTTGCATT
GGCTGTTAGCCTCACTTCATCACACAATGACCTCGTGCAGCTGGAAATCAACGAGGAAGACAATATTCT
TAAGAAAACAAGAGACGCTGTAATAACTCTACAGGGAGTTTTGGAAGGGAAGCATGATCAATCCATTA
CAGAAACCCTCGACCAATACCACGTTTCTAAGGATAGATGGTGGTCGGCTCTATATAAAATGATCGAAG
GTCTTGCCGTGATGAGTAGGCTGGCTCGAAATAGGCTTCATCAATCAGAATATGAATATCAAGACAAA
ACAATTGACTGGGATCATGTTGCCATGTAA

locus_24210_WAT1 homologue

ATGGGGTTGACTTCATGTTTAACAGAAGCTTTGCCTTATGCAGCCATGGTGTTTGAAGAAATTCTTCTG
GTTGGTTTAACTACACTCACCAAAAATGCCTTATCCAAAGGAATGGACCGCTATGCCTTTATGTTTTATG
CTAATGTGCTTGCTACTCTTATCCTCTTTCCTCTTTCTTTCATCTTCAAAAGAAAGGAAATTCCTCCTTTGA
GCTTCTCTCTTCTTGGCAAGTTTTTCTTCATAAGCTTGATTGGGGTTACAATTAACCAGAATGTTATATAT
GCTGGACTAGATTACAGCTCTCCCACACTTTTATCTGCCTTGAGGAACTTAATCCCAGCTTTCACCTTTTT
GCTTGCTGTTTTATTCAGGATGGAAATTTTGAATTTGAGAAGCTCAAGAAGTCAGATCAAGATTATAGG
AACATTGGTGACAATTTTAGGAGCATTAACTATAACCTTTTATAATGGTTTTCCAATTGGGTATTCAGGA
GATGAATCCTCATTGAAAAGATTTCTTGAATCACCAAACAATTTCTCTGCATTATTTTCAACACAAAGAA
ATTGGGCAATTGGAGGCCTCTGTTTTGTTATTTCAGCTTTATGCATTGCACTATGGATAACAGCTCAGGG
AGCAATTGTTAAGGAGTATCCATCAGAGTTAACCATTGTGGCATACTATTTCTTGTTTGGGACAATTCAA
TCTGCTGCAGTCTCTTTAGTTGGTGAAAGACATGAACCAAATCCTTGGACGATAAGGACAAATATTGAG
CTCCTCTCAGTTACCTGCTCAGCTTTAATTGGAAGTGTGATGGTATTTTTGTTGGATATATGGTGCATAC
ACAAAAAGGGACCTGTATTTGCAGCCATGTTCAAGCCTCTAGGCATTGCTATTGCTGCAATTGCTGGTG
TCATCTTTCTTGGCGAAACTCTCTACATTGGAACCATTATAGGATCAATTGTAATAATAATAGGTTTCTAT
GGAGTAATGTGGGCAAAGGCAAAAGAGAATGAAGATTTGTTAGATGATTATCTACCTTCATCTTCAAA
GAAGATGCCTTTGTTGCAAAATCTCAGAGATGATGTATAG

locus_1770 amino acid permease

ATGACAATTGGAAGCCAAAAACCAGCCAAAGAAAGGAAGTCCAGGAAGAGCAAAAAAGTTGTGGTTC
ATGAAAAATCACCCCTTTTGCCCACTAAGCATGAGGAGGATGCTGGGTTTGATGAGTTTGATGGAGCTT
CCTTCAGTGGGGCAGTTTTTAATCTGTCGACCACTATTGTTGGTGCTGGAATCATGGCCTTGCCAGCTAC
TATGAAAGTGCTGGGCCTTATTCCTGGGATTGCTGCTATCATCTTTATGGCATTCTTGACAGATGCTTCA
ATTGAGTTGTTGATTAGGTTTAGTAGAGCTTCTAAATCAGTTTCTTATGGAGGTGTTATGGGGGATGCC
TTTGGGAAGTATGGTCGCATGTTTCTGCAAATATGTGTTCTAGTAAACAACATAGGCGTCCTTGTTGTAT
ACATGATTATCATAGGTGATGTGCTTTCTGGAACAACATCAAGTGGGACTCACCATGCTGGTGTCTTGG

AAGGATGGTTTGGAGTACACTGGTGGAATGGACGTTTTTTTGTTCTCCTAATCACTACCCTTGGCATCTT
TGCACCTTTGGCAAGTCTAAAGCGCATTGATTCTTTGAGGTTCACATCTGCATTATCAGTTGCACTAGCT
GTGGTCTTCCTAGTTATTACGGTGGGCATTACAGTTTTCAAACTGATCAATGGAACAGTGTTGATGCCT
AGACTACTTCCTGATGTGCATGATCTGACATCATTCTTTAAACTCTTCACTGTGGTTCCTATCCTTGTCAC
TGCTTACATCTGTCACTACAATGTTCACTCGATAGAAAATGAACTTGAAGACAACACCCAGATACAAGC
TGTTGTACGGACCTCACTTGCTCTGTGCTCATCTGTATATGTGATGACCAGCCTCTTTGGATTCCTCCTAT
TTGGTGATGCAACTCTTGACGATGTCCTTTCCAACTTTGACTCTAATCTGGGGATTCCTTATGGCTCCCT
GCTTAATGATGCCGTTCGCATTAGCTATGCTGCACACTTGATGCTCGTCTTCCCCATTGTGTTCTATCCGC
TGAGGTTGAACTTGGATGGCCTTCTCTTTCCGTCTGCTAGGCGCCTGACATCTGACAATCTGAGGTTCG
CATTATTAAGCATCAGTCTCATTGCCCTCATCTTTCTGGGTGCTAATTTCATACCCAGCATCTGGGACGCT
TTCCAATTTACTGGAGCAACTGCTGCTGTTTGCATTGGATTCATCTTCCCTGCTGCCATTACTCTCAGGG
ATCGATATGGTATAGCTACAAAGAGGGACAAGATCCTATGCATATTCATGATCGTCCTTGCTGTATTCTC
AAACGCGGTGGCCATATATAGTGATGCATATGCACTGTTCAAGAAAAATGCATCACCCAGCAAATGA

Glutathione *S*-transferases

locus_7197_GST1

ATGGGTACAGAAAGCATAGAGAGAAAGAGAAAAAGAGATCAAGAAGAAACTGAAAAAATGGCTGAA
GAAAAGAGGGTTAAATTGCTGGGATTTTGGTCAAGTGCTTATAGTATAAGAGTAGATTTAGCACTCAAT
GCAAAGGGAATTCCATACGAATATCAAGATGAAGATCTGAGGAACAAAAGTGAATTGCTTCTGAAAAC
AAACCCAATTCACAAAAAAGTACCAGTTTTGATTCATAATGGAATTCCCGTTGCTGAATCTCTTATCGTC
CTCGAATACATCGATGATACTTGGCCTCAATCTCCCAAATTTGTCCCTGAGGATCCTCTCTTGAAGGCCA
AGATTCGTTTCTGGTCACACTTTGTCCACTTTCAGCTGTTCGAGAGCATGCTTCCAATAGCACTAACAGC
AGGGGAAGCACAAGAAAAGGCATTGGAGACATTTTATGGTAATCTAAAGACAATTGAAGATGGTACCA
AAGATCTTTTTGAAGATGGAAAGCCATTAATTGATGAAAACAACTTGAACTTGTTGGACATAATAATAT
GGTCAGTTCTTGGTCCTCACAAATCATTTGAAGAAGCTCTTGATTTGAAGCTTTTGGATCCTAAAAAATA
CCCATTTTTACACTCATGGGTTGGTGCCATTAACGGACTCAAGGCGGCAAAAAAGGCAATCCCAGATCA
TAATACTATGGTTACCCATCTTAAGCATTTCAGGCAAATGGCTCTCCAGCGCTCCGCACCCTCCTAA

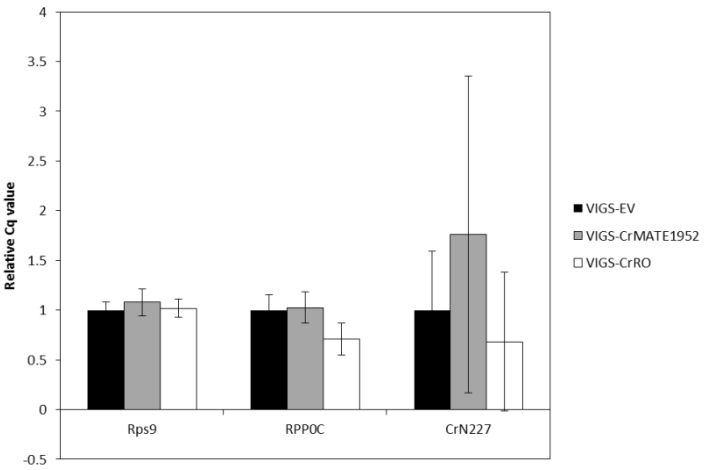
locus_6502_GST2

ATGGCGAACGAGAAAGGAAAACTTAAGCTTCTTGGCTATTGGTTTAGCGCTTACTGCAAGAGGGTTGA
ACTAGCCCTTAATGCTAAAGGAATACCCTTTGAGTACCAAGAAGAAGATTTAACTAACAAGAGTGAATT
ACTCTTGAAATCTAATCCAGTCAAAAAGGTTCCTGTGGTTCTCCACAACGGGACACCCATATGTGAATCC
CTCATAATTCTTGAGTATATTGATGAAACTTGGACTAATTCTCCAAAACTTTTACCTGAGGATTCTGCTAC
TAGATCCAACTTTCGCTTTTGGGCAAGCTTCTTCCATATCCAGCTATTCGATAGCCTATATCCAGCGGTG
ACATCTGAGGGAGAGGAACAAGAAAAGGCCTTTTCCGTTTACTCTGCAAATCTGAAAACGGTTGAAGA
AGGAATTGCCGCCGAGTTTTTCGGCAGCGGCGGCGGACGAGAGGAGATAAAGTCGGAGGAGAATTTG
AACTTGGTAGATGTTGTCATATGGTCAGTTTTGGGTTGGTATAAAGCATTTGAATCAATAGAATTGAAC
ATTTTGGATGCTGACAAGTATCCATTAACATACTCATGGGTTACTAAGCTCAACAATCTGTCTGTTGTTA
AGAAAACAGCAGCTGATACTGCTAAAGTCGCCTCTGAAATCAACTTCTTCCGTAACCTCGGTCTCCAGA
AGGTCGCTGCTCATGCATGA

locus_2878_GST3

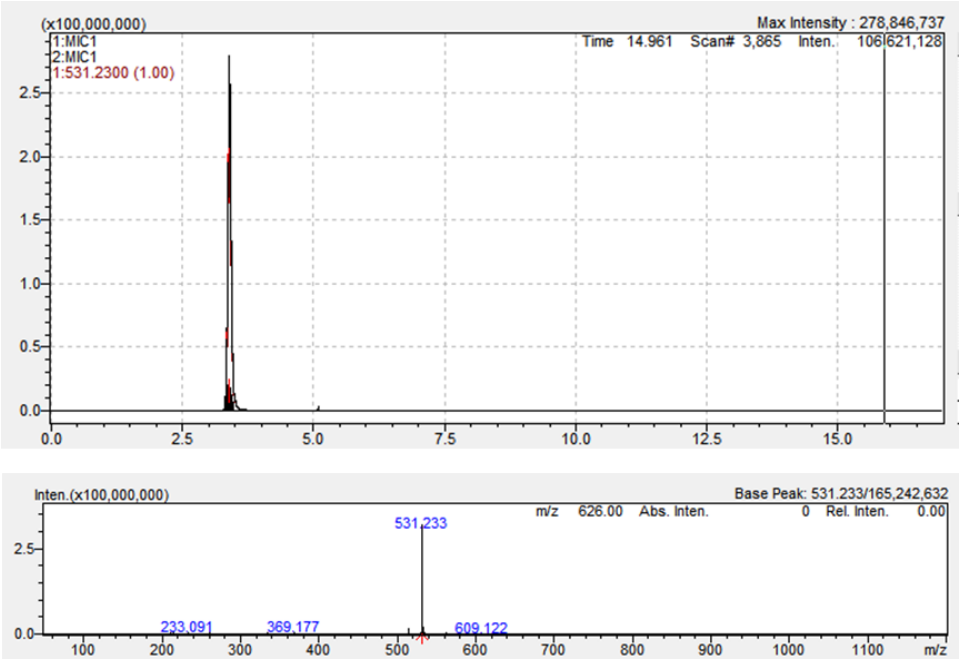
ATGGCTGGAGAATTTGTTCTATTGGATACTTATGTGAGTATGTTTGGAATGAGGGCAAGAGTTGCTTTG
AATGAAAAGGGTTTGAATTACGAGTACAAAGAAGAGAATTTGGGCAGCAAAAGCCCTCTGCTTCTTCA
GATGAACCCAATTCATAAGAAAATCCCTGTTTTGATTCATAATGGAAAACCTATTTGTGAGTCCCTTATT
ATTGTTCAGTATTTAGATGAAATTTGTCCTGATAAGAATCCATTGTTGCCTACTGATCCTTATGAGAGAG
CTCAAGCTCGCTTTTGGGCTGATTATATTGACAAGAAGGTATACGATGCTGGAAGAAAAATATGGACA
ACAAAAGGGGAAGAGCAAGAAGCAGCCAAGAAAGAGTTCATTGAAATCCTCAAAACTTTGGAAGGTG
TACTTGGGGATAAACCATACTTTGGTGGTCAAGAATTTGGATTTGTGGATGTGGCTTTGATCCCATTTTA
CAGTTGGTTCCATGCCTATGAGACATCTGGAAATTTGAAGATAGAGCCAGAAGTTCCAAAACTAGTGG
CATGGGCCAAGAAGTGCATGCAAAGGGACAGTGTGGCAAAGGGACTTGCTGATCCTCACAAAGTTTAT
GAATTTATTTTGTCCATGAAGAAGAGGTTGGGAATTGAGTGA

9.6 qPCR reference genes

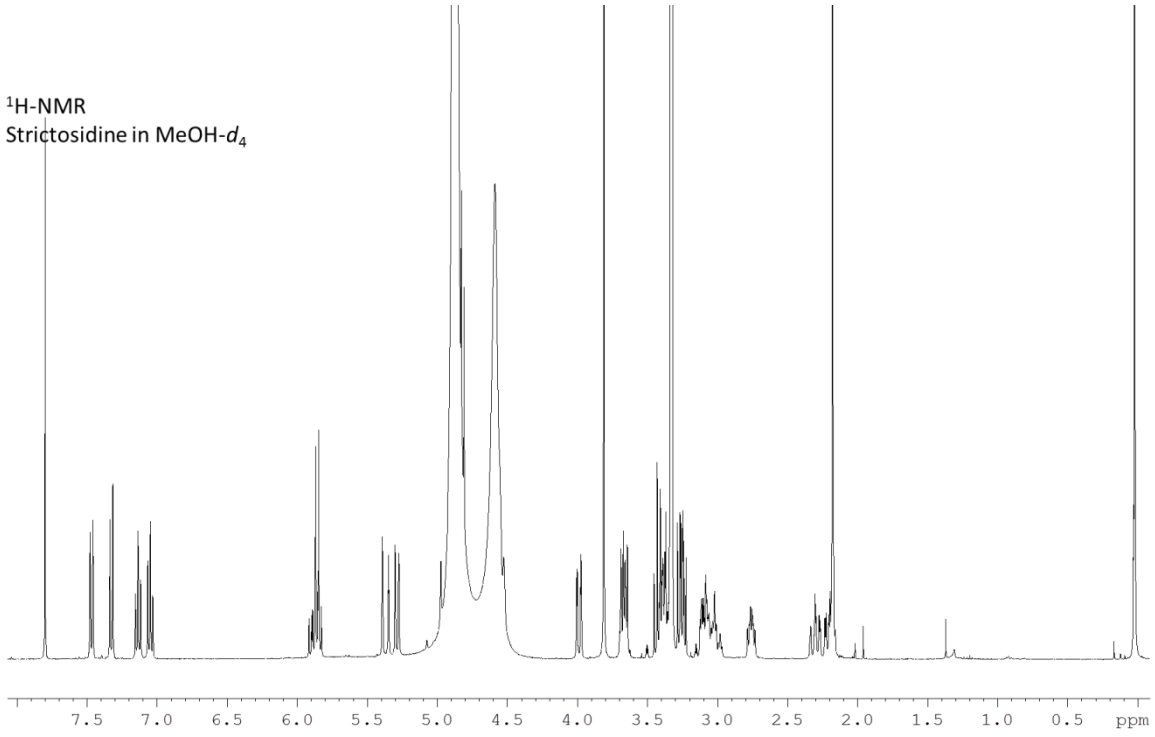


Relative expression level of the reference genes Rps9, RPP0C and CrN227 against VIGS-EV, VIGS-CrMATE1952 and VIGS-CrRO tissue. As can be seen the reference gene and primers proposed by (Pollier et al., 2014), CrN227, resulted in greater variation between VIGS experiments than the previously characterised Rps9 and RPP0C primers that have been used in *C. roseus* VIGS experiments in our hands. Therefore Rps9 or RPP0C were preferentially used for data normalisation in the qPCR experiments. All data shown are mean ± SEM.

9.7 Purification of strictosidine



LC-MS chromatogram and mass spectra of purified strictosidine



Proton NMR of purified strictosidine