Transient expression for engineering triterpenoid diversity in plants

James Ashley Reed

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

John Innes Centre

Norwich Research Park

Norwich

Norfolk

United Kingdom

Abstract

The triterpenes are one of the largest and most structurally diverse classes of plant specialised metabolites, with important commercial, agronomical and medical potential. However the structural complexity and low abundance of these compounds in nature presents significant challenges for exploiting this diversity. In recent years, advances in our understanding of triterpene biosynthesis in various plant species have greatly facilitated the ability to produce these compounds in other hosts. Plant-based host systems hold great promise because they may provide substrates, cofactors and subcellular compartmentalisation mechanisms facilitating engineering production of complex triterpenoids. Agroinfiltration is a rapid and scalable process that enables transient expression of biosynthetic genes in amenable plants such as *Nicotiana benthamiana*. This technique holds great promise for selective production and modification of triterpenes, but studies in this area have been limited.

In this PhD thesis, transient expression is used to produce and oxidise triterpenes in *N. benthamiana.* Triterpene yields are improved through expression of rate-limiting genes in the mevalonate pathway (Chapter 3). An enzymatic 'toolkit' is established from a collection of previously characterised triterpene biosynthetic genes, allowing production of milligram quantities of a diverse array of simple and oxidised triterpenes in *N. benthamiana*. Furthermore, combinatorial biosynthesis is utilised to engineer production of novel oxidised triterpenes *in planta* (Chapter 4). A set of oxidised derivatives of β-amyrin are purified and evaluated for antiproliferative and anti-inflammatory activity in human cell lines, yielding insights into structural features underlying these activities in humans (Chapter 5). Finally, *N. benthamiana* is used for functional screening of cytochrome P450s implicated in the biosynthesis of agronomically important antifungal triterpene glycosides (avenacins) in oat, leading to identification of new P450s required for disease resistance (Chapter 6). This work highlights the utility of *N. benthamiana* as triterpene engineering platform and provides a basis for future studies exploring triterpene structure-activity relationships.

Acknowledgements

I would like to express a huge thanks to my supervisors, Prof Anne Osbourn at JIC and Prof Maria O'Connell in the School of Pharmacy at UEA both for giving me the opportunity to take on such a unique and fulfilling project, as well as for always giving me fantastic support and advice needed throughout. I would also like to extend additional thanks to Anne for giving me so many opportunities over the last few years to embrace my creativity during this PhD and to learn and develop in areas outside this project.

I would also like to say an additional thank you to the members of the Osbourn lab, both current and former for making the last four years such an enjoyable experience. Without your help and talents many aspects of this project would not have been possible. In particular I would like to say a huge thank you to Aymeric Leveau, Tessa Moses, Thomas Louveau and Michael Stephenson both for help in practical and scientific matters but also your support through the tougher times, particularly in the last year.

I am also hugely grateful to Lionel Hill and Paul Brett within the metabolite services team at JIC for always taking the time to provide excellent help and advice with analytical chemistry. Particularly, without Paul's insight into the many complexities of GC-MS this thesis would have turned out very differently.

Finally I would like to thank my family, without whose encouragement and good humour I would not be here. In addition, my partner Rhiannon's family, for all the support they've provided during the last four years and the years before. Lastly and most importantly thanks to Rhiannon for her always unwavering support at home and constant encouragement throughout this PhD.

List of common abbreviations

AsbAS1	-	*Avena strigosa* β-amyrin synthase
CA	-	Cycloartenol
CP	-	Chloroplast
CPMV	-	Cowpea mosaic virus
DD-II	-	Dammarenediol-II
DMAPP	-	Dimethylallyl diphosphate
DOS	-	Dioxidosqualene
EI	-	Electron-impact
ELISA	-	Enzyme-Linked ImmunoSorbent Assay
EpDM	-	Epoxydammarane
EpHβA	-	12,13β-epoxy,16β-hydroxy-β-amyrin
ER	-	Endoplasmic Reticulum
ESI	-	ElectroSpray Ionisation
FPP	-	Farnesyl diphosphate
FPS	-	Farnesyl diphosphate synthase
GC-MS	-	Gas Chromatography-Mass Spectrometry
GFP	-	Green Fluorescent Protein
GGPP	-	Geranylgeranyl diphosphate
GPP	-	Geranyl diphosphate
GPPS	-	Geranyl diphosphate synthase
HMGR	-	3-hydroxy,3-methylglutaryl-CoA reductase
HT	-	Hypertranslatable
IPP	-	Isopentenyl diphosphate

IS	-	Internal Standard
Keap1	-	Kelch-like ECH-associated protein 1
LC	-	Liquid Chromatography
LPS	-	Lipopolysaccharide
MEP	-	2-C-methyl-D-erythritol 4-phosphate
MVA	-	Mevalonate
NFκB	-	Nuclear Factor Kappa B
NMR	-	Nuclear Magnetic Resonance
Nrf2	-	Nuclear Factor Erythroid 2 Related Factor 2
OS	-	Oxidosqualene
OSC	-	Oxidosqualene cyclase
P450	-	Cytochrome P450
rDA	-	Retro Diels-Alder
RT-PCR	-	Reverse Transcriptase PCR
SAD1	-	*Avena strigosa* β-amyrin synthase
SAD2	-	*Avena strigosa* CYP51H10
SQE	-	Squalene epoxidase
SQS	-	Squalene synthase
T-DNA	-	Transfer DNA
tHMGR	-	truncated HMGR
TIC	-	Total Ion Chromatogram
TLC	-	Thin Layer Chromatography
TMS	-	Trimethylsilyl
TNFα	-	Tumour Necrosis Factor alpha

Contents

Abstract I

Acknowledgements II

List of common abbreviations III

Contents V

List of Figures X

List of Tables XIII

Chapter 1 – General Introduction 1

1.1 – Plant specialised metabolism 2

1.2 – Terpenoids: functions in nature and human uses 3

1.3 – Triterpenes: Importance and bioactivity. 5

1.4 – Triterpenoid biosynthesis in plants 8

1.4.1 – Oxidosqualene cyclisation generates diversity in triterpene scaffolds 11

1.4.2 – Oxidation allows further diversification of triterpenes 13

1.4.3 – Clustering aids identification of new triterpene oxidases 16

1.5 – Engineering diversity in triterpene biosynthesis 22

1.5.1 – Combinatorial biosynthesis 22

1.5.2 – Protein engineering augments combinatorial biosynthesis 25

1.6 – Triterpene production in heterologous hosts 26

1.6.1 – *Nicotiana benthamiana* is an emerging host for producing plant specialised metabolites 28

1.6.2 – Realising *N. benthamiana* as a (tri)terpene production platform 29

1.7 – PhD overview: 31

Chapter 2 – General materials and methods 32

2.1 – General methods 33

2.1.1 – Gateway cloning 33

2.1.2 – Microbiological techniques 37

2.1.3 – Agroinfiltration 38

2.1.4 – General considerations for purification of compounds 42

2.1.5 – GC-MS programs: 44
2.1.6 – Common reagents and solutions 44
Chapter 3 – Improving triterpene production in *N. benthamiana* 46
3.1 – Introduction 47
3.1.1 – Engineering early pathway steps in triterpene production 48
3.1.2 – The importance of compartmentalisation in terpene biosynthesis 51
3.1.3 – Aims 52
3.2 – Results and discussion 53
3.2.1 – Production of β-amyrin in *N. benthamiana* 53
3.2.2 – Improving production of β-amyrin in *N. benthamiana* with MVA genes 54
3.2.3 – Combination of tHMGR and SQS further improves β-amyrin yields 57
3.2.4 – SQE allows production of alternative cyclisation products 58
3.2.5 – Oxidation of the β-amyrin scaffold in *N. benthamiana*. 60
3.2.6 – Exploring alternative subcellular compartmentalisation for triterpene production. 62
3.2.7 – Plastid-synthesised squalene is unavailable for the cytosolic biosynthesis of β-amyrin. 66
3.3 – Conclusions and perspectives 69
3.3.1 – Discussion of results 69
3.3.2 – Investigating other candidate enzymes 70
3.3.3 – Further optimising production 71
3.4 – Materials and methods for Chapter 3 72
3.4.1 - Searching the oat 454 database for MVA pathway genes 72
3.4.2 – Cloning and generation of expression constructs 72
3.4.3 – Generating constructs for chloroplast-targeting work 73
3.4.4 – Primers 75
3.4.5 – Infiltrations 76
3.4.6 – Extraction and quantification of β-amyrin in plant leaves 77
3.4.7 – Dispersive liquid-liquid microextraction 78
3.4.8 – Scaled production and purification of EpHβA 78

3.4.9 – Expression and analysis of chloroplast-targeted *FPS* and *SQS* 79
Chapter 4 – Engineering triterpene diversity in *N. benthamiana*..... 80
4.1 – Introduction 81
4.1.1 – Combinatorial biosynthesis of triterpenes in *N. benthamiana* 81
4.1.2 – Analysis of triterpenes by mass spectrometry 81
4.2 – Aims 83
4.3 – Results and discussion. 84
4.3.1 – Part 1: Establishing a toolkit for functionalisation of β-amyrin 84
4.3.2 – Part 2 – Combinatorial biosynthesis in *N. benthamiana* 91
4.3.3 – Incorporating CYP51H10 mutant enzymes into the toolkit. 101
4.3.4 – Combinatorial biosynthesis of CYP51H10 mutant enzymes 103
4.3.5 – Part 3 – Beyond β-amyrin: Diversifying triterpene production in *N. benthamiana* 106
4.4 – Conclusions 110
4.5 – Materials and Methods for Chapter 4 111
4.5.1 – Cloning toolkit enzymes 111
4.5.2 – Primer sequences 113
4.5.3 – Testing the toolkit enzymes in *N. benthamiana* 113
4.5.4 – Combinatorial biosynthesis 113
4.5.5 – Extraction and analysis of leaf samples 114
4.5.6 – Production, analysis and purification of 11-oxo,24-hydroxy-β-amyrin and 11-oxo,24-acetoxy-β-amyrin. 114
4.5.7 – Production and purification of 11-oxo-oleanolic acid 115
4.5.8 – NMR spectroscopy 116
Chapter 5 – Investigating structural features underlying triterpenoid bioactivity ... 119
5.1 – Introduction 120
5.2 – Aims 122
5.3 - Results and discussion 123
5.3.1 - Obtaining a series of oxidised β-amyrin derivatives 123
5.3.2 – Investigating anti-proliferative activity 125

5.3.3 – Utilising combinatorial biosynthesis to further investigate structure-function relationships 128
5.3.4 – Investigating the anti-inflammatory effects of oxidised β-amyrin compounds. 130
5.3.5 – Results summary: insights into structural features underlying activity 133
5.4 – Conclusions and perspectives 136
5.5 – Materials and Methods for Chapter 5 139
5.5.1 – Production and purifications of compounds from *N. benthamiana* 139
5.5.2 – Preparation of compounds for assays 141
5.5.3 – Culture of human cell lines 142
5.5.4 – MTS antiproliferation assay 143
5.5.5 – Anti-inflammatory assay 143
Chapter 6 – Identification of P450s required for avenacin biosynthesis in oat 146
6.1 – Introduction 147
6.1.1 – Saponins in *Avena* species 147
6.1.2 – Structure and biosynthesis of avenacins 147
6.1.3 – Approaches for discovering new *Sad* gene candidates. 149
6.1.4 – Aims 151
6.2 – Results and discussion 152
6.2.1 – Identification of new avenacin P450 candidates 152
6.2.2 – Expression of P450 candidates in *N. benthamiana* 154
6.2.3 – CYP72A475 is synonymous with the *Sad6* locus 158
6.2.4 – A CYP94D family member lies next to *CYP72A475* 162
6.2.5 – Product profiling of *CYP94D65* 164
6.3 – Conclusions and perspectives 169
6.3.1 – P450s involved in avenacin biosynthesis. 169
6.4 – Materials and Methods for Chapter 6 172
6.4.1 – Identification of P450 candidates 172
6.4.2 – Growth and harvesting of *A. strigosa* seedlings 172
6.4.3 – RNA extraction from oat seedlings and cDNA synthesis 172

6.4.4 – Cloning of candidate *A. strigosa* P450s 173
6.4.5 – Cloning Primers 173
6.4.6 – Infiltration, extraction and analysis of *N. benthamiana* leaves 174
6.4.7 – LC-MS analysis of oat mutants 174
6.4.8 – RT-PCR profiling 174
6.4.9 – qPCR analysis 175
6.4.10 – BLAST searches of the *A. strigosa* genome 176
6.4.11 – Extraction and GC-MS analysis of oat roots 176
Chapter 7 – General Discussion 178
7.1 – Introduction 179
7.2 – *N. benthamiana* as a platform for triterpene production 179
7.3 – Further enhancing triterpene production in *N. benthamiana* 180
7.4 – Expanding triterpene diversity 182
7.4.1 – Engineering triterpene glycosides 183
7.4.2 – Expanding the enzyme collection further 184
7.5 – Understanding structural features behind biological activity 184
References 188
Supplementary Figures 205
Supplementary Tables 232
Appendix 1 – Publications 237

List of Figures

Figure 1-1: Examples of terpenes found in the plant kingdom and their functions 4
Figure 1-2: Functions of triterpenoids 6
Figure 1-3: The MVA and MEP pathways 10
Figure 1-4: Cyclisation of oxidosqualene to form sterols and triterpenes 12
Figure 1-5: Activity of various triterpene oxidases on the β-amyrin scaffold 14
Figure 1-6: Oxidases of non-β-amyrin triterpene scaffolds 16
Figure 1-7: Clustering of OSCs and P450s 18
Figure 1-8: Combinatorial biosynthesis of triterpenoids 23
Figure 1-9: β-amyrin oxidases which are also functional on other pentacyclic scaffolds 24
Figure 1-10: Using mutant enzymes in combinatorial biosynthesis 26
Figure 1-11: Synthetic biology for engineering plant specialised metabolites 27
Figure 1-12: Strategies to increase terpenoid yield in *N. benthamiana* 30
Figure 2-1: Infiltration of *N. benthamiana* by hand 40
Figure 2-2: Vacuum infiltration of *N. benthamiana* 41
Figure 3-1: Metabolic engineering for increased terpenoid yields in *N. benthamiana* 48
Figure 3-2: Triterpene biosynthesis 49
Figure 3-3: Production of β-amyrin in *N. benthamiana* 54
Figure 3-4: Manipulation of the MVA pathway in *N. benthamiana* 56
Figure 3-5: Phenotypes of leaves following agroinfiltration 57
Figure 3-6: Effects of expressing multiple MVA pathway genes in *N. benthamiana* 59
Figure 3-7: SQE allows for select production of dioxidosqualene (DOS) in *N. benthamiana* 60
Figure 3-8: tHMGR enhances production of 12,13β-epoxy, 16β-hydroxy-β-amyrin (EpHβA) 61
Figure 3-9: Chloroplast targeting of FPS and SQS in N. benthamiana 63
Figure 3-10: Chloroplast targeting of triterpene biosynthetic genes 65
Figure 3-11: Impact of chloroplast production of squalene on cytosolic β-amyrin synthesis 67
Figure 4-1: EI fragmentation of β-amyrin 82
Figure 4-2: Oxidation of β-amyrin by CYP51H10 84
Figure 4-3: Oxidation of β-amyrin by CYP716A12 86
Figure 4-4: Oxidation of β-amyrin by CYP93E1 87
Figure 4-5: Oxidation of β-amyrin by CYP88D6 88

Figure 4-6: Oxidation of β-amyrin by CYP72A63, CYP72A65 and CYP72A154 89
Figure 4-7: Summary of the major products produced by the β-amyrin oxidases expressed in the present study 91
Figure 4-8: Schematic of combinatorial compounds 94
Figure 4-9: Production of 11-oxo,24-hydroxy-β-amyrin and acetylated derivatives 97
Figure 4-10: Transesterification of 24-hydroxy-β-amyrin 99
Figure 4-11: Combinatorial biosynthesis of CYP716A12 and CYP88D6 100
Figure 4-12: CYP51H10 model and mutant product profiles 102
Figure 4-13: Combinatorial products formed using CYP51H10 mutants 104
Figure 4-14: Production of diverse triterpene scaffolds in *N. benthamiana* 107
Figure 4-15: Production of ginsenosides in *N. benthamiana* 109
Figure 5-1: Synthetic triterpenes from β-amyrin 121
Figure 5-2: Structures of oxidised forms of β-amyrin investigated in this study 123
Figure 5-3: Schematic of the MTS assay 126
Figure 5-4: Additional products tested for antiproliferative activity 129
Figure 5-5: TNFα production in LPS-stimulated THP-1 cells incubated with triterpenes 132
Figure 5-6: Dose-dependent TNFα production following incubation with different triterpene concentrations 133
Figure 5-7: Mechanisms of glycyrrhetinic acid anti-inflammatory activity 135
Figure 5-8: Conjugate addition of C-ring enone triterpenes 135
Figure 5-9: Possible future test molecules based on the results of this study 137
Figure 5-10: Schematic of the T-DNA portion of pAGM4723 139
Figure 6-1: Avenacins in oat 148
Figure 6-2: Biosynthesis of avenacins in oat 150
Figure 6-3: Analysis of the expression profiles of candidate P450 genes 154
Figure 6-4: Analysis of *N. benthamiana* leaf extracts by GC-MS 156
Figure 6-5: LC-MS analysis of oat roots 159
Figure 6-6: ESI mass spectra for compounds in root extracts of *A. strigosa sad* mutants 160
Figure 6-7: Analysis of *CYP72A475* in oat mutants #825 and #1243 162
Figure 6-8: Clustering and coexpression of novel avenacin biosynthetic genes ... 163
Figure 6-9: Activity of *CYP94D65* towards β-amyrin 165
Figure 6-10: Formation of putative C-23 oxidised EpHβA in *N. benthamiana* 167
Figure 6-11: Formation of novel products through transient expression of *CYP94D65* with *CYP51H10* and *CYP72A475* 168

Figure 6-12: Phylogenetic tree of cytochrome P450s implicated in triterpenoid biosynthesis 170
Figure 6-13: Combinations of oat P450s expressed in this study and the predicted products 171

List of Tables

Table 1-1: P450s involved in triterpene biosynthesis 19
Table 2-1: A list of commonly used primer sequences 36
Table 2-2: List of plasmids used in the present thesis 36
Table 2-3: List of antibiotics used for culturing bacteria 38
Table 2-4: Flash chromatography programs 43
Table 2-5: List of triterpene standards 44
Table 3-1: Reported yields of various terpenes produced in *N. benthamiana* 47
Table 3-2: Primer sequences as used in Chapter 3 75
Table 4-1: β-amyrin oxidases used in this study 85
Table 4-2: Summary of P450 combinations 92
Table 4-3: Summary of combinations involving CYP51H10 mutants 103
Table 4-4: Primer sequences as used in Chapter 4 113
Table 4-5: NMR assignments for 11-oxo,24-hydroxy-β-amyrin 117
Table 5-1: Triterpenes tested in the present study 124
Table 5-2: Antiproliferative effects of oxidised triterpenes 127
Table 5-3: Antiproliferative activity of oxidised triterpenes 130
Table 5-4: Compounds used for the MTS assays and ELISAs 142
Table 6-1: P450s identified in this study 153
Table 6-2: Primers used to clone candidate P450s from *A. strigosa* 173
Table 6-3: RT-PCR thermal cycle parameters 175
Table 6-4: Primers used for RT-PCR expression profiling of P450s *A. strigosa* 175
Table 6-5: Primers used for q-PCR expression profiling of *A. strigosa* P450s 176

Chapter 1:

General introduction

1.1 – Plant specialised metabolism

Plants have adapted to the constraints of a sessile lifestyle through chemical innovation. They produce a huge range of specialised metabolites that are likely to have important ecological functions. For humans, these specialized metabolites are an important source of therapeutics such as painkillers, anti-cancer and anti-parasitic compounds [1-3]. Because of the chemical complexity of these compounds, plants are still the primary source. Unfortunately, many of the source species are difficult to cultivate, slow-growing and provide only low yields of the desired product. These factors create significant supply issues and high market prices for many of the most highly valued chemicals such as taxol, vincristine and artemisinin. Alternative production methods such as plant cell or hairy root cultures have seen only limited success and have often proven difficult to scale [4].

Knowledge of the enzymes responsible for biosynthesis of metabolites of interest offers the possibility of engineering production in more tractable species. Recent advances in sequencing technologies are now rapidly making this possibility a reality. Consequently we are entering a new epoch in production of high-value compounds from plants in heterologous hosts [5-7]. Furthermore, as the number of characterised pathways increases, this presents the exciting opportunity to tap into the diversity in plant specialised metabolism to engineer new compounds. Success in this endeavour will require not only the elucidation of the relevant pathways, but also the development of suitable host systems for their production.

In the present PhD thesis, *Nicotiana benthamiana* is explored as a host system for the production and purification of triterpenes, a therapeutically and commercially important class of plant specialised metabolites. This introductory chapter focuses on the background to this work. The importance of triterpenes for both plants and humans is first discussed. Current understanding of the biosynthesis of triterpenoids is also described with a key focus on scaffold generation and oxidation by cytochrome P450s. This chapter further considers how to address the demand for structurally diverse triterpenes through discovery of new enzymes, combinatorial biosynthesis and engineering enzyme function.

1.2 – Terpenoids: functions in nature and human uses

Terpenoids (or isoprenoids) are the most diverse class of plant secondary metabolic products, with in excess of 46,000 different plant terpenoids reported [8]. They form a hugely structurally diverse class of natural products built from condensation of basic units of 5-carbon isoprene diphosphate precursors [9]. The classes of terpenoids found in plants include monoterpenes (C-10), sesquiterpenes (C-15), diterpenes (C-20), triterpenes (C-30) and tetraterpenes (C-40 - commonly called carotenoids). Terpenes are associated with a diverse array of essential biological functions. Small linear terpenoids serve as membrane-anchoring tails for proteins as well as components of the respiratory and photosynthetic electron transport chains [9, 10]. Others such as sterols have essential structural roles in regulating membrane fluidity, while carotenoids function as photosynthetic accessory pigments [9]. Terpenoids also have key plant developmental roles as they are necessary for biosynthesis of phytohormones, including brassinosteroids, gibberellins, strigolactones, cytokinins and abscisic acid [9, 11].

In addition to these primary functions, terpenoid biosynthesis in plants has been recruited towards specialised functions in defence, survival and communication. For example, bitter terpenoids may serve as general antifeedant agents against a range of pests and herbivores [12, 13]. Others display highly specific physiological interactions with target organisms, such as the potent insecticidal limonoid triterpene azadiractin [14] and phytoecdysteroids, which mimic insect moulting hormones [15]. Some terpenoids may serve to confer a competitive advantage against other plants. The diterpene momilactone B is a known allelopathic agent that is produced in the roots of rice and that inhibits the growth of *Echinochloa* weed species [16]. Low molecular-weight terpenoids are also volatile and fulfil a number of communicative roles between plants and animals [17-19]. Hence, beyond their roles in primary metabolism, terpenes serve vital roles in mediating interactions between plants and their environments and ensuring survival (**Figure 1-1**).

For humans, terpenoids have great economic value. Plant essential oils associated with strong flavours and aromas are rich in terpenoids such as menthol, limonene, pinene and geraniol (**Figure 1-1**). These are widely used as fragrances and flavourings as well as in cleaning agents and cosmetics [20]. Natural rubber is a terpenoid polymer of huge industrial importance, and is used in the manufacture of over 40,000 products [21]. Furthermore, the hydrocarbon nature of many common terpenes makes them a potential future source of renewable biofuels [22, 23]. Other

terpenes are important medicines. Examples include the sesquiterpene artemisinin, an important anti-malarial therapeutic agent for which extensive efforts to engineer production in microbial hosts have been undertaken to meet global demand [24]. The diterpene taxol is also a widely used chemotherapeutic agent because of its ability to stabilise microtubules and disrupt mitosis in mammalian cells [3]. (**Figure 1-1**).

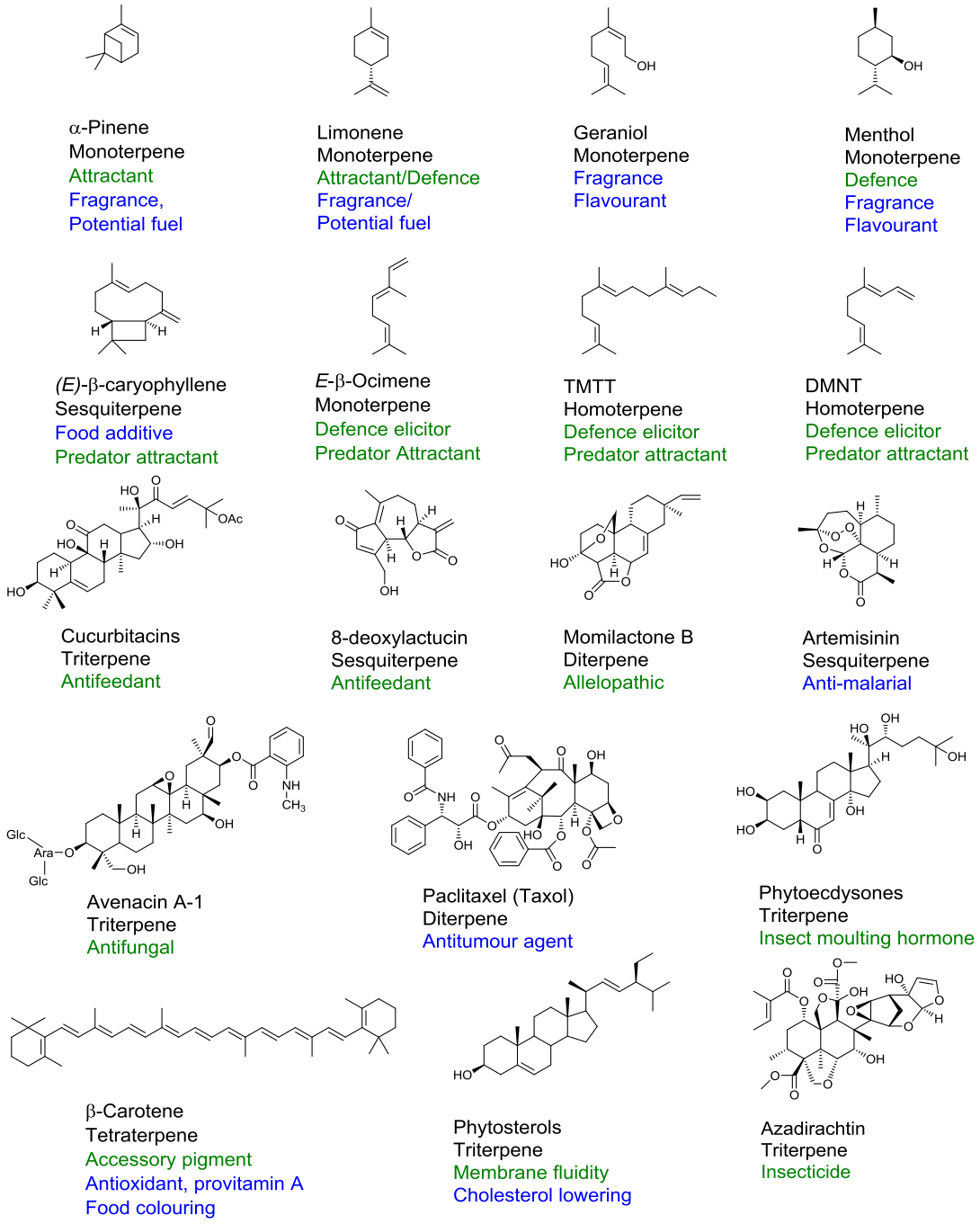


Figure 1-1: Examples of terpenes found in the plant kingdom and their functions. The name and class of molecules are given in black. Known functions of the compounds in nature are given in green, and uses by humans in blue.

1.3 – Triterpenes: Importance and bioactivity.

The triterpenes are C-30 members of the terpenoid family. Simple triterpenoids are amongst the oldest known biomolecules and have been isolated from sediments as old as 2.7 billion years [25]. The sterols are the best known examples of the triterpenes. Sterols are essential in eukaryotes, regulating membrane fluidity and serving as steroid hormone precursors. In plants, as well as serving these essential functions, triterpenes have diversified to fulfil more specialised roles. The term triterpene encompasses all C-30 terpenoids but is often used to refer to the structurally diverse, non-steroidal class of compounds usually associated with specialised roles. This thesis will focus on these non-steroidal C-30 compounds, which are hereafter referred to as triterpenes.

Triterpenes are likely to have a number of different functions *in planta*. Simple triterpenes are lipophilic molecules and are known to accumulate in the waxy cuticle covering the surface of aerial organs of plants [26], where they may have a structural role in maintaining the flexibility of the cuticle [27]. Other simple triterpenes have roles in growth and development. For example in the model legume *Lotus japonicus*, the pentacyclic triterpene lupeol plays a role in suppressing nodulation [28] (**Figure 1-2**) In oat, accumulation of elevated levels of β-amyrin in mutants results in altered cell specification and short super-hairy roots [29] (**Figure 1-2**). Similarly, altered expression of the genes for thalianol and marneral biosynthesis result in significant morphological effects in *Arabidopsis thaliana* [30-32]. The molecular mechanisms underlying these various phenomena are currently poorly understood and it remains to be seen whether these examples are indicative of a more widespread role for triterpenes in plant development.

In several cases triterpenes have been shown to have roles in plant defence, particularly triterpene glycosides (also known as saponins). Many triterpene glycosides are known to have allelopathic, antifungal, insecticidal, piscicidal or molluscicidal activity [33-37]. Such compounds may have potential as environmentally-sustainable pest-control agents. For example, the ability of triterpene glycosides to kill snails at low concentrations has applications for controlling parasite-carrying snails in regions where synthetic alternatives would be too expensive [37]. Even in modern agriculture, oats are often sown in rotation to control fungal disease outbreaks in wheat. The roots of oats produce anti-fungal triterpenes, avenacins, which have been implicated in the control of soil-borne diseases such as take-all [38].

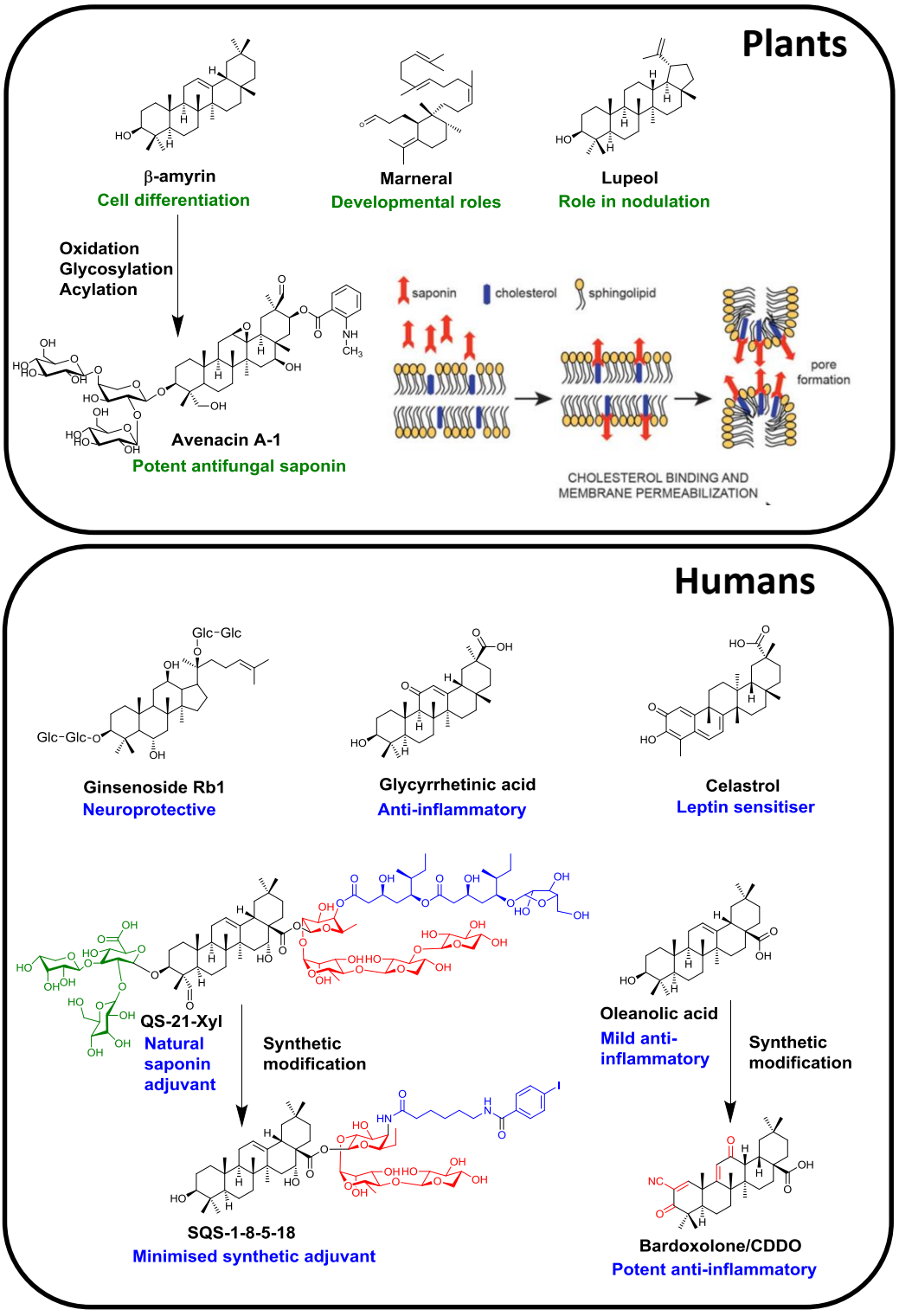


Figure 1-2: Functions of triterpenoids. **Top panel**: The simple triterpenes β-amyrin, marneral and lupeol have been shown to influence development in *Avena strigosa, Arabidopsis thaliana* and *Lotus japonicus*, respectively. Triterpene glycosides (saponins) such as the avenacins often show detergent-like properties, allowing them to insert into cell membranes and complex with sterols, leading to the formation of pores and subsequent cell lysis (Figure adapted from Moses et al. [39]) **Bottom panel**: Triterpenes and saponins have a number of important biological activities in humans. Synthetic derivatives of these have allowed for their therapeutic activities to be improved.

Triterpene glycosides are frequently amphipathic due to the coupling of the hydrophobic backbone to the highly polar sugar moieties. Thus many saponins have surfactant activity. Indeed, the name saponin derives from 'sapo', the Latin for soap. Extracts of saponin-rich plant species have traditionally been used as natural soaps [36]. The bioactivity of saponins is often linked with this property. For example, numerous studies have shown that saponins can complex with membrane sterols, resulting in the formation of pores and channels that permeabilise cell membranes (reviewed extensively by Augustin et al [35]) (**Figure 1-2**). Many saponins are haemolytic and also toxic to a range of human cell lines [40]. Removal of one or more sugars can abolish this activity [41]. Indeed, a number of specialised fungal pathogens of plants are capable of detoxifying saponins by enzymatic removal of sugars [42, 43].

Many traditional medicines have triterpenes and saponins amongst their active constituents. *Panax* (ginseng) species collectively accumulate >150 structurally diverse ginsenosides [44]. Both the saponins and their aglycones are reported to have a wide variety of beneficial properties including anti-cancer, immunostimulatory and anti-inflammatory activities [44-46]. Likewise, glycyrrhizin is a triterpene saponin produced by liquorice and used as a sweetening agent. The aglycone of glycyrrhizin, glycyrrhetinic acid, is known to interact with glucocorticoid signalling and has been associated with the anti-inflammatory activity of liquorice [47]. Many other species are also known to produce triterpenes with potent anti-inflammatory activity [48].

Beyond 'traditional' uses, there is growing interest in development of triterpenes as new therapeutic agents. For example, the pentacyclic triterpene celastrol was recently identified as an effective anti-obesity agent in mice due to its ability to restore leptin sensitivity [49]. Other triterpenes with cytoprotective functions have been considered as possible adjuvants in cancer therapy to help negate the toxicity of traditional chemotherapies [50]. Saponins also show promise for use as vaccine adjuvants [51-53]. One such saponin from *Quillaja saponaria*, QS-21, is a complex molecule featuring both glycosyl and acyl substituents. QS-21 has been the subject of extensive efforts to understand its mode-of-action, enhance stability and lessen side effects. These studies have led to the development of a semi-synthetic 'minimal' adjuvant with reduced toxicity, good immunogenicity and better synthetic accessibility [54-56] (**Figure 1-2**).

Other triterpenes have also been the subject of extensive synthetic efforts to improve their therapeutic profiles. Bardoxolone (or CDDO) is a synthetic triterpene derived from oleanolic acid. This molecule was generated using an organic chemistry-based approach that aimed to enhance the natural anti-inflammatory action of oleanolic acid. Bardoxolone is a remarkably potent inducer of the anti-inflammatory cytoprotective Keap1-Nrf2 pathway, and is several thousand-fold more effective at inhibiting production of the pro-inflammatory mediator nitric oxide than the parent compound [57, 58] (**Figure 1-2**).

QS-21 and Bardoxolone serve as excellent examples of the potential of triterpenoids for human health. The synthetic development of triterpenes as therapeutic agents using conventional chemistry depends on access to supply of the relevant scaffold and availability of functional groups for modification on this scaffold at key positions. Poor availability of triterpene scaffolds, stereochemical complexity and the problems of selective chemical oxidation have restricted such approaches. Hence the vast majority of triterpene structural diversity remains unexploited. This underscores the need to understand triterpene biosynthesis in nature and identify the enzymes responsible for scaffold formation and oxidation.

1.4 – Triterpenoid biosynthesis in plants

Two distinct pathways exist in nature for the biosynthesis of terpenoids. The first of these is the mevalonate (MVA) pathway, which uses acetyl-CoA as a starting molecule and is found in most eukaryotes, archaea and some bacteria [59]. In contrast, most bacteria have an alternative (MEP or DOXP) pathway, which utilises pyruvate and D-glyceraldehyde-3-phosphate. Despite the significant differences in their biochemistry, both pathways culminate in the production of the same 5-carbon isomers, namely isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These are the basic building blocks for all terpenes [11, 60]. Condensation of a single unit of IPP and DMAPP yields the C-10 molecule geranyl diphosphate (GPP). Incorporation of additional units of IPP yields successively longer chains, including C-15 farnesyl diphosphate (FPP) and C-20 geranylgeranyl diphosphate (GGPP) (**Figure 1-3**).

Plants are unusual in possessing both the MVA and MEP pathways, a feature that is believed to be a consequence of endosymbiotic acquisition of the cyanobacterial-like ancestor to modern-day chloroplasts [60]. Concordantly, the two pathways are

spatially separated; the MEP pathway operates within plastids, while the MVA pathway operates mostly within the cytosol, but also the endoplasmic reticulum (ER) and peroxisomes [11, 60]. The presence of two pathways underlines the importance of terpenoid biosynthesis in the plant kingdom and may enable optimal use of the carbon resources available for growth and defence [11]. The maintenance of both of these pathways throughout evolution suggests that they are functionally non-redundant and it follows that they produce clearly distinct types of terpene end products. GPP and GGPP are synthesised in the plastids, and give rise to the monoterpenes (GPP), diterpenes and carotenoids (GGPP). In contrast, FPP is synthesised from the MVA pathway in the cytosol and is used for biosynthesis of sesquiterpene and triterpenes. GGPP may also be formed in the cytosol for protein prenylation [11] (**Figure 1-3**).

Transformation of these linear substrates into cyclic derivatives is an important mechanism underlying the enormous diversity seen in the terpenes [61, 62]. Monoterpenes, sesquiterpenes and diterpenes are derived directly from diphosphorylated substrates. In contrast, for triterpene biosynthesis, two molecules of FPP first undergo head-to-head condensation to form the C-30 hydrocarbon squalene with loss of the diphosphate groups [63]. In bacteria, squalene may be cyclised directly by the action of squalene-hopene cyclases [64]. In eukaryotes however, squalene is generally further oxidised prior to cyclisation by squalene epoxidase (SQE) to form 2,3-oxidosqualene (OS). This epoxide gives rise to the characteristic C-3 oxygen functionality seen in sterols and triterpenes.

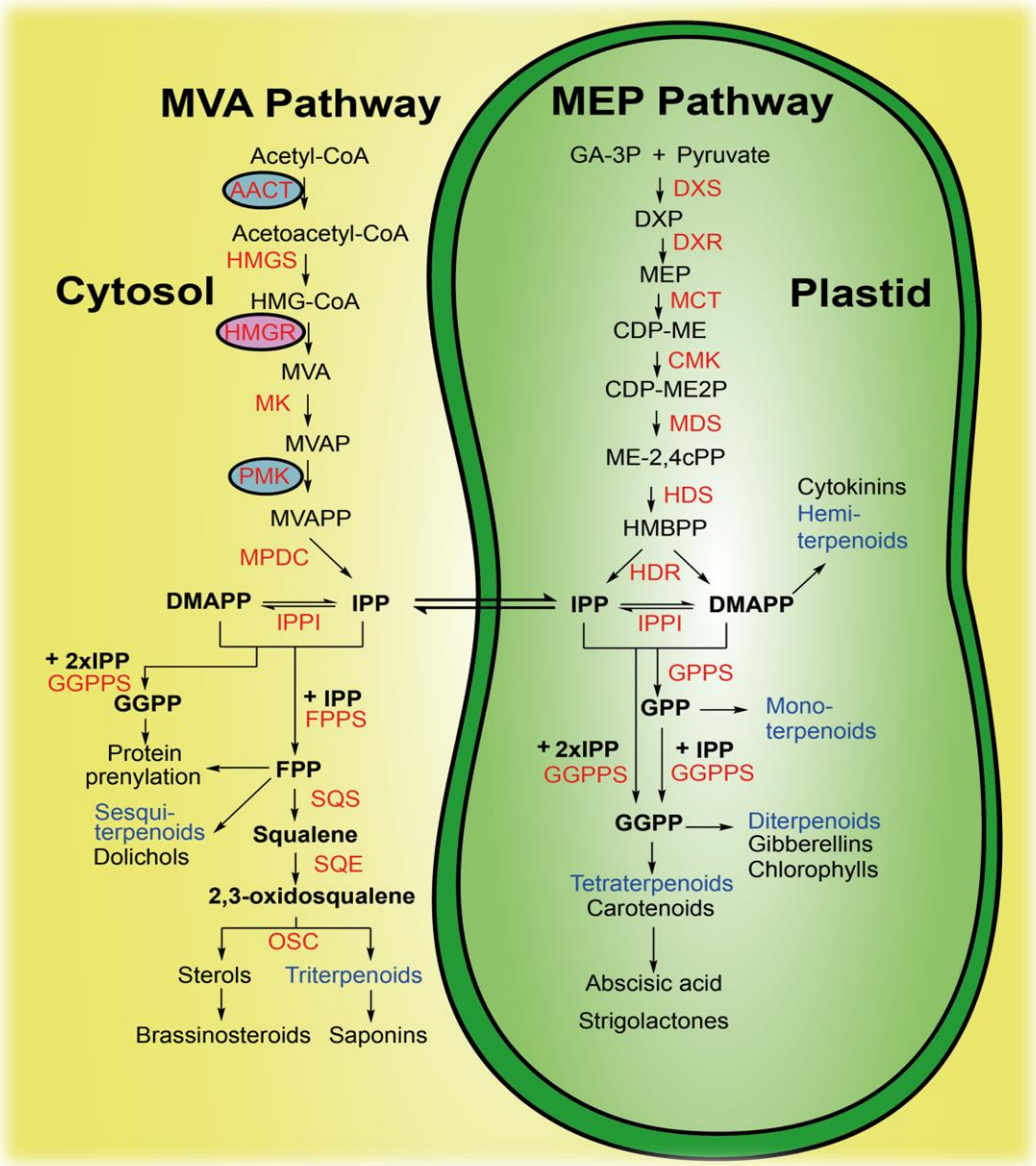


Figure 1-3: The MVA and MEP pathways. The MEP pathway is in the plastids while the MVA pathway is based primarily in the cytosol. In *A. thaliana,* HMGR is localised to the endoplasmic reticulum (pink bubble), while AACT and PMK are believed to be peroxisomal (blue bubbles)[11]. Trafficking of IPP and DMAPP occurs between compartments (double-headed arrows). **Compound abbreviations**: **MVA Pathway:** CoA, Coenzyme A; HMG, 3-hydroxy-3-methylglutaryl-CoA; MVA, mevalonate; MVAP, Mevalonate-5-phosphate; MVAPP, Mevalonate-5-diphosphate. **MEP pathway:** GA-3P, D-glyceraldehyde-3-phosphate; DXP, 1-deoxy-D-xylulose 5–phosphate; MEP, 2-C-Methyl-D-erythritol 4-phosphate; CDP-ME, 4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol; CDP-ME2P, 2-Phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol; ME-2,4-cPP, 2-C-Methyl-D-erythritol ,2,4-cyclodiphosphate; HMBPP, 4-Hydroxy-3-methylbut-2-enyl-diphosphate; **Downstream:** IPP, isopentenyl diphosphate, DMAPP – dimethylallyl diphosphate, GPP – geranyl diphosphate, FPP – farnesyl diphosphate; GGPP, geranylgeranyl diphosphate. **Enzyme abbreviations**: **MVA Pathway:** AACT, Acetyl-CoA *C*-acetyltransferase; HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MK, MVA kinase; PMK, Phospho-MVA kinase; MPDC, Diphospho-MVA kinase. **MEP pathway:** DXS, DXP synthase; DXR, DXP reductoisomerase; MCT, MEP cytidylyltransferase; CMK. CDP-ME kinase; MDS, ME-2,4cPP synthase; HDS, HMBPP synthase; HDR, HMBPP reductase; **Downstream:** IPPI, IPP isomerase; GPPS, GPP synthase; FPPS, FPP synthase; GGPPS, GGPP synthase; SQS, Squalene synthase; SQE, Squalene epoxidase; OSC, Oxidosqualene cyclase.

1.4.1– Oxidosqualene cyclisation generates diversity in triterpene scaffolds

Oxidosqualene cyclases (OSCs) are the enzymes that cyclize OS to triterpenes. Triterpene cyclisation is initiated through protonation of the terminal OS epoxide by a conserved aspartic acid residue within the motif DCTAEA [65]. Protonation opens the epoxide and leads to the formation of cationic carbon species [62]. Double bonds within the substrate are brought into proximity to the cation by conformational reorganisation within the active site of the OSC. This leads to a cascade of nucleophilic attacks that result in ring formation. These cyclised intermediates may undergo further ring expansions and methyl- and hydride shifts to yield the final product [66, 67]. The degree of cyclisation can vary greatly and may result in the formation of anything from mono- to pentacyclic products [62, 68]. In some cases, rings may be reopened once formed, leading to the production of seco- triterpenes [69]. Termination of the reaction often occurs by deprotonation resulting in the formation of a characteristic double bond, but can also arise through water capture [68]. The latter results in the formation of triterpene diols such as dammarenediol-II, a commercially important scaffold found in the ginsenosides (**Figure 1-4**).

These diverse mechanisms can give rise to in excess of 100 different cyclic products [66, 68]. However only two such products, cycloartenol and lanosterol, are required for sterol production in eukaryotes. The former is produced exclusively by plants. Formation of sterols occurs through a highly specific cyclisation route involving the formation of a 6,6,6,5 tetracyclic protosteryl cation intermediate. This intermediate adopts a chair-boat-chair (C,B,C) conformation in the A, B and C rings respectively (**Figure 1-4**). Several of the major pentacyclic triterpene scaffolds found in the plant kingdom (such as β-amyrin and lupeol), are also formed through a 6,6,6,5 tetracyclic cationic intermediate known as the dammarenyl cation. In contrast to the protosteryl cation, however, the dammarenyl cation adopts a chair-chair-chair (C,C,C) conformation and undergoes further ring expansion and generation to form the final pentacyclic products [62, 68, 70] (**Figure 1-4**). The protosteryl and dammarenyl cations provide a useful mechanistic distinction between the biosynthesis of sterols and common triterpenes. It must be noted however that other common ‘triterpene’ scaffolds which function in specialised metabolism (ie cucurbitadienol) are also formed through the protosteryl cation [62]. Finally, given the mechanistic flexibility of OS cyclisation, it follows that OSCs are often multifunctional and produce a variety of products. An extreme example of this is the *A. thaliana* OSC baruol synthase BARS1, which makes a total of 23 different cyclization products [71].

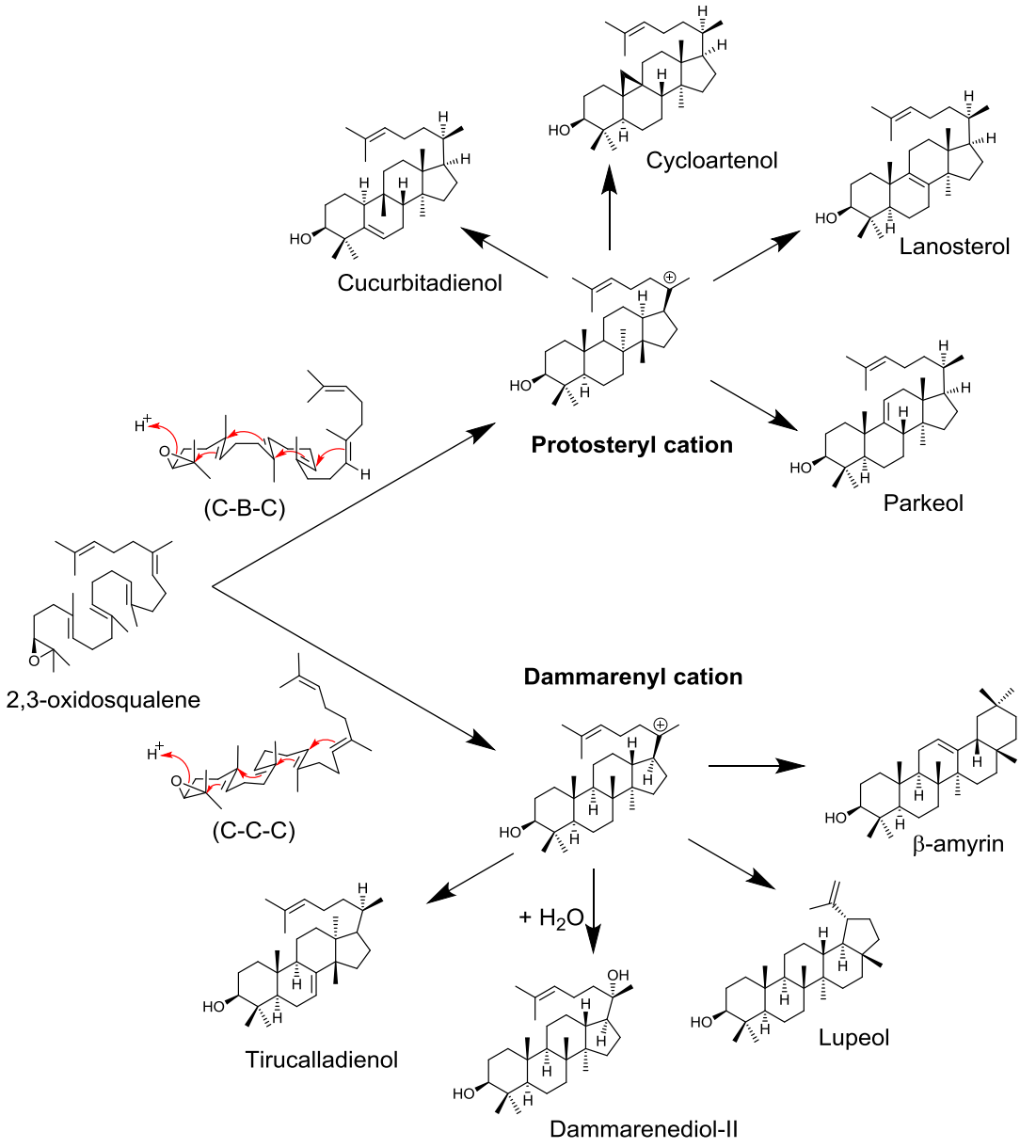


Figure 1-4: Cyclisation of oxidosqualene to form sterols and triterpenes. The substrate adopts different conformations in the active site of the OSC prior to cyclisation. Two common conformations include the chair-boat-chair (C-B-C) and chair-chair-chair (C-C-C). Protonation results in a series of electrophilic attacks which generates cationic intermediates including the tetracylic protosteryl- and dammarenyl cations, respectively. The subsequent rearrangements of these intermediates define the final products. The protosteryl cation gives rise to the sterols including lanosterol and cycloartenol but can also form other products including parkeol and cucurbitadienol which may serve as scaffolds for specialised metabolites. The dammarenyl cation can give rise to tetracyclic products or undergo further ring expansion to form pentacylic triterpenes including β-amyrin and lupeol.

It is particularly remarkable that despite the extensive diversity available in OS cyclisation, the majority of the saponins characterised from plants are based on only a very limited repertoire of eleven scaffolds [72]. This would imply that the vast majority of scaffolds are largely unrepresented in plant extracts and that a valuable

source of bioactive compounds may be unutilised. Significant structural diversity has been found in the products of OSCs even from well-studied species such as *A. thaliana* [62]. However many of these are present only at very low levels *in planta*. Therefore identification of the OSCs that synthesise these diverse products and their expression in heterologous systems represents an important strategy for accessing this metabolic diversity.

1.4.2 – Oxidation allows further diversification of triterpenes

Oxidation of the scaffold is an integral feature of triterpene biosynthesis and is critical for generating structural diversity [73]. Oxidation imparts significant effects on the polarity, solubility and reactivity of the backbone. Importantly, it further allows functionalisation of chemically-inert C-H bonds within the triterpene scaffold, so enabling further tailoring, frequently by glycosylation or acylation. Enzymatic oxidation of triterpene scaffolds can also facilitate fundamental structural changes such as bond cleavage [74, 75] and bridge formation [76]. A good example of this is seen in the limonoid triterpenes, which can undergo such extensive oxidation that they barely resemble their tetracyclic precursors [14].

The principle enzymes responsible for triterpene oxidation are the cytochrome P450s (P450s). These enzymes catalyse the regio- and stereoselective oxidation of their substrates. P450s are monoxygenases, splitting molecular oxygen to incorporate one atom into the substrate while reducing the other to form water [77]. In a large number of cases this results in formation of a hydroxylated product:

$$RH + O_2 + 2e^- + 2H^+ \rightarrow ROH + H_2O$$

This process is catalysed by a central iron atom that is part of a haem cofactor within the P450 active site. The mechanism requires electrons donated from NADPH via a partner CYP450 reductase (CPR) enzyme [78]. P450s can catalyse a remarkable range of unusual reactions [79, 80]. In triterpenes, the most common modifications involve hydroxylation, although subsequent rounds of additional oxidation can also occur to give carbonyl or carboxylic acid moieties (**Table 1-1**).

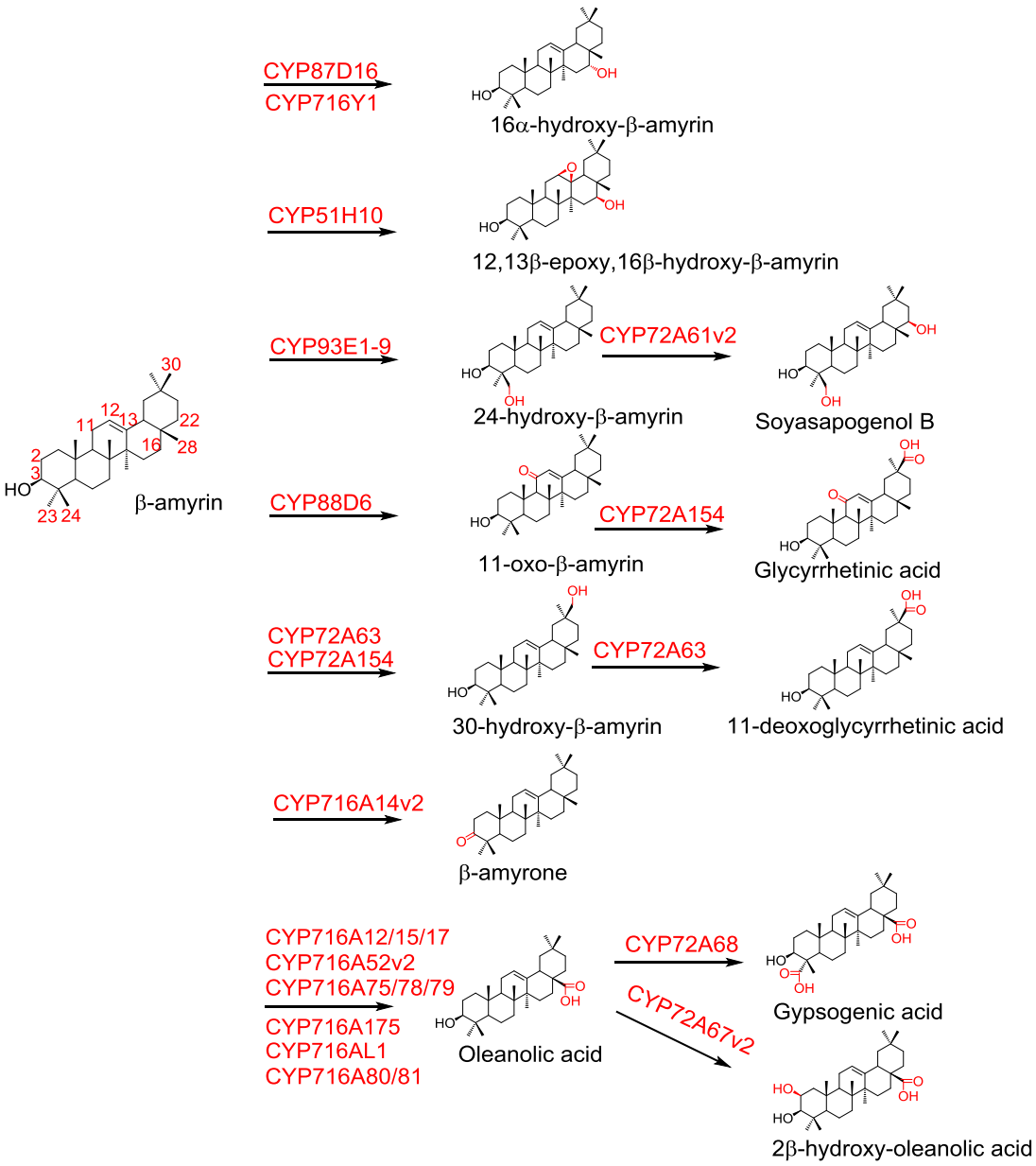


Figure 1-5: Activity of various triterpene oxidases on the β-amyrin scaffold. The numbering on β-amyrin (left) indicates the carbons which are oxidised by P450s. For simplicity, only the activity upon β-amyrin is shown, except in the cases where an oxidised form of β-amyrin is needed for full activity.

P450s are one of the largest and diverse enzyme superfamilies, representing approximately 1% of the genome of higher plants [73, 81]. P450s show a great deal of flexibility in primary sequence, which underlies the catalytic diversity and substrate tolerance of these enzymes. P450 categorisation is based on sequence similarity to other P450s. At the broadest level, P450s are grouped into clans. P450s within each clan are believed to derive from a single ancestral gene. There are a total of eleven clans in land plants, ten of which are present in angiosperms

[81]. Clans are further divided into families and subfamilies; P450s are assigned to these based on values of approximately 40 and 55 percent amino acid sequence identity to their nearest neighbours [82]. There is significant variation in the number of families within a single clan. For example, in angiosperms, clans CYP51, CYP74, CYP97, CYP710, CYP711 and CYP727 each contain only a single family. By contrast, the CYP71, CYP72, CYP85 and CYP86 clans contain multiple families [81]. P450s within the single-family clans are often functionally associated with primary metabolism such as sterol or hormone biosynthesis. By contrast, P450s from multi-family clans are more commonly associated with specialised plant metabolic pathways [73]. Accordingly, most triterpene-oxidising P450s are found within the latter group [83] (**Table 1-1**). However despite these generalisations, members of clans normally associated with primary metabolic functions may still be recruited towards roles in specialised metabolism. For example, the CYP51 clan is found across all biological kingdoms and its members have ancient conserved functions in 14α demethylation of sterols [84]. In grasses however, the CYP51H subfamily has arisen from the wider plant CYP51G sterol demethylases and has been recruited to secondary metabolic role. In oats, CYP51H10 is a multifunctional triterpene (β-amyrin) oxidase required for the biosynthesis of avenacins [85, 86]. Examples such as this underscore the importance of experimental verification of P450 function.

Great progress in identifying triterpene-oxidising P450s has been made in recent years and the number continues to grow steadily. The most common substrate for these enzymes is the ubiquitous pentacyclic triterpene β-amyrin. Collectively these P450s oxidise eleven carbons around this scaffold (**Figure 1-5** and **Table 1-1**). Oxidases for other less common but commercially-important triterpene scaffolds are also becoming known. These include the dammarenediol-II oxidases CYP716A47 and CYP716A53v2, which give rise to two of the major branches of bioactive ginsenosides in species of ginseng [44, 87, 88] (**Figure 1-6A**). Similarly, a number of P450s have recently emerged which are functional on cucurbitadienol [89, 90], a tetracyclic triterpene widely distributed throughout the cucurbitaceae and responsible for bitterness in these species [91] (**Figure 1-6B**).

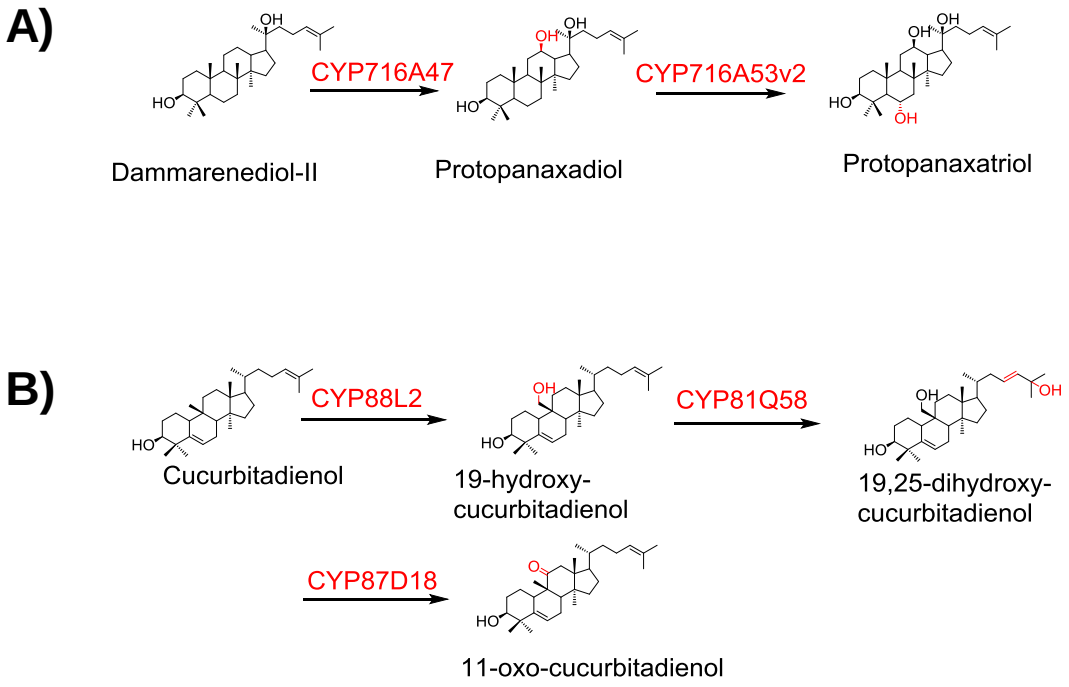


Figure 1-6: Oxidases of non-β-amyrin triterpene scaffolds. **A)** Dammarenediol-II is a tetracyclic triterpene which can be oxidised to protopanaxadiol (PPD) and protopanaxatriol (PPT) in ginseng species through the action of CYP716A47 and CYP716A53v2. **B)** Cucurbitadienol is a tetracyclic triterpene scaffold widely distributed in the cucurbitaceae. P450s have been identified which oxidise this scaffold in cucumber (CYP88L2 and CYP81Q58) as well as *Siratia grosvenorii* (CYP87D20).

1.4.3 – Clustering aids identification of new triterpene oxidases

In recent years, it has become apparent that the genes for many plant specialised metabolic pathways are organised into biosynthetic gene clusters [83]. Triterpenes appear to be no exception to this, and a number of OSCs have been found with nearby P450s (and other tailoring enzymes) which are coregulated and functionally compatible when expressed in heterologous hosts. Examples of this kind have been found in the genome of diverse plant species including *Arabidopsis thaliana* [30, 32, 74, 92], *Lotus japonicus* [93], cucumber [89, 92] and oat [94] (**Figure 1-7**). As a result, this has huge implications for the discovery of new triterpene biosynthetic enzymes, allowing new pathways to be elucidated through genome mining or sequencing bacterial artificial chromosomes. Recently, analysis of the genomic region within 50kb of terpene synthases highlighted a frequent non-random physical 'pairing' of OSCs with P450 enzymes in the genomes of flowering plants [92]. This study also revealed common families of P450s which pair with OSCs (**Figure 1-7B**). For example, in dicots, members of the CYP716 family were commonly found proximal to OSCs [92]. This fits remarkably well with the list of triterpene oxidases

characterised to date, with CYP716 members shown to oxidise a variety of triterpene backbones (**Table 1-1**). Further members of this family (and others) can therefore be expected to be characterised in future. Knowledge of common families of triterpene oxidases may also prove useful for species for which genomic sequence data is not available, allowing identification of candidates from transcriptomic data [95].

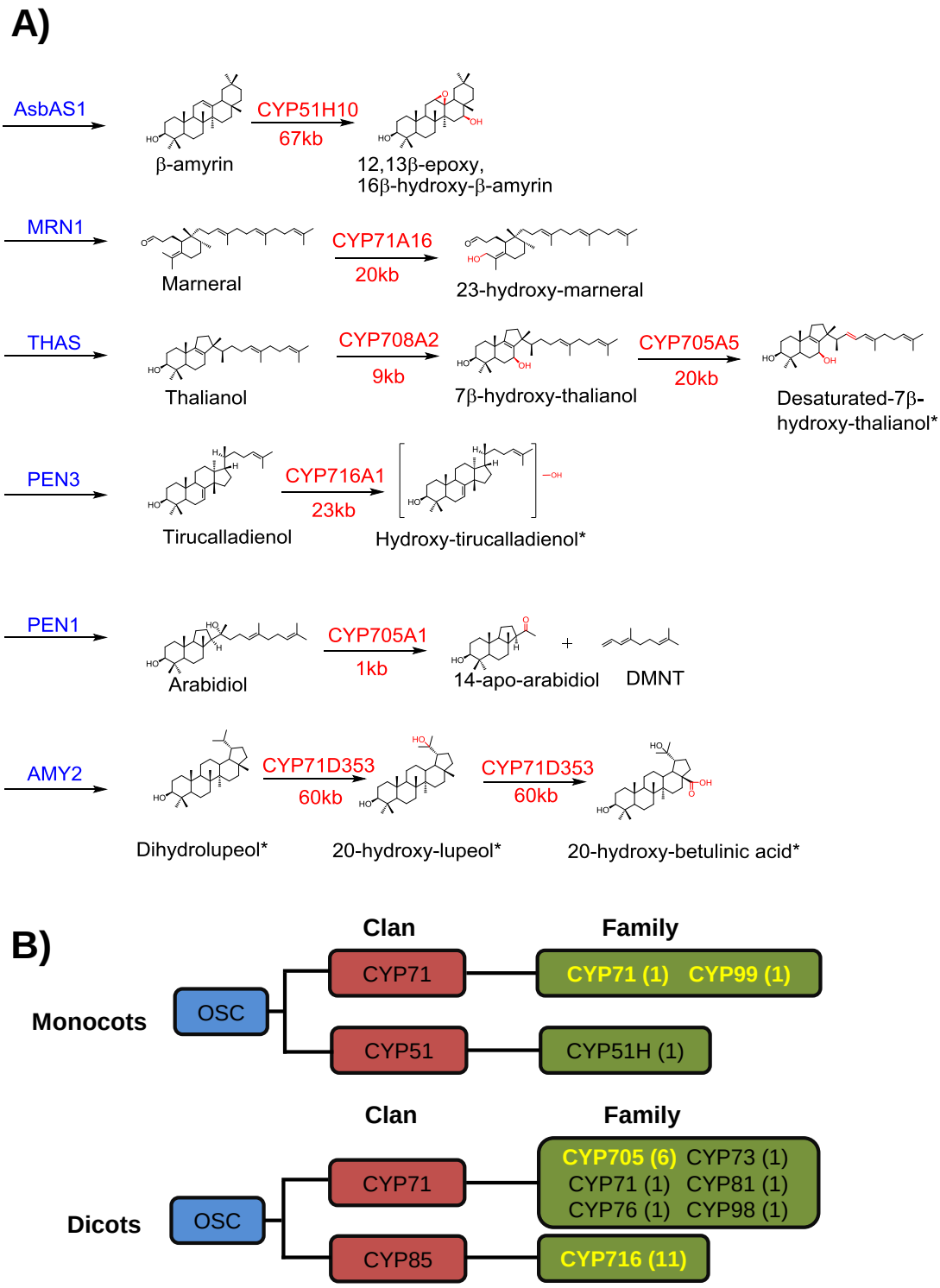


Figure 1-7: Clustering of OSCs and P450s. **A)** Examples of triterpene oxidases which are known to be clustered with oxidosqualene cyclases. OSCs are named in blue and P450s are named in red. The approximate distance between the relevant OSC and the P450s are given in kb below the P450 name. **B)** Summary of common OSC-P450 pairs found from a recent analysis of five monocot and twelve dicot species [92]. The P450 clans are indicated in red boxes, while the individual families are given in green boxes. Numbers in brackets indicate the numbers of each pair found within the analysis. Yellow indicates the most abundant pairings. Figure adapted from Boutanaev et al [92].

Clan	Gene	Substrate	Modification	Genbank ID	Ref
CYP51	AsCYP51H10	β-amyrin	C12,13β-epoxidation, C-16β hydroxylation	ABG88961	[85, 86]
CYP71	AtCYP71A16	Marnerol	C-23 hydroxylation	NM_123623	[30, 74]
	LjCYP71D353	Dihydrolupeol*	C-20 hydroxylation C-28 oxidation (3 step) *	KF460438	[93]
	AtCYP705A1	Arabidiol	C-15-16 cleavage	NM_117621	[74, 75]
	AtCYP705A5	7β-hydroxythalianol	C-15-16 desaturation*	NM_124173	[32]
	CsCYP81Q58	19-hydroxy-cucurbitadienol	C-25 hydroxylation	KM655856	[89, 92]
	GmCYP93E1	β-amyrin/ Sophoradiol	C-24 oxidation (3-step)	AF135485	[96]
	MtCYP93E2	β-amyrin	C-24 hydroxylation	DQ335790	[97]
	GuCYP93E3	β-amyrin	C-24 hydroxylation	AB437320	[98]
	AhCYP93E4	β-amyrin	C-24 oxidation (3-step)	KF906535	[99]
	CaCYP93E5	β-amyrin	C-24 oxidation (3-step)	KF906536	[99]
	GgCYP93E6	β-amyrin	C-24 oxidation (3-step)	KF906537	[99]
	LcCYP93E7	β-amyrin	C-24 hydroxylation	KF906538	[99]
	PsCYP93E8	β-amyrin	C-24 oxidation (3-step)	KF906539	[99]
	PvCYP93E9	β-amyrin	C-24 oxidation (3-step)	KF906540	[99]

Table 1-1: P450s involved in triterpene biosynthesis. * denotes putative identity. ** denotes where other products were also observed with unconfirmed identity. Initials preceding gene name are species as follows: Aa - *Artemesia annua*, Ah - *Arachis hypogaea*, As - *Avena strigosa*, At - *Arabidopsis thaliana*, Bf - *Bupleurum falcatum*, Bv - *Barbarea vulgaris*, Ca - *Cicer arietinum*, Cq – *Chenopodium quinoa*, Cr - *Catharanthus roseus*, Cs - *Cucumis sativus*, Gg - *Glycyrrhiza glabra*, Gm - *Glycine max*, Gu - *Glycyrrhiza uralensis*, Lc - *Lens culinaris*, Lj - *Lotus japonicus*, Md - *Malus domestica*, Ml - *Maesa lanceolata*, Mt - *Medicago truncatula*, Pg - *Panax ginseng*, Ps - *Pisum sativum*, Pv - *Phaseolus vulgaris,* Sg - *Siratia grosvenorii*, Vv - *Vitis vinifera*.

Clan	Gene	Substrate	Modification	Genbank ID	Ref
CYP72	GuCYP72A154	β-amyrin/ 11-oxo-β-amyrin	C-30 hydroxylation/ C-30 oxidation (3-step)**	AB558153	[100]
	MtCYP72A61v2	24-hydroxy-β-amyrin	C-22β hydroxylation	AB558145	[101]
	MtCYP72A63	β-amyrin	C-30 oxidation (3-step)	AB558146	[100, 101]
	MtCYP72A67	Oleanolic acid	C-2β hydroxylation	AB558149	[102]
	MtCYP72A68v2	Oleanolic acid	C-23 oxidation (3-step)	AB558150	[101]
CYP85	AtCYP708A2	Thalianol	C-7β hydroxylation	NM_180822	[32, 74]
	AtCYP716A1	Tirucalladienol	Hydroxylation*	NM_123002	[92]
	MtCYP716A12	α/β-amyrin/ Lupeol	C-28 oxidation (3-step)	FN995113	[97]
	AaCYP716A14v2	α/β/δ-amyrin/ lupeol	C-3 oxidation (to ketone)	KF309251	[95]
	VvCYP716A15	α/β-amyrin/ Lupeol	C-28 oxidation (3-step)	AB619802	[97]
	VvCYP716A17	β-amyrin	C-28 oxidation (3-step)	AB619803	[97]
	PgCYP716A47	Dammarenediol-II	C-12β hydroxylation	JN604536	[103]
	PgCYP716A52v2	β-amyrin	C-28 oxidation (3-step)	JX036032	[87]
	PgCYP716A53v2	Protopanaxadiol	C-6α hydroxylation	JX036031	[88]
	MlCYP716A75	β-amyrin	C-28 oxidation (3-step)	KF318733	[76]

Table 1-1 (continued): P450s involved in triterpene biosynthesis. * denotes putative identity. ** denotes where other products were also observed with unconfirmed identity. Initials preceding gene name are species as follows: Aa - *Artemesia annua*, Ah - *Arachis hypogaea*, As - *Avena strigosa*, At - *Arabidopsis thaliana*, Bf - *Bupleurum falcatum*, Bv - *Barbarea vulgaris*, Ca - *Cicer arietinum*, Cq – *Chenopodium quinoa*, Cr - *Catharanthus roseus*, Cs - *Cucumis sativus*, Gg - *Glycyrrhiza glabra*, Gm - *Glycine max*, Gu - *Glycyrrhiza uralensis*, Lc - *Lens culinaris*, Lj - *Lotus japonicus*, Md - *Malus domestica*, Ml - *Maesa lanceolata*, Mt - *Medicago truncatula*, Pg - *Panax ginseng*, Ps - *Pisum sativum*, Pv - *Phaseolus vulgaris,* Sg - *Siratia grosvenorii*, Vv - *Vitis vinifera*.

Clan	Gene	Substrate	Modification	Genbank ID	Ref
CYP85	CqCYP716A78	β-amyrin	C-28 oxidation (3-step)	KX343075	[104]
	CqCYP716A79	β-amyrin	C-28 oxidation (3-step)	KX343076	[104]
	BvCYP716A80	α/β-amyrin/ Lupeol	C-28 oxidation (3-step)**	KP795926	[105]
	BvCYP716A81	α/β-amyrin/ Lupeol	C-28 oxidation (3-step)**	KP795925	[105]
	MdCYP716A175	α/β-amyrin/ Lupeol/ Germanicol	C-28 oxidation (3-step)**	EB148173	[106]
	CrCYP716AL1	β-amyrin	C-28 oxidation (3-step)	JN565975	[107]
	BfCYP716Y1	α/β-amyrin	C-16α hydroxylation	KC963423	[108]
	MlCYP87D16	β-amyrin	C-16α hydroxylation	KF318734	[76]
	SgCYP87D18	Cucurbitadienol	C-11 oxidation (2-step)	AEM42986	[90]
	GuCYP88D6	β-amyrin	C-11 oxidation (2-step)	AB433179	[98]
	CsCYP88L2	Cucurbitadienol	C-19 hydroxylation	KM655860	[89]

Table 1-1 (continued): P450s involved in triterpene biosynthesis. * denotes putative identity. ** denotes where other products were also observed with unconfirmed identity. Initials preceding gene name are species as follows: Aa - *Artemesia annua*, Ah - *Arachis hypogaea*, As - *Avena strigosa*, At - *Arabidopsis thaliana*, Bf - *Bupleurum falcatum*, Bv - *Barbarea vulgaris*, Ca - *Cicer arietinum*, Cq – *Chenopodium quinoa*, Cr - *Catharanthus roseus*, Cs - *Cucumis sativus*, Gg - *Glycyrrhiza glabra*, Gm - *Glycine max*, Gu - *Glycyrrhiza uralensis*, Lc - *Lens culinaris*, Lj - *Lotus japonicus*, Md - *Malus domestica*, Ml - *Maesa lanceolata*, Mt - *Medicago truncatula*, Pg - *Panax ginseng*, Ps - *Pisum sativum*, Pv - *Phaseolus vulgaris,* Sg - *Siratia grosvenorii*, Vv - *Vitis vinifera*.

1.5 – Engineering diversity in triterpene biosynthesis

Characterisation of the genes and enzymes in triterpene biosynthesis permits reproduction of these pathways in heterologous hosts. At the same time, given the difficulties in chemical synthesis of the triterpenes, there is particular interest in how these enzymes can be employed to access novel products. Integrating enzymes from different pathways into a common host is a powerful means to achieve this goal. This 'combinatorial' biosynthesis is becoming an increasingly popular approach for producing an array of both extant and new-to-nature terpenes in different hosts.

1.5.1 – Combinatorial biosynthesis

For triterpenes, combinatorial biosynthesis has been demonstrated through the combination of common scaffolds such as β-amyrin with tailoring enzymes including P450s and glycosyltransferases. Selectively oxidised forms of β-amyrin can be predictably produced by the combined action of pairs of β-amyrin oxidases [101]. Taking this a step further, triterpene glycosides have also been produced by combining OSCs, P450s and glycosyltransferases [108]. In this latter example, the C-3 glucoside of echinocystic acid was made in yeast through a combination of enzymes from four different plant species (**Figure 1-8**). As the repertoire of triterpene biosynthetic enzymes continues to expand, the assembly of a diverse collection of enzymes offers significant potential for the bespoke biosynthesis of triterpenes of interest. The success of this approach will be bolstered by the fact that numerous triterpene oxidases are known with overlapping functions, but with differences in efficiency or activity towards individual substrates [99, 100]. Thus several routes may be possible towards the same product and the exact choice of enzymes may be tailored for optimising yields.

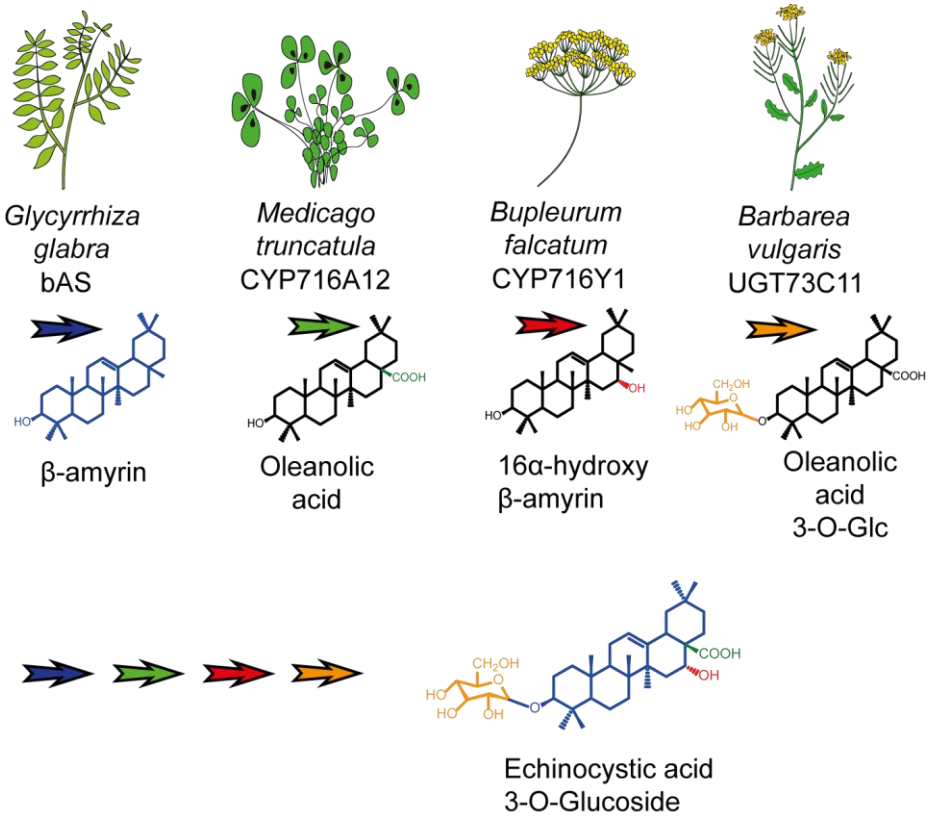


Figure 1-8: Combinatorial biosynthesis of triterpenoids. Combining enzymes from different organisms which work on a common scaffold (i.e. β-amyrin), can allow for predictable design of triterpenoids **Abbreviations:** bAS – β-amyrin synthase, Glc – Glucoside.

Combinatorial biosynthesis relies on the 'promiscuity' of enzymes in specialised metabolism. As a result, enzymes may be active towards substrates which differ from that encountered in their native host pathway. Hence, scaffolds for which few native tailoring enzymes are known may still be functionalised by enzymes from other pathways. For example, members of the CYP716 family have been shown to have activity towards a range of pentacyclic backbones (**Figure 1-9A**) [95, 97, 108]. In these examples, the activity of the P450s on other scaffolds was identical to that on β-amyrin, due to the similarities between the scaffolds. However other studies have hinted that some triterpene oxidases might have activity towards markedly different substrates. For example the *A. thaliana* CYP716A1 has been shown to be capable of oxidising both the tetracyclic triterpene tirucalladienol [92] as well as the pentacyclic triterpenes lupeol and α- and β-amyrin [109] (**Figure 1-9B**).

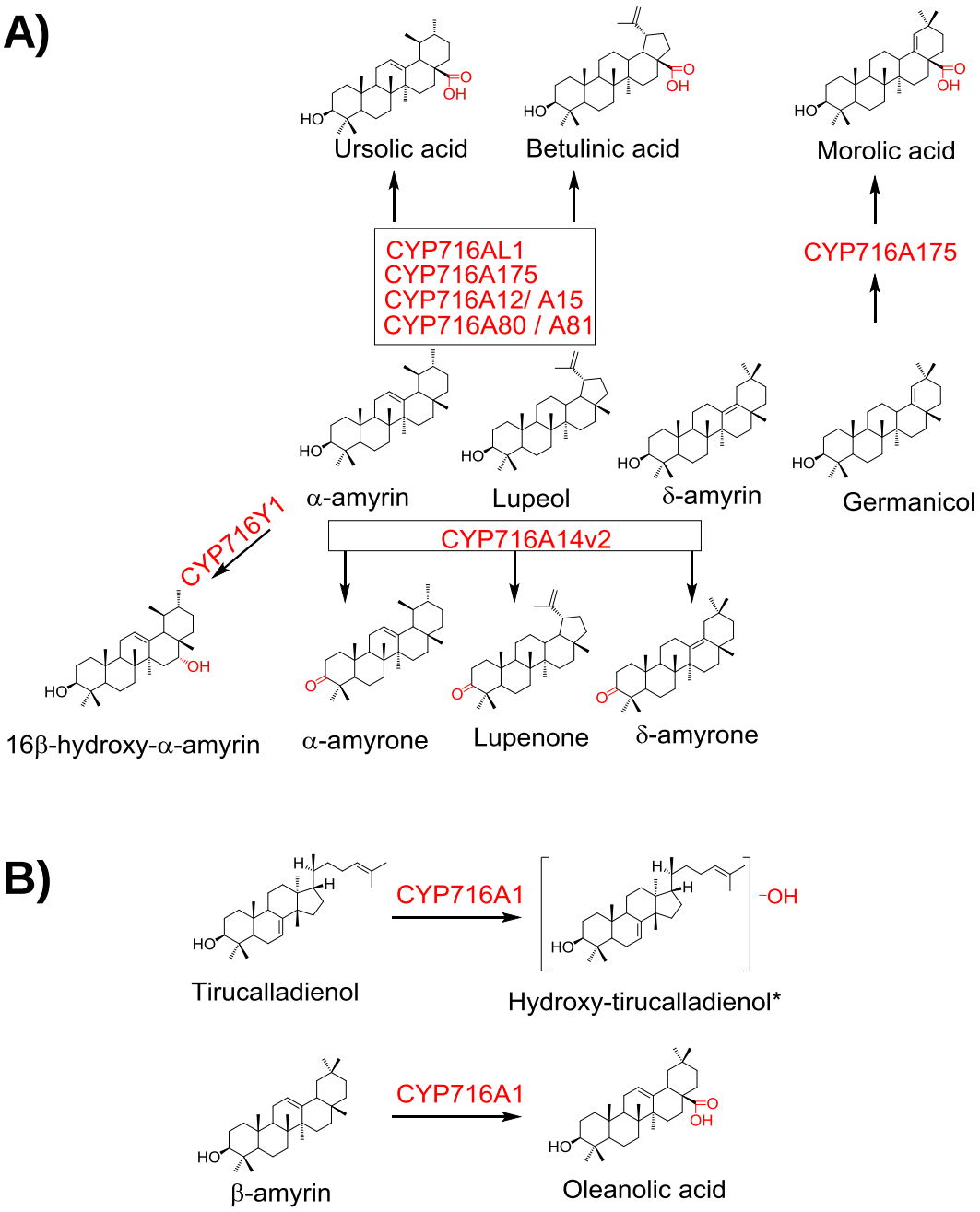


Figure 1-9: β-amyrin oxidases which are also functional on other pentacyclic scaffolds. **A)** A number of triterpene oxidases have been found to be capable of oxidising a range of pentacyclic scaffolds in addition to β-amyrin (see **Figure 1-5**). These scaffolds are structurally similar and the activity of the P450 is retained in each case. **B)** Triterpene oxidases may also be active on structurally dissimilar scaffolds. The CYP716A1 enzyme acts as a C-28 oxidise towards a number of pentacyclic scaffolds (β-amyrin only shown). The enzyme also oxidises tirucalladienol, which lacks an equivalent methyl group at the same position, thus the regiospecific oxidations will differ depending on substrate.

At present, a systematic approach to combinatorial biosynthesis of triterpenes is lacking, however it is reasonable to assume that a high success rate will be found. Some excellent recent examples have showcased how effective a systematic combinatorial biosynthesis approach can be for generating diterpene diversity [110-112]. In one study, the *A. thaliana* CYP701A3 was shown to oxidise an array of 15 out of 22 diverse diterpene scaffolds, with different regioselectivity depending on the type of substrate [113]. Remarkably, this enzyme normally operates as an *ent*-kaurene oxidase (KO) in the biosynthesis of essential gibberellins. This suggests that even enzymes from primary metabolism might be successfully recruited for combinatorial biosynthesis. The obvious triterpene counterpart would be sterol oxidases, although the success of such an approach remains to be seen. Alternatively, P450s are also known to operate in the biosynthesis of sterol-derived specialised metabolites such as the steroidal glycoalkaloids and might also be applied for triterpene biosynthesis [114].

1.5.2 – Protein engineering augments combinatorial biosynthesis

Enzymes such as terpene cyclases and P450s show a high degree of functional plasticity and changes to the active site of these enzymes can have a dramatic effect on both product and substrate specificity. This provides further opportunity to engineer triterpene diversity. For example, mutant cyclases producing 'new-to-nature' diterpenes [115] have been successfully integrated into a combinatorial biosynthesis approach [111] (**Figure 1-10A**). Alternatively, while combinatorial biosynthesis is possible thanks to the promiscuity of enzymes in specialised metabolism, generally the activity of such enzymes towards novel substrates is not optimal. This can result in poor product yields and non-specific reactions which would impede creation of a multistep pathway. Hence protein engineering efforts can also be used to enhance the specific product yield following the initial combinatorial screen. This principle has likewise been demonstrated for diterpene P450s, allowing for substantially improved yields of specific products [110] (**Figure 1-10B**). In nature, enzymes catalysing sequential steps in a metabolic pathway may also interact physically. Such protein complexes or 'metabolons' are expected to efficiently channel products through the pathway and have recently been demonstrated for cyanogenic glycoside production in *Sorghum* [116]. Assembly of synthetic metabolons could present a complementary strategy to the above approaches [117].

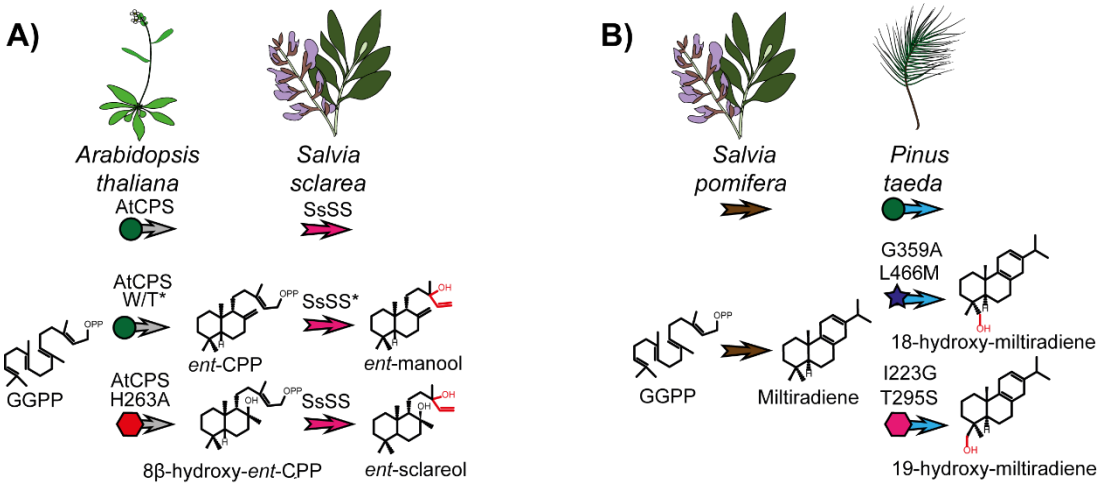


Figure 1-10: Using mutant enzymes in combinatorial biosynthesis **A)** Mutant terpene cyclases may serve as a source of novel scaffolds for combinatorial biosynthesis. In the example here, an *A. thaliana ent*-copalyl diphosphate synthase (AtCPS) H263A mutant produces the new-to-nature diterpene 8β-hydroxy-*ent*-CPP. Both wild-type and mutant products can be further cyclised by the *Salvia sclarea* sclareol synthase (SsSS). NB: while the SsSS enzyme is active on *ent*-CPP as shown, the specific *ent*-CPP synthase used in the study by Jia et al was ZmCPS2/An2 from *Zea mays* [111]. **B)** Mutagenesis can be used to enhance the product specificity and activity of enzymes following combinatorial biosynthesis. In this example, the specificity and activity of the *Pinus taeda* abietadiene oxidase (PtAO) towards the non-native diterpene substrate miltiradiene could be improved through modelling and targeted mutagenesis. The type and combination of mutations allowed for different product specificities.

1.6 – Triterpene production in heterologous hosts

A common method of accessing existing and novel plant metabolites is through expression of their biosynthetic enzymes in heterologous host systems. The metabolic engineering of biological host systems for production of valuable natural products is frequently referred to as 'synthetic' biology. A core principle of synthetic biology is the application of engineering principles to biological systems including a 'design, build, test' approach to optimise production [118, 119]. Ultimately, synthetic biology approaches aim to establish a series of well-defined, standardised biological 'parts' which can be integrated to achieve high level production of the desired molecule. The potential of synthetic biology to deliver high-value specialised metabolites is highlighted by numerous examples such as the flagship semi-synthetic production of the antimalarial sesquiterpene lactone, artemisinin (**Figure 1-11**). This project saw the development of a yeast-based platform for artemisinin production by transfer of the biosynthetic genes from the native producing plant *Artemisia annua.* Successful utilisation of this microbial host required the overexpression of the entire MVA pathway to increase precursor supply, as well as optimisation of enzyme stoichiometry and expression of relevant coenzymes to

avoid build-up of toxic intermediates, thereby improving productivity and host cell viability [24]. The final strains were capable of producing 25 g per L of artemisinic acid, which can be subsequently photochemically converted to artemisinin [6, 24].

More recently, engineering of other highly complex specialised metabolic pathways from plants into heterologous hosts has been achieved. These include engineering yeast to produce the monoterpene indole alkaloid strictosidine (**Figure 1-11**), a precursor for anticancer drugs normally derived from the Madagascan periwinkle *Catharanthus roseus* [5]. Yeast has also been engineered for production of opioids, an important class of alkaloid painkillers normally only attainable from poppy (**Figure 1-11**) [7]. These efforts required incorporation of in excess of ten enzymes to catalyse the necessary chemical transformations to the desired metabolites from common precursors. Additionally enzymes from animals, plants and yeast, mutant and fusion enzyme variants and host gene knockouts were all employed to improve the titre of the final products. These endeavours showcase the enormous potential of synthetic biology to deliver complex natural products which may be limiting in nature.

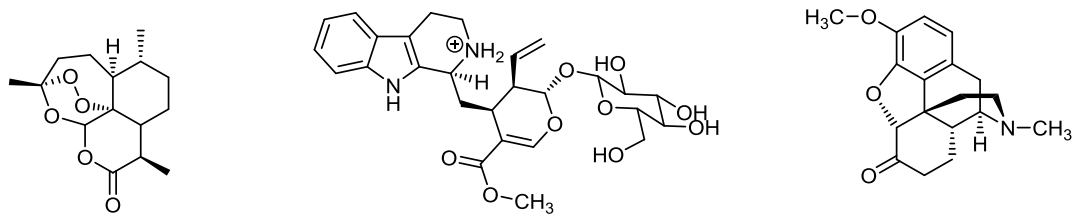


Figure 1-11: Synthetic biology for engineering plant specialised metabolites. The structures of some important plant specialised metabolites which have been engineered in heterologous hosts are shown. The native host producing species is listed below each product.

Choosing an appropriate host producer is a key factor in successful production of the desired compound. Suitable hosts should be well-characterised, fast-growing and amenable to genetic manipulation. Bacteria such as *Escherichia coli* are common platforms for hosting a variety of biosynthetic pathways, particularly from other bacterial species. However these lack cellular compartmentalisation and post-translational processing capability which can impede reconstruction of pathways from eukaryotic systems. For instance, P450s normally localise to the endoplasmic reticulum (ER) and consequently often function poorly in *E. coli*. Hence eukaryotic

systems such as yeast have become the preferred production hosts for high-value plant terpenoids such as artemisinin [24].

For triterpenoids, yeast is also the platform most-used for validating the function of OSCs and their associated P450s. Furthermore, this host has shown promise for the production of simple synthetic triterpene glycosides [108]. However an open question is whether yeast will prove to be suitable for the production of more complex triterpenes such as saponins. Saponins frequently feature uncommon sugars and acyl moieties [72] which are not products of endogenous yeast metabolism and would require further development of yeast strains. Furthermore, the inherent antimicrobial activity of many saponins may also impose difficulties on efficient production. These factors warrant the parallel exploration of the potential of other host organisms.

1.6.1 – *Nicotiana benthamiana* is an emerging host for producing plant specialised metabolites

As photosynthetic organisms, plants have excellent potential for the environmentally-sustainable production of high-value products. Furthermore, plants share common pathways, codon usage, subcellular compartments and enzyme cofactors, thereby facilitating transfer of metabolic pathways from one plant species to another. Indeed tobacco-based systems have also proven to be viable platforms for artemisinin production [120, 121].

N. benthamiana is a fast-growing, wild relative of tobacco which has grown in popularity in recent years as a heterologous expression host for numerous plant metabolites. One of the major reasons for this is due to the amenability of this species to transient gene expression through agroinfiltration. This technique involves infiltration of *Agrobacterium tumefaciens* into the interstitial spaces of *N. benthamiana* leaves. The bacteria then transfer a section of DNA (T-DNA) carrying the gene(s) of interest from a binary vector to the plant cell where they are translated. If these genes encode biosynthetic enzymes, this technique can be used for production of specialised metabolites. To-date, transient expression in *N. benthamiana* has been used to produce a range of molecules derived from other plant species including seco-iridoids [122], glucosinolates [123, 124] and cyanogenic glucosides [125]. A range of terpenes have likewise been successfully produced in this system, including monoterpenes [126, 127], diterpenes [112, 128-130], sesquiterpenes [131-134] and triterpenes [85, 93, 95, 105].

1.6.2 – Realising *N. benthamiana* as a (tri)terpene production platform

In order to enhance the capacity of *N. benthamiana* for terpene production, a number of strategies can be employed to augment the yield of the desired product. In the first instance, agroinfiltration can be used to manipulate the host plant's metabolism to overcome metabolic bottlenecks and loss of intermediates to other pathways. Such strategies include expression of yield-boosting enzymes [128, 132] and silencing of competing pathways [131]. Targeting certain proteins to non-native subcellular compartments such as mitochondria or chloroplasts might also be employed to increase yields [127, 129, 131, 132] (**Figure 1-12**).

Molecular techniques such as Golden Gate cloning facilitate generation of single vectors containing multiple genes [135]. This allows for entire biosynthetic pathways to be expressed from a single strain of *A. tumefaciens*. Furthermore, vector construction can incorporate suites of well-characterised promoters, terminators, localisation signals and UTRs [136]. Hence fine-tuned control over individual enzymes in a pathway is possible and may be utilised for optimising production (**Figure 1-12**). In parallel to optimisation of the molecular components, the agroinfiltration process itself has proven to be highly scalable which offers the opportunity to apply this technique for mass production of high-value chemicals (**Figure 1-12**). Companies such as Medicago (www.medicago.com) are capable of large-scale vacuum infiltration of *N. benthamiana* for the production of virus-like particles. The development of Leaf Systems at the Norwich Research Park in England (leafsystems.com) will similarly allow for scaled production and purification of both proteins and small metabolites in the near future.

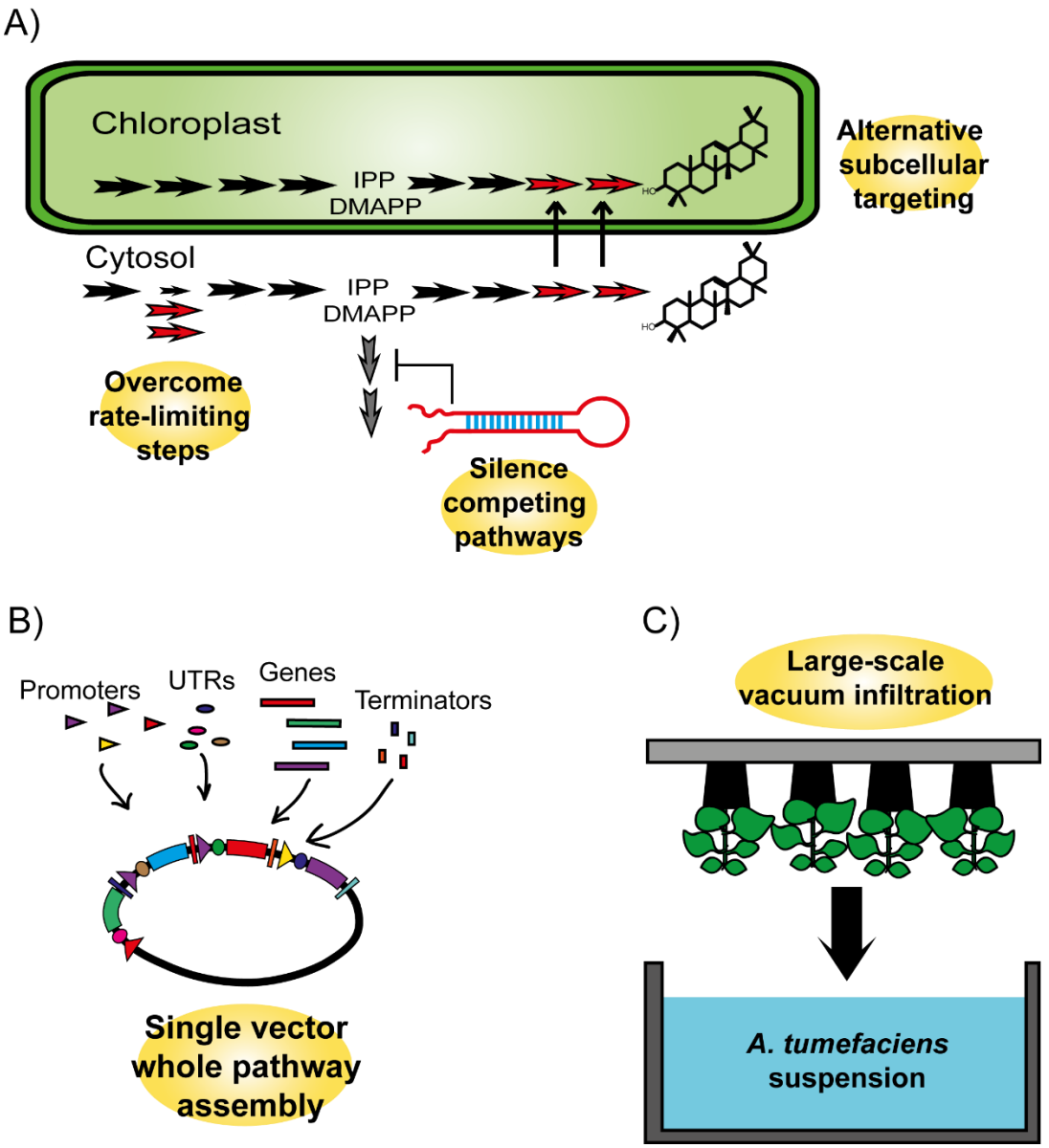


Figure 1-12: Strategies to increase terpenoid yield in *N. benthamiana* **A)** Manipulating the host plant metabolism can be achieved by overcoming endogenous rate-limiting steps or silencing competing pathways. Targeting the enzymes for the product of choice to alternative subcellular locations such as chloroplasts (or mitochondria) may also be used. **B)** Entire pathways can be encoded in a single vector which allows for the relative expression of individual proteins to be altered. **C)** Agroinfiltration can also be scaled through the use of techniques such as vacuum infiltration which allows for much higher throughput than infiltration by hand.

1.7– PhD overview:

The number of characterised triterpene synthases, oxidases and glycosyltransferases has expanded greatly in recent years. This presents for the first time an opportunity to harness the vast structural diversity of this important class of molecules. While *N. benthamiana* has been investigated as a platform for production of other classes of terpenes, at present no systematic investigation has been conducted into the value of this plant as a host for triterpene production. This PhD thesis attempts to address this. In the following chapters, the system is first developed to enhance triterpene yields (Chapter 3). Following this, the possibility of using *N. benthamiana* as a general triterpene production platform is explored, resulting in synthesis of a range of both extant and previously-unreported triterpenes (Chapter 4). Furthermore, scaled production and purification of a series of triterpene analogues from *N. benthamiana* is performed to begin a systematic structure-function study of antiproliferative and anti-inflammatory activity (Chapter 5). Finally, the system is used to screen and identify a number of previously-unreported triterpene oxidases involved in avenacin biosynthesis (Chapter 6). Collectively this work suggests that *N. benthamiana* is likely to be an excellent system for the rapid production and purification of a diverse array of triterpenes.

Chapter 2:

General materials and methods

2.1 – General methods

2.1.1 – Gateway cloning

2.1.1.1 – Entry clones

For generating Gateway Entry clones, pDONR207 was used in all cases unless otherwise specified. Gateway cloning was performed by incubation of the insert DNA (purified PCR product or other vector) and pDONR207 in equal ratios (approximately 150ng each) in a total volume of 8µL of Tris-EDTA (TE) buffer (10mM Tris (pH 8.0 with HCl) and 1mM EDTA.). A further 2µL of BP clonase II (Invitrogen) was added and vortexed briefly. The enzyme reaction was incubated for approximately 1 hour at room temperature. Following this, 5µL of the BP clonase reaction was added to chemically competent *E. coli* for transformation as detailed in **section 2.1.2.1**. The remaining 5µL was kept aside at room temperature for 24 hours and used in the event that the first transformation was unsuccessful. Positive cells were selected for on LB agar plates plus 50µg/ml gentamicin (pDONR207). Colony PCR was performed to screen colonies (**section 2.1.1.3**) and positive colonies were cultured overnight (37°C, 200rpm) in 10mL liquid LB (plus 50µg/ml gentamicin). Plasmids were recovered using a QIAprep spin miniprep kit (Qiagen) and sent for sequencing using attL-1/R-2 primers and where necessary, additional custom primers as detailed in **Table 2-1**.

2.1.1.2 – Expression clones

To generate pEAQ-*HT*-DEST1 vectors carrying the gene of interest, Gateway cloning was performed by incubation of the Entry vector (usually a pDONR207-derived construct) with pEAQ-*HT*-DEST1 in equal ratios (approximately 150ng each) in 8µL TE buffer. A further 2µL of LR clonase II was added to the reaction and incubated at room temperature for a minimum of 6 hours. 5µL of the LR clonase reaction was used to transform chemically competent *E. coli* as in **section 2.1.2.1**. The remaining 5µL was kept aside at room temperature for up to 24 hours and reused if the first transformation was unsuccessful. Positive transformants were selected for on LB agar kanamycin plates plus 50µg/ml kanamycin. It was generally not necessary to perform colony PCR. Finally, a single colony was picked from the plate and grown overnight (37°C, 200rpm) in 10mL liquid LB media (plus 50µg/ml kanamycin).Plasmids were purified using a QIAprep spin miniprep kit (Qiagen) and used to transform *A. tumefaciens* LBA4404 (**section 2.1.2.3**).

2.1.1.3 – Colony PCR

Individual colonies were labelled for identification and a small stab of the colony was taken using a 10µL pipette tip and resuspended in 30µL distilled water by pipetting up and down. 1µL of the suspension was used as the template in the PCR reaction as follows:

Reagent	Per sample
Template	1µL
attL1/2 primers (10µM each)	1µL
2x GoTaq Green master mix	5µL
H2O	3µL
Total	10µL

For thermal cycling, the following parameters were used:

	Step	Temp (°C)	Time
	Denaturation	98	30 sec
30 cycles	Denaturation	98	10 sec
30 cycles	Annealing	55	10 sec
30 cycles	Extension	72	1 min/kb
	Extension	72	5 minutes

Following the PCR reaction, 5µL was analysed on a 1% agarose gel. Positive colonies were identified as those which had a band corresponding to the size of the expected insert (plus ~200bp due to the attL primer binding sites).

2.1.1.4 – DNA electrophoresis

For analysis of DNA by agarose gel electrophoresis, an agarose gel was made with addition of usually 1:99 w:v agarose (Sigma) to an appropriate volume of TAE buffer (Formedium) with approximately 0.01% ethidium bromide (1mg/mL solution). Gels were run at constant 120 V for approximately 20-40 minutes depending on the size of the gel. A 2-log ladder (NEB) was included Gels were visualised under UV light.

2.1.1.5 – Sequencing

Sequencing was performed by Eurofins (http://www.eurofinsgenomics.eu/). Purified plasmid templates were sent for sequencing. Sequencing primers are given in **Table 2-1**.

2.1.1.6 - 2-step Gateway PCR:

For the first step, 10µL PCR reactions were setup as follows:

Reagent	Per sample
HF Buffer (5x)	2µL
dNTPs (10mM)	0.2µL
Primers (10mM)	2µL
Template	0.5µL
Phusion polymerase	0.1µL
H_2O	5.2µL
Total	10µL

Tubes were sealed and thermal cycling was performed as below:

	Step	Temp	Time
10 cycles	Denaturation	98°C	30 sec
	Denaturation	98°C	10 sec
	Annealing	varied	10 sec
	Extension	72°C	1 min/kb
	Extension	72°C	3 minutes

For the second step, master mix (40µL/sample) was made as shown below and pipetted into the 10µL PCR tube above.

Reagent	Per sample
Buffer (5x)	8µL
dNTPs (10mM)	0.8µL
attB1/2 primers (10mM)	4µL
Phusion polymerase	0.4µL
H_2O	26.8µL
Total	40µL

Thermal cycling was performed as shown below:

	Step	Temp	Time
	Denaturation	98°C	30 sec
10 cycles	Denaturation	98°C	10 sec
	Annealing	45°C	10 sec
	Extension	72°C	1 min/kb
20 cycles	Denaturation	98°C	10 sec
	Annealing	55°C	10 sec
	Extension	72°C	1 min/kb
	Extension	72°C	3 min

2.1.1.7 – Common primers

	Name	Sequence
Sequencing Primers	attL1-F	TCGCGTTAACGCTAGCATGGATCTC
	attL2-R	ACATCAGAGATTTTGAGACACGGGC
	M13-F	GAGCGGATAACAATTTCACACAGG
	M13-R	AGGGTTTTCCCAGTCACGACGTT

	Name	Adapter sequence
Gateway adapter sequences	attB1F-1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA
	attB2R-1	GGGGACCACTTTGTACAAGAAAGCTGGGTA
	attB1F-2	GGGGACAAGTTTGTACAAAAAAGCAGGCT
	attB2R-2	GGGGACCACTTTGTACAAGAAAGCTGGGT
	attB1Fs-1	AAAAAGCAGGCTTA
	attB2Rs-1	AGAAAGCTGGGTA

Table 2-1: A list of commonly used primer sequences. **Top:** Sequencing of Gateway Entry vectors was commonly performed using the attL1/2 F/R primers, while M13 F/R primers were commonly used for TOPO-based vector sequencing. **Bottom:** Gateway adapter sequences added to the 5' end of cloning primers. attB1Fs and attB2Rs primers are truncated versions of the full-length Gateway attB primers and were used for the 2-step PCR as detailed in **section 2.1.1.6**.

2.1.1.8 – Plasmids

Name	Vector Type	Supplier	Selection
pDONR207	Gateway Entry	Invitrogen	Gentamicin
pENTR-D-TOPO	Gateway Entry	Invitrogen	Kanamycin
Zero Blunt TOPO	Non-Gateway cloning	Invitrogen	Kanamycin
pEAQ-*HT*-DEST1	Gateway Destination	Lomonossoff lab, JIC	Kanamycin

Table 2-2: List of plasmids used in the present thesis.

2.1.2 – Microbiological techniques

2.1.2.1 – Transformation of chemically-competent *E.coli*

Plasmid DNA was prepared as detailed in specific sections. *E. coli* was stored as 50µL aliquots of cells at -80°C. For transformation, cells were removed from the freezer and transferred to ice to thaw. DNA was incubated with the cells on ice for approximately 30 minutes before heat shock in a water bath set to 42°C for 20-30 seconds. Cells were transferred back to ice for a further 2 minutes before addition of 200µL of media (LB or SOC). Cells were then incubated for 1 hour at 37°C at 200rpm. Finally cells were removed from the incubator and plated onto LB agar plates with appropriate antibiotics for selection of positive transformants. Plates were incubated at 37°C overnight.

2.1.2.2 – Preparation of chemically competent A. tumefaciens

A glycerol stock culture (-80°C) of *A. tumefaciens* LBA4404 was streaked onto LB agar plates containing rifampicin and streptomycin (50µg/mL and 100µg/mL, respectively). These were grown overnight at 28°C in a standing incubator. A second overnight culture was performed in 50mL LB media (plus antibiotics as above) at 28°C, 200rpm. The following morning, the cultures were placed on ice for 10 minutes before centrifugation at 4°C (4,500g) for 5 minutes. The supernatant was discarded and cells were resuspended in 1ml ice cold $CaCl_2$ (20mM). The cells were again centrifuged as before and resuspended in cold 20mM $CaCl_2$. Finally, cells were aliquoted in 50µL batches and flash frozen in liquid N_2 before transfer to the -80°C freezer.

2.1.2.3 – Transformation of *A. tumefaciens* LBA4404

Chemically competent *A. tumefaciens* strain LBA4404 were taken from the -80°C freezer, thawed on ice (in microcentrifuge tubes) and incubated with 1-200ng of the pEAQ-*HT*-DEST1 vector carrying the gene of interest for approximately 30mins. The tubes were transferred to liquid N_2 for 30 seconds before removal and thawing at room temperature. 200µl of SOC medium was added to each tube and the tubes were transferred to the 28°C shaking incubator at 200rpm for 2-4 hours. After this, tubes were centrifuged briefly and the bacterial suspension was transferred to LB agar plates containing rifampicin, kanamycin and streptomycin (50/50/100µg per mL respectively). The plates were stored in the 28°C standing incubator for three days until colonies could clearly be seen. After this period, a single colony was transferred to 5mL liquid LB media containing the same three antibiotics/concentrations as above and transferred to the 28°C shaking incubator

overnight. Around 1mL of the culture was then transferred to a labelled microcentrifuge tube containing 200µL glycerol and stored at -80°C. These freezer stocks were subsequently used for the culturing of the *Agrobacteria* for future infiltrations.

2.1.2.4 – Antibiotics

Antibiotics were made as detailed in **Table 2-3**. The same concentrations were used in both liquid and solid (agar) LB media.

Antibiotic	Stock	Final	Dilution factor
Gentamicin	50mg/mL in H_2O	50µg/mL	1/1000
Kanamycin	50mg/mL in H_2O	50µg/mL	1/1000
Streptomycin	100mg/mL in H_2O	100µg/mL	1/1000
Rifampicin	50mg/mL in dimethylformamide	50µg/mL	1/1000
Chloramphenicol	30mg/mL in Ethanol	30µg/mL	1/1000
Ampicillin	100mg/mL in H_2O	100µg/mL	1/1000

Table 2-3: List of antibiotics used for culturing bacteria. All antibiotic stocks were stored at 1mL aliquots in the specified solvent at -20°C.

2.1.3 – Agroinfiltration

2.1.3.1 – Growth of *N. benthamiana*

N. benthamiana plants were grown under greenhouse conditions set to 25°C with 16 hour lighting. Supplementary lighting was used during the winter months. Seedlings were sown in F1 compost (Levington) and grown for 2 weeks before transfer to individual cells containing F2 compost (Levington). Plants were grown for an additional 3-4 weeks before infiltration.

2.1.3.2 – Preparation of *A. tumefaciens* for agroinfiltration

A. tumefaciens cultures stored as glycerol stocks were streaked onto LB agar plates containing appropriate rifampicin, kanamycin and streptomycin (50/50/100µg per mL respectively) and transferred to a standing incubator at 28°C for 24 hours. A sterile pipette tip was used to transfer a small amount of the bacterial streak to liquid media (usually 10mL) with antibiotics as above. This was incubated overnight in the 28°C shaking incubator (200rpm). Depending on the application and volume required, a second subculture (up to 1.5L per strain for vacuum infiltration, but typically <200mL for hand infiltrations) was started by transferring the contents of the first culture to fresh LB media with antibiotics as above.

For preparation of the infiltration suspension, the cultures were centrifuged at 4000 x *g* to pellet the cells and the supernatant was discarded. The cells were resuspended in an small volume (~5mL per 50mL culture) of MMA solution and incubated for 1 hour.

To measure the optical density, a spectrophotometer was set to 600nm and blanked using MMA solution. The bacterial suspension was diluted (usually 1 in 10) in MMA in a 1mL cuvette to bring the suspension to within the accurate range of the spectrophotometer and measured. The reading was multiplied by the dilution factor to get the actual density of the cultures and recorded. For adjusting culture density to the desired OD_{600} the following equation was used:

$$\frac{\text{Current } OD_{600} \text{ x current suspension volume (mL)}}{\text{Final } OD_{600}}$$

The figure used for 'final OD_{600}' was ***n*** x 0.2 (where ***n*** is the number of strains to be coinfiltrated). This ensured that when mixed in equal ratios, each strain would be at a density of OD_{600} = 0.2, as required for ensuring plant cells received a copy of all genes. The figure from the equation above gave the total volume (mL) of suspension needed to reach the desired final OD_{600} and MMA solution was added to reach this volume. Appropriately diluted suspensions of individual strains were then combined in equal ratios.

2.1.3.3 – Agroinfiltration of *N. benthamiana*

Strains of *A. tumefaciens* were prepared as in **section 2.1.3.2**. For hand infiltration of *N. benthamiana* plants, expanded leaves of five-week-old plants were first perforated with a pipette tip (10µl) (**Figure 2-1A-C**). The holes were used to aid infiltration of the *A. tumefaciens* suspensions from the underside of the leaf using a 5mL needleless syringe (**Figure 2-1D-F**). Unless otherwise indicated, the whole leaf was infiltrated. Following infiltration, plants were left under the same greenhouse conditions for an additional five days. Harvesting and downstream processing of leaves is detailed in specific sections.

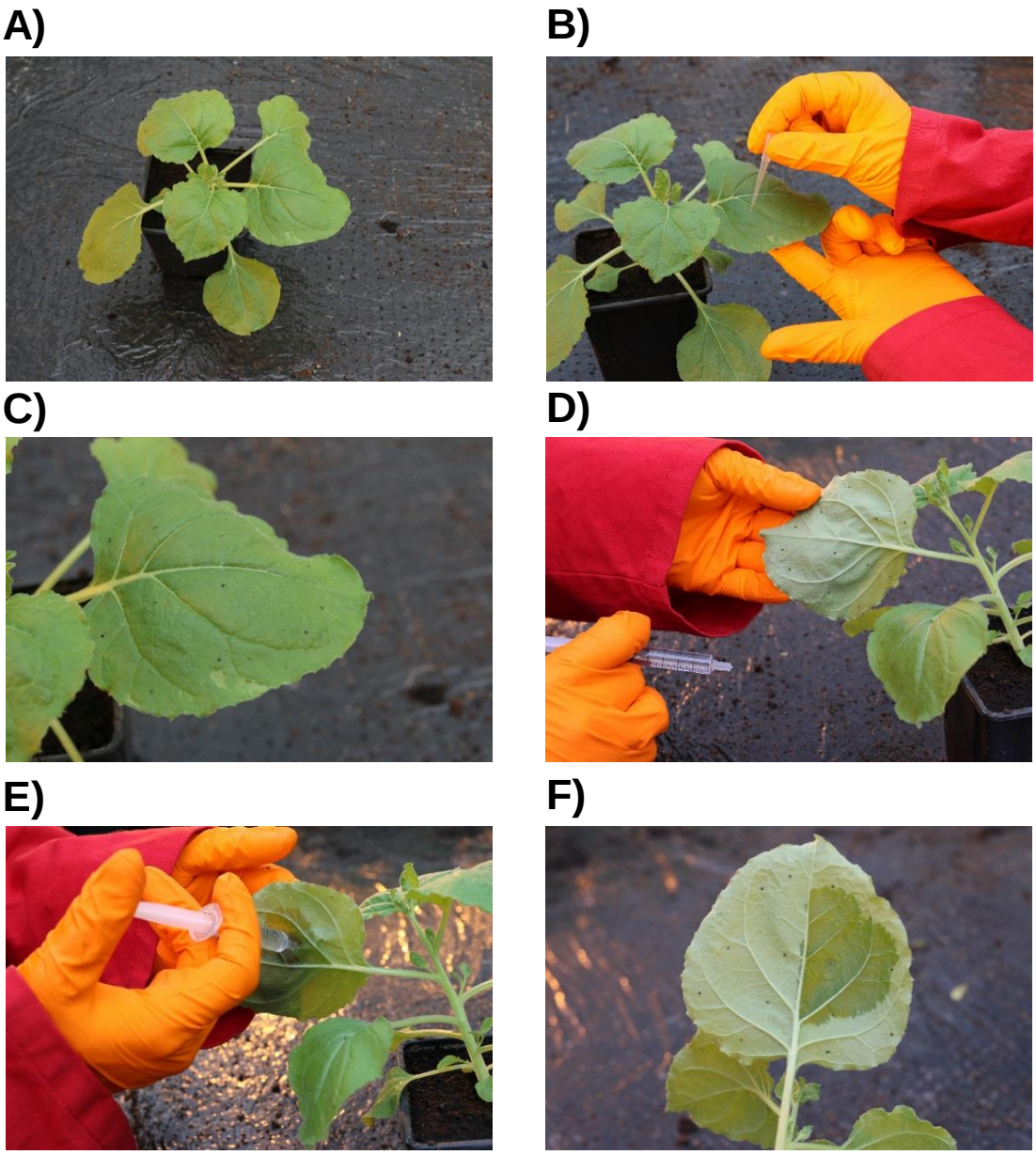


Figure 2-1: Infiltration of *N. benthamiana* by hand **A)** A representative 5-week-old *N. benthamiana* plant grown under greenhouse conditions. **B)** Perforation of a leaf using a pipette tip. This does minimal damage to the leaf and makes the infiltration process much easier **C)** Multiple perforations are usually required for whole leaf infiltration **D)** The underside of the leaf is used for the infiltration. **E)** Infiltration using a 1mL syringe. **F)** The infiltrated area can be visualised easily as a darkened patch on the underside of the leaf. This process is usually repeated until the whole leaf is infiltrated.

An infiltrator for vacuum agroinfiltration of *N. benthamiana* was built by Dr Bastiaan Brouwer and contained within the greenhouse housing the plants. Briefly, this consists of a chamber connected to a vacuum pump via an intermediary vacuum reservoir (**Figure 2-2A**). A 10L removable bath fits in the chamber and is filled with the appropriate agroinfiltration suspension. A bespoke acrylic holder was designed

to hold up to four 5-6-week-old *N. benthamiana* plants simultaneously in 9cm pots (**Figure 2-2B**). The plants were positioned in the holder and inverted into the filled bath so that the aerial parts of the plants were completely submerged (**Figure 2-2B-C**). For infiltration, the vacuum was applied to ~26mmHg and held for approximately one minute to draw the air out of the interstitial leaf space. The vacuum reservoir was then isolated and the chamber re-equilibrated with the external atmosphere over 20-30 seconds to allow the bacterial suspension to be drawn into the leaves. Generally this process efficiently infiltrates all of the submerged leaves. Plants were carefully removed from the chamber and returned to their former positions in the greenhouse and left for 5 days.

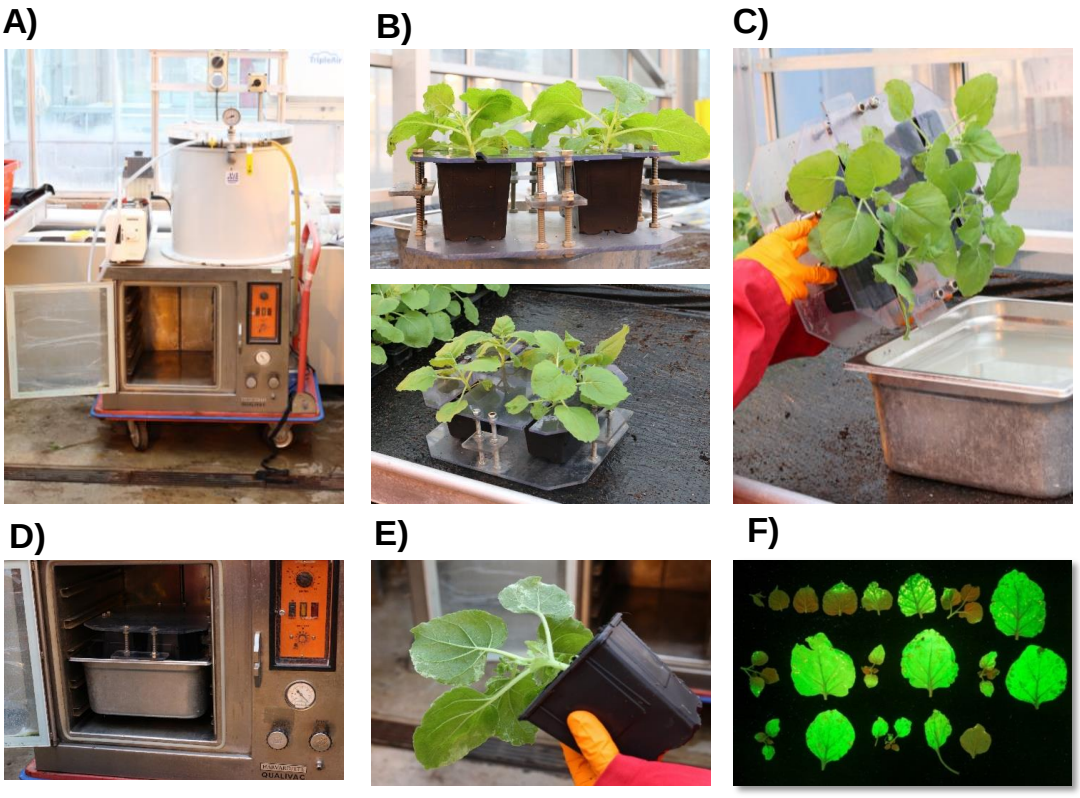


Figure 2-2: Vacuum infiltration of *N. benthamiana*. **A)** The infiltrator as used in the present thesis. The vacuum chamber is located at the base and the vacuum pump sits atop the chamber. An intermediary reservoir vacuum increases the speed at which the necessary pressure can be achieved. **B)** Plants are loaded into an acrylic holder which can accommodate four plants simultaneously. **C)** The plants are inverted within the holder into the bath containing ~10L of agrobacterium suspension. **D)** The submerged plants are loaded into the vacuum chamber for infiltration. **E)** A plant after infiltration, the efficiency of infiltration can be seen by the darker infiltrated areas. **F)** A vacuum-infiltrated plant expressing *GFP* five days after infiltration as seen under UV. Leaves are arranged in approximate order of age, (youngest-oldest, left to right). The youngest leaves (top left) are not expressing GFP as these have emerged in the days following infiltration. Picture F) kindly provided by Dr Bastiaan Brouwer.

2.1.4 – General considerations for purification of compounds

Leaf material was harvested five days after infiltration either by hand or vacuum as described in **section 2.1.3**, taking efforts to ensure that only the producing parts of the plants were sampled. Leaves were frozen at -80°C and lyophilised prior to use to remove water. Dried leaves were ground in liquid N_2 using a pestle and mortar and saponified for 2 hours at 65°C. Following saponification, the ethanol was removed by rotary evaporation and the products were extracted by partitioning multiple times with ethyl acetate. The ethyl acetate was dried and the extract was subject to flash chromatography to isolate the compound of interest. Following flash chromatography, decolourisation and recrystallization were frequently performed to remove minor impurities.

2.1.4.1 – Flash chromatography

All flash chromatography steps were performed using a Biotage Isolera One. Samples were either wet loaded in toluene or dry loaded on silica gel. Flash chromatography was performed under normal phase conditions using a mobile phase of hexane (weak) and ethyl acetate (strong). Details of the various programs are given in **Table 2-4**.

2.1.4.2 – Decolourisation.

Compounds were dissolved in several mL of ethanol and decolourisation was performed by adding a small amount of activated charcoal powder (Sigma). This was incubated for several minutes at room temperature with intermittent agitation. The suspension was then transferred to a column made from a Pasteur pipette packed with diatomaceous earth (Celite, Sigma) between two thin layers of cotton wool and pre-equilibrated with ethanol. The suspension was passed through the column and the subsequent eluent was collected in a round bottom flask. The column was washed multiple times with ethanol to ensure maximal recovery of compound and this was likewise collected. Finally the collected solution was dried using rotary evaporation.

2.1.4.3 – Flash chromatography programs

Program	Strong solvent	Col Vols	Program	Strong solvent	Col Vols
P-1	6%	2	**P-8**	15%	15
	6-100%	10		15-20%	10
	100%	7		20%	15
P-2	0-50%	60		20-100%	5
P-3	5%	2		100%	5
	5-100%	18	**P-9**	6%	2
	100%	4		6-50%	6
P-4	20%	20		50-100%	1
	20-100%	3		100%	5
	100%	5	**P-10**	10%	1
P-5	20%	15		10-80%	10
	30%	10		80%	2
	30-100%	5	**P-11**	7%	1
	100%	5		7-70%	10
P-6	5%	2		70%	2
	5-100%	15	**P-12**	7%	1
	100%	5		7-60%	10
P-7	5%	1		60%	2
	5-40%	15	**P-13**	0-100%	20
	40%	2		100-100%	40
	40-100%	2			
	100%	10			

Table 2-4: Flash chromatography programs. The solvent system was always hexane (weak) and ethyl acetate (strong). The volume of solvents are represented in column volumes (col vols) due to differences in column size. Details of these columns are given in the relevant sections. Flow rate was always the maximum recommended by the manufacturer (Biotage).

2.1.4.4 – Recrystallization

Two flasks containing ethanol and water were placed in a ~90°C water bath. Compounds were dissolved in a separate vessel in a minimal amount of ethanol (90°C). The flask was kept at 90°C and hot water was slowly added dropwise, allowing any precipitate to re-dissolve fully between drops. This was repeated until the solution became slightly turbid, at which point the flask was removed from the water bath and left at room temperature covered with a small glass plate without disturbing for ~30 mins. The flask was further transferred to ice for ~30 minutes to fully precipitate compounds. The recrystallized suspension was transferred to a 15mL centrifuge tube and centrifuged briefly before removal of the supernatant. Crystals were then dissolved in an appropriate solvent (usually dichloromethane: ethyl acetate 1:1 and transferred to a vial before evaporation under N_2 followed by lyophilisation overnight.

2.1.5 – GC-MS programs:

2.1.5.1 – 31 minute program

The GC system used was an Agilent 7890B fitted with a Zebron ZB5-HT Inferno column). 1µL of the sample was injected into the GC inlet (250°C) using a pulsed splitless method (30psi pulse pressure). The oven temperature gradient began at 170°C (held 2 min), then ramped to 290°C at 6°C/min and held for 4mins. Finally the oven temperature was increased to 340°C at a rate of 10°C/min. The final time was 31mins. For analysis of mass spectra, the solvent delay was set to 15mins. The detector was set to scan mode over a range from m/z 60-800.

2.1.5.2 – 20 minute program

GC-MS analysis was performed using the GC-MS and column as described above. 1µl of each sample was injected into the GC inlet (250°C) using a pulsed splitless mode (30 psi pulse pressure). The oven temperature program was set to 170°C for 2 mins, followed by a ramp to 300°C at 20°C/min. This was held at 300°C for a further 11.5 min to give a final run time of 20 minutes. Electron-impact mass spectrometry was performed using an Agilent 5977A Mass Selective Detector, set to scan mode from 60-800 mass units. The solvent delay was set to 8 minutes. For quantification, the detector was set to 7.2 scans/sec.

2.1.6 – Common reagents and solutions

2.1.6.1 – Triterpene standards

Compound	Supplier
Squalene	Sigma
2,3-Oxidosqualene	Sigma
β-Amyrin	Extrasynthese
Coprostanol (internal std)	Sigma
DD-II	Dr R. Thimmappa
EpDM	Dr R. Thimmappa
Erythrodiol	Extrasynthese
Glycyrrhetinic acid	Extrasynthese
Hederagenin	Apin
Oleanolic acid	Extrasynthese
Protopanaxadiol	Extrasynthese
Protopanaxatriol	Extrasynthese
Cycloartenol	Sigma

Table 2-5: List of triterpene standards as used to verify products in this thesis. Compounds from Dr Ramesha Thimmappa are as detailed in Salmon et al [137].

2.1.6.2 – MMA solution:

10mM 2-(N-morpholino) ethanesulphonic acid buffer (MES, Sigma) pH5.6 (KOH)
10mM $MgCl_2$
150µM 3’5’-dimethoxy4’acetophenone (Acetosyringone, Sigma).

2.1.6.3 – Saponification solution

The saponification solution consisted of ethanol:H_2O:KOH mixed 9:1:1 (v:v:w). The KOH was first fully dissolved in water 1:1 (w:v) before addition of ethanol. Where the internal standard (coprostanol, Sigma) was added, the saponification solution volume was first accurately measured using a measuring cylinder or volumetric flask before addition of the standard.

Chapter 3:

Investigating strategies for improving triterpene production in *N. benthamiana*

Acknowledgements:

All of the work is this chapter is the author's own. Constructs for tHMGR and CYP51H10 were provided by members of the Osbourn lab. Constructs for GFP were kindly provided by the Lomonossoff group.

3.1 – Introduction

Triterpenes are a structurally diverse class of specialised metabolites with great medical and agronomic potential [34, 35, 39, 48, 57]. However exploitation of these compounds is often hindered by their limited availability from natural sources and problems in isolating single products from extracts containing mixtures of structurally similar compounds [46]. Chemical synthesis of triterpenoids is hampered by their structural and stereochemical complexity, with total synthesis of even simple scaffolds requiring numerous synthetic steps [138, 139]. Collectively, these problems limit the sourcing and modification of triterpenoids to exploit their full potential. Engineering triterpene production in heterologous hosts opens up alternative routes to access triterpenoids, thereby addressing the above problems. Knowledge of the genes and enzymes responsible for triterpenoid biosynthesis in nature allows their transfer to suitable expression hosts for the biosynthesis of high value compounds. *Nicotiana benthamiana*, a wild relative to tobacco, is one such host. *N. benthamiana* has gained popularity in recent years as a platform for reconstituting plant natural product pathways by transient expression [140, 141]. A number of studies have investigated strategies for enhancing the capacity of this system for biosynthesis of various terpenoids, primarily monoterpenes, sesquiterpenes and diterpenes (**Figure 3-1**). *N. benthamiana* shows potential for production of triterpenes [85, 93, 105]. However at present, little effort appears to have been made to optimise triterpene production in this system.

Compound	Class of molecule	Quoted yield	Reference
Geraniol	Monoterpene	27 µg/g FW	[142]
Geraniol	Monoterpene	129 µg/g FW	[127]
Amorphadiene	Sesquiterpene	6.2 mg/kg FW	[132]
Artemisinic acid diglucoside	Sesquiterpene	39.5 mg/kg FW*	[132]
Costunolide	Sesquiterpene	60 ng/g FW	[133]
Valencene	Sesquiterpene	0.70 µg/g/24h	[131]
Casbene	Diterpene	227 ng/cm^2	[128]
Cembratrienol	Diterpene	2500 ng/cm^2	[128]
Isopimaric acid	Diterpene	45–55 µg/g DW	[129]
Levopimaradiene	Diterpene	100 ng/cm^2	[128]
EpHβA	Triterpene	1.18 mg/g DW ǂ	[85]

Table 3-1: Reported yields of various terpenes produced in *N. benthamiana* Abbreviations: FW, fresh weight; DW, dry weight. NB: Yields are given as reported. It is not always possible to compare directly between studies due to differences in the units of reported yield. *The glycosylation of artemisinic acid results from endogenous *N. benthamiana* activity and corresponds to 16.6mg/kg FW free artemisinic acid. **ǂ** Indicates isolated yield.

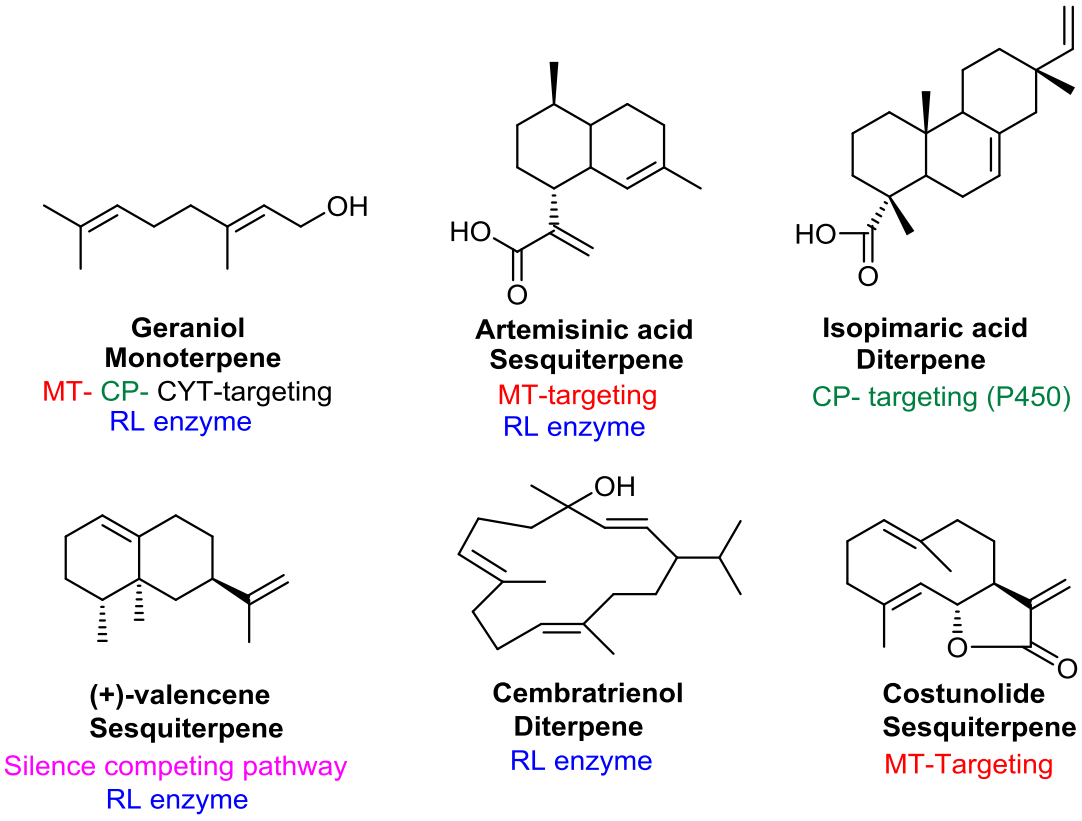


Figure 3-1: Metabolic engineering for increased terpenoid yields in *N. benthamiana* [127-129, 131-133]. Abbreviations: MT – mitochondria, CP – chloroplast, CYT - cytosol RL – expression of a rate-limiting enzyme.

3.1.1 – Engineering early pathway steps in triterpene production

Successful production of specialised metabolites in heterologous systems often benefits from overcoming the regulatory mechanisms within the host cell which might limit precursor availability. A common strategy is the overexpression of one or more 'rate-limiting' enzymes. Triterpene scaffolds such as β-amyrin are synthesised from acetyl-CoA via the mevalonate (MVA) pathway (**Figure 3-2**). The key regulatory enzyme in the MVA pathway is 3-hydroxy, 3-methylglutaryl-coenzyme A reductase (HMGR), which catalyses the irreversible conversion of HMG-CoA to mevalonate [60]. Both transcriptional and post-translational mechanisms exist to fine-tune HMGR activity based on both local feedback and global energy levels and highlights the central role that this enzyme plays in regulating flux through the MVA pathway [143, 144]. Importantly, post-translational regulation of HMGR activity suggests that increased expression of this enzyme may not be sufficient by itself to overcome this yield-limiting step. Hence overcoming the post-translational mechanisms regulating HMGR activity is an important strategy for increasing the yields of MVA pathway-derived terpenes.

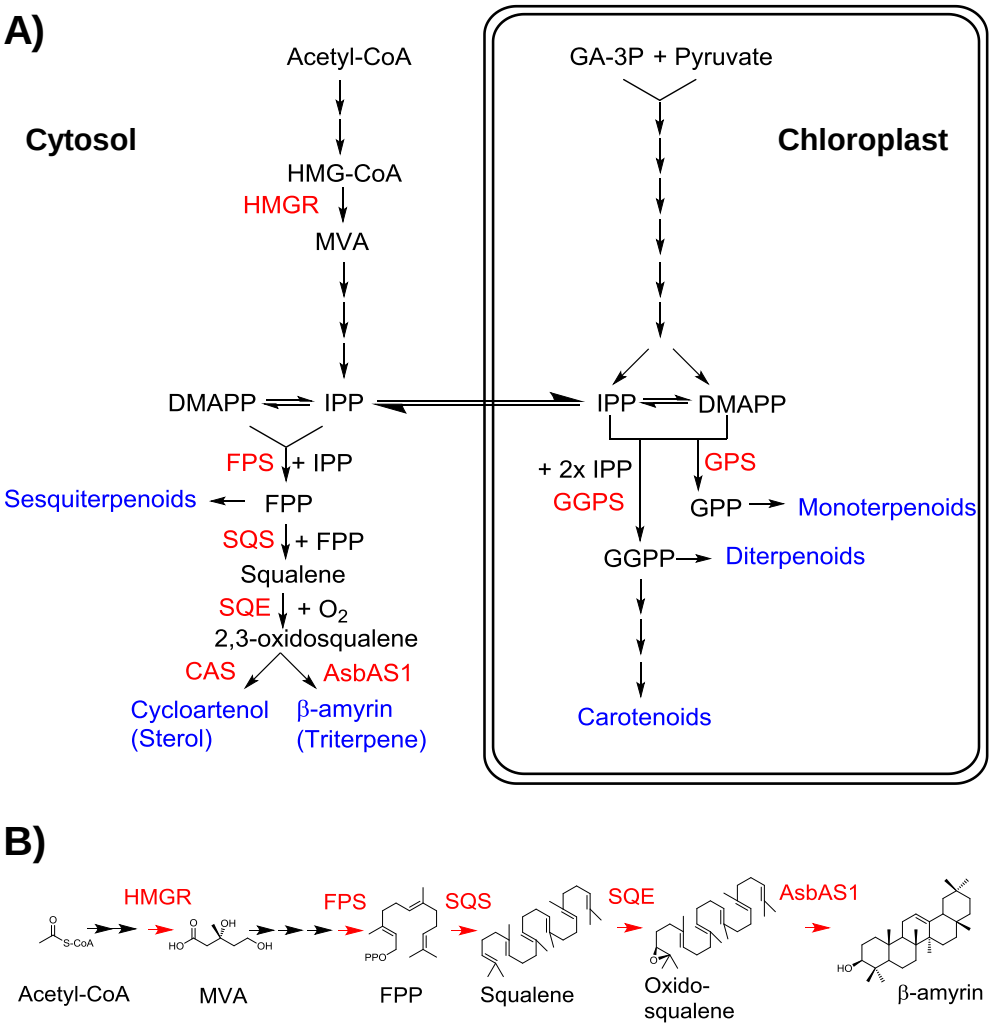


Figure 3-2: Triterpene biosynthesis **A)** MVA and MEP pathways within the plant cell. Enzymes are indicated in red. **Compound abbreviations**: CoA, Coenzyme A; HMG, 3-hydroxy-3-methylglutaryl; MVA, mevalonate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; GA-3P, D-Glyceraldehyde-3-phosphate; GPP, geranyl diphosphate; GGPP, geranylgeranyl diphosphate. **Enzyme abbreviations**: HMGR, HMG-CoA reductase; FPS, FPP synthase; SQS, Squalene synthase; SQE, Squalene epoxidase; CAS, cycloartenol synthase; AsbAS1, *Avena strigosa* β-amyrin synthase; GPS, Geranyl diphosphate synthase; GGPS, Geranylgeranyl diphosphate synthase. **B)** Biosynthesis of β-amyrin from acetyl-CoA. The structures of key intermediates relevant to this study are shown.

The N-terminal membrane-spanning region of HMGR is dispensable for catalytic activity, but has a key regulatory role [60]. Phosphorylation and likely ubiquitination of this region serve to repress plant HMGR activity [145, 146]. Removal of the N-terminal region can therefore alleviate regulatory control of the enzyme. Indeed expression of an N-terminal-truncated version of HMGR (tHMGR) has become a common strategy for boosting MVA pathway yields. Expression of tHMGR has been applied for enhancing triterpene and sterol content in yeast [108, 147, 148] and in stable transgenic lines of tobacco and rubber [149, 150]. Similarly, tHMGR has been used in transgenic tobacco to enhance the production of sesquiterpenes [120], which are also derived from the MVA pathway (**Figure 3-2**). This strategy is also applicable for transient expression and has been used for enhancing sesquiterpene yields in *N. benthamiana* [131, 132]. Therefore it is important to investigate how this may likewise be used to increase triterpene production in *N. benthamiana.*

Although the importance of HMGR within the MVA pathway is clear, the latter enzymes downstream of the universal terpene intermediates IPP and DMAPP may also be relevant targets for enhancing triterpene yields (**Figure 3-2**). Studies have demonstrated that overexpression of FPP synthase (FPS) [151] or squalene synthase (SQS) [152, 153] may improve triterpene yields. Furthermore, there is growing evidence for a regulatory role of squalene epoxidase (SQE) [154, 155] and this enzyme has demonstrated value as a target for metabolic engineering in yeast [148, 156]. Again, the value of these enzymes for improving triterpene production in *N. benthamiana* remains to be investigated.

3.1.2 – The importance of compartmentalisation in terpene biosynthesis

The realisation that plants have two subcellular isoprenoid biosynthetic pathways has prompted interest in the way plants synthesise and regulate production of the various classes of terpenes [60] (**Figure 3-2**). Although the cytosolic MVA and plastidic MEP pathways both produce the universal C5 terpene substrates IPP and DMAPP, the assembly of these into higher order terpenes generally follows a compartment-specific pattern. Production of sesquiterpenes and triterpenes results from the exclusive synthesis of the C-15 FPP in the cytosol, whereas plastidial production of monoterpenes, diterpenes and carotenoids (which are tetraterpenes) occurs through the production of C10 GPP and C20 GGPP intermediates [157].

In recent years, a number of studies have demonstrated that it is possible to synthesise sesquiterpenes in other compartments, often with dramatically increased yields. For example, redirection of FPS and a patchoulol synthase to the chloroplasts of tobacco plants resulted in approximately 10-fold increase in patchoulol content when compared to control lines expressing cytosolic FPS and PTS [158]. Similarly, targeting sesquiterpene cyclases to mitochondria has been demonstrated to be an effective strategy in both transgenic [120] and transient expression [132, 133]. Enhanced production of the triterpene precursor squalene has also been achieved in transgenic tobacco by targeting FPS and SQS to the chloroplasts [159]. However this strategy does not appear to have been applied to the production of cyclised triterpenes. These studies highlight the importance of considering alternative subcellular compartments in strategies to optimise terpene biosynthesis.

3.1.3 – Aims

Metabolic engineering is an important practice for improving yields of target metabolites in biological production systems. *N. benthamiana* has significant potential as a host system for the production of triterpenes, however to date no efforts appear to have been made to optimise triterpene production in this host. Therefore this chapter aims to systematically evaluate strategies for enhancing triterpene production in *N. benthamiana* through transient expression. The first approach investigates a series of possible rate-limiting enzymes in the MVA pathway which could be used to enhance triterpene yields. The second approach explores the potential of exploiting alternative subcellular sources of triterpene precursors by redirecting triterpene biosynthetic enzymes to the chloroplasts. These approaches provide important insight into transient expression strategies which may be employed for enhanced triterpene production in *N. benthamiana*.

3.2 – Results and discussion

3.2.1 – Production of β-amyrin in *N. benthamiana*

To assess the efficacy of the strategies employed herein for improving triterpene production in *N. benthamiana*, a candidate triterpene was first considered as a 'read-out' through quantification by gas chromatography mass spectrometry (GC-MS). The simple triterpene β-amyrin has previously been produced successfully by transient expression of the *Avena strigosa* (oat) β-amyrin synthase (AsbAS1) in *N. benthamiana* leaves [85]. This study utilised a cowpea mosaic virus (CPMV)-based expression system [160] whereby the *AsbAS1* gene was inserted between 5' and 3' untranslated regions (UTRs) of the CPMV RNA-2 present within a binary vector transfer DNA (T-DNA). The 5' UTR is modified resulting in enhanced translation of mRNA and high level protein accumulation [160]. This 'hypertranslatable' (CPMV-*HT*) system has since been developed into the pEAQ vector series [161]. In particular, the pEAQ-*HT*-DEST1 is a Gateway-compatible cloning system and contains a P19 silencing suppressor within the T-DNA [161]. This facilitates generation of constructs and prevents the need to coinfiltrate a separate P19-expressing strain of *A. tumefaciens*. This system was therefore chosen for the present work.

The *AsbAS1* coding sequence was cloned into the pEAQ-*HT*-DEST1 vector and the resulting construct used to transform *A. tumefaciens* strain LBA4404. A second *A. tumefaciens* strain carrying the green fluorescent protein gene (*GFP*) in pEAQ-*HT*-DEST1 was also generated for use as a control for the *N. benthamiana* infiltration experiments (**methods 3.4.2.1**). These strains were infiltrated into the leaves of 5-week-old *N. benthamiana* plants (**methods 3.4.5.1**). After 5 days, the leaves were harvested and extracts analysed by GC-MS. A novel peak was detected in the leaves infiltrated with the *AsbAS1* strain which could be identified as β-amyrin. This was not present in the *GFP*-infiltrated leaves (**Figure 3-3**). Since β-amyrin was readily detectable in extracts from *AsbAS1*-treated leaves and absent from the control treatment it was considered to be a suitable molecule to use for quantification in subsequent experiments directed at optimising the system.

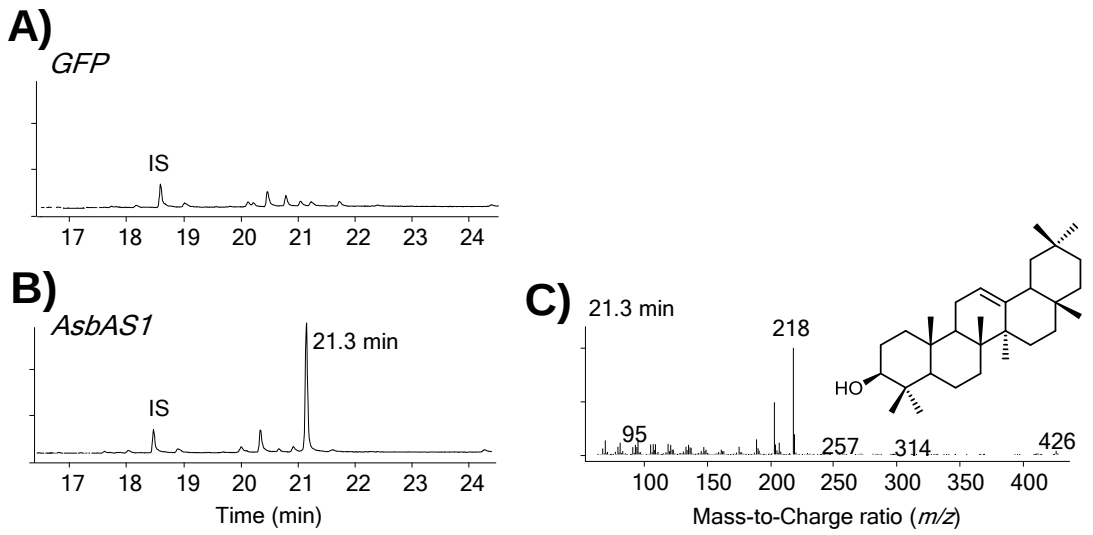


Figure 3-3: Production of β-amyrin in *N. benthamiana*. GC-MS analysis of extracts from *N. benthamiana* leaves expressing **A)** *GFP* only or **B)** *AsbAS1*. A mass spectrum for the peak at 21.3 (β-amyrin) is shown on the right. This matched previously reported mass spectral data [162] IS = internal standard (coprostanol)

3.2.2 – Improving production of β-amyrin in *N. benthamiana* with MVA genes

The effects of expression of *AsbAS1* with upstream MVA pathway enzymes was then investigated, since boosting the expression of rate-limiting enzymes may enhance β-amyrin levels. Coding sequences for HMGR, FPS, SQS and SQE were identified in an oat root tip transcriptome database by searching with sequences from characterised enzymes in *A. thaliana* [29] (**Figure 3-2**). Details of the oat isoforms are given in **Supplementary Table 1.** Briefly, only a single full-length isoform was present in the 454 database for each of the *FPS, SQS* and *SQE* genes. Two full-length isoforms of *HMGR* were found, plus a third for which only part of the sequence was present (**Supplementary Figure 1**). It is possible that other isoforms of the four enzymes identified here may be present within the oat genome but not expressed in the root tips.

Next, full-length clones of *FPS*, *SQS* and *SQE* were cloned from *A. strigosa* root tip cDNA (**methods 3.4.2.2**). For *HMGR*, the N-terminus serves a regulatory role and removal of this region has been reported to increase the activity of the protein [150]. Therefore a truncated form of HMGR (tHMGR) was generated by using an internal forward primer to eliminate the first 5' 417 bp in the coding sequence of contig 17852. A V139M substitution was also introduced to initiate protein translation at this point, resulting in a 1275 bp construct (**methods 3.4.2.3**). Following cloning and

sequence verification, the constructs were transferred to pEAQ-*HT*-DEST1 vectors and transformed into individual strains of *A. tumefaciens* strain LBA4404.

Next, transient expression of these MVA pathway genes with *AsbAS1* was performed in *N. benthamiana.* This was achieved by co-infiltration of two strains of *A. tumefaciens,* each carrying a single expression construct and mixed in equal ratios. A total inoculum density of *A. tumefaciens* OD_{600} of 0.4 was used giving a final OD_{600} of 0.2 per strain (**methods 3.4.5.2**). The combinations used were *tHMGR/AsbAS1, FPS/AsbAS1, SQS/AsbAS1* and *SQE/AsbAS1.* The *GFP*-carrying strain was used as a negative control (*GFP/GFP*) and also in place of MVA pathway gene expression strains for the AsbAS1 control (*GFP/AsbAS1*). An additional *AsbAS1/AsbAS1* control was also included for comparison. As before, 5-week-old *N. benthamiana* plants were infiltrated with the various strains, using 3 separate plants per condition as biological replicates. Following harvesting, three disks were cut from each leaf and analysed separately, to give a total of 9 samples per condition tested. Extracts were made from the leaf disks and subjected to GC-MS analysis using an external standard to quantify β-amyrin (**methods 3.4.6**).

Analysis of the control samples showed that β-amyrin was undetectable in the *GFP/GFP*-expressing plants, while the production of β-amyrin in *GFP/AsbAS1* plants was estimated at 1.1 mg/g dw (**Figure 3-4A**). No significant difference was found between the *AsbAS1/AsbAS1*-treated leaves compared to the *GFP/AsbAS1* control. Similarly, analysis of leaf extracts from plants infiltrated with *FPS/AsbAS1* and *SQE/AsbAS1* combinations showed no significant difference in β-amyrin levels compared to the *GFP/AsbAS1* leaves. In contrast, the *SQS/AsbAS1* combination resulted in accumulation of 2.6 mg/g dw β-amyrin, representing a 2.4-fold increase relative to the *GFP/AsbAS1* treatment. Larger increases were seen for the *tHMGR*/*AsbAS1* treatment, which yielded 4.5mg/g dw. This is an increase in excess of 4-fold compared to the *GFP/AsbAS1* control. These results suggest that the availability of the precursor supply is likely to be limiting triterpene production in *N. benthamiana.* In accordance with this, GC-MS analysis of *tHMGR/GFP* plants revealed the presence of large peaks corresponding to squalene and 2,3-oxidosqualene in these samples. These precursors were not detectable in extracts from the *GFP/GFP*-expressing controls (**Figure 3-4B**). Finally, a small increase in cycloartenol was also observed, indicating that expression of tHMGR also impacts sterol biosynthesis (**Figure 3-4B**).

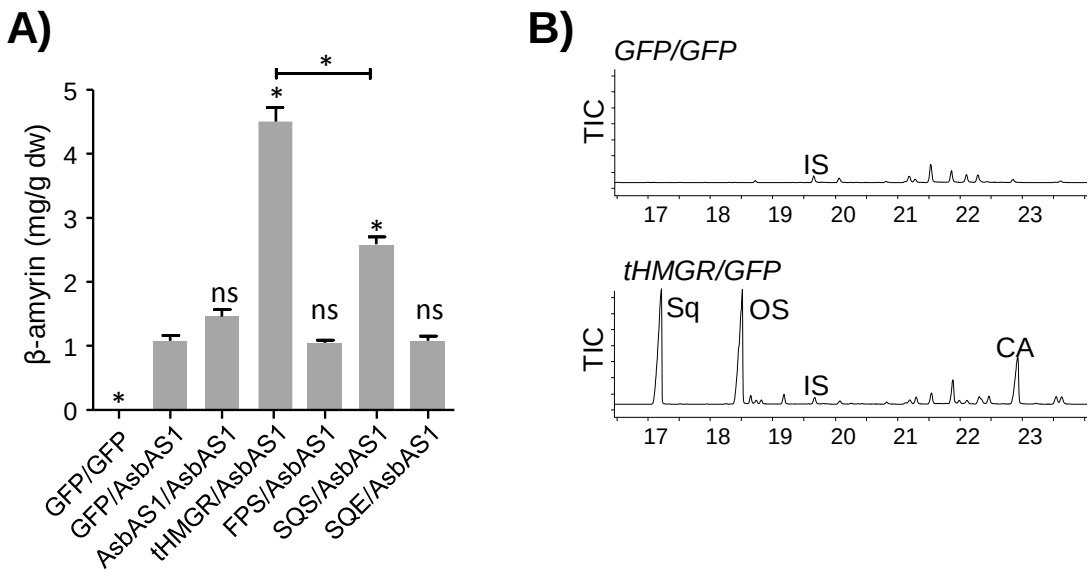


Figure 3-4: Manipulation of the MVA pathway in *N. benthamiana* **A)** Quantification of β-amyrin content in *N. benthamiana* leaves expressing *AsbAS1* with combinations of MVA genes. *GFP*-expressing leaves are also included as a negative control. Statistical significance against the *GFP/AsbAS1* control is shown (an additional comparison between *tHMGR/AsbAS1* and *SQS/AsbAS1* is also shown) where * = $p<0.05$ and ns = not significant (methods 3.4.6) **B)** GC-MS total ion chromatograms (TIC) for leaves expressing *tHMGR* in the absence of coexpressed AsbAS1. Leaves expressing *GFP* are shown as a control. Peak labels are squalene (Sq), 2,3-oxidosqualene (OS) and cycloartenol (CA). IS = internal standard (coprostanol).

A striking observation was that the infiltrated *N. benthamiana* leaves showed clear phenotypic effects in response to some of the treatments. While there were no obvious effects observed for the leaves infiltrated with the *GFP*-carrying strain, infiltration with the *AsbAS1*-expressing strain resulted in mild chlorosis (**Figure 3-5**). While *FPS/GFP* and *SQS/GFP*-treated plants showed little or no signs of chlorosis, the symptoms markedly increased in severity with the *SQS/AsbAS1* treatment, but not with *FPS/AsbAS1* (**Figure 3-5**). The elevated levels of β-amyrin in the *SQS/AsbAS1* samples could indicate that β-amyrin is responsible for the toxic effects. At the same time however, toxic effects were seen in all *SQE*- and *tHMGR*-expressing leaves, regardless of the presence of *AsbAS1* (**Figure 3-5**). Therefore the toxicity might be related to the protein itself, or (particularly with tHMGR) effects on other pathways, such as sterol biosynthesis.

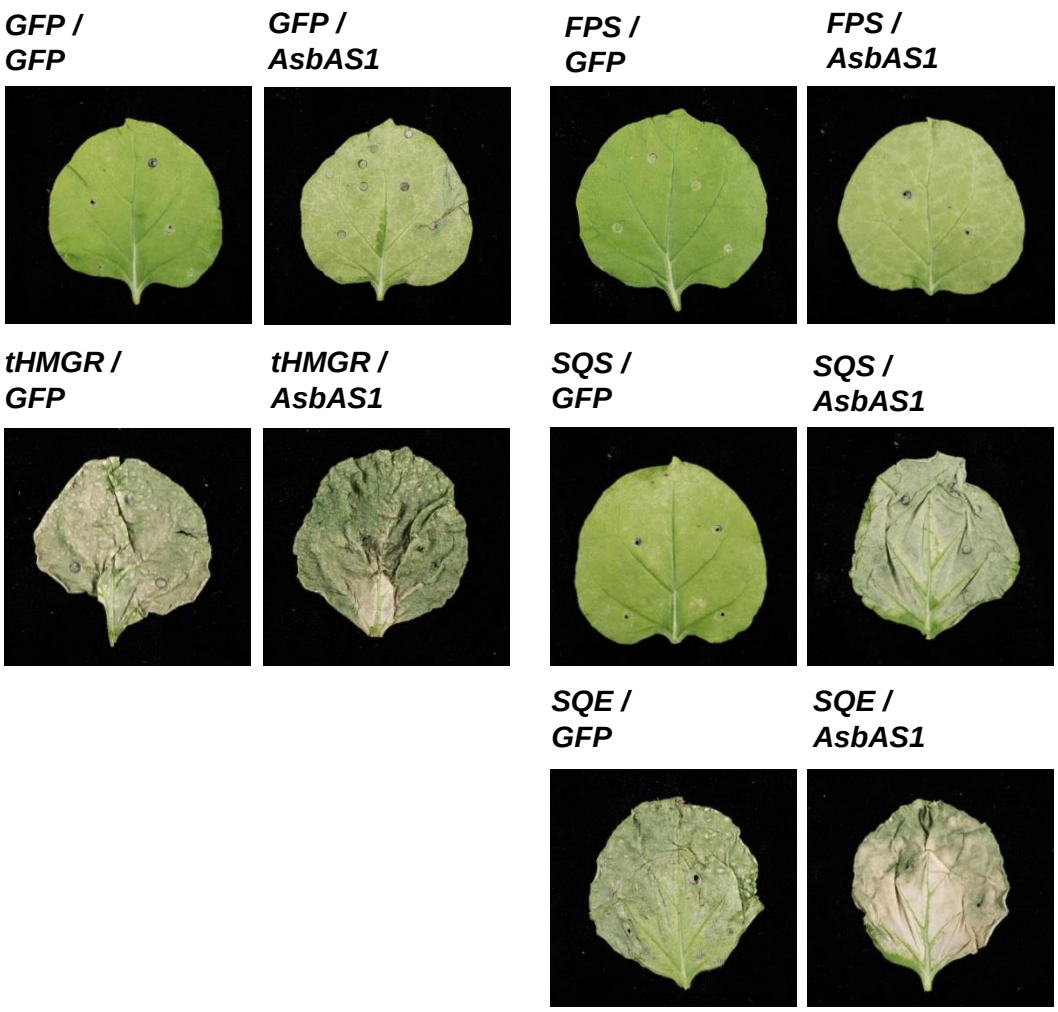


Figure 3-5: Phenotypes of leaves following agroinfiltration. The genes encoded by the various strains are given above the relevant photographs. Photographs were taken five days after infiltration.

3.2.3 – Combination of tHMGR and SQS further improves β-amyrin yields

Of the four enzymes tested, tHMGR gave the most substantial increase in β-amyrin. SQS gave a modest increase in β-amyrin while FPS and SQE did not improve yields. However it is possible that these latter genes could afford further yields if used in conjunction with tHMGR. To investigate this, a second set of infiltrations were performed. In these experiments, two strains of *A. tumefaciens*, one containing the *tHMGR* expression construct and the other the *AsbAS1* construct were co-infiltrated along with a third strain carrying expression constructs for either *GFP*, *FPS*, *SQS* or *SQE*. For coinfiltration of three strains, the total OD_{600} of the infiltrate was adjusted to 0.6 (**methods 3.4.5.3**).

GC-MS analysis of the leaf extracts revealed that leaves infiltrated with the strains carrying *tHMGR/FPS/AsbAS1* produced similar levels of β-amyrin to control

tHMGR/GFP/AsbAS1 leaves. Similar results were observed for the *tHMGR/SQE/AsbAS1* combination. In contrast, the *tHMGR/SQS/AsbAS1* treatment gave a modest (1.4-fold) but significant increase in β-amyrin relative to the controls (**Figure 3-6A**). Thus co-expression of SQS may be useful in further increasing flux towards triterpene biosynthesis.

3.2.4 – SQE allows production of alternative cyclisation products

The lack of any discernible impact on β-amyrin accumulation following infiltration with strains carrying *FPS* or *SQE* could imply that these steps are not rate-limiting for triterpene production in *N. benthamiana* leaves. An alternative possibility is that the proteins are simply non-functional in this system. However, surprisingly, examination of the total ion chromatogram from GC-MS analysis of *tHMGR/SQE/AsbAS1* leaves revealed a new peak at 26.1 min which was not seen in *tHMGR/GFP/AsbAS1* controls (**Figure 3-6B**). Given that the novel compound was specific to leaves infiltrated with the *SQE*-carrying strain; this compound could arise from the cyclisation of dioxidosqualene, rather than oxidosqualene.

Squalene has two chemically identical termini arising from head-to-head condensation of FPP. This allows for epoxidation by SQE at either end of the molecule to form 2,3-oxidosqualene (OS). Further oxidation by SQE results in 2,3(*S*):22(*S*),23-dioxidosqualene (DOS) (**Figure 3-6C**). Accordingly, *N. benthamiana* leaves infiltrated with the *tHMGR*-carrying strain accumulate abundant amounts of both squalene and OS (**Figure 3-7**). In contrast, leaves coinfiltrated with both *tHMGR*- and *SQE*-carrying strains resulted in total consumption of squalene and appearance of a large peak that is likely to be DOS (**Figure 3-7**). Cyclisation of DOS by *AsbAS1* results in the product 20*S*,24*S*-epoxydammarane (epDM) [137] (**Figure 3-6C**). Analysis of the novel product that accumulated in *N. benthamiana* revealed that it had the same retention time and mass spectrum as that of an epDM standard (**Figure 3-6B**). Collectively, these results indicate that the *A. strigosa* SQE is indeed functional, if not rate-limiting for β-amyrin biosynthesis.

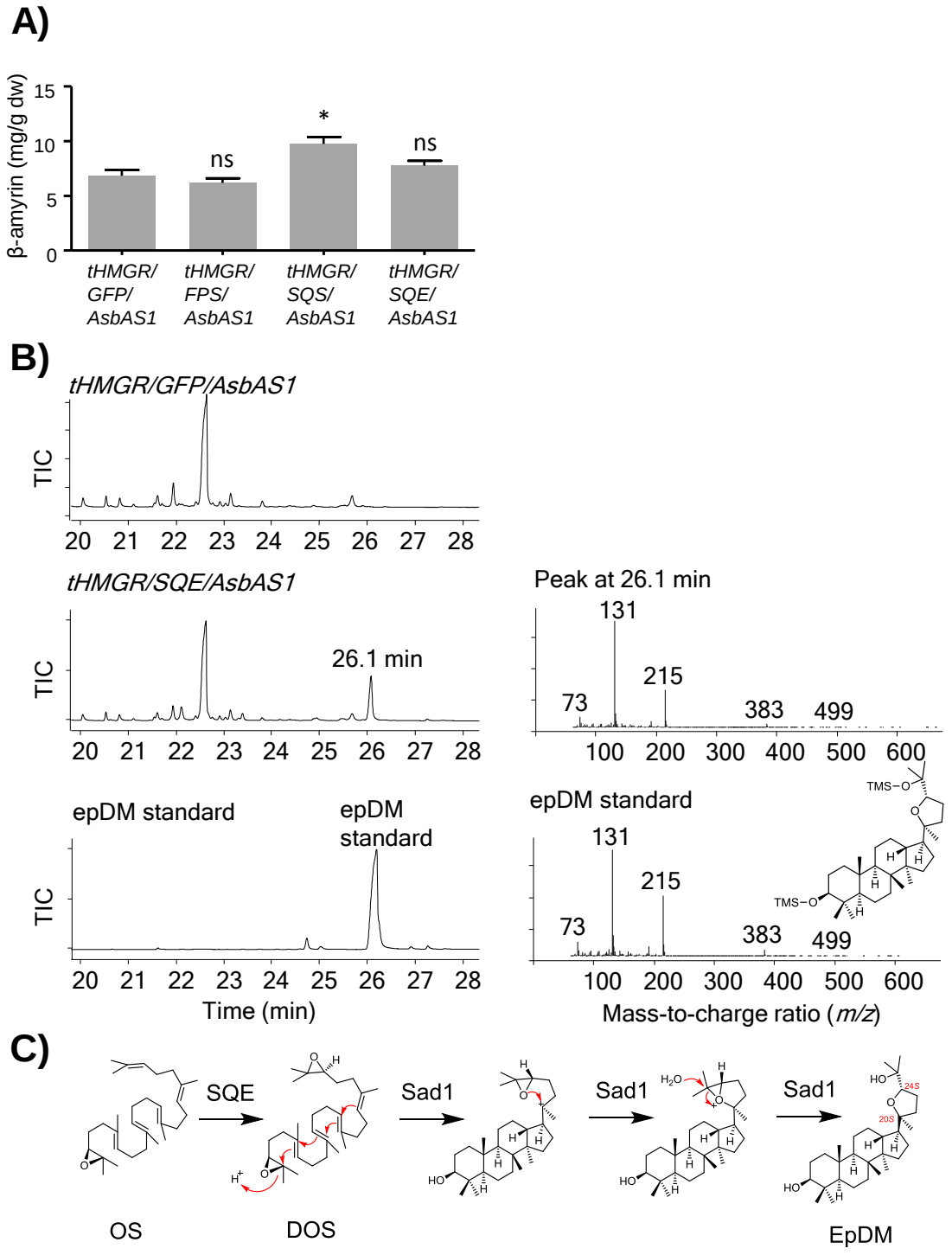


Figure 3-6: Effects of expressing multiple MVA pathway genes in *N. benthamiana.* Leaves were infiltrated with strains carrying *tHMGR* plus *AsbAS1* and either GFP, FPS, SQS or SQE. **A)** GC-MS quantification of leaf estimated β-amyrin content represented as mg/g dry weight. Statistical significance against the *tHMGR/GFP/AsbAS1* control is indicated where * = $p<0.05$ and ns = not significant (methods 3.4.6). **B)** GC-MS total ion chromatograms (TIC) of leaves expressing *tHMGR/GFP/AsbAS1* (top) or *tHMGR/SQE/AsbAS1* (middle). The *tHMGR/SQE/AsbAS1* plants feature a novel product at 26.1 min (the mass spectrum is shown to the right). The new product has an identical retention time and mass spectrum to epoxydammarane (epDM) **C)** Proposed mechanism of dioxidosqualene cyclisation by AsbAS1 to form epDM as described in Shan et al [163].

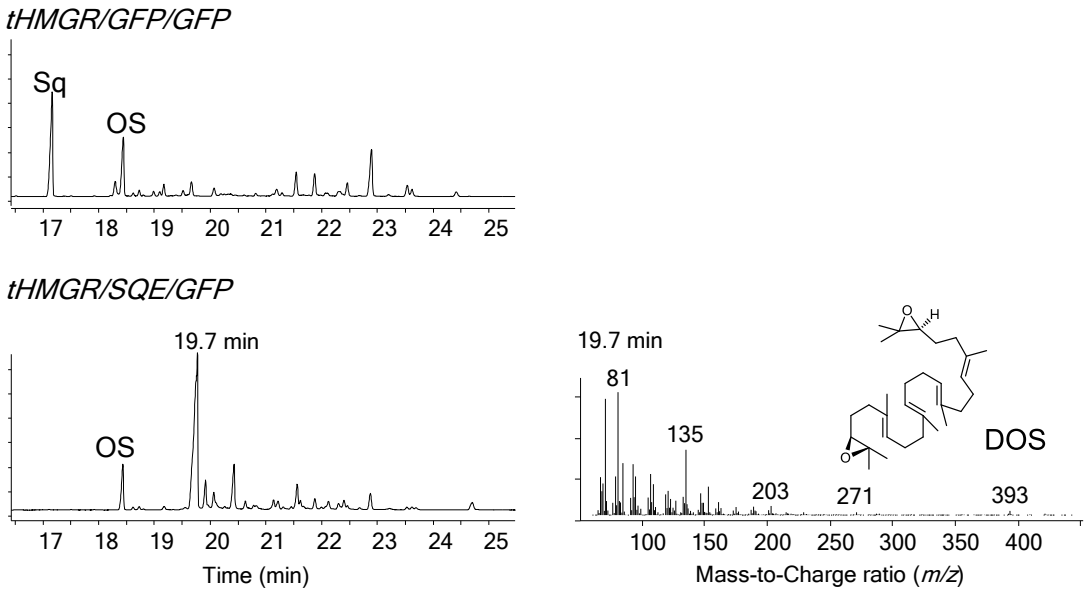


Figure 3-7: SQE allows for select production of dioxidosqualene (DOS) in *N. benthamiana.* GC-MS total ion chromatograms (TIC) of leaves expressing either *tHMGR/GFP/GFP* (top) *or tHMGR/SQE/GFP* (bottom). The large peak at 19.7min accumulating in *tHMGR/SQE/GFP* samples has a mass spectrum consistent with DOS (right) [164].

3.2.5 – Oxidation of the β-amyrin scaffold in *N. benthamiana*.

The results presented in **section 3.2.2** demonstrated that the use of tHMGR substantially increases β-amyrin levels in transient expression in *N. benthamiana*. Previously, it has been shown that coexpression of AsbAS1 with the *A. strigosa* cytochrome P450 CYP51H10 in *N. benthamiana* results in accumulation of the oxidised β-amyrin product 12,13β-epoxy,16β-hydroxy-β-amyrin (EpHβA) [85]. Therefore, an investigation was carried out to determine if increased yields of the oxidised product could be obtained upon coinfiltration of strains carrying *AsbAS1* and *CYP51H10* with *tHMGR*.

The full length sequence of *CYP51H10* was amplified by PCR from a plasmid template, cloned into the pEAQ-*HT*-DEST1 vector and transformed into *A. tumefaciens* LBA4404 (**methods 3.4.2.1**). Strains of *A. tumefaciens* carrying the pEAQ-*HT*-DEST1 expression constructs for GFP, tHMGR, AsbAS1 and CYP51H10 were cultured before combining in equal ratios to give suspensions of *tHMGR/AsbAS1/CYP51H10* and *GFP/AsbAS1/CYP51H10,* each at a total density (OD_{600}) of 0.6 (0.2 per strain) (**methods 3.4.5.4**). These were infiltrated into *N. benthamiana* leaves and harvested after 5 days. GC-MS analysis of leaf extracts

confirmed that levels of the CYP51H10 product EpHβA were also significantly increased upon coinfiltration of the strain carrying *tHMGR* (**Figure 3-8**).

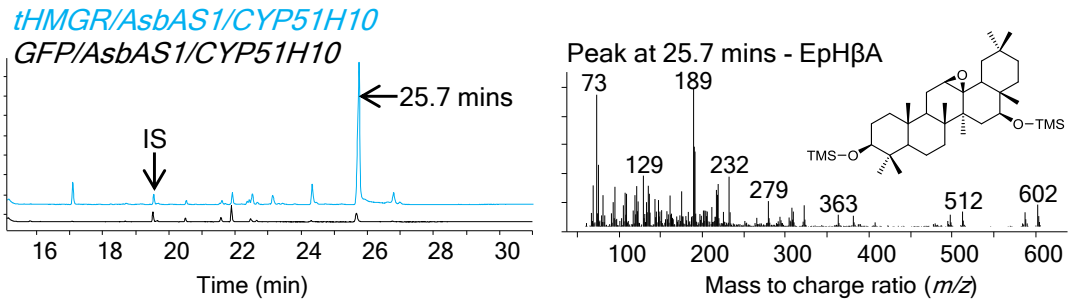


Figure 3-8: tHMGR enhances production of 12,13β-epoxy, 16β-hydroxy-β-amyrin (EpHβA). GC-MS total ion chromatograms (TIC) of *N. benthamiana* leaves expressing either *GFP/AsbAS1/CYP51H10* (black) or *GFP/AsbAS1/CYP51H10* (blue). Chromatograms are normalised to the internal standard (IS - coprostanol). A mass spectrum for the product at 25.7min is shown on the right and is consistent with a silylated form of EpHβA[85].

Previously, a yield of 1.17 mg EpHβA per gram dry leaf material was reported upon expression of AsbAS1 and CYP51H10 [85]. To investigate how use of tHMGR translated into isolable yields of EpHβA, a large-scale infiltration of approximately 70 plants with *A. tumefaciens* strains carrying *tHMGR*, *AsbAS1* and *CYP51H10* was performed. After five days the leaves were harvested (110g fresh weight) and lyophilised to yield 20g dry leaf material. After saponification and metabolite extraction, flash chromatography was used to isolate the CYP51H10 product from the bulk of the leaf extract. Further recrystallization to enhance the purity of the product yielded 79 mg of EpHβA crystals, equating to ~4mg/g dw (**methods 3.4.8**). This is around 3.4-fold higher than previously reported [85]. It should be noted that differences in constructs and extraction/purification techniques might be responsible for some of the differences between the two yields. Nevertheless, the difference in yields is in a similar range to the observed difference in β-amyrin content between *GFP/AsbAS1* and *tHMGR/AsbAS1* plants as predicted by GC-MS (**Figure 3-4A**). Together, these results demonstrate the potential of using tHMGR in transient expression experiments for enhancing yields of both simple and oxidised triterpenes in *N. benthamiana*.

3.2.6 – Exploring alternative subcellular compartmentalisation for triterpene production.

In recent years, a number of studies have highlighted how yields of monoterpene and sesquiterpenes can differ dramatically depending on the subcellular compartment in which they are synthesised [120, 127, 133]. The MEP and MVA pathway genes show large tissue- and development-specific variations in expression in *A. thaliana* [11]. In particular, in photosynthetic leaf tissue, it has been shown that components of the plastidial MEP pathway are highly expressed [11]. Consequently, it is possible that higher levels of triterpene production in *N. benthamiana* leaves might be attainable by redirecting the triterpene biosynthetic machinery to the chloroplast. High level production of the triterpene precursor squalene has been successfully produced in chloroplasts of transgenic tobacco, but it is unknown whether this strategy is applicable for production of cyclised triterpenes [159].

To investigate this, biosynthesis of squalene in *N. benthamiana* chloroplasts was first attempted. To this end, the *A. strigosa* FPS and SQS enzymes were modified by N-terminal fusion of a 58-residue chloroplast signal peptide (CP) from the *A. thaliana* RuBisCO small subunit 1A protein (**Figure 3-9A**). Furthermore, the SQS protein contains a small C-terminal hydrophobic anchor believed to aid localisation of this enzyme to the endoplasmic reticulum. As this is non-essential for activity, truncated forms of SQS (tSQS) were also generated for both chloroplast-targeted and cytosolic forms by removal of the final 21 residues (**methods 3.4.3**). This gave a total of six constructs, which were transferred to pEAQ-*HT*-DEST1 vectors. Six new strains of *A. tumefaciens* were generated by transformation with the two *FPS* (*FPS* and *CP-FPS*) and four *SQS* (*SQS, tSQS, CP-SQS* and *CP-tSQS*) pEAQ constructs (**methods 3.4.9**).

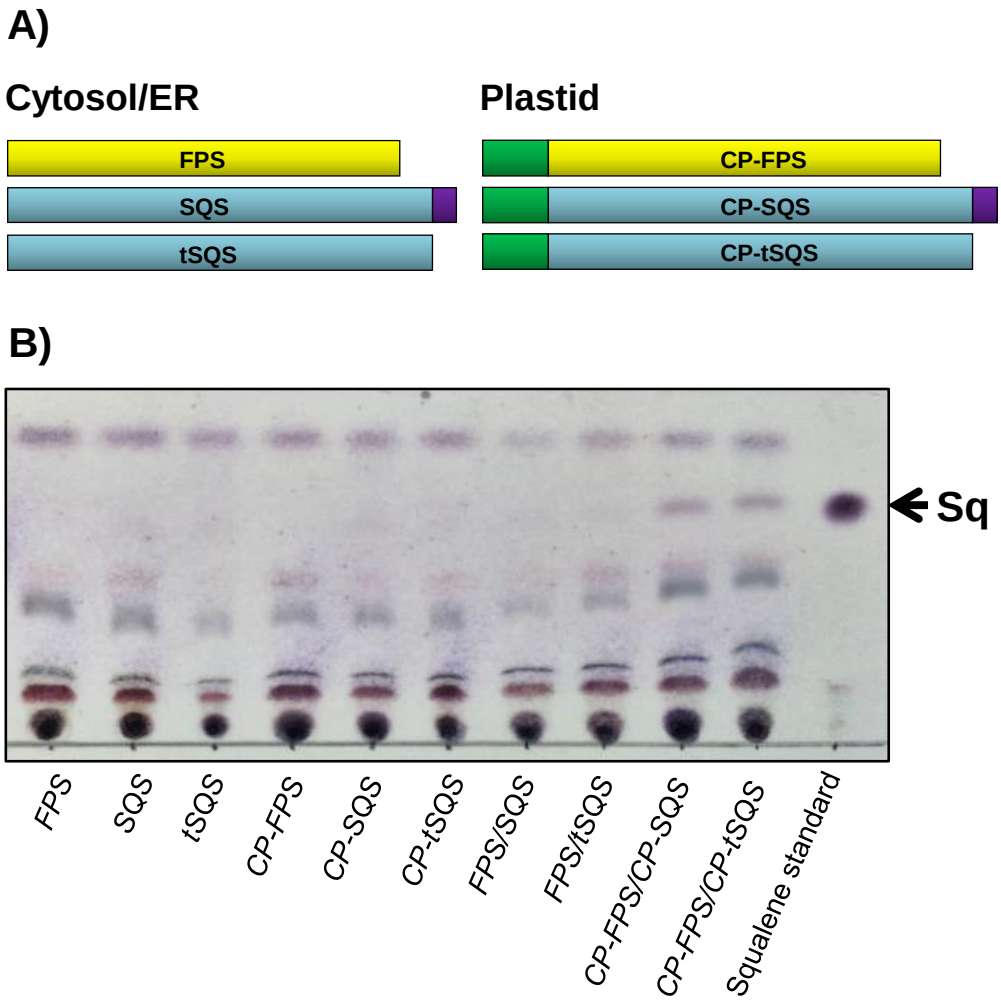


Figure 3-9: Chloroplast targeting of FPS and SQS in N. benthamiana. **A)** Schematic of cytosolic and chloroplast-targeted (CP) FPS and SQS constructs. The chloroplast-targeting signal peptide is indicated in green. Two SQS variants were made either with (SQS) or without (tSQS) the native ER-retention peptide (purple). **B)** TLC analysis of *N. benthamiana* leaf extracts after 5 days expressing various combinations of genes. A squalene standard is included in the right-most lane.

The six *A. tumefaciens* strains were infiltrated into *N. benthamiana* either individually or in pairs of cytosolic (*FPS*/*SQS* and *FPS/tSQS*) and plastid-targeted (*CP-FPS/CP-SQS* and *CP-FPS/CP-tSQS*) combinations. After five days, leaf extracts were screened by thin layer chromatography (TLC) (**methods 3.4.9.1**). TLC analysis revealed a distinct band co-migrating with a squalene standard in extracts of leaves expressing chloroplast-targeted FPS and SQS (*CP-FPS/CP-SQS* and *CP-FPS/CP-tSQS*). By contrast this was not seen in the extracts of leaves expressing pairs of cytosolic (*FPS/SQS* and *FPS/tSQS*), or any of the individually-expressed

genes (**Figure 3-9B**). This suggested that squalene was indeed more abundant only in the leaves expressing both chloroplast-targeted FPS and SQS.

To further confirm this result, additional tests were performed in triplicate and leaf extracts were analysed by GC-MS (**methods 3.4.9.2**). Large peaks with identical mass spectra and retention times to a squalene standard were present in the extracts from leaves expressing *CP-FPS/CP-SQS* or *CP-FPS/CP-tSQS* (**Figure 3-10A,** *CP-FPS/CP-SQS* not shown for clarity). In accordance with the results of the TLC, squalene was more abundant in *CP-FPS/CP-SQS* and *CP-FPS/CP-tSQS* than in the cytosolic *FPS/SQS* samples. Furthermore, squalene was approximately 2-fold more abundant in *CP-FPS/CP-tSQS*-expressing plants, compared to *CP-FPS/CP-SQS* plants. This suggests that removal of the ER-anchoring peptide from the chloroplast-targeted SQS permits better enzyme activity (**Figure 3-10B**).

Further examination of the chromatograms from leaves expressing the cytosolic *FPS/SQS* revealed prominent peaks corresponding to OS and cycloartenol (**Figure 3-10**). This implies that the squalene produced in the cytosol is further consumed by endogenous *N. benthamiana* SQE and cycloartenol synthase. By contrast, OS and cycloartenol were only minor constituents of the extracts from chloroplast-targeted *CP-FPS/CP-SQS* and *CP-FPS/CP-tSQS.* This suggests that the squalene in these samples is generally unavailable to the endogenous downstream enzymes, as might be expected from biosynthesis in the chloroplast.

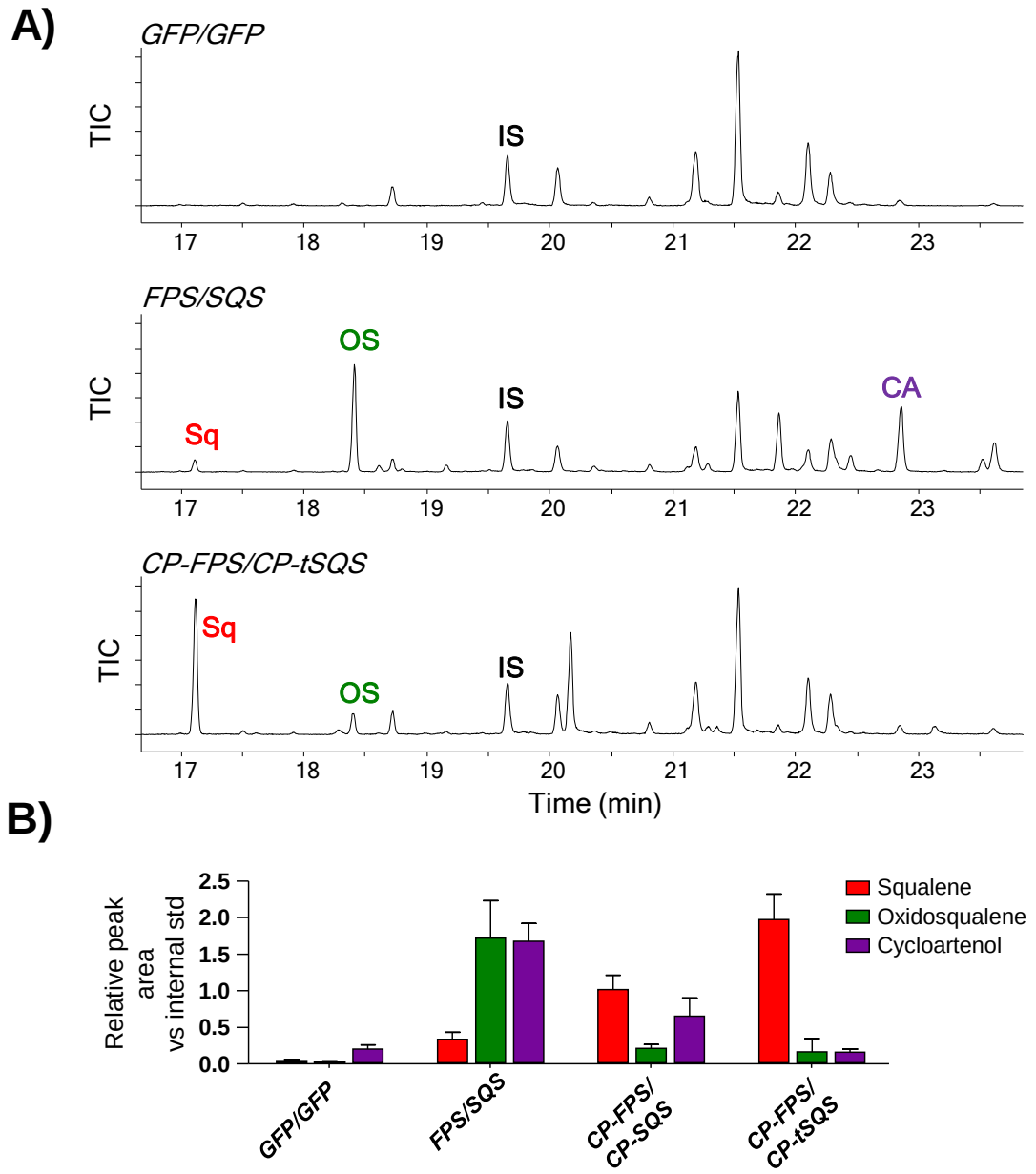


Figure 3-10: Chloroplast targeting of triterpene biosynthetic genes. A) Representative GC-MS chromatograms of *N. benthamiana* leaves expressing *GFP/GFP*, *FPS/SQS* or *CP-FPS/CP-tSQS*. Peaks labels are squalene (Sq), oxidosqualene (OS), Cycloartenol (CA). Chromatograms are normalised to the Internal standard (IS - coprostanol). B) Histogram of the relative amounts of squalene, oxidosqualene and cycloartenol produced in leaves expressing cytosolic or plastid-targeted FPS and SQS. Bars represent the average peak area of each compound divided by the peak area of the internal standard. Error bars show standard deviation.

3.2.7 – Plastid-synthesised squalene is unavailable for the cytosolic biosynthesis of β-amyrin.

As the results of the previous section implied that squalene was synthesised in the chloroplasts of *N. benthamiana*, an attempt was subsequently made to synthesize β-amyrin by targeting four enzymes (FPS, SQS, SQE and AsbAS1) to this compartment. However this approach was unsuccessful and did not result in β-amyrin production (data not shown).

Instead, a preliminary investigation was performed to look into the effect of synthesising FPP and squalene in the chloroplast with simultaneous expression of the native-localised (cytosolic-facing ER membrane) AsbAS1 protein. Here, different areas of a single leaf were infiltrated with combinations of *AsbAS1, AsbAS1* plus *CP-FPS* or *AsbAS1*, *CP-FPS* and *CP-tSQS*. A control *GFP*-expressing strain was used in place of *CP-FPS* and *CP-tSQS* as appropriate. This was repeated for three leaves. Disks were harvested from the relevant sections of leaves after 5 days and the subsequent extracts were analysed by GC-MS (**methods 3.4.9.2**).

GC-MS analysis of samples revealed that samples expressing *CP-FPS and AsbAS1* gave comparable levels of β-amyrin to an *AsbAS1*-only control. Upon further introduction of *CP-tSQS* however, a large decrease in β-amyrin was observed compared to an *AsbAS1*-only control. At the same time, a peak corresponding to squalene was also observed in these samples (**Figure 3-11**). This would seem to support the findings in the previous section (**3.2.6**) suggesting that squalene produced via this route is not available to enzymes outside of the chloroplast. Furthermore, the observed drop in β-amyrin content implies that the production of squalene in chloroplasts effectively sequesters a source of carbon which might otherwise be utilised for β-amyrin production. By extension, this suggests that the MEP pathway may contribute to the production of β-amyrin upon transient expression of *AsbAS1*.

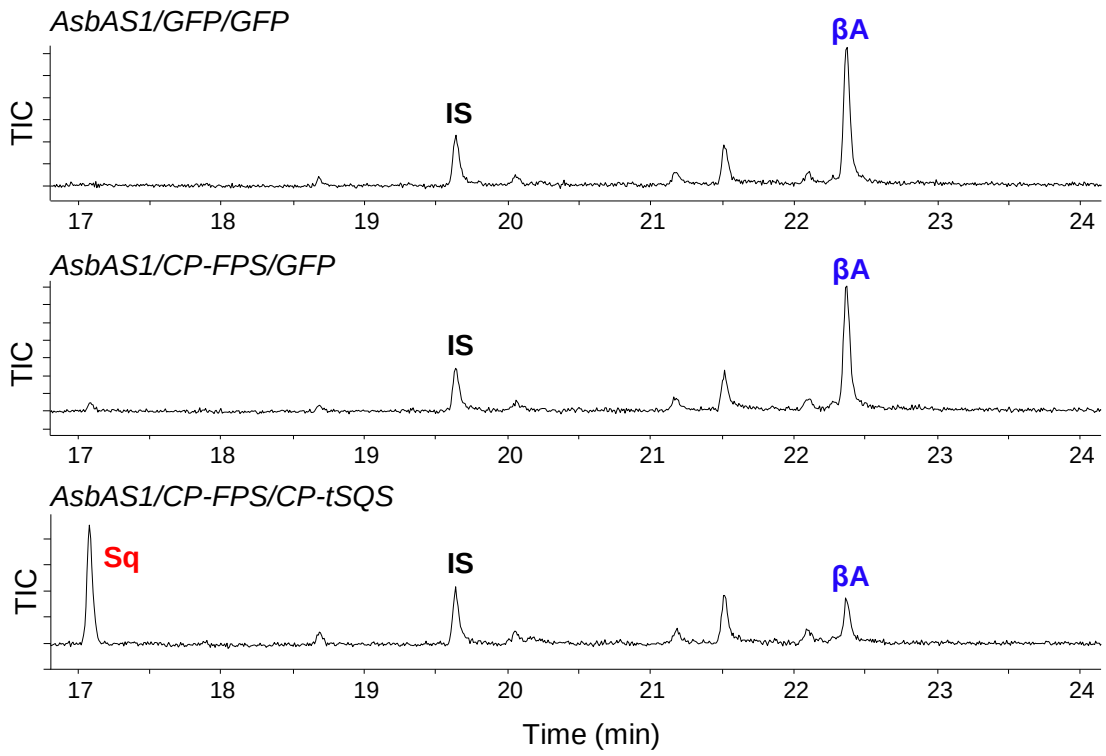


Figure 3-11: Impact of chloroplast production of squalene on cytosolic β-amyrin synthesis. Representative GC-MS total ion chromatograms (TIC) of *N. benthamiana* leaves expressing *AsbAS1/GFP/GFP* (top) *AsbAS1/CP-FPS /GFP* (middle) and *AsbAS1/CP-FPS/CP-tSQS* (bottom). Peak labels are Sq = Squalene, IS = Internal standard βA = β-amyrin.

It is widely accepted that a trafficking system exists for shuffling prenyl diphosphate precursors between various compartments of the cell [60]. Transport appears to predominantly involve IPP/DMAPP and (to a lesser extent) GPP and FPP, but not GGPP [165]. Supporting this, transient expression of a GPP synthase (GPPS) in *N. benthamiana* has been found to boost the yields of the monoterpene geraniol even when GPPS and geraniol synthase were targeted to different subcellular compartments [127]. In the present study, the simultaneous expression of *AsbAS1* and *CP-FPS* did not appear to either enhance or be detrimental to β-amyrin production. Given that CP-FPS was apparently necessary for efficient activity of CP-tSQS (section 3.2.6) it is expected that this enzyme is functional. Therefore it is possible that while FPP was produced in the chloroplast, it was capable of being transported to the cytosol while squalene was not. In future, microscopy could be employed to provide further evidence that the proteins used herein are indeed localised to the chloroplast as expected. Fusion of GFP to the RuBisCO targeting peptide would provide a simple means of achieving this as has been previously demonstrated [127].

Further investigation into strategies to manipulate triterpene biosynthesis in plant subcellular compartments is worth pursuing. The investigation here into squalene production in chloroplasts suggests that the MEP pathway may contribute to triterpene biosynthesis in *N. benthamiana* leaves. With this in mind it will be particularly interesting to see whether overexpression of the rate-limiting MEP pathway genes such as DXS could also further enhance triterpene production as has previously been shown [166]. Transient expression offers an excellent means to rapidly explore a variety of different approaches to metabolic engineering in plants. The outcome of these studies may ultimately have important applications for engineering stable transgenic plant lines.

3.3 – Conclusions and perspectives

3.3.1 – Discussion of results

The expression of transgenes following agroinfiltration is a transient process. Therefore maximising the metabolic output of the host plant during the post-infiltration period is vital to realise the capacity of *N. benthamiana* for triterpene production. Here two strategies were explored to this end. In the first instance, expression of rate-limiting enzymes in the native triterpene biosynthetic MVA pathway was undertaken. Secondly, redirecting triterpene biosynthetic enzymes to the plastid was explored.

Investigation of MVA pathway genes showed that of the four possible rate-limiting enzymes tested, tHMGR was most effective at increasing yields of β-amyrin in *N. benthamiana* when coexpressed with AsbAS1. Furthermore, coexpression of tHMGR was applicable for enhancing the isolation of an oxidised form of β-amyrin. The yield of EpHβA was increased by more than 3-fold compared to previous reports (section 3.2.5).

Of the other MVA pathway genes tested, SQS also gave a small increase in β-amyrin yields, while FPS and SQE by contrast did not give an appreciable increase in β-amyrin content. A similar scenario was observed upon combination of tHMGR and AsbAS1 with FPS, SQS or SQE, suggesting that coexpression of both tHMGR and SQS might be useful in future for further boosting product yields (section 3.2.3). It is unlikely that the failure of FPS and SQE to enhance β-amyrin content under any condition is simply due to inactivity of the enzymes, as subsequent tests showed that SQE expression allowed for the production of DOS (section 3.2.4) while FPS expression in the chloroplast was necessary for maximal SQS activity (section 3.2.6). Therefore the most likely explanation is that FPS and SQE activity are not limiting for triterpene biosynthesis under the conditions tested.

It will be worthwhile testing other forms of HMGR in *N. benthamiana* to see if differences in activity can be observed. Analysis of the *A. strigosa* 454 database revealed an additional full-length HMGR isoform which was not tested here. Transient expression of HMGR from other species has also been successfully used for sesquiterpene production in *N. benthamiana.* Coexpression of an *A. thaliana* tHMGR with a (mitochondrially-targeted) amorphadiene synthase (ADS) resulted in more than 300-fold increases over ADS alone [132]. This would seem to be a dramatic contrast to the 4-fold increase reported here, however at the same time the

maximum reported yields of amorphadiene were below the yields of triterpenes in this study. For example, the capacity of *N. benthamiana* amorphadiene production was estimated by GC-MS at 27mg per kg fresh weight [132]. The isolated yield of EpHβA in this study (79mg per 110g fresh weight), equates to 718mg per kg fresh weight. It is worth reiterating that this latter value represents an isolated yield of purified product rather than a predicted value. Inevitably the purification process results in product losses during chromatography and recrystallization as performed here. As a result, these comparisons are most likely a conservative estimate of the differences in yield.

Given their shared biosynthetic pathway (**Figure 3-2**), why the capacity in *N. benthamiana* for triterpene biosynthesis might seemingly be higher than for sesquiterpenes is an interesting question. It has previously been suggested that the endogenous *N. benthamiana* SQS may outcompete the sesquiterpene synthases for FPP. Accordingly RNAi silencing of SQS was shown to be an effective strategy for enhancing sesquiterpene yields [131]. In the present study, expression of tHMGR resulted in readily detectable quantities of squalene and 2,3-oxidosqualene without expression of AsbAS1. Although small increases could be observed in endogenous sterols, the remaining abundance of these immediate β-amyrin precursors would indicate that there is relatively little competition for AsbAS1 with the endogenous *N. benthamiana* OSCs. Hence, *N. benthamiana* may prove to be a highly favourable system for production of triterpenes, although further tests with a range of triterpene synthases will be required to see if this generally applicable.

3.3.2 – Investigating other candidate enzymes

Although the current study focussed on four genes for enhancing triterpene yields in *N. benthamiana*, continued investigation of other potential yield-enhancing enzymes is highly worthwhile. For example it was recently shown that isopentenyl phosphate kinase (IPK), an enzyme forming a core part of the archaeal (but not eukaryotic) MVA pathway remains active in plants, where it serves to regulate availability of IPP from a pool of isopentenyl phosphate (IP) [167]. Likewise, other genes within the MVA pathway such as HMG-CoA synthase (HMGS) have been implicated as potential targets for metabolic engineering [168, 169]. Finally, as discussed in **section 3.2.7**, it is also worth further investigating the impact of manipulating key steps in the MEP pathway.

3.3.3 – Further optimising production

The present study took a metabolic engineering approach to enhance triterpene content in *N. benthamiana*. However in conjunction with optimisation at the molecular level, it is important not to overlook environmental factors. For example, the plants used in these experiments were grown under greenhouse conditions, which results in significant seasonal variations in the morphology and growth rate of the plants. Although this has not been formally investigated, this variability might be expected to have a significant impact on the production capacity. As a result, optimisation of lighting, temperature and humidity conditions would be merited to both further improve yields and ensure batch-to-batch consistency.

In the present study, tens of milligrams of oxidised triterpene were produced by hand using a syringe to infiltrate 70 plants over multiple weeks. In order to translate this to larger preparative scales, it is important to develop practical techniques to increase the infiltration throughput. Recently, a vacuum chamber was developed in the Osbourn group for scaling production of triterpenes in *N. benthamiana* (Dr Bastiaan Brouwer, personal communication). This process involves inverting the entire aerial parts of the plants into a solution of *A. tumefaciens* and can accommodate four plants simultaneously per run. One run takes approximately five minutes and thus allows in excess of 100 plants to be infiltrated in a single afternoon. Use of this vacuum infiltrator coupled with the tHMGR enzyme described here has allowed for the production of a gram of β-amyrin from 459 plants (Dr M. Stephenson, personal communication). β-Amyrin has a value of £24.50 per milligram (based on Sigma-Aldrich 2016 prices). This offers the possibility for isolating useful quantities of costly metabolites for downstream applications such as synthetic chemical diversification and bioactivity screens. These possibilities are explored in subsequent chapters.

3.4 – Materials and methods for Chapter 3

3.4.1- Searching the oat 454 database for MVA pathway genes

The Ensembl Plants (http://plants.ensembl.org/) Biomart database was used to obtain amino acid sequences from known *A. thaliana HMGR*, *FPS*, *SQS* and *SQE* genes from gene models given by Vranova et al [11]. These sequences were used as queries to search the *A. strigosa* root tip 454 transcriptome database using the tBLASTn search tool with a cutoff value of e^{-10}.

3.4.2 – Cloning and generation of expression constructs

3.4.2.1 – Cloning of *AsbAS1*, *CYP51H10* and *GFP*

AsbAS1 and *CYP51H10* were amplified by PCR from existing vectors as described in Geisler et al [85]. Cloning of *CYP51H10* was performed by Prof Robert Minto using primers AsbAS1-attB1-F/2-R and CYP51H10-attB1-F/2-R (**Table 3-2**). Gateway cloning was performed as described in **section 2.1.1**. The GFP-expressing pEAQ-*HT*-DEST1 construct was kindly provided by the Lomonossoff group at JIC. Transformation of *A. tumefaciens* LBA4404 with the pEAQ-*HT*-DEST1 constructs is described in **section 2.1.2.3**.

3.4.2.2- Cloning oat *FPS*/*SQS*/*SQE*

The oat *FPS*, *SQS* and *SQE* genes were cloned from oat root tip cDNA (cDNA provided by Dr Thomas Louveau). Primers were designed to incorporate a small 5' overhang to allow optional restriction enzyme-based cloning (which was not used in this study). The primers for *FPS*, (FPS-SacI-F/FPS-NotI-R), *SQS* (SQS-HindIII-F/SQS-XbaI-R) and *SQE* (SQE-SacI-F/SQE-XbaI-R) are given in (**Table** 3-2). *FPS* and *SQS* were amplified by using Q5 polymerase (NEB) and *SQE* was amplified with Phusion polymerase (NEB). PCR products were purified by gel extraction using a QIAquick gel extraction kit (Qiagen). Purified products were ligated into a TOPO vector using a Zero Blunt TOPO PCR cloning kit (Invitrogen) according to the manufacturer's instructions. Following ligation, the vector was used to transform chemically competent *E. coli* TOP-10 cells (*FPS* and *SQS*) or subcloning efficiency DH5α cells (*SQE*) with Kanamycin selection (50µg/mL). Constructs were sequenced with M13 primers (**Table 2-1)** to verify the correct insert sequence.

To further generate Gateway Entry clones of these genes, *FPS*, *SQS*, and *SQE* were amplified by PCR from the relevant TOPO vectors as described above. Primers were designed to carry 5' full-length attB sites adjacent to the gene-specific sequence. Primer sequences for *FPS* (FPS-attB1-F/2-R), *SQS* (SQS-attB1-F/2-R)

and *SQE* (SQE-attB1-F/2-R) are given in **Table 3-2**. A truncated form of *SQS*, (*tSQS*) was amplified using SQS-attB1-F and tSQS-attB2-R. Purified PCR products were then transferred into pDONR207, sequenced with attL1-F/attB2-R (**Table 2-1)** and further transferred into pEAQ-*HT*-DEST1 (protocol as described in **section 2.1.1**).

3.4.2.3 - Cloning oat *tHMGR*

Cloning the truncated form of HMGR (contig17852), was carried out by Dr Aymeric Leveau. Oat root cDNA was used as a template for PCR, *tHMGR* was amplified using t-AsHMGR-17852-F-GW and AsHMGR-17852-R-GW (**Table 3-2**). The forward primer contained an additional 5' CACC overhang to allow directional cloning into the Gateway-compatible Entry vector pENTR-D-TOPO (Invitrogen). PCR products were purified and ligated into pENTR-D-TOPO according to the manufacturer's instructions. *E. coli* TOP-10 cells were transformed and selected for on kanamycin (50µg/mL) media. Positive colonies were sequenced to verify the insert.

The pENTR-D-TOPO and pEAQ-*HT*-DEST1 vector both utilise kanamycin selection, preventing efficient selection of positive colonies. Therefore the *HMGR*-containing pENTR-D-TOPO vectors were first digested using *NruI* to cut the vector backbone in the NPT-II gene. These were then run on a 1% agarose gel to separate the linearised and undigested plasmid. The products were excised from the gel and purified with a QIAquick gel extraction kit (Qiagen). The purified gel products were cloned into pEAQ-*HT*-DEST1 and transformed into *A. tumefaciens* LBA4404.

3.4.3 – Generating constructs for chloroplast-targeting work

3.4.3.1 – *CP-FPS*-pDONR207

To generate a chloroplast-targeted form of FPS, the *FPS* gene was fused to the first 174bp of the *A. thaliana RuBisCO* small subunit 1A gene (*AT1G67090*) by a fusion PCR-based method. The 174bp chloroplast-targeting sequence was first amplified from *Arabidopis thaliana* leaf cDNA (Columbia ecotype – kindly supplied by Dr Hans Nützmann) using RuB-attB1 and RuB-FPS-R (**Table 3-2**). This gave a product containing a 5' attB1 sequence for Gateway cloning and 3' 18nt which were complementary to the first 18nt of the oat *FPS* gene. Similarly the oat *FPS* was amplified from the *FPS*-TOPO construct (**Section 3.4.2.2**) with FPS-RuB-F and FPS-attB2 primers (**Table 3-2**). This resulted in a product featuring 5' 18nt complementary to the 3' end of the above 174bp chloroplast targeting sequence

and 3' attB2 site for Gateway-based cloning. Together the overlap between the two PCR products contained 36nt of complementary sequence.

Following amplification, the two products were analysed by gel electrophoresis before excision and purification (QIAquick gel extraction kit, Qiagen). The two products were pooled and 2ng of this mix was used as a template for a second PCR reaction with the RuB-attB1 and FPS-attB2 primers (**Table 3-2**). The resulting product was purified and inserted into the pDONR207 with a BP clonase reaction to generate *CP-FPS*-pDONR 207. *E. coli* TOP-10 cells were transformed with gentamicin selection. The resulting clones were sequenced (attL1-F/attB2-R **Table 2-1**) and found to contain the correct *CP-FPS* fusion product. These were then cloned into pEAQ-*HT*-DEST1 and transformed into *A. tumefaciens* LBA4404.

3.4.3.2 – *CP-SQS*- and *CP-tSQS-pDONR207*

For the different forms of *CP-SQS*, a similar strategy was used to *CP-FPS*. The RuBisCo chloroplast targeting sequence (with 3' SQS overhang) was amplified from *CP-FPS*-pDONR207 (**Section 3.4.3.1**) using RuB-attB1s-F and RuB-*SQS*-R (**Table** 3-2). The full-length *SQS* (with 5' RuBisCO overhang) was amplified from the *SQS*-TOPO plasmid (**Section 3.4.2.2**) using *SQS*-RuB-F and either *SQS*-attB2s-R (**Table 3-2**). The two PCR products were analysed by gel electrophoresis, before excision and purification (QIAquick gel extraction kit, Qiagen). The purified fragments were pooled and adjusted to 2ng. These were further used as a template for PCR to amplify *CP-SQS* or *CP-tSQS* using RuB-attB1 and SQS-attB2-R or tSQS-attB2-R, respectively (**Table 3-2**). The resulting products were purified and inserted into pDONR207 (section 2.1.1.1) and sequenced using attL1-F/attB2-R primers (**Table 2-1**). After sequencing, the constructs were cloned into pEAQ-*HT*-DEST1 (**section 2.1.1.2**) and used to transform *A. tumefaciens* LBA4404 (**section 2.1.2.3**).

3.4.4 – Primers

	Name	Adapter/5'sequence	Gene-specific sequence
TOPO cloning	FPS-SacI-F	GAGCTC	ATGGCGGCGGCAGCGGTGAAC
	FPS-NotI-R	GCGGCCGC	CTACTTCTGCCTCTTGTAGATC
	SQS-HindIII-F	AAGCTT	ATGGGGGCGCTGTCG
	SQS-XbaI-R	TCTAGA	TCACTTGGCGTACAGTACACC
	SQE-SacI-F	GAGCTC	ATGGCTGAGGTCGCCGCC
	SQE-XbaI-R	TCTAGA	TTAGAACTCCGCTTCCGGAGG
p-ENTR-D-TOPO cloning	t-AsHMGR-17852-F-GW	CACC	ATGGCGCCCGAGA-AAATGCCCGAG
	AsHMGR-17852-R-GW		TCMGCAGGCGATC-TTGGACATGTCC
Gateway cloning	FPS-attB1-F	**attB1F-1**	ATGGCGGCGGCAGCGGTG
	FPS-attB2-R	**attB2R-1**	CTACTTCTGCCTCTTGTAG
	SQS-attB1-F	**attB1F-1**	ATGGGGGCGCTGTCGCGG
	SQS-attB2-R	**attB2R-1**	TCACTTGGCGTACAGTACACC
	SQE-attB1-F	**attB1F-1**	ATGGCTGAGGTCGCCGCC-
	SQE-attB2-R	**attB2R-1**	TTAGAACTCCGCTTCCGGAGG
	tSQS-attB2-R	**attB2R-1**	TCACCTATACTTTGACTTTTC
	AsbAS1-attB1-F	**attB1F-1**	ATGTGGAGGCTAACAATAGGTG
	AsbAS1-attB2-R	**attB2R-1**	TCAGCTCTTAATCGCAAGAAGTCG
	CYP51H10-attB1-F	**attB1F-2**	ATGGACATGACAATTTGCGTCGT
	CYP51H10-attB2-R	**attB2R-2**	TTAGTTTGCAGGCATACGACATC
Chloroplast targeting	RuB-attB1	**attB1F-1**	ATGGCTTCCTCTATGCTC
	RuB-FPS-R	CACCGCTGCCGCCGCCAT	CACCTGCATGCAGTTAACTC
	FPS-RuB-F	GTTAACTGCATGCAGGTG	ATGGCGGCGGCAGCGGTGAAC
	FPS-attB2	**attB2R-1**	CTACTTCTGCCTCTTGTAG-
	RuB-attB1s-F	**attB1Fs-1**	ATGGCTTCCTCTATGCTCTC
	RuB-SQS-R	CGACAGCGCCCCCAT	CACCTGCATGCAGTTAAC
	SQS-RuB-F	GTTAACTGCATGCAGGTG	ATGGGGGCGCTGTCG
	SQS-attB2s-R	**attB2Rs-1**	TCACTTGGCGTACAGTACACC
Sequencing	AsbAS1-seq-F		CCACATTTTCTTCCGATTCA
	AsbAS1-seq-R		CCAACAAGGTCAGCGGATA

Table 3-2: Primer sequences as used in Chapter 3. Gene-specific sequences are given in the column on the right, while adapter sequences incorporated into the 5' end of the primer are given in the middle column. Adapter sequences given in bold refer to the sequences in **Table 2-1**.

3.4.5 – Infiltrations

3.4.5.1 – Infiltration and analysis of *AsbAS1* in *N. benthamiana*

Strains of *A. tumefaciens* carrying *AsbAS1* or *GFP* were cultured and prepared as detailed in **section 2.1.3.2.** An inoculum density of OD_{600} = 0.2 was used to infiltrate leaves of 5-week-old *N. benthamiana* (**section 2.1.3.3**). Leaves were harvested after 5 days before being freeze dried. Leaf material was weighed and saponification solution containing internal standard was added to give 0.4µg coprostanol per mg dry leaf. Saponification was performed at 65°C for 2 hours before partitioning with hexane. The hexane extract was analysed directly by GC-MS using the program detailed in **section 2.1.5.1.**

3.4.5.2 – Coexpression of single MVA genes with *AsbAS1*

Infiltration suspensions of *A. tumefaciens* were made according to **section 2.1.3.2** using an OD_{600} = 0.4. The two upper-most fully-expanded leaves of five-week old *N. benthamiana* plants were infiltrated as described in **section 2.1.3.3**. Three plants were infiltrated for each test condition (biological replicates). After five days, leaves were harvested. For the upper-most leaf from each plant, three 12mm diameter disks were cut using a #7 cork borer and stored as separate technical replicates, giving a total of nine samples per test condition. Leaf disks were stored in 1.5mL microcentrifuge tubes and flash-frozen in liquid N_2 before storage at -80°C. The lower leaves from representative plants were used for photography (Andrew Davis, JIC). Extraction and analysis was performed as detailed in **section 3.4.6.**

3.4.5.3 – Coexpression of *tHMGR* and *AsbAS1* with other MVA genes

Infiltration of leaves with *tHMGR* and *AsbAS1* and either *FPS, SQS, SQE* or *GFP* was performed as in **section 3.4.5.2** with the exception that the OD_{600} = 0.6. Extraction and analysis was performed as detailed in **section 3.4.6.** The chromatogram in **Figure 3-6B** is from analysis of the same samples used for quantification in **Figure 3-6A**, but with the exception that GC-MS program was performed according to **2.1.5.1.**

3.4.5.4 – Coexpression of *tHMGR/AsbAS1/CYP51H10*

Infiltration of leaves with *tHMGR, AsbAS1* and *CYP51H10* was performed as in **section 3.4.5.2** with the exception that the OD_{600} = 0.6. Extraction was performed as in **section 3.4.7** with derivatisation. GC-MS analysis was performed as in **section 2.1.5.1.**

3.4.6 – Extraction and quantification of β-amyrin in plant leaves

Leaf disks were removed from the -80°C freezer and directly lyophilised for 24 hours. Individual dried disks were weighed using a microbalance (Sartorius) to the nearest 10µg and weights recorded. The leaf disks were further transferred to 2mL autosampler vials before extraction and analysis as detailed below.

For quantification of β-amyrin in *N. benthamiana* leaf disks, a ten-step calibration series of β-amyrin was made from 50µg down to 0.098µg (these values represent dry weight of β-amyrin per vial). This was achieved by making a 0.5mg/mL solution of β-amyrin in ethanol and performing 2-fold serial dilutions starting with 100µl (50µg). The β-amyrin dilution series was dried under N_2. Next a saponification solution (**section 2.1.6.3**) was made containing the internal standard (IS) coprostanol (Sigma) to 10µg/mL. 500µl of this was pipetted to each sample (leaf disks and dried β-amyrin calibration standards) to give a total of 5µg of IS per sample. Liquid handling steps for β-amyrin calibration and pipetting of saponification solution were performed using a Gerstel MultiPurpose Sampler (MPS) to minimise pipetting errors.

Samples were heated in an oven at 65°C for 2 hours to saponify. Following this 250µl of distilled water was added per sample and mixed briefly followed by addition of 500µl of hexane. Samples were each vortexed vigorously twice for 5-10 seconds and allowed to stand for 5 minutes to partition. Finally, 100µl of hexane from each sample was taken and transferred to a new autosampler vial. These were dried using a Genevac EZ-2 evaporator. Finally the dried samples were resuspended in 50µl of Tri-Sil Z (Sigma) and heated at 70°C to derivitise. Samples were analysed by GC-MS according to **section 2.1.5.2**.

Data analysis was performed using the Agilent Masshunter Quantification software. Quantification of β-amyrin was performed using the ion at *m/z* 218 while quantification of the internal standard was performed using the ion at *m/z* 370. The external standard was used to determine the relative response factor of β-amyrin and quantify β-amyrin in the leaf samples. Finally estimates of dry leaf β-amyrin content were determined by dividing the quantification values by the leaf disk weight. Statistical tests were performed using GraphPad Prism software (version 5). A one way ANOVA was performed with post-hoc analysis performed by Tukey's multiple comparison test.

3.4.7 – Dispersive liquid-liquid microextraction

For extraction of sterols and triterpenes from *N. benthamiana*, a protocol was developed by Prof Robert Minto (Indiana University-Purdue University, Indianapolis, USA**)** using dispersive liquid-liquid microextraction (DLLME) ([170]). A saponification solution was prepared according to **section 2.1.6.3** and the internal standard (coprostanol 1mg/mL stock in DMSO) was added to 1.6µg per mL. Lyophilised leaf disks from *N. benthamiana* were weighed and recorded before adding the saponification mixture to 250µl/mg dry leaf disk weight (equivalent 0.4µg coprostanol per mg dry leaf). Samples were then heated at 65°C for two hours with intermittent vortexing. After incubation, 125µl of the saponified extract was transferred to a second tube containing 900µl 60mM HCl, plus a small amount of bromophenol blue as a pH indicator. A colour change from orange to blue was observed upon addition of the saponified extract. The samples were vortexed briefly, before adding 400µl of a 4:46 Chloroform:Acetonitrile solution. Samples were vortexed again for approximately five seconds before being centrifuged at full speed for 30 seconds to partition a lower yellow organic phase and an upper blue aqueous phase. The lower phase (approximately 60µl) was carefully extracted using a 200µl pipette. This could either be analysed directly by GC-MS, or first derivitised using 50µL Tri Sil-Z (Sigma) and heated 70°C for 30mins.

3.4.8 – Scaled production and purification of EpHβA

Three separate cultures of *A. tumefaciens* strain LBA4404 carrying *tHMGR*, *AsbAS1* and *CYP51H10* were grown as per the protocol in **section 2.1.3.2**. These were resuspended in MMA to an OD_{600} = 0.6 and combined in equal ratios. The bacterial suspension was used to infiltrate approximately 70 *N. benthamiana* plants by hand using a 5mL syringe. After 5 days the infiltrated leaves were harvested (109.6 fresh weight) and lyophilised to yield 20.4g dry weight. Extraction of the EpHβA product was performed as in **section 2.1.4**. The crude extract was dissolved in toluene prior to loading on a SNAP KP-Sil 25g silica gel column (Biotage). A flash chromatography gradient was setup as in program P-1 (**Table 2-4**) collecting 9mL fractions. Fractions were analysed by TLC and GC-MS and the fractions containing the most concentrated EpHβA product were combined and dried by rotary evaporation. The solid was dissolved in dichloromethane and adsorbed onto silica gel by rotary evaporation. A second flash chromatography step was performed as per program P-2 (**Table 2-4**) using a SNAP Ultra 10g column (Biotage). The fractions were again monitored by TLC and GC-MS and the purest EpHβA-containing fractions were combined and dried to give a yellow solid weighing

approximately 100mg. This crude solid was decolourised and recrystallized according to **sections 2.1.4.2** and **2.1.4.4** to give a total of 79mg of the EpHβA product.

3.4.9 – Expression and analysis of chloroplast-targeted *FPS* and *SQS*

N. benthamiana leaves were infiltrated with cytosolic (*FPS,SQS,tSQS*) and chloroplast-targeted (*CP-FPS,CP-SQS,CP-tSQS*) constructs as described in the main text. Infiltrations were performed according to **section 2.1.3.3** using a single leaf per combination (for TLC analysis) or in triplicate (for GC-MS analysis). After 5 days the leaves were harvested, ground and saponification solution was added (**section 2.1.6.3**) before heating at 65°C for 2 hours. Hexane was used to extract the metabolites for analysis by TLC (**section 3.4.9.1**).

3.4.9.1 – TLC analysis of leaves expressing chloroplast-targeted constructs

TLC analysis was performed using 10cm x 5cm aluminium-backed silica gel plates (Sigma). For analysis of squalene production in chloroplasts, extracts of *N. benthamiana* leaves were run using a mobile phase of ethyl acetate:hexane 3:97. For development of TLC plates, a solution of p-anisaldehyde:sulphuric acid:acetic acid (1:1:48 v:v:v) was used. The plates were developed using a TLC heater plate (Cameg) at 130°C for approximately 30 seconds.

3.4.9.2 – GC-MS analysis of leaves expressing chloroplast-targeted constructs

Extraction of the leaves was performed according to **section 3.4.7**. The samples were not derivatised. GC-MS analysis was performed according to **section 2.1.5.1**. For quantification, data analysis was performed using Agilent MassHunter Quantitative Analysis software. The products were identified based on their retention time (rt) and qualifier ion as follows (squalene rt=17.07 min, m/z=69; oxidosqualene rt=18.37 min, m/z=69; coprostanol rt=19.63 min m/z=373; β-amyrin rt = 22.37 mz=218; cycloartenol rt=22.86 m/z=393). Total ion peak areas were obtained by extracting 1000 mass units either side of the qualifying ion. Peak areas of the compounds of interest were divided by the peak area of the internal standard.

Chapter 4:

Engineering triterpene diversity in *N. benthamiana*

Acknowledgements:

NMR experiments and interpretation were performed by Dr Michael Stephenson. IT-TOF experiments were performed by Dr Lionel Hill. All other work is the author's own. The following people are gratefully acknowledged for provision of various constructs used to create the triterpene toolkit: Prof Alain Goossens, Prof Robert Minto, Prof Kalliope Papadopoulou, Dr Tessa Moses, Dr Aymeric Leveau and Dr Ramesha Thimmappa.

4.1 – Introduction

Cytochrome P450s are important enzymes that expand the chemical diversity of natural products through stereo- and regio-specific oxidation [73]. A growing number of triterpene-oxidising P450s have been characterised in recent years [171] (**Table 1-1**). Expression of these P450s with triterpene synthases in heterologous systems offers an attractive means to produce and selectively oxidise triterpenes in a manner which is not currently possible by other methods. This method thus has important applications for exploring the structural features underlying triterpene bioactivity and also facilitates further modification by synthetic chemistry.

4.1.1 – Combinatorial biosynthesis of triterpenes in *N. benthamiana*

Transient expression of triterpene biosynthetic enzymes in *N. benthamiana* has the potential to enable production of both simple and oxidised triterpenes within a matter of days. In addition, this technique allows the coexpression of multiple genes through simultaneous infiltration of different *A. tumefaciens* strains, each containing a different expression construct. This approach can therefore be used to rapidly assemble different combinations of biosynthetic enzymes providing a powerful system for screening enzyme compatibility [112].

Amongst the triterpene oxidases described above, a large number are known to oxidise β-amyrin (**1**) [171] (**Table 1-1**). This triterpene is one of the most ubiquitous scaffolds found in plants [62] and is found at the heart of a number of medicinally and commercially important triterpene saponins and synthetic derivatives [39, 57]. Therefore combinatorial expression of β-amyrin oxidising P450s in *N. benthamiana* offers the potential to engineer novel derivatives of β-amyrin through combinatorial biosynthesis [101, 108].

4.1.2 – Analysis of triterpenes by mass spectrometry

Gas chromatography with mass spectrometry (GC-MS) is a powerful technique for analysis of simple triterpenes. This technique relies on separation of the molecules in the gas phase and offers excellent resolving power. Mass spectrometry is frequently performed using electron impact (EI) ionisation. The fragmentation spectra of β-amyrin (oleanane) triterpenoids by EI mass spectrometry are well documented [162, 172-174]. Briefly, the breakdown of β-amyrin predominantly proceeds via a retro-Diels-Alder (rDA) cleavage mechanism due to the presence of the C-12-13 double bond in the C-ring. This results in the production of a characteristic ion at *m/z* 218 which corresponds to the diene fragment (C*DE rings - where C* represents the partial C-ring) of the molecule (**Figure 4-1**). This

breakdown mechanism is generally conserved even with additional substituents within the C/D/E rings. Thus the shift of the *m/z* 218 ion and subsequent fragmentation can inform as to the nature of the modification. Similarly, the ABC* rings are also represented by a characteristic ion at *m/z* 279 (or *m/z* 207 if β-amyrin is not silylated). This fragment is derived from hydrogen transfer rather than rDA cleavage and is a minor constituent of the spectrum [162, 174]. Likewise, the presence of other substituents on this portion of the molecule is often indicated by the loss of *m/z* 279 [173]. Thus mass spectra can be regarded as unique 'fingerprints' enabling structural identification of novel oleanane compounds.

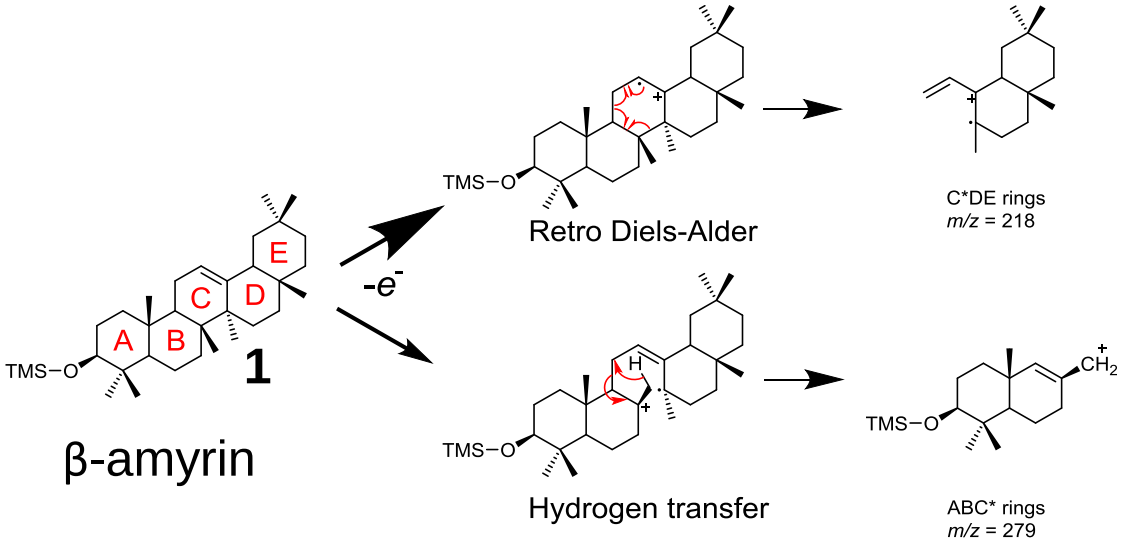


Figure 4-1: EI fragmentation of β-amyrin (**1**). The predominant mechanism of fragmentation proceeds via retro Diels-Alder cleavage (top). The radical cation from the C*DE ring fragment yields a characteristic ion at *m/z* 218. A less common ABC* fragment can also be generated from hydrogen transfer and gives an ion at *m/z* 279.

4.2 – Aims

The ability of P450s to selectively functionalise relatively inert C-C and C-H bonds makes them attractive tools for engineering natural product diversity. Significant progress has been made in recent years in identifying P450s participating in triterpene biosynthesis, providing an opportunity to produce and selectively functionalise triterpene scaffolds for various downstream applications. This chapter aimed to explore the capacity for engineering structurally diverse triterpenes in *N. benthamiana*. In particular, here a 'toolkit' of characterised β-amyrin oxidases from different plant species is established and evaluated for function in the *N. benthamiana* transient expression system. Systematic 'combinatorial' expression of both known and novel mutant oxidases is also employed for the generation of previously unreported novel β-amyrin derivatives. Finally, the potential of this system for production and oxidation of other triterpene scaffolds is demonstrated.

4.3 – Results and discussion.

4.3.1 – Part 1: Establishing a toolkit for functionalisation of β-amyrin

Chapter 3 demonstrated that enhanced production of β-amyrin in *N. benthamiana* could be achieved through coexpression of a truncated, feedback-insensitive form of 3-hydroxy-3-methylglutaryl-CoA reductase (*tHMGR*) with the oat β-amyrin synthase *AsbAS1*. In turn, increased yields of an oxidised form of β-amyrin (12,13β-epoxy,16β-hydroxy-β-amyrin (EpHβA)) were also achieved by coexpression of the oat cytochrome P450 CYP51H10 with tHMGR and AsbAS1 (**Figure 4-2**).

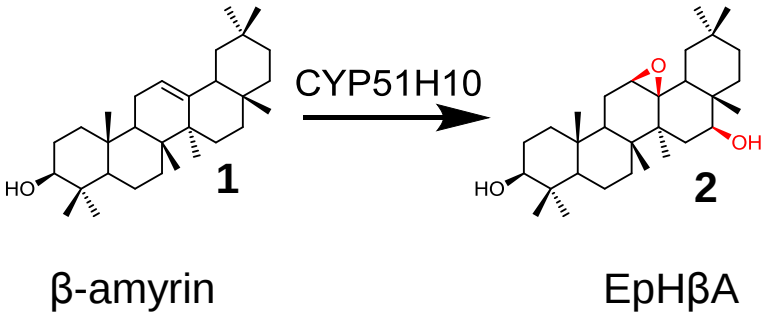


Figure 4-2: Oxidation of β-amyrin by CYP51H10 to form 12,13β-epoxy,16β-hydroxy-β-amyrin (EpHβA).

To investigate the potential of the system for production of other β-amyrin derivatives, coexpression of *AsbAS1* with other characterised triterpene-modifying P450s was carried out. A set of P450s that have previously been shown to oxidise the β-amyrin backbone directly were selected for functional analysis in the plant transient expression system. Using a Gateway-based cloning strategy (www.thermofisher.com), a series of Entry vectors were generated, each containing a P450-encoding gene (**Table 4-1**) (**methods 4.5.1**). These were each subsequently transferred to the pEAQ-*HT*-DEST1 binary vector and transformed into *A. tumefaciens* strain LBA4404 [161]. The individual P450 expression strains were then used to infiltrate *N. benthamiana* leaves along with strains carrying *tHMGR* and *AsbAS1.* After five days, leaves were harvested and GC-MS analysis of leaf extracts was performed (**methods 4.5.3**). Where authentic standards were available for the expected compounds, these were included for comparison.

Species	Name	Position	Product	Accession	Ref
Avena strigosa	CYP51H10	C-12/13 C-16	12,13β-epoxy-16β-hydroxy-β-amyrin	ABG88961	[85]
Medicago truncatula	CYP716A12	C-28	Oleanolic acid	FN995113	[97, 175]
Glycine max	CYP93E1	C-24	24-hydroxy-β-amyrin	AF135485	[96]
Glycyrrhiza uralensis	CYP88D6	C-11	11-oxo-β-amyrin	AB433179	[98]
Medicago truncatula	CYP72A63	C-30	11-deoxo-glycyrrhetinic acid	AB558146	[100]
Glycyrrhiza uralensis	CYP72A154	C-30	30-hydroxy-β-amyrin	AB558153	[100]
Medicago truncatula	CYP72A65	C-30	30-hydroxy-β-amyrin	XM_ 003628019.2	This study*
Avena strigosa	CYP51H10-I471M	C-12/13	12,13β-epoxy-β-amyrin	N/A	This study**
Avena strigosa	CYP51H10-A354L	C-16	16β-hydroxy-β-amyrin	N/A	This study**

Table 4-1: β-amyrin oxidases used in this study. * CYP72A65 activity is known from personal communication with the Goossens lab, Ghent, Belgium. ** Osbourn lab, unpublished.

4.3.1.1 – Production of oleanolic acid by CYP716A12

CYP716A12 is a C-28 oxidase from *Medicago truncatula* that catalyses the three-step oxidation of β-amyrin (**1**) to oleanolic acid via alcohol (erythrodiol) and aldehyde (oleanolic aldehyde) intermediates [97, 176] (**Figure 4-3A**). When *CYP716A12* was coexpressed with *AsbAS1* and *tHMGR*, obvious peaks were evident at 24.3 min and 25.1 min that matched the retention time and mass spectra of the authentic erythrodiol (**3**) and oleanolic acid (**5**) standards, respectively (**Figure 4-3B**). Of the two new peaks, oleanolic acid was by far the most abundant compound, indicating efficient conversion to the end product. This was consistent with the substantial decrease in β-amyrin relative to the controls (**Figure 4-3B**).

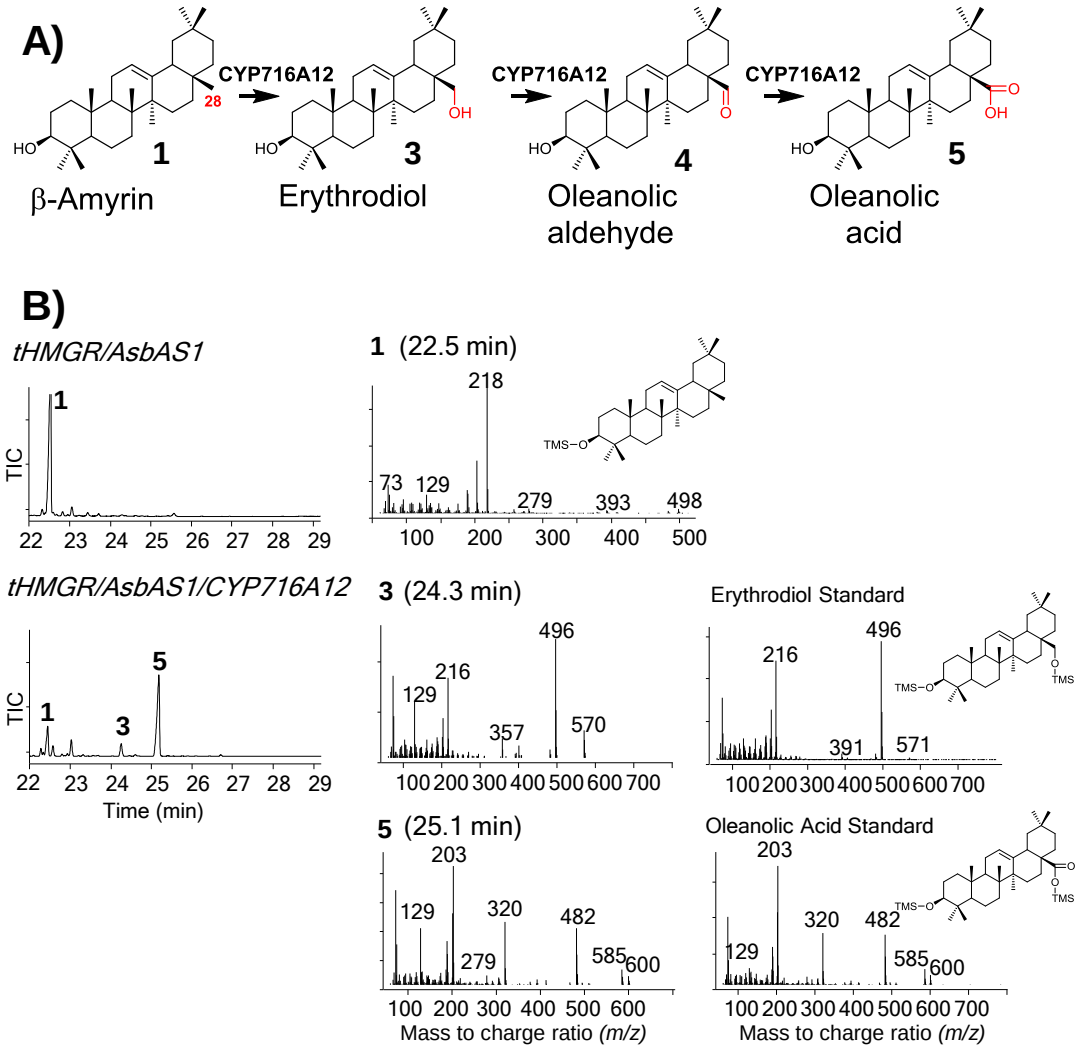


Figure 4-3: Oxidation of β-amyrin by CYP716A12. **A)** Schematic for stepwise C-28 oxidation of β-amyrin to oleanolic acid. **B)** GC-MS data from coexpression of *tHMGR/AsbAS1/CYP716A12* in *N. benthamiana*. Mass spectra for the numbered peaks are shown to the right of the relevant chromatograms. Mass spectra for erythrodiol and oleanolic acid standards are also given.

4.3.1.2 – Production of 24-hydroxy-β-amyrin by CYP93E1

CYP93E1 is a C-24 oxidase involved in soyasaponin biosynthesis in *Glycine max* [96]. Although predominantly a C-24 hydroxylase, it has also been shown to be capable of further oxidation at the C-24 position to the aldehyde and carboxylic acids [99] (**Figure 4-4A**). Coexpression of *tHMGR/AsbAS1/CYP93E1* resulted in a large peak at 24.0 min which had a GC-MS profile that matched a previously published spectrum for 24-hydroxy β-amyrin (**6**) [96, 99]. As observed with *CYP716A12*, the conversion of β-amyrin by CYP93E1 appeared to be highly efficient (**Figure 4-4B**). In addition to 24-hydroxy β-amyrin (**6**), numerous additional products could be detected. One such peak at 25.3 min was identified as the C-24 carboxylic acid (**8**) as previously reported [99]. The identity of the other peaks is not known. These may be extraction or GC-MS-related artefacts (**Supplementary Figure 3**). However the major product by far was 24-hydroxy-β-amyrin, as expected (**6**).

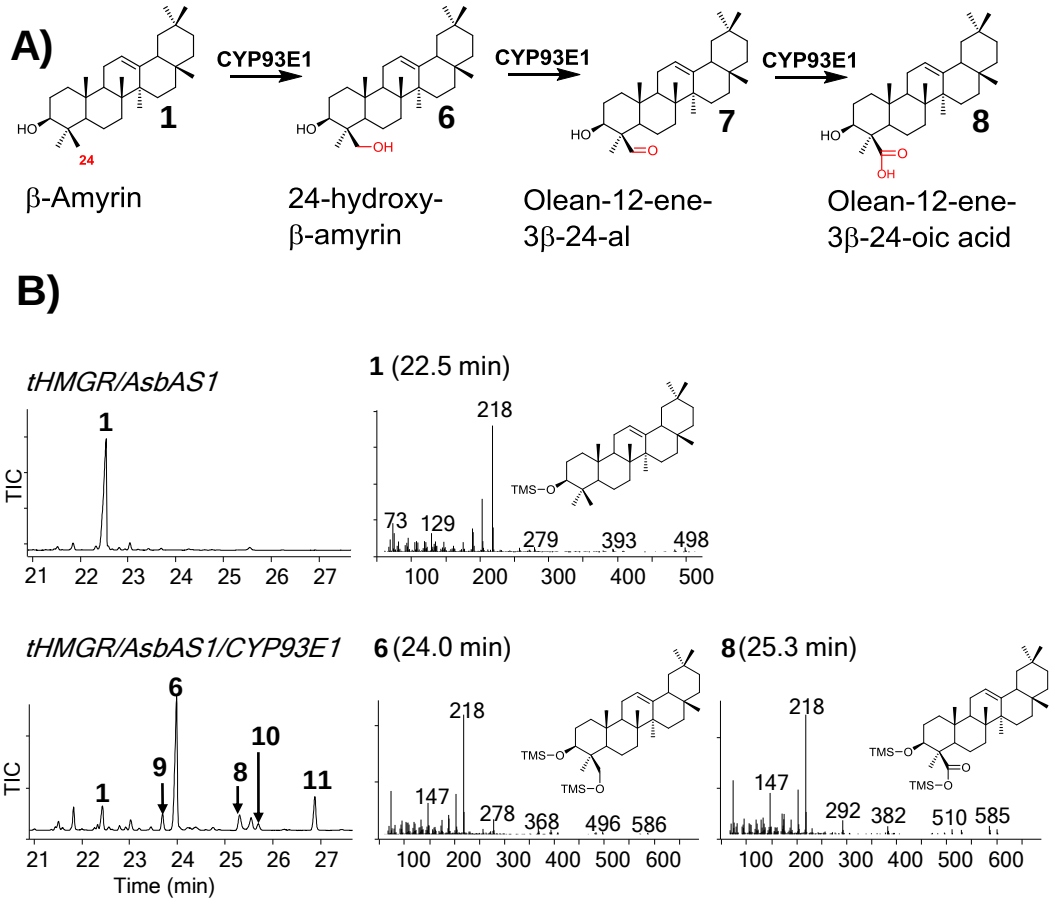


Figure 4-4: Oxidation of β-amyrin by CYP93E1. **A)** Schematic for the step C-24 oxidation of β-amyrin by CYP93E1. **B)** GC-MS data of extracts of *N. benthamiana* leaves expressing *tHMGR/AsbAS1/CYP93E1* plus controls. Mass spectra for numbered peaks are shown to the right of the chromatograms. Mass spectra for peaks 9-11 not shown above are given in **Supplementary Figure 3**.

4.3.1.3 – Production of 11-oxo-β-amyrin by CYP88D6

CYP88D6 is involved in glycyrrhizin biosynthesis in *Glycyrrhiza uralensis* and is responsible for the two-step C-11 oxidation of β-amyrin to 11-oxo-β-amyrin (**13**) via 11α-hydroxy-β-amyrin (**12**) (**Figure 4-5A**) [98]. GC-MS analysis of *tHMGR/AsbAS1/CYP88D6*-expressing leaves revealed a large peak at 25.6 min with a mass spectrum consistent with that of 11-oxo-β-amyrin [98] (**13**) (**Figure 4-5B**). A second peak at 22.4 mins (**15**) is expected to be a derivative of 11-oxo-β-amyrin containing an additional trimethylsilyl (TMS) group due to tautomerisation of the C-11 carbonyl (**13**) (**Supplementary Figure 4A**). Re-analysis of the sample without derivatisation yielded a single major product with the expected mass for underivitised 11-oxo-β-amyrin (440) (**Supplementary Figure 4B**). This confirms that the previous product **15** is likely to be an artefact of the derivatisation process. The identities of the other peaks, including **14** and **16,** are not known (**Supplementary Figure 4A**).

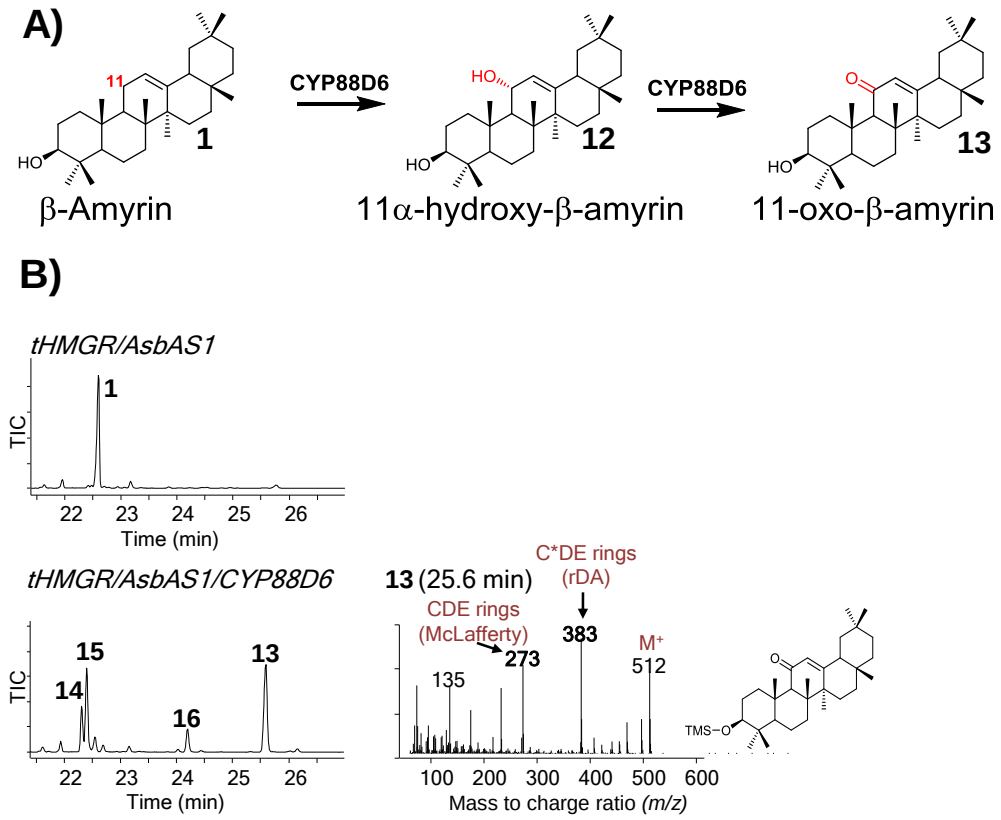


Figure 4-5: Oxidation of β-amyrin by CYP88D6. **A)** Schematic for the 2-step C-11 oxidation of β-amyrin to 11-oxo-β-amyrin. **B)** GC-MS total ion chromatograms (TIC) from leaves expressing *tHMGR/AsbAS1/CYP88D6* plus controls. Mass spectra for 11-oxo-β-amyrin (**13**) is shown to the right. 11-oxo-β-amyrin and derivatives undergo McLafferty rearrangement as well as retro Diels-Alder (rDA) cleavage to give the fragments indicated. A schematic for this is shown in **Supplementary Figure 5**. Spectra for peaks 14-16 not shown here are given in **Supplementary Figure 4A.**

4.3.1.4 – Production of 30-hydroxy-β-amyrin by CYP72A63 and CYP72A65

CYP72A63 and CYP72A154 are C-30 oxidases capable of acting on β-amyrin directly, as well as on numerous oxidised forms of this scaffold [100, 101] (**Figure 4-6A**). Analysis of *N. benthamiana* leaves expressing *tHMGR/AsbAS1/CYP72A63* revealed a new peak present at 25.6 mins corresponding to 30-hydroxy-β-amyrin (**17**) [100]. In contrast to other P450s tested so far, CYP72A63 showed relatively poor substrate conversion, with most of the β-amyrin remaining unmodified (**Figure 4-6B**).

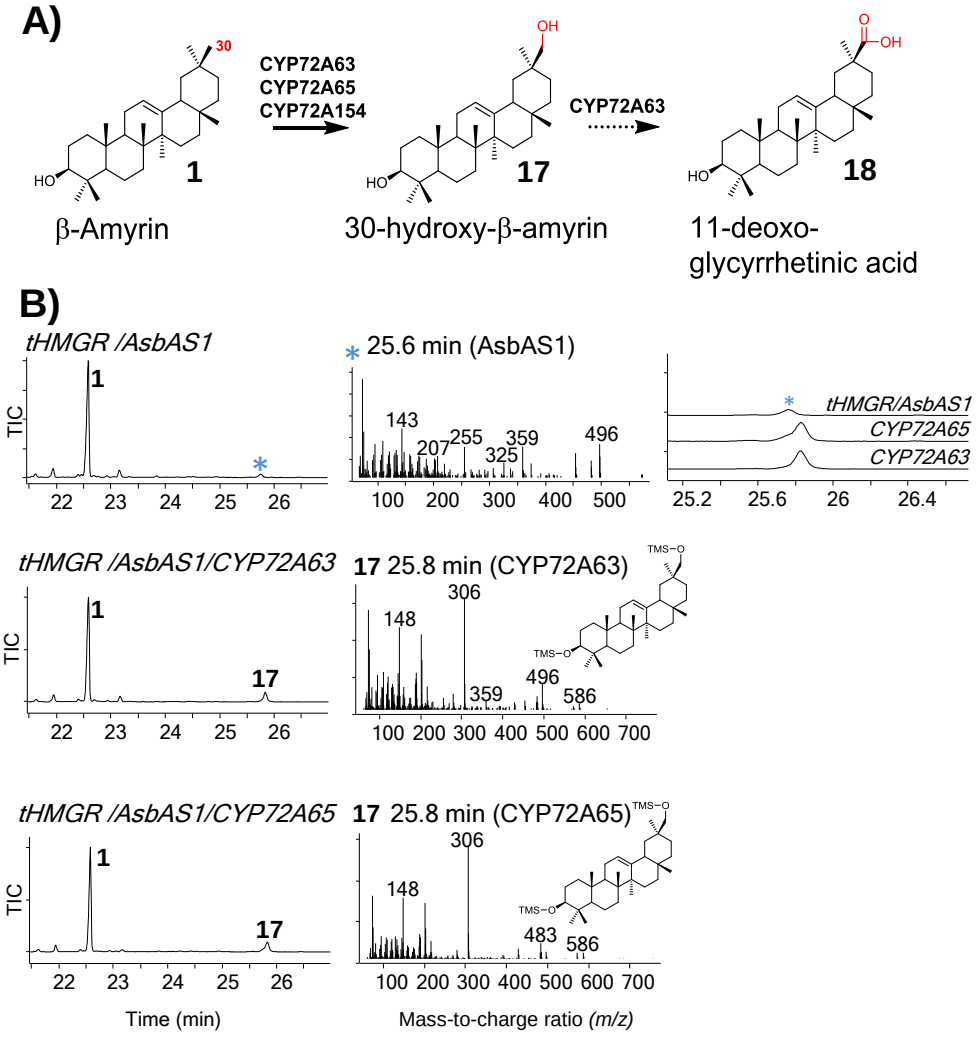


Figure 4-6: Oxidation of β-amyrin by CYP72A63, CYP72A65 and CYP72A154. **A)** Schematic for the step oxidation of β-amyrin at C-30. CYP72A65 and CYP72A154 are only known to perform the first hydroxylation step. The dashed arrow indicates an undefined oxidation step. **B)** GC-MS total ion chromatograms (TIC) for plants expressing *tHMGR/AsbAS1/CYP72A63* and *tHMGR/AsbAS1/CYP72A65* (normalised to internal standard, coprostanol, not visible above). 30-hydroxy-β-amyrin has a close retention time to another product in the AsbAS1-only control sample marked with an asterisk (shown in close detail on the upper right). Mass spectra for these products are shown to the right.

A second C-30 oxidase - CYP72A154 - was also tested but did not show any detectable activity towards β-amyrin (not shown). When previously tested in yeast, this enzyme was suggested to prefer 11-oxo-β-amyrin as a substrate (**13**) [100]. However the combination of this enzyme with CYP88D6 (*tHMGR/AsbAS1/ CYP88D6/CYP72A154*) likewise failed to reveal any obvious conversion of 11-oxo-β-amyrin (data not shown). Thus CYP72A154 may be inactive or poorly expressed in *N. benthamiana*. Finally, the test was repeated with a closely related P450 from *M. truncatula* which is 95% similar at the protein level and has also been found to have C-30 oxidase activity in yeast (Accession XM_003628019.2) herein called CYP72A65 (Prof A. Goossens, Dr. T. Moses, personal communication). This had comparable activity to CYP72A63, resulting in production of a small amount of 30-hydroxy-β-amyrin (**Figure 4-6B**).

4.3.1.5 – Part 1 Summary

These results demonstrate that selective oxidation of β-amyrin at the C-11, C-24, C-28 and C-30 positions is possible by coexpressing a series of triterpene-oxidising P450s with *AsbAS1* in *N. benthamiana*. Expression of *CYP716A12, CYP88D6* or *CYP93E1* (with *tHMGR* and *AsbAS1*) resulted in conversion of the majority of the β-amyrin to oxidised products. Thus, as demonstrated previously for CYP51H10 (Section 3.2.5), substrate availability appears to be the main factor limiting product yields. In contrast, CYP72A63 and CYP72A65 only converted a relatively small proportion of the available β-amyrin (**Figure 4-6**). In these cases it may be that β-amyrin is not the optimal substrate, or that these enzymes are poorly expressed or inactive in *N. benthamiana*. **Supplementary Table 2** shows the various β-amyrin-derived compounds produced in section 4.3.1 and key ions within their mass spectra with the corresponding predicted fragment.

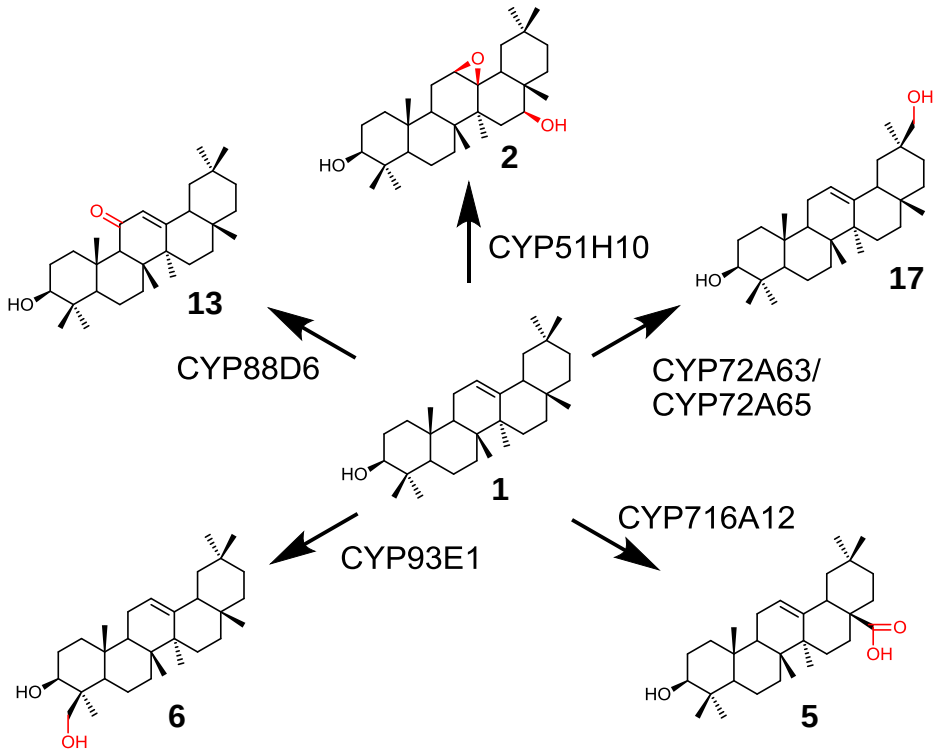


Figure 4-7: Summary of the major products produced by the β-amyrin oxidases expressed in the present study.

4.3.2 – Part 2 – Combinatorial biosynthesis in *N. benthamiana*

The P450s evaluated in the present study collectively oxidise six different positions on the β-amyrin backbone (**Figure 4-7**). This opens up opportunities to investigate the compatibility of these enzymes when expressed in combination with the aim of producing 'hybrid' oxidised forms of β-amyrin. To explore this possibility, pairwise combinations of the P450s were tested by transient coexpression in *N. benthamiana.* All combinations included *tHMGR* and *AsbAS1.* After five days the leaves were harvested and extracts were analysed by GC-MS. Compounds were identified by comparison of the chromatograms of the pairwise combinations with those for control plants expressing *tHMGR*, *AsbAS1* and the single respective P450s (**methods 4.5.4**).

Analysis of the GC-MS data revealed that, remarkably, new products could be detected for almost every combination tested. From a total of ten possible pairwise combinations including *CYP51H10*, *CYP716A12*, *CYP93E1*, *CYP88D6* and *CYP72A65,* all were found to yield new products. These are summarised below in **Table 4-2**. For simplicity, most chromatograms are provided as supplementary data, also detailed in **Table 4-2**.

	CYP93E1 (C-24)	CYP72A65 (C-30)	CYP88D6 (C-11)	CYP51H10 (C-12/13/16)
CYP716A12 (C-28)	2 peaks (19-20) Supplementary Figure 7A	1 peak (21) Supplementary Figure 7B	3 peaks (43-45) Figure 4-11	2 peaks (50-51) Supplementary Figure 14B
CYP93E1 (C-24)	→	3 peaks (24-26) Supplementary Figure 8B	6 peaks (37-42) Figure 4-9	9 peaks (27-35) Supplementary Figure 9
CYP72A65 (C-30)	→	→	2 peaks (22-23) Supplementary Figure 8A	1 peak (36) Supplementary Figure 12
CYP88D6 (C-11)	→	→	→	4 peaks (46-49) Supplementary Figure 14A

Table 4-2: Summary of P450 combinations. The number of new peaks identified in the sample is given, with the numbers in brackets used to assign these peaks. The chromatograms are generally included as supplementary data, in the box. Although two C-30 oxidases were characterised above, CYP72A63 was the least active of the two and failed to give peaks when combined with either CYP716A12 or CYP51H10 (see **Supplementary Figure 6A**). Furthermore, combinations of CYP72A63 with CYP88D6 or CYP93E1 resulted in only trace amounts of new products (**Supplementary Figure 6B and C**). CYP72A63 is functionally equivalent to CYP72A65, and these trace products were found to be identical to these in the equivalent CYP72A65 samples where they were far more abundant. Therefore for simplicity, only the combinations with CYP72A65 are included here.

Four of the ten combinations resulted in putative products that have been previously reported and EI mass spectral data are available for comparison. These included 24-hydroxy oleanolic acid (**19**) (4-epi hederagenin) from CYP93E1 and CYP716A12 [101]; 30-hydroxy oleanolic acid (or queretaroic acid - **21**) from CYP72A65 and CYP716A12 [101, 173]; 11-oxo-30-hydroxy-β-amyrin (**23**) [100] from CYP72A65 and CYP88D6 and 24,30-dihydroxy-β-amyrin (**24**) from CYP72A65 and CYP93E1 [101]. These structures are shown in **Figure 4-8**. **Supplementary Table 3** presents a summary of these compounds and their fragmentation patterns.

The remaining combinations contained products for which reference EI spectra were not available and several of the anticipated products were expected to be previously unreported. However, the identity of products from three combinations could be reasonably assumed by analysis of their mass spectra. The major product

derived from combination of CYP51H10 and CYP93E1 had a mass spectrum consistent with 12,13β-epoxy, 16β,24-dihydroxy-β-amyrin (**29** - **Supplementary Figure 9**)**.** A very minor peak in this sample was also identified as 12,13β-epoxy,16β-hydroxy-β-amyrin (**28**). The combination of CYP51H10 and CYP72A65 resulted in a product consistent with 12,13β-epoxy, 16,30-dihydroxy-β-amyrin (**36**). Finally, the combination of CYP93E1 and CYP88D6 gave a novel product which had a fragmentation pattern consistent with 11-oxo-24-hydroxy-β-amyrin (**37**). This latter combination of CYP93E1 and CYP88D6 was chosen for purification and is detailed below in **section 4.3.2.1**. Work is currently underway in the Osbourn lab to determine the identity of the remaining structures. The putative structures are shown below in **Figure 4-8** and a breakdown of their fragmentation is given in **Supplementary Table 4**.

The major products in the remaining three combinations were difficult to identify with any certainty. These included products from coexpression of CYP88D6 with CYP716A12, and from CYP51H10 coexpressed with either CYP716A12 or CYP88D6. The CYP88D6 and CYP716A12 combination was chosen for purification to determine the identity of the product and is described below in **section 4.3.2.3**. Work is ongoing to determine the structure of the products from the latter CYP51H10 combinations, however trace parent ions were visible in the spectra of the new products which allowed for putative assignment of their identities. For combination of CYP51H10 with CYP88D6, the most abundant product was assigned as 11α-hydroxy,12,13β-epoxy,16β-hydroxy-β-amyrin (**47**). A second product was assigned as 11-oxo-12,13β-epoxy,16β-hydroxy-β-amyrin (**49**). A third minor product in the sample had a mass spectrum which could be interpreted as 11-oxo,16β-hydroxy-β-amyrin (**48**). For the combination of CYP51H10 and CYP716A12, the most abundant novel peak was assigned as 12,13β-epoxy oleanolic acid (**50**). These putative structures are given in **Figure 4-8**.

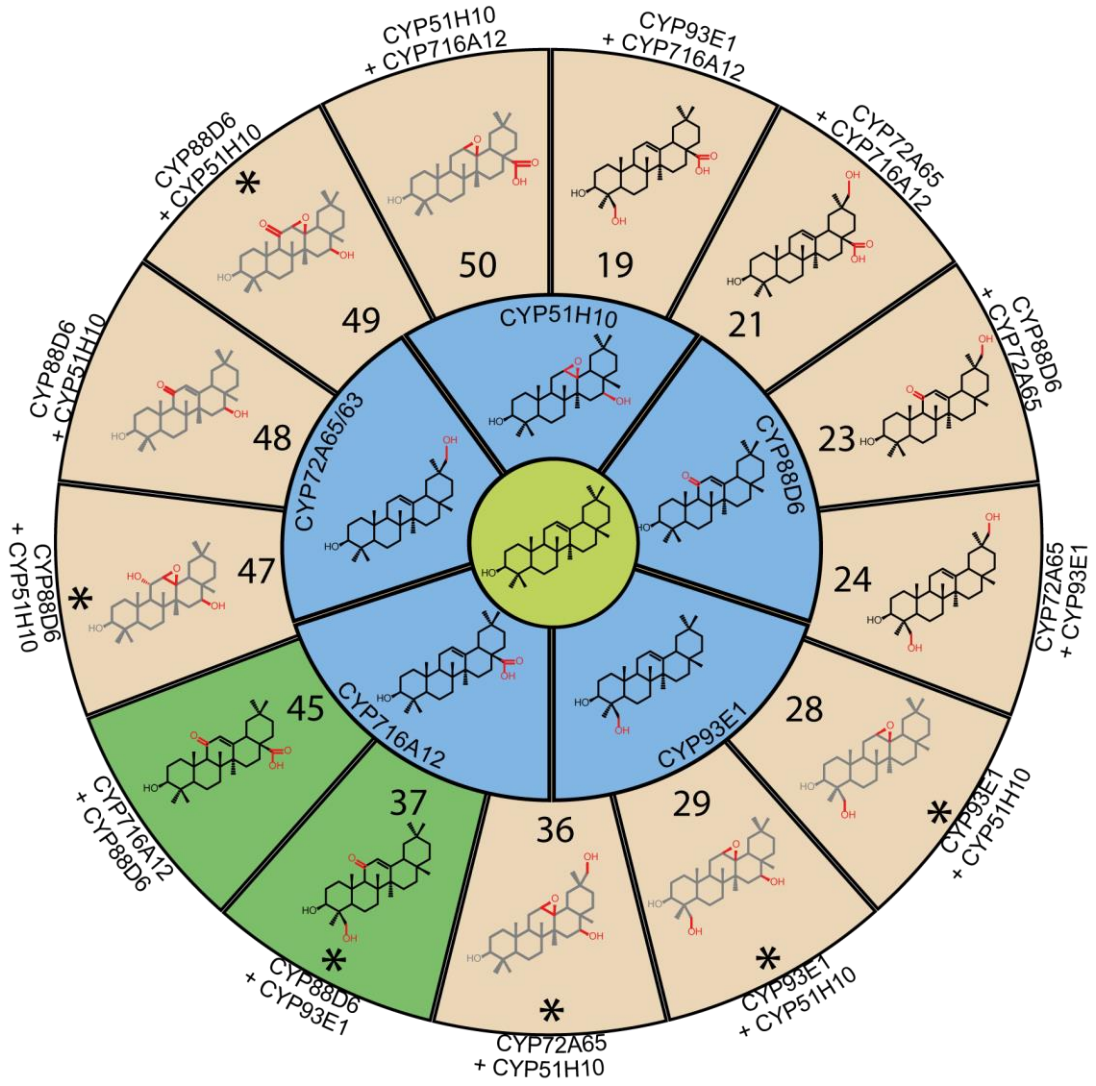


Figure 4-8: Schematic of combinatorial compounds identified in the present study (shown in numerical order clockwise from top). β-amyrin is shown in the centre and the major oxidised product for each individual P450 is represented in the inner blue circle. Combinatorial products are shown in the outer circle. Compounds drawn in grey are putative structures. Those shown in black are known structures based either on comparison with mass spectra from other studies, or from purification and NMR analysis in this work. The compounds for which NMR data is given in this study are represented with a green background. Finally, asterisks indicate those structures that do not appear to have previously been reported.

4.3.2.1– Determining the combinatorial products of CYP93E1 and CYP88D6

In addition to the products identified above, several of the combinations resulted in the formation of other products which were often not possible to assign an identity based on mass spectra. In many cases these products were only minor, however the combination of *tHMGR/AsbAS1/CYP88D6/CYP93E1* resulted in an unknown product (**42**) which was of a similar abundance to the major product putatively identified as 11-oxo,24-hydroxy-β-amyrin (**37**) (**Figure 4-9**). The mass spectrum of

this compound indicated that it was closely related to 11-oxo-24-hydroxy-β-amyrin (**37**) and 11-oxo-β-amyrin (**13**), due to the presence of prominent shared ions such at *m/z* 273 and 232 (**Figure 4-9B**). However the product had a putative parent mass of 570 which could not obviously be accounted for by the known activity of CYP88D6 and CYP93E1. Therefore, a large-scale infiltration of *tHMGR/AsbAS1/CYP88D6/CYP93E1* was performed in order to confirm the identity of 11-oxo,24-hydroxy-β-amyrin (**37**) and determine the identity of the unknown product (**methods 4.5.6**).

Extracts were taken from leaves infiltrated with *tHMGR/AsbAS1/ CYP88D6/CYP93E1* and flash column chromatography was performed to isolate two series of fractions containing the putative 11-oxo, 24-hydroxy-β-amyrin (**37**)**,** and the unknown compound (**42**) respectively (**methods 4.5.6**).

Preliminary proton (^{1}H) NMR analysis was performed on the isolated compound **37**. This revealed the presence of a singlet which integrated to a single hydrogen atom at δ5.59ppm. This suggested the presence of an alkene hydrogen atom between two quaternary carbons, as would be expected from the lone hydrogen at C-12 resulting from incorporation of a ketone moiety at C-11. This hydrogen atom appears as a triplet in β-amyrin due to coupling with the two protons at C-11. The eight methyl groups in β-amyrin are visible as singlets in the ^{1}H spectrum. Only seven such singlets were observed in the new compound **37**, which indicated that one of the methyl groups had been modified, as would be expected by incorporation of a primary alcohol at C-24. Further evidence for this was seen by the appearance of doublets at δ3.33 and δ4.21 ppm, which were found to be geminal hydrogens by DEPT-edited HSQC. These were expected to be the remaining C-24 hydrogens, with the downfield shift characteristic of a geminal oxygen atom. The NMR spectra for β-amyrin and compound **37** are provided as **Supplementary Figure 10** and **Supplementary Figure 11 respectively**.

As minor impurities were visible within the sample containing **37**, this was further purified by semi-preparative HPLC. The purified product then underwent further NMR analysis allowing full assignment of the compound (**methods 4.5.8.1**). The identity of compound **37** was consistent with 11-oxo, 24-hydroxy-β-amyrin.

The semi-purified unknown product **42** was further analysed by LC-MS and found to have a mass of 498 (**Figure 4-9C**) (**methods 4.5.6.1**). Thus addition of a single trimethylsilyl (TMS) group during derivatisation for GC-MS would be sufficient to account for the previously observed mass at *m/z* 570. Furthermore, while simple

triterpenes such as β-amyrin show poor UV absorbance, **42** showed strong absorption at ~250nm, suggesting the presence of the C-ring α,β-unsaturated ketone (**Figure 4-9C**). Taken together, this suggested that **42** featured the 11-oxo functionality and a single hydroxyl group available for silylation (likely the C-3 hydroxyl) and suggested the compound was modified at the C-24 position. Given these observations, a possible identity for **42** was suggested to be 11-oxo,24-acetoxy-β-amyrin (**Figure 4-9B, right**). If so, it should be possible to hydrolyse the esterified C-24 group to recover the expected product, 11-oxo,24-hydroxy-β-amyrin. To test this possibility, saponification of the semi-purified **42** was performed. Thin-layer-chromatography (TLC) analysis of the saponified sample revealed a new spot which was not present in the pre-saponified extract and co-migrated with 11-oxo, 24-hydroxy-β-amyrin (**37**) (**Figure 4-9D**). The presence of 11-oxo, 24-hydroxy-β-amyrin in the saponified extract was further confirmed by GC-MS (data not shown). This demonstrated that compound **42** could be hydrolysed by saponification to form 11-oxo, 24-hydroxy-β-amyrin.

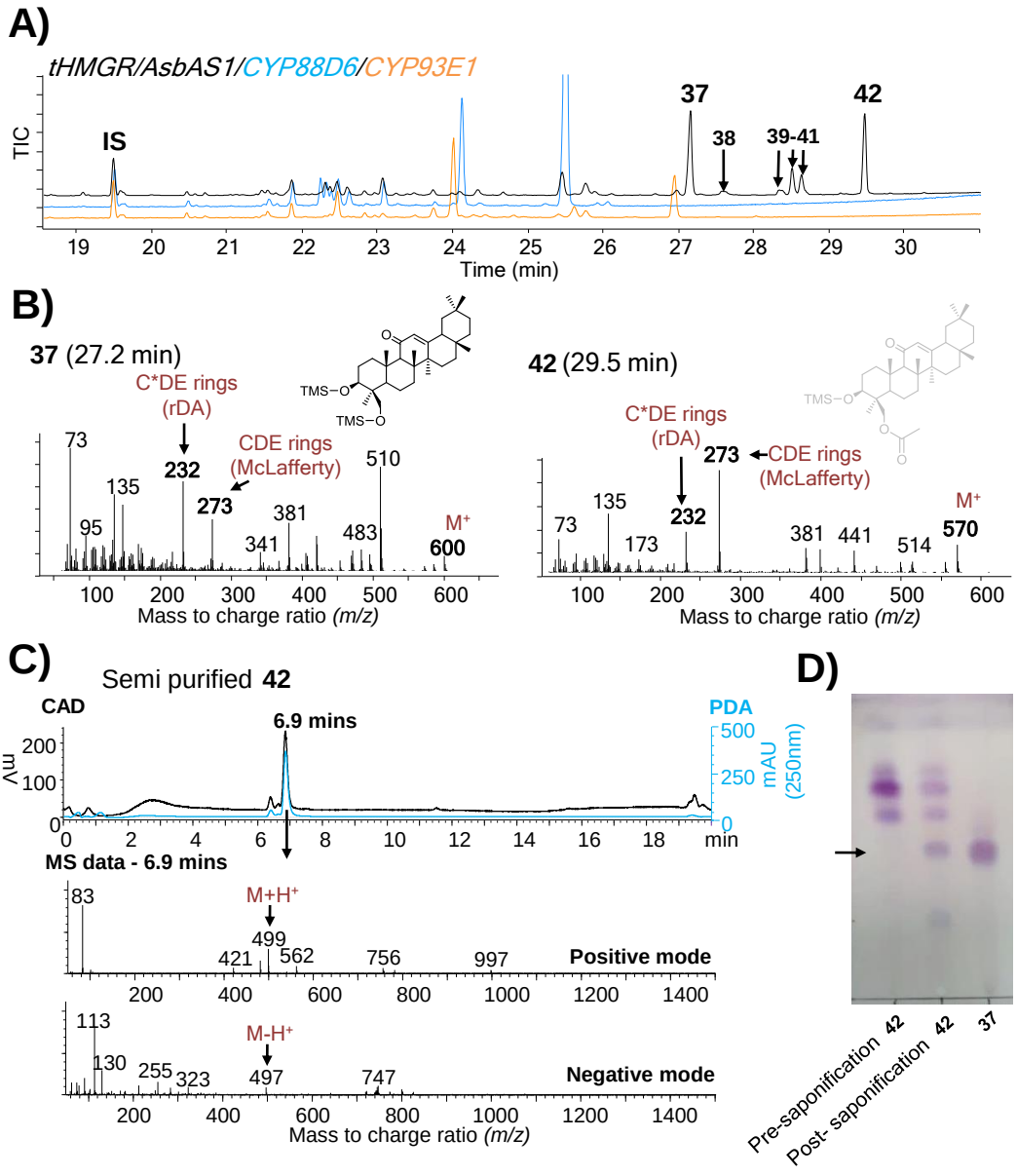


Figure 4-9: Production of 11-oxo,24-hydroxy-β-amyrin and acetylated derivatives. **A)** GC-MS total ion chromatogram (TIC) of *N. benthamiana* leaves expressing *tHMGR, AsbAS1, CYP93E1* and *CYP88D6*. **B)** Mass spectra for the major products **37** (left) and **42** (right). Structures of the compounds are given above, with the putative product 11-oxo,24-acetoxy-β-amyrin given in grey. Labels above the ions indicate likely fragments as detailed in **Supplementary Table 4**. Mass spectra for peaks 38-41 not shown here are given in **Supplementary Figure 13**. **C)** LC analysis of the putative 11-oxo,24-acetoxy-β-amyrin. Charged aerosol detector (CAD - black) and Photodiode array (PDA - blue) (250nm) data are overlaid (top) with mass spectral (MS) data for the product at 6.9 mins shown below. **D)** TLC analysis of the semi-purified extract of 11-oxo,24-acetoxy-β-amyrin before (left) and following hydrolysis by saponification (middle). The purified 11-oxo,24-hydroxy-β-amyrin **37** is in the lane on the right. The arrow denotes the appearance of the new spot co-migrating with this compound after saponification of the semi-purified extracts containing **42**.

4.3.2.2 – Acetylated compounds are caused by transesterification during extraction

The presence of an acetyl group results in a compound 30 mass units lower than would be predicted from a TMS group at the same position. This therefore allowed for putative identification of other acetylated compounds in other combinations, most commonly in samples from leaves expressing *CYP93E1*. These compounds are given above the mass spectra in the relevant supplementary figures. Acetylated triterpenes could result from modification by *N. benthamiana.* However given that saponification was routinely performed during processing of the leaf material it was somewhat surprising that these should still be abundant in the final extracts. Therefore, it was considered possible that these products were an extraction artefact. Saponification is a useful means of processing leaf material prior to organic phase extraction allowing removal of unwanted chlorophylls and fats. However the saponification solution (ethanol, water and KOH 9:1:1) is miscible with a range of organic solvents preventing liquid-liquid partitioning. To overcome this, following saponification the solution was routinely evaporated prior to direct addition of ethyl acetate (**methods 4.5.5**). However the residual KOH might catalyse transesterification, with transfer of the acetyl group from ethyl acetate onto the primary alcohol at C-24.

To test this possibility, two disks were taken from a *tHMGR/AsbAS1/CYP93E1* leaf. One was treated with the previous extraction method detailed above, while the other was immersed directly into ethyl acetate (**methods 4.5.5.1**). GC-MS analysis of the samples revealed that while 24-hydroxy-β-amyrin (**6**) could be detected in both samples, a peak corresponding to a putative 24-acetoxy-β-amyrin (**11**) was present only in the sample which had undergone saponification (**Figure 4-10A**). This strongly suggests that the extra products were formed by transesterification (**Figure 4-10B**).

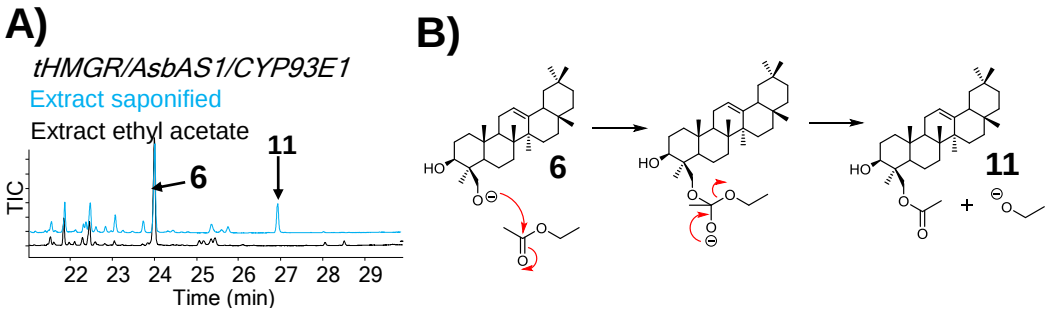


Figure 4-10: Transesterification of 24-hydroxy-β-amyrin upon ethyl acetate extraction **A)** GC-MS total ion chromatogram (TIC) analysis of leaf disks from a single *N. benthamiana* leaf expressing *tHMGR/AsbAS1/CYP93E1*. Saponification followed by ethyl acetate extraction (blue) may cause transesterification of 24-hydroxy-β-amyrin (**6**) resulting in the product **11**. This is not observed in leaves extracted with ethyl acetate directly (black). **B)** A possible mechanism for the base-catalysed transesterification reaction between 24-hydroxy-β-amyrin and ethyl acetate is shown.

4.3.2.3 – Identification of 11-oxo oleanolic acid from combination of CYP88D6 and CYP716A12.

The second product chosen for purification was compound **44**, found in *tHMGR/AsbAS1/CYP88D6/CYP716A12* leaf extracts. Based on the activity of the two P450s, this combination was expected to result in the formation of 11-oxo-oleanolic acid. However none of the novel peaks could be obviously identified as such from their EI mass spectrum. The most abundant product (**44**) had a putative molecular mass at *m/z* 496, which does not correspond to the expected mass for 11-oxo oleanolic acid (MW = 614) or potential intermediates (**Figure 4-11A**).

A large-scale *tHMGR/AsbAS1/CYP88D6/CYP716A12* infiltration was performed and the leaves were processed and metabolites were extracted (**methods 4.5.7**). Unexpectedly, GC-MS analysis of this crude extract revealed a previously unobserved peak at 28.3 min (**45**) which showed the correct parent mass (614) for 11-oxo-oleanolic acid (**Figure 4-11B**). However, further attempts to isolate **45** failed. Despite this, it was possible to isolate compound **44**. The purified compound was analysed by ion trap time-of-flight (IT-TOF) mass spectrometry to get an accurate mass and predict a molecular formula (**methods 4.5.7.2**). **44** was determined to have a mass of 493.3283, consistent with the sodium adduct of a compound with the molecular formula $C_{30}H_{46}O_4$ (1.13ppm). Surprisingly, this is the expected mass of 11-oxo oleanolic acid (**Figure 4-11C**), which suggested that this was the identity of **44**. Further NMR analysis of the purified compound **44** was also consistent with 11-oxo oleanolic acid (**methods 4.5.8.2**). GC-MS analysis of the purified 11-oxo-oleanolic acid showed that the purified compound likewise appeared

predominantly as peak **44** (data not shown). The mass differences between **44** and **45** would imply the loss of the C-28 carboxylic acid group (**Figure 4-11D**). This may point to pyrolytic degradation of the compound within the GC inlet. It is possible that in future this artefact may be alleviated by the use of a cooled injection system.

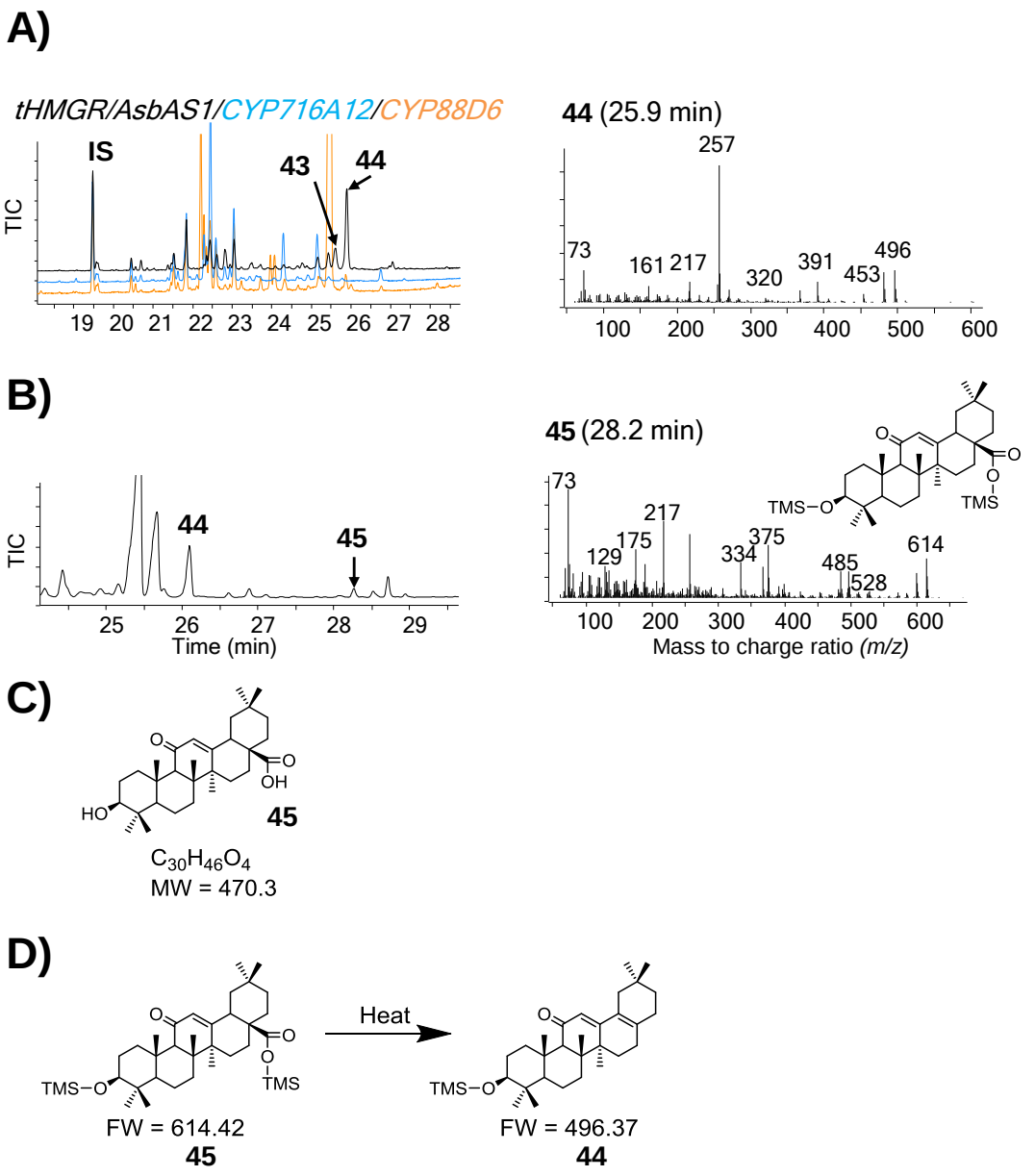


Figure 4-11: Combinatorial biosynthesis of CYP716A12 and CYP88D6. **A)** GC-MS Total ion chromatogram (TIC) of *N. benthamiana* leaves expressing *tHMGR/AsbAS1/CYP716A12/CYP88D6* (left). Chromatograms for controls samples are shown in blue and orange with the combinatorial sample in black. The mass spectrum (right) of the major product at 28.2 mins does not correspond to the expected spectrum for 11-oxo oleanolic acid. Mass spectra for compound **43** is shown in **Supplementary Figure 13B**. **B)** GC-MS analysis of the large-scale infiltration of *tHMGR/AsbAS1/CYP716A12/CYP88D6* (left) revealed minor amounts of a novel product which has the expected fragmentation of 11-oxo-oleanolic acid. **C)** Structure of 11-oxo-oleanolic acid **D)** Pyrolytic breakdown of 11-oxo-oleanolic acid may occur in the GC inlet and be responsible for the product with a parent mass of *m/z* = 496.

4.3.3 – Incorporating CYP51H10 mutant enzymes into the toolkit.

CYP51H10 is a multifunctional P450 that catalyses both the hydroxylation and epoxidation of β-amyrin, forming 12,13β-epoxy,16β-hydroxy-β-amyrin [85] (EpHβA (**2**) **Figure 4-2**) No other triterpene oxidases have been reported to catalyse either of these modifications, which makes CYP51H10 an important enzyme for triterpene engineering. However selective incorporation of each of these two groups individually is not currently possible. If these two activities could be separated this would enable introduction of the C-12,13 epoxide and C-16β hydroxyl functionalities separately, thus enhancing the triterpene engineering toolkit.

Previously, *in silico* modelling was performed to identify a series of 15 residues lining the active site of CYP51H10 and which are predicted to be functionally important [85]. To expand upon this work, three of these residues were considered for mutagenesis to determine the impact of the various mutations on product specificity. These include an isoleucine at position 471 and a pair of alanine residues at positions 353 and 354. These residues lie in close proximity within the active site and have been suggested to play important roles in determining the size of the active site as well as interacting with β-amyrin and a putative oxidised intermediate (16β-hydroxy-β-amyrin) en route to EpHβA [85] (**Figure 4-12A**). Saturation mutagenesis experiments were performed at each of these three residues in the Osbourn lab by Prof Robert Minto. The CYP51H10 mutants were transiently expressed in *N. benthamiana* with AsbAS1 and screened by GC-MS. This resulted in the identification of a series of mutants with altered product profiles. Amongst these, mutants of CYP51H10 were identified that each perform only one of the two oxidations normally carried out by the wild type enzyme. Mutant CYP51H10-I471M produces predominantly 12,13β-epoxy-β-amyrin (**52**), while retaining the ability to produce a small amount of the wild-type product (**2**) (**Figure 4-12B**). Conversely, CYP51H10-A354L produces 16β-hydroxy-β-amyrin (**53**), which lacks the epoxide group found in the wild-type product (**Figure 4-12B**). Mass spectra for the two mutant products are shown in **Figure 4-12B**. A scheme for the EI fragmentation of CYP51H10 wild-type and mutant products is given in **Supplementary Figure 17**. These mutants therefore allow bifurcation of the CYP51H10 wild type activity and represent useful additions to the toolkit for selective functionalisation of the β-amyrin scaffold.

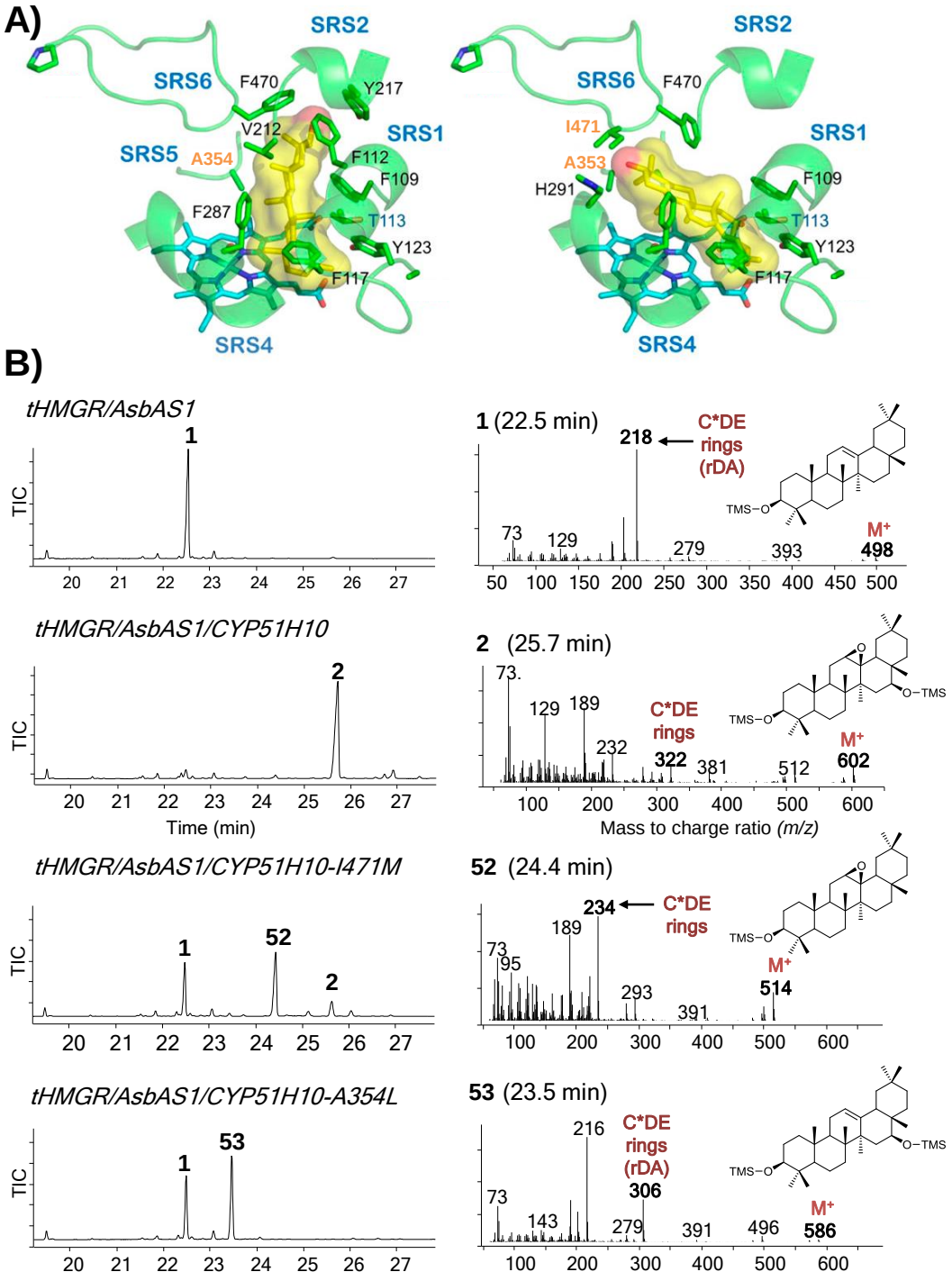


Figure 4-12: CYP51H10 model and mutant product profiles. **A)** A model of the active site of CYP51H10 is shown with putative precatalytic binding modes of β-amyrin (left) and 16β-hydroxy-β-amyrin (right). Residues lining the active site cavity are labelled within their respective substrate recognition sites (SRS). The three residues selected for mutagenesis are indicated with orange text. Figure adapted from Geisler et al [85]. **B)** GC-MS profiles of CYPH10 mutants. CYP51H10-I471M acts on β-amyrin to form 12,13β-epoxy-β-amyrin (**52**), while CYP51H10-A354L forms 16β-hydroxy-β-amyrin (**53**). GC-MS analysis of extracts from *N. benthamiana* leaves expressing *CYP51H10* and the mutants are shown. Total ion chromatograms (TIC) are given on the left hand side with EI mass spectral data and structure of the major new product given to the right. Key ions in the mass spectra of the mutant products are highlighted. Further details of this mechanism are given in **Supplementary Figure 17**.

4.3.4 – Combinatorial biosynthesis of CYP51H10 mutant enzymes

Further combinatorial experiments were performed with CYP51H10 mutants in combination with the other β-amyrin oxidases, *CYP716A12*, *CYP88D6*, *CYP93E1* and *CYP72A65*. Each combination was also coinfiltrated with *tHMGR* and *AsbAS1* as before (**section 4.3.2**). The results are summarised in **Table 4-3**.

CYP51H10 Mutant	CYP88D6	CYP93E1	CYP716A12	CYP72A65
A354L	1 peak (48) Supplementary Figure 15A	2 peaks (54-55) Supplementary Figure 15B	-	1 peak? (56) Supplementary Figure 15C
I471M	-	-	1 peak (50) Supplementary Figure 16A	2 peaks (58-59) Supplementary Figure 16B

Table 4-3: Summary of combinations involving CYP51H10 mutants. The number of new peaks seen by GC-MS is given along with their numbered assignments.

4.3.4.1 – Combinatorial biosynthesis using CYP51H10-A354L

Combination of CYP51H10-A354L with CYP88D6 accumulated a single new peak with the same retention time and mass spectrum as the putative 11-oxo, 16β-hydroxy-β-amyrin (**48**) as previously identified by combination of the wild-type CYP51H10 and CYP88D6. Coexpression of CYP51H10-A354L with CYP93E1 resulted in detection of a new product identified as 16β,24-dihydroxy-β-amyrin (**54**). Only trace amounts of this compound could be found from expression of CYP93E1 with the wild-type CYP51H10 (not shown). The putative structures for these products are given in **Figure 4-13**.

A peak was found in the combination of CYP51H10-A354L and CYP72A65 with a spectrum which would be expected of 16β,30-dihydroxy-β-amyrin (**56**). However this peak had the same retention time and a similar mass spectrum to a peak in the CYP51H10-A354L control (**57**) (**Supplementary Figure 15**). Therefore the uniqueness of this peak to the combinatorial sample could not be ascertained. In future, an alternative GC-MS program or column might allow separation of these products. Finally, no new products were observed upon coexpression of CYP51H10-A354L with CYP716A12.

4.3.4.2 – Combinatorial biosynthesis of CYP51H10-I471M.

Combination of CYP51H10-I471M with CYP716A12 gave a single novel peak which was found to have an identical retention time and mass spectra to the product **50** in

the corresponding wild-type CYP51H10 and CYP716A12 sample. Coexpression of CYP51H10 with CYP72A65 gave a novel peak which was assigned an identity of 12,13β-epoxy, 30-hydroxy-β-amyrin (**58**). This product had not been identified in leaves expressing CYP72A65 with the wild-type CYP51H10. Hence here, the use of CYP51H10-I471M allows access to compounds which are unobtainable with the wild-type enzyme. In contrast, no new products were observed with the combination of CYP51H10-I471M with CYP88D6 or CYP93E1 (not shown).

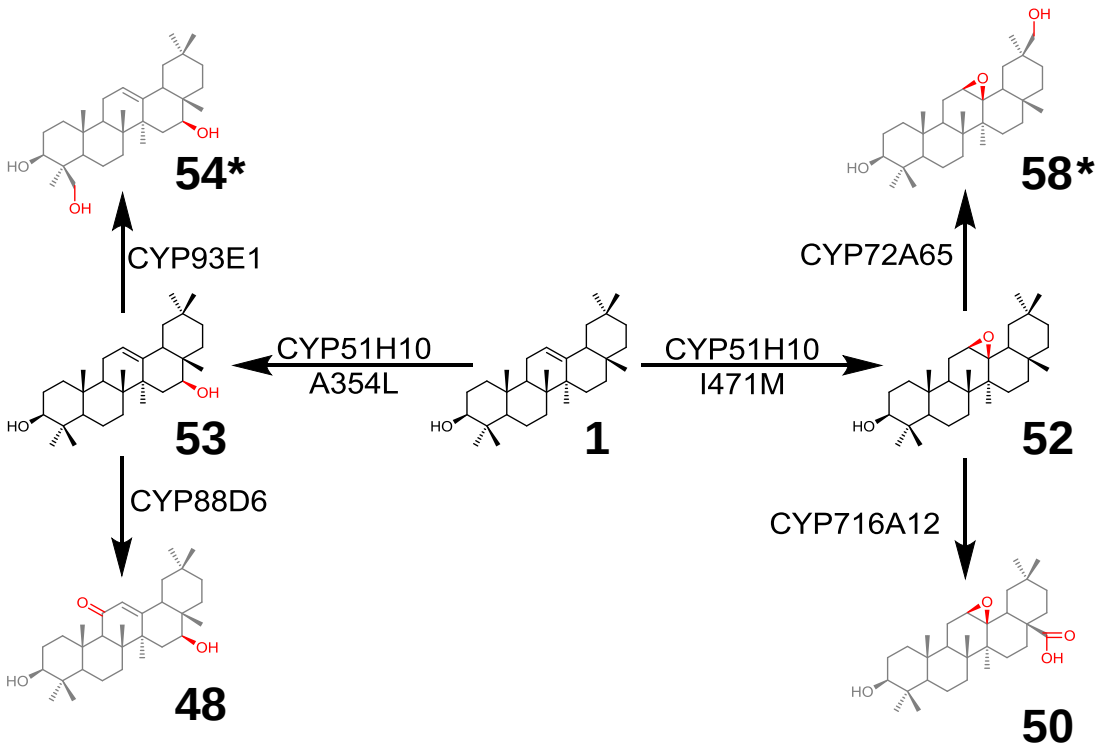


Figure 4-13: Combinatorial products formed using CYP51H10 mutants. The putative structures for the combinatorial products are indicated in grey. Compounds marked with an asterisk were present in only trace amounts or undetectable in the equivalent combinations with wild type CYP51H10. These two putative compounds do not appear to have previously been reported in the literature. Please note that the CYP51H10 mutants are shown first acting on β-amyrin for simplicity, however the order of oxidations is unknown.

4.3.4.3 – Part 2 summary

One of the ideals of combinatorial biosynthesis is the rational design of molecules from selection of available enzyme components. When systematically expressed together *in planta*, impressively almost all pairs of β-amyrin oxidases resulted in the formation of new products. This suggests that these enzymes may indeed be used interchangeably for directing the biosynthesis of custom triterpenes.

The identity of the new compounds was predicted from analysis of their EI mass spectra resulting in putative assignment of 13 compounds (**Figure 4-8**). The identity of four of these was confirmed by comparison to other studies, while two of the novel compounds, 11-oxo oleanolic acid (**45**) and 11-oxo,24-hydroxy-β-amyrin (**37**) were verified by purification and NMR. 11-oxo oleanolic acid has been isolated from a number of plant species [177-179] however it does not appear to have been reported from *G. uralensis* or *M. truncatula* from which the two P450s (CYP88D6 and CYP716A12 respectively) are derived. By contrast 11-oxo, 24-hydroxy-β-amyrin does not appear to have been reported previously, although the C-24 and C-11 modifications are both present in the liquorice saponin G2 (24-hydroxy glycyrrhizin) from *G. uralensis* [180, 181]. Therefore it is possible that this compound may be present at low levels in this species as an intermediate, however extraction from the natural source is unlikely to be a viable option. This demonstrates the utility of this approach to obtain otherwise inaccessible metabolites.

Work is currently ongoing in the Osbourn lab to verify the identity of the remaining compounds which are all derived from combinations of CYP51H10 (structures given in grey in **Figure 4-8**). CYP51H10 is presently the only reported triterpene oxidase from monocots and is involved in the biosynthesis of antifungal triterpene glycosides in oat (avenacins – see Chapter 6) [85, 86]. Aside from the C12,13β epoxide and C-16β hydroxyl groups incorporated by CYP51H10, the avenacins are also oxidised at the C-21, C-23 and C-30 positions [182]. Thus, none of the regiospecific modifications introduced by CYP716A12 (C-28), CYP88D6 (C-11) and CYP93E1 (C-24) are featured in the avenacins. Furthermore, use of two CYP51H10 mutants allowed access to compounds which accumulated either in trace amounts (**54**) as well as products which were not detected at all in the wild-type (**58**) (**Figure 4-13**). This provides substantial potential for engineering new antifungal saponins and for performing in-depth structure-function investigations.

Finally, there remains significant potential for further combinatorial biosynthesis of β-amyrin, both with other β-amyrin oxidases which were not tested here, as well as other P450s which oxidise derivatives of β-amyrin. For example CYP72A67 and CYP72A68v2 are inactive on β-amyrin directly, but oxidise oleanolic acid [101, 102]. These have also been found to be functional in *N. benthamiana* (**Supplementary Figure 18** and **Supplementary Figure 19**) and could be incorporated into combinatorial biosynthesis utilising three or more P450s.

4.3.5 – Part 3 – Beyond β-amyrin: Diversifying triterpene production in *N. benthamiana*

The triterpene β-amyrin was the focus of the present study due to the number of characterised oxidases available, however β-amyrin represents only a single scaffold amongst more than 100 possible OSC cyclisation products [62]. Numerous other scaffolds are commonly found throughout the plant kingdom and are further derivatised in specialised pathways [39, 72]. The type of scaffold and the ways in which it is tailored may be an important contributor to bioactivity [57]. As new pathways continue to be characterised for non-β-amyrin scaffolds [89, 114, 183], further opportunities to apply combinatorial biosynthesis to these other scaffolds will be made possible.

To explore the capacity to produce other triterpene scaffolds in *N. benthamiana,* a series of OSCs were chosen which synthesise a variety of triterpene scaffolds commonly found in higher plants [72]. These included an α-amyrin synthase from *Malus domestica* (MdOSC1)*,* a lupeol synthase from *Lotus japonicas* (LjOSC3) and a cucurbitadienol synthase from *Cucumis sativus* (CsCPQ)*.* These OSCs were cloned into pEAQ-*HT*-DEST1 and transformed into *A. tumefaciens* for expression in *N. benthamiana* (**Supplementary Table 5**). GC-MS analysis of extracts from leaves expressing the various OSCs with *tHMGR* confirmed that all of the expected products could be detected in abundance (**Figure 4-14)**. This therefore suggests that *N. benthamiana* is a suitable platform for production of a range of triterpene scaffolds.

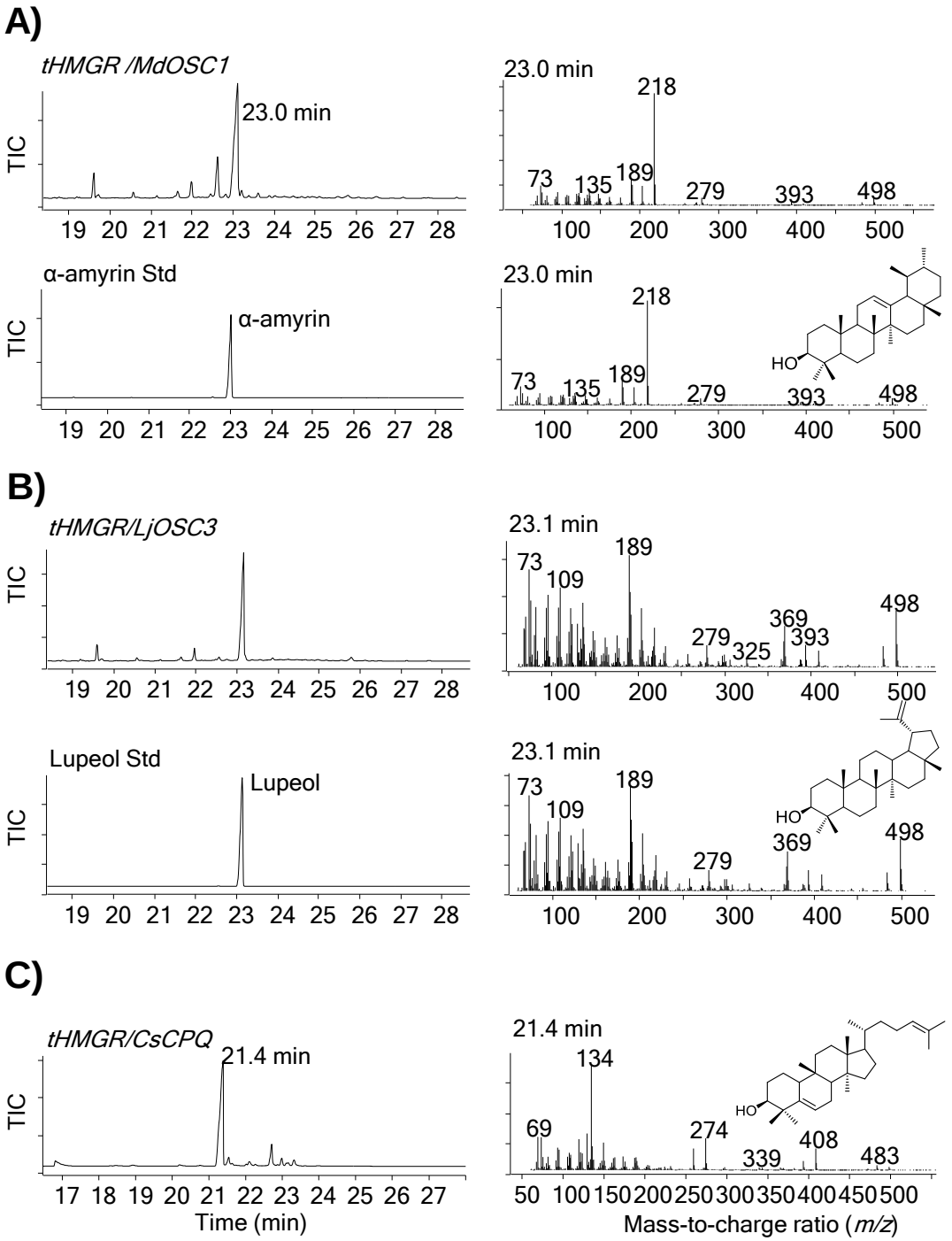


Figure 4-14: Production of diverse triterpene scaffolds in *N. benthamiana* through coexpression of *tHMGR* and various OSCs. GC-MS total ion chromatograms (TIC) of the samples are shown on the left with mass spectra for the peaks indicated on the right. **A)** Production of α-amyrin by expression of *Malus* domestica α-amyrin synthase (MdOSC1). **B)** Production of lupeol by expression of *Lotus japonicus* Lupeol synthase (LjOSC3). **C)** Production of cucurbitadienol by expression of *Cucumis sativus* cucurbitadienol synthase (CsCPQ). Standards of α-amyrin and lupeol are shown is shown below the relevant chromatograms and mass spectra. Details of the genes used here are given in **Supplementary Table 5**.

4.3.5.1 – Production of bioactive triterpenes from ginseng in *N. benthamiana*

Recently an AsbAS1-S728F mutant was described which produces alternative cyclisation products, resulting in the predominant production of the tetracyclic triterpene (20*S*)-dammarenediol-II (**60**) (DD-II) [137]. DD-II is the core scaffold found in dammarane-type ginsenosides, a class of biologically-active triterpene saponins produced by *Panax* species [45, 46]. DD-II can be oxidised at the C-12β position by CYP716A47 to form protopanaxadiol (PPD) (61) and this can undergo further C-6α oxidation by CYP716A53v2 to form protopanaxatriol (PPT) (62) [88, 103] (**Figure 4-15**). PPD and PPT form the aglycones for two major branches of pharmaceutically-active ginsenosides [44].

The significant interest in these scaffolds prompted an investigation into the potential for production of these ginsenosides in *N. benthamiana*. Therefore, the genes encoding the AsbAS1-S728F mutant and two *P. ginseng* P450s were cloned into pEAQ-*HT*-DEST1 and transformed into *A. tumefaciens* (**methods 4.5.1**). The various strains were infiltrated in different combinations in *N. benthamiana* along with a strain carrying *tHMGR.* Infiltration densities were maintained at a total OD_{600} of 0.8 through use of a strain carrying *GFP* in the absence of other strains (**methods 4.5.3**). GC-MS analysis of *tHMGR/AsbAS1-S728F/GFP/GFP* leaf extracts revealed a large peak corresponding to a (20*S*)-dammarenediol-II standard (**Figure 4-15**). Further inclusion of the strain carrying CYP716A47 resulted in a major new peak which matched a standard of PPD, indicating that the major product of the mutant AsbAS1 enzyme was a suitable substrate for this enzyme (**Figure 4-15**). CYP716A53v2 was inactive directly against the DD-II backbone as previously reported [88], but in combination with *tHMGR*, *AsbAS1-S728F* and *CYP716A47*, a peak corresponding to PPT was observed. Together, this demonstrates that high-value ginsenoside aglycones can be efficiently produced in high yield in *N. benthamiana.*

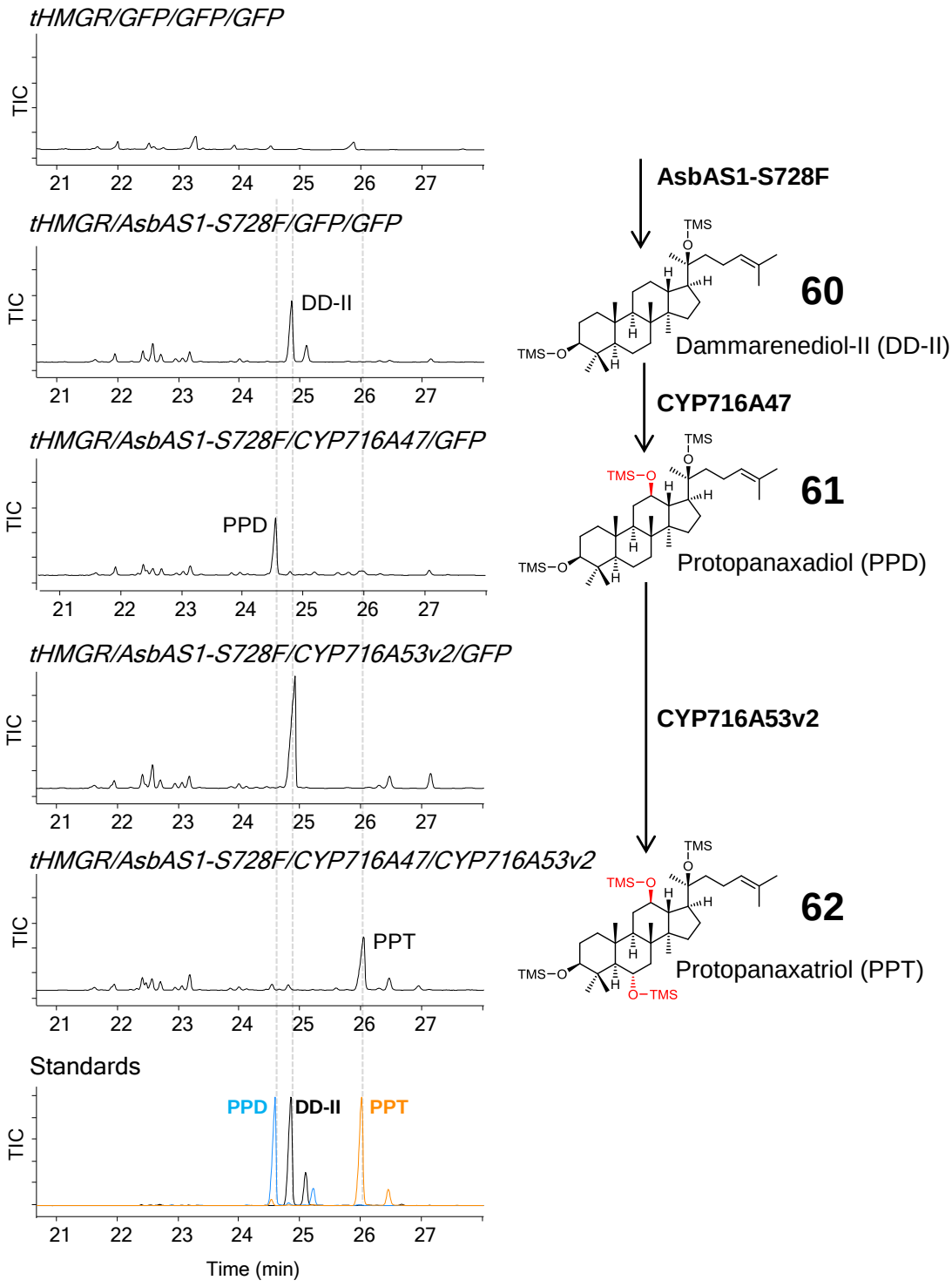


Figure 4-15: Production of ginsenosides in *N. benthamiana.* (**Left**) GC-MS total ion chromatograms (TIC) of extracts of *N. benthamiana* leaves expressing *tHMGR* with *AsbAS1-S728F* and combinations of the two P450s, *CYP716A47* and *CYP716A53v2*. Standards of DD-II, PPD and PPT are shown at the bottom. Mass spectra for the standards are given in **Supplementary Figure 21**. (**Right**) AsbAS1-S728F cyclises oxidosqualene to 20*S*-dammarenediol-II (DD-II)**.** DD-II is oxidised at the C-12β position to protopanaxadiol (PPD) by CYP716A47. CYP715A53v2 is inactive directly against DD-II but oxidises PPD at the C-6α position to protopanaxatriol (PPT). Details of the enzymes are given in **Supplementary Table 5**.

Panax species are highly prized for their medicinal value; however they are slow growing and threatened in the wild by overhunting. More than 150 unique ginsenosides are known derived from a number of backbones and displaying different patterns of glycosylation [44]. Hence extracts of ginseng species are complex and sophisticated analytical and purification methods may be required for the isolation of individual compounds [184]. Expression of specific enzymes in *N. benthamiana* therefore offers the ability for select production of individual ginsenosides. Recently, a variety of glycosides of PPD and PPT were successfully produced through transient expression of *P. ginseng* glycosides in *N. benthamiana* recently (Dr Thomas Louveau, Mr Charlie Owen, (Osbourn lab), personal communication). Therefore there is particularly exciting potential for further combinatorial biosynthesis in this area.

4.3.5.2 – Part 3 summary

The above work demonstrates that *N. benthamiana* holds significant potential as a production platform for a diverse range of triterpenes besides β-amyrin. Although a relatively small range of common scaffolds were used here as proof of concept, the system has also proven to be suitable for synthesis and modification of uncommon scaffolds, such as thalianol (**Supplementary Figure 20**)**.** There is now significant capacity to expand combinatorial biosynthesis approaches to different scaffold-oxidase combinations [97, 108]. The ability to coinfiltrate different strains of *A. tumefaciens* from the toolkit will greatly facilitate these studies and work is currently underway in the Osbourn lab to explore this potential.

4.4 – Conclusions

Here, *N. benthamiana* is demonstrated as a versatile production host for engineering a diverse collection of triterpenes. In particular, numerous oxidised derivatives of the pentacyclic triterpene β-amyrin were produced by coexpression of the β-amyrin synthase AsbAS1 with previously-characterised β-amyrin oxidases from different plant species. Further combinatorial biosynthesis with these P450s showed that almost all of the P450 pairings were compatible, resulting in the production of hybrid oxidised forms of β-amyrin. This work has significant future potential for production of ‘customised’ triterpenes and saponins. This will permit systematic studies of important structure-function relationships with regards to biological activity. This is further explored in Chapter 5.

4.5 – Materials and Methods for Chapter 4

4.5.1 – Cloning toolkit enzymes

4.5.1.1 – *CYP72A65, CYP72A68v2, CYP716A12, CYP93E1, CYP88D6, CYP716A47* and *CYP716A53v2*

A number of the P450s were kindly supplied by Prof Alain Goossens (VIB Ghent). These included *CYP72A65*, *CYP72A68v2*, *CYP716A12*, *CYP93E1*, *CYP88D6*, *CYP716A47* and *CYP716A53v2*. Of these, *CYP72A65*, *CYP72A68* and *CYP716A12* were supplied in pEAQ-*HT*-DEST1, allowing direct transformation of A. tumefaciens LBA4404. *CYP93E1*, *CYP88D6*, *CYP716A47* and *CYP716A53v2* were supplied in the Gateway-compatible yeast expression vector pAG423GAL-1 vectors. To transfer the genes in pAG423GAL-1 vectors into the pDONR207 Entry vector, a BP reaction was performed followed by transformation of subcloning efficiency *E.coli* DH5α. Colonies were screened by PCR and sequenced with attL1-F/attB2-R primers (**Table 2-1**) to verify the presence and integrity of the insert. Finally, the genes were transferred to pEAQ-*HT*-DEST1 via an LR reaction and the resulting constructs were used to transform *A. tumefaciens* LBA4404.

4.5.1.2 – *CYP51H10* mutants

Strains of *A. tumefaciens* LBA4404 carrying mutant constructs *CYP51H10-A354L* and *CYP51H10-I471M* were kindly supplied by Prof Robert Minto (Indiana University-Purdue University, Indianapolis, USA).

4.5.1.3 – AsbAS1-S728F

A strain of LBA4404 carrying *AsbAS1-S728F* was kindly supplied by Dr Ramesha Thimmappa.

4.5.1.4 – *MdOSC1, CYP72A63, CYP72A154, CYP72A67* and *LjOSC3*

Synthesised versions of *MdOSC1*, *CYP72A63* and *CYP72A154* were supplied by Dr Aymeric Leveau in the Golden Gate level 0 [135] vectors PL0M-SC-cMdOSC1, pL0M-SC-cCYP72A63 and pL0M-SC-cCYP72A154 respectively. *LjOSC3* was kindly provided by Prof Kalliope Papadopoulou. The coding sequences for these genes were amplified from these vectors using PCR (Phusion polymerase, NEB), primers were MdOSC1-attB1-F/2-R, 72A63-attB1-F/2-R, 72A154-attB1-F/2-R and LjOSC3-attB1-F and LjOSC3-attB2-R (**Table 4-4**).

CYP72A67 was cloned by PCR from *M. truncatula* root cDNA. cDNA was kindly supplied by Dr Sonali Roy (John Innes Centre). Phusion polymerase and primers

72A67-attB1-F/2-R were used to amplify the CYP72A67 coding sequence (**Table 4-4**).

The resulting PCR products for the genes above were purified (QIAquick PCR purification kit, Qiagen) and cloned into pDONR207. Colonies were screened by PCR and sequenced using attL1-F/2-R primers **Table 2-1**. OSCs were additionally sequenced with primers in **Table 4-4** Following this, purified Entry vectors containing the P450s was transferred to the pEAQ-*HT*-DEST1 vector and used to transform *A. tumefaciens* LBA4404.

4.5.1.5 – THAS, CYP708A2 and CYP705A5

The *THAS, CYP708A2* and *CYP705A5* genes were cloned from *A. thaliana* root cDNA (kindly provided by Dr Hans Nützmann) using the primers THAS-attB1s-F, THAS-attB2s-R, CYP708A2-attB1s-F, CYP708A2-attB2s-R, CYP705A5-attB1s-F, CYP705A5-attB2s-R (**Table 4-4**). These primers contained half-length Gateway attB sites, requiring a 2-step Gateway PCR to introduce the full-length attB sites. CYP705A5 was successfully amplified in this manner and was subsequently used to perform a BP reaction with pDONR207.

For THAS and CYP708A2, a single 30-cycle PCR was performed using the same primers as above (Phusion polymerase, NEB) resulting in the full-length product flanked by truncated Gateway attB sites. These PCR products were purified (QIAquick PCR purification kit, Qiagen) and inserted into a Zero Blunt TOPO PCR cloning kit (Invitrogen) according to the manufacturer's instructions. This vector was then used as a template for a second PCR using the full-length attB primers attB1F-1 and attB2R-1 (**Table 2-1**). The resulting PCR product was likewise purified and inserted into pDONR207. Finally, the genes were transferred to pEAQ-*HT*-DEST1 vector and used to transform *A. tumefaciens* LBA4404.

4.5.2 – Primer sequences

	Name	Adapter/5'sequence	Gene-specific sequence
Gateway cloning	MdOSC1-attB1-F	attB1F-1	ATGTGGAAGATCAAGTTTGGAGAG
	MdOSC1-attB2-R	attB2R-1	TCAAGCGATCTTTTTGATAGGCAAC
	72A63-attB1-F	attB1F-1	ATGGAAGTGTTTATGTTTCCCACAG
	72A63-attB2-R	attB2R-1	TTACAGTTTATGCAAAATGATGCTTG
	72A154-attB1-F	attB1F-1	ATGGATGCATCTTCCACACCAGG
	72A154-attB2-R	attB2R-1	TTACAGTTTATGCAGAATGATGGG
	LjOSC3-attB1-F	attB1F-1	ATGTGGAAGTTGAAGGTAGCAGAAGG
	LjOSC3-attB2-R	attB2R-1	TCATGCACATAAGACTCTGCGACGG
	72A67-attB1-F	attB1F-1	ATGGAAGCATCATTGGCC
	72A67-attB2-R	attB2R-1	TTATGCTTTCACTTTGCGTAG
	THAS-attB1s-F	attB1Fs-1	ATGTGGAGGCTGAGAACTGG
	THAS-attB2s-R	attB2Rs-1	TTAAGGGAGGAGACGTCGCAG
	CYP708A2-attB1s-F	attB1Fs-1	ATGGATCGACGGGGCAAG
	CYP708A2-attB2s-R	attB2Rs-1	CTAGAGTGACTGGGAAATCTTG
	CYP705A5-attB1s-F	attB1Fs-1	ATGGCATCAATGATCACTGTTG
	CYP705A5-attB2s-R	attB2Rs-1	TTAAGTGTTTAGGTTTCGAGGAAC
Gene-specific sequencing primers	THAS_seq_F		GCCACAGTGTTATGTTCTG
	THAS_seq_R		CCGTACATTCTACATACC
	MdOSC1		GTTCTGCTACTGTAGGTTGAC
	LjOSC3		CTGTCGCTTAGTTTACATGCC

Table 4-4: Primer sequences as used in Chapter 4. The gene-specific sequences are given on the right, with 5' adapters for Gateway cloning indicated in the column to the left and detailed in **Table 2-1**.

4.5.3 – Testing the toolkit enzymes in *N. benthamiana*

Strains of *A. tumefaciens* carrying the relevant constructs were cultured and prepared as detailed in **section 2.1.3.2**. Infiltrations were performed as in **section 2.1.3.3**. Control infiltrations were also performed without P450s (*tHMGR* and OSC) or without the OSC (*tHMGR* and P450) to ensure that new products were dependent on the presence of β-amyrin (not shown in main text). In these cases, a strain expressing *GFP* was used in place of the missing strain.

4.5.4 – Combinatorial biosynthesis

Strains of *A. tumefaciens* were cultured and prepared as detailed in **section 2.1.3.2**. Strains were made to a total OD_{600} of 0.8 and four strains carrying *tHMGR, AsbAS1* and two P450s were mixed in equal ratios. For single P450 controls, an equivalent volume of MMA buffer was used in place of the missing strain. Infiltrations were performed as in **section 2.1.3.3**.

4.5.5 – Extraction and analysis of leaf samples

Following leaf harvesting at 5 days, a 16mm (#11) cork borer was used to cut a single disk from each leaf. Disks were stored at -80°C before being lyophilised overnight. Disks were weighed using a microbalance (Sartorius) before transfer to a 1.5mL centrifuge tube. Next a saponification solution was made containing an internal standard (coprostanol, Sigma) to 10µg/mL. The saponification solution was added to each disk based on weight (100µL/mg) giving a final internal standard concentration of 1mg/g dw. Saponification was performed in a water bath at 75°C for 1 hour in sealed tubes. Tubes were then opened and incubated for an additional hour at 75°C to evaporate ethanol. Tubes were removed from the water bath and allowed to cool before addition of 300µL ethyl acetate and vortexing briefly. Finally, 500µL of water was added before vortexing again. Samples were centrifuged for 1 minute at full speed (16,000g) before removal of the upper organic phase. This was transferred to a 2mL autosampler vial and dried under N_2. 50µL Tri-Sil Z (Sigma) was added per sample and heated at 70°C for 30 minutes before addition of an equal volume of ethyl acetate. GC-MS was performed as in **section 2.1.5.1**.

4.5.5.1 – Comparing extraction methods for investigating transesterification

Two leaf disks from a single leaf expressing *tHMGR/AsbAS1/CYP93E1* were taken using a cork borer. One was saponified (without use of an internal standard) and extracted as per the method detailed above (**section 4.5.5**). The other was immersed directly into 500µL ethyl acetate in a 1.5mL microcentrifuge tube and shaken at 1000rpm at room temperature for 2 hours.

4.5.6 – Production, analysis and purification of 11-oxo,24-hydroxy-β-amyrin and 11-oxo,24-acetoxy-β-amyrin.

Large volume (~1.5L) cultures of strains of *A. tumefaciens* carrying *tHMGR, AsbAS1, CYP88D6* and *CYP93E1* were prepared (**section 2.1.3.2**) and used to infiltrate 112 plants by vacuum infiltration (**section 2.1.3**). These were harvested after five days (154g fresh weight) to give 18.5g dry leaves. Extraction was performed as in **section 2.1.4.** The crude extract was adsorbed onto silica gel and loaded onto a SNAP KP Sil 25g column (Biotage). Flash chromatography was performed according to program P-3 (**Table 2-4**). Fractions were analysed by GC-MS and those expected to contain 11-oxo,24-hydroxy-β-amyrin were pooled. The fractions containing the unknown compound were also pooled. Fractions containing 11-oxo,24-hydroxy-β-amyrin were dry loaded onto silica gel and a second chromatography step was performed using a SNAP Ultra 10g column (Biotage) (Program P-4). The absorption at 256nm was used to direct fraction collection of 11-

oxo,24-hydroxy-β-amyrin. Once dried, these gave a white crystalline powder which weighed 41mg.

For 11-oxo-24-acetoxy-β-amyrin, pooled fractions were dry loaded in silica gel onto a SNAP Ultra 10g column (Biotage) and run according to program P-5. All fractions were collected, and GC-MS and TLC analysis was used to determine those containing 11-oxo-24-acetoxy-β-amyrin.

4.5.6.1 – LC-MS analysis of 11-oxo,24-acetoxy-β-amyrin

Analysis of the semi-purified 11-oxo,24-acetoxy-β-amyrin was performed using a Shimadzu Nexera X2 installed with a Kinetex 2.6µ XB-C18 100Å 20 x 2.1mm column. A mobile phase of methanol and water (with 0.1% formic acid) was run (300µl/min) from 70-100% methanol over 17 minutes. The sample was monitored by MS (Shimadzu LC-2020 dual source MS), PDA (Shimadzu SPD-M20A) and CAD (Thermo Dionex Corona Veo RS).

4.5.6.2 – Saponification of 11-oxo, 24-acetoxy-β-amyrin

Approximately 100µg of the crude extract containing 11-oxo,24-hydroxy-β-amyrin was dissolved in 500µL of saponification mix and saponified as in **section 4.5.5.** TLC analysis was performed using a mobile phase of hexane:ethyl acetate (1:1) and visualised as previously described in **section 3.4.9.1.**

4.5.7 – Production and purification of 11-oxo-oleanolic acid

Large volume (~1.5L) cultures of strains of *A. tumefaciens* carrying *tHMGR, AsbAS1, CYP716A12* and *CYP88D6* were made (**section 2.1.3.2**)**.** This suspension was used to inoculate 141 plants by vacuum infiltration (**section 2.1.3**). These were harvested after 5 days (702g fresh leaves) and lyophilised giving 70g dry material. Extraction of the compound was performed (**section 2.1.4).** The extract was adsorbed onto silica gel and loaded onto a Biotage SNAP Ultra 50g column. Chromatography was performed according to program P-6 (**Table 2-4**). Fractions were analysed by GC-MS, those containing the products of interest were pooled and dried. This was further dry-loaded in silica gel onto a Biotage SNAP10 ultra cartridge and chromatography performed according to program P-7 (**Table 2-4**). Finally the extracts were dry loaded onto another SNAP10 ultra cartridge (Biotage) and run according to program P-8. This gave 46mg of a semi-pure extract which was further purified by HPLC as below.

4.5.7.1 – Semi-preparative HPLC purification of 11-oxo-oleanolic acid.

Semi-preparative HPLC was performed using a Shimadzu Nexera X2 system installed with a Phenomenex Luna 5µm C18 100Å 250x10mm column. The mobile phase was a methanol/water mix with 0.1% formic acid. A 4mL/minute gradient was set up from 70-100% methanol over 14mins and held for a further 6 minutes. The gradient was returned back to 70% methanol over 1 minute before equilibration over a further 4 minutes. 11-oxo-oleanolic acid typically eluted between 14-15mins. Fraction collection was directed by PDA monitoring absorbance at 252nm.

4.5.7.2 – Accurate mass measurement of 11-oxo-oleanolic acid

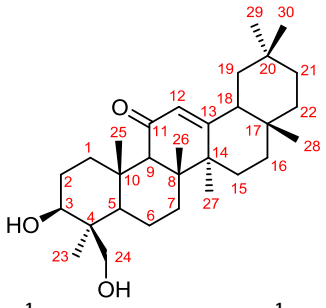
HPLC-purified 11-oxo-oleanolic acid was made to a 10µM solution in methanol. The sample was subject to flow injection analysis using a Shimadzu-IT-TOF. Analysis was performed by Dr Lionel Hill using ESI in positive mode (4.5kV). The nebulising gas flow was set to 1.5L/min, CDL temperature was 250°C and heat block was set to 300°C. The instrument was calibrated using trifluoroacetate according to the manufacturer's instructions.

4.5.8 – NMR spectroscopy

NMR analysis was performed by Dr Michael Stephenson. Spectra were recorded in fourier transform mode using a 400MHz nominal frequency (^{1}H NMR) and 100MHz (^{13}C NMR). Chemical shifts (δ) are recorded in parts per million (ppm), referenced to either a residual solvent peak or the tetramethylsilane internal standard as indicated. Coupling constants are reported in hertz. **Abbreviations**: br, broad; s, singlet; d, doublet; m, multiplet; dd, doublet of doublets; td, triplet of doublets; dt, doublet of triplets; br, broad; gem, geminal; ovlp, overlapping.

4.5.8.1 – 11-oxo,24-hydroxy-β-amyrin

Prior to NMR, semi-preparative HPLC purification of 4.2mg 11-oxo, 24-hydroxy-β-amyrin was performed by Dr Michael Stephenson using the same method as in **section 4.5.7.1**. NMR was performed in deuterated pyridine (pyridine-*d*5) and referenced to the most downfield residual solvent peak (^{1}H δ 8.74, ^{13}C δ 150.35). Full assignment of the compound is given in **Table 4-5**.



Atom number	^{13}C δ (ppm)	^{13}C Multiplicity	^{1}H δ (ppm)	Integration	^{1}H Multiplicity	J_{HH} (Hz)	COSY	HMBC
11	200.03	C	/	/	/	/	/	/
13	170.62	C	/	/	/	/	/	/
12	128.88	CH	5.80	1	s	/	/	9, 14, 18
3	80.33	CH	3.70	1	dt	11.8, 4.4	2, COH	/
24	64.98	CH_2	[3.78] [4.78]	[1] [1]	[dd] [dd]	[10.5, 8.4] [10.8, 1.9]	gem, C24-OH	3, 23
9	62.67	CH	2.54	1	s	/	/	8, 10, 11, 25, 26
5	56.43	CH	1.01	1	m (ovlp)	/	6	/
18	48.17	CH	2.14	1	m (ovlp)	/	/	/
8	46.08	C	/	/	/	/	/	/
19	45.75	CH_2	[1.67] [0.99]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	/	/
14	44.09	C	/	/	/	/	/	/
4	44.02	C	/	/	/	/	/	/
1	40.08	CH_2	[3.23] [1.22]	[1] [1]	[dt] [m (ovlp)]	[13.3, 3.5] [/]	[2] [/]	/
10	37.86	C	/	/	/	/	/	/
22	37.18	CH_2	[1.42] [1.23]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	/	/
21	35.03	CH_2	[1.40] [1.11]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	/	/
7	33.75	CH_2	[1.64] [1.33]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	/	/
29	33.52	CH_3	0.91	3	s	/	/	19, 20, 21, 30
17	32.97	C	/	/	/	/	/	/
20	31.55	C	/	/	/	/	/	/
28	29.21	CH_3	0.85	3	s	/	/	16, 17, 18, 22
2	28.94	CH_2	[2.22] [1.98]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	[3] [3]	/
15	27.13	CH_2	[1.71] [1.07]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	/	/
16	26.96	CH_2	[2.01] [0.86]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	/	/
23	24.08	CH_3	1.60	3	s	/	/	3, 4, 5, 24
30	23.95	CH_3	0.87	3	s	/	/	19, 20, 21, 29
27	23.95	CH_3	1.36	3	s	/	/	8, 13, 14, 15
26	19.17	CH_3	1.13	3	s	/	/	7, 8, 9, 14
6	18.82	CH_2	[1.77] [1.52]	[1] [1]	[m (ovlp)] [m (ovlp)]	/	[/] [5]	/
25	17.79	CH_3	1.36	3	s	/	/	1, 5, 9, 10
COH	/	/	6.73	1	d	4.7	3	/
C24-OH	/	/	5.39	1	dd	8.2, 2.2	24	/

Table 4-5: NMR assignments for 11-oxo,24-hydroxy-β-amyrin in pyridine *d5*. Carbon atom numbering is as detailed in the structure above the table. Abbreviations are as detailed above in **section 4.5.8** and specific details of the method are detailed in **section 4.5.8.1**

4.5.8.2 – 11-oxo-oleanolic acid

11-oxo-oleanolic acid was analysed in deuterated DMSO (DMSO *d*6) – referenced to the residual solvent peak at 2.50ppm. ^{1}H NMR spectrum was consistent with previously-reported literature [185]. δ 12.40 (1H, brs), 5.44 (1H, s), 4.31 (1H, d, *J*= 5.1), 3.01 (1H, td, *J*=11.4, 5.1), 2.85 (1H, dd, *J*= 13.7, 4.1), 2.60 (1H, dt, *J*= 13.4, 3.3), 2.30 (1H, s), 2.14-0.69 (18H, m) 1.34 (3H, s), 1.01 (3H, s), 0.91 (3H, s), 0.90 (3H, s), 0.89 (3H, s), 0.87 (3H, s), 0.68 (3H, s). Abbreviations are as detailed in **section 4.5.8.**

Chapter 5:

Investigating structural features underlying triterpenoid bioactivity

Acknowledgements:

The author performed purification of the compounds and the antiproliferative assays on HL60s. Antiproliferative assays on MCF-7, M202 and MIA PaCa-2 cells were performed by Margo Hollamby and Roderick Li at the University of East Anglia. The ELISA assays were performed by the author with assistance from Tyler Wooldridge. All data interpretation was performed by the author.

5.1 – Introduction

Natural products have proven to be highly important for development of pharmaceutical products [186]. Triterpenes have a long history of medicinal use [187] and are believed to be the major bioactive constituents of medicinal plants such as ginseng [44] and liquorice [47, 180]. Simple triterpenes are also abundant in commonly consumed fruits and herbs [188] and have a diverse range of activities, showing particular promise as anti-cancer and anti-inflammatory agents [48, 189].

Inflammation is an important physiological process that occurs in response to tissue damage and infection. This process involves the localised release of cytotoxic components such as reactive oxygen species (ROS) [190, 191]. These cytotoxins are beneficial for killing potential pathogens; however they can also cause tissue damage. Consequently, inflammation is typically a transient process that subsides to allow for subsequent tissue repair. Prolonged aberrant inflammation is a hallmark of a number of diseases including arthritis and asthma, and can be a causative factor in the development of malignancies [192, 193]. Thus therapeutics aimed at targeting both the inflammatory processes and the cancer cells themselves are a promising avenue for drug development. Triterpenes have frequently been shown to down-regulate pro-inflammatory targets and upregulate cellular detoxification mechanisms, as well as showing selective toxicity towards cancer cells. These compounds therefore offer significant promise as a safe and effective class of therapeutic agents for the treatment of inflammatory disorders and malignancies [48, 50, 187, 189].

Most naturally occurring triterpenoids are not considered potent enough to be deployed therapeutically. Hence there is a need to understand the structural features underlying bioactivity and improve upon the therapeutic potential of these molecules. Although some efforts have been made in this area [40, 194, 195], detailed understanding is lacking. The chemical complexity and vast structural diversity of naturally occurring triterpenes exacerbates this problem and the lack of availability of suites of closely related structural variants has hampered a systematic approach. Currently, the most detailed understanding of structure-function relationships for triterpenes has resulted from synthetic derivatisation of common starting scaffolds. For example the synthetic triterpene 2-cyano-3,12,-dioxooleana-1,9(11)-dien-28-oic acid (CDDO - hereafter referred to by its generic name bardoxolone) was developed from the naturally occurring β-amyrin-derived triterpene oleanolic acid, which has weak anti-inflammatory and anti-cancer activity,

using a synthetic chemistry approach [57] (**Figure 5-1**). Bardoxolone and other related synthetic analogues were found to be highly potent anti-inflammatory agents capable of upregulating the 'phase 2' antioxidant response [196] and inhibiting pro-inflammatory mediators, including nuclear-factor-kappa-B (NFκB) [197].

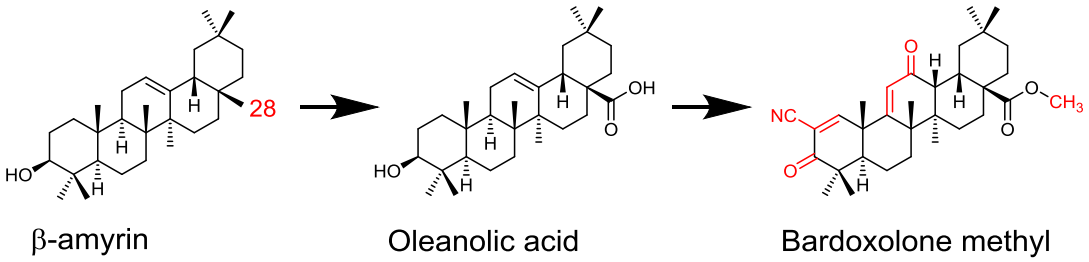


Figure 5-1: Synthetic triterpenes from β-amyrin. The structures of the simple triterpene scaffold β-amyrin, its C-28 acid (oleanolic acid) and the synthetic derivative bardoxalone methyl are shown. The synthetic modifications to oleanolic acid are highlighted in red.

Stage III clinical trials of bardoxolone methyl for the treatment of chronic kidney disease and type 2 diabetes were ultimately terminated due to an increased risk of heart failure. Later analyses determined a key subpopulation of patients who were 'at risk', including those who had previously been treated for heart conditions [198, 199]. This allowed future trials to avoid this group of patients. Further phase II trials in this area are now underway in Japan (ClinicalTrials.gov identifier - NCT02316821). Bardoxolone also shows potential as a treatment for pulmonary arterial hypertension with phase II trials currently underway in the US (NCT02036970). The clear potential of bardoxolone methyl has prompted significant interest in the further development of triterpenes as therapeutics.

Considerable advances have been made in our understanding of triterpene biosynthesis over the last few years, including for several medicinally important plant species [88, 98, 100, 103]. This provides a unique opportunity to utilise these enzymes to produce and selectively functionalise a range of structurally distinct scaffolds in a manner which is currently difficult or impossible using synthetic chemistry. In Chapter 4, triterpene oxidases from various plant species were expressed in *N. benthamiana* to give a series of oxidised derivatives of the common triterpene scaffold, β-amyrin. This enzymatic toolkit opens up opportunities to deconstruct the complexity of naturally occurring triterpenes and build a systematic

understanding of the structural features underlying bioactivity as well as providing novel scaffolds for chemical derivatisation.

5.2 – Aims

Many triterpenoids are known to have anti-inflammatory and anti-proliferative activity. However the structural features that govern these activities are poorly understood. Recent advances in understanding the biosynthesis of triterpenoids in plants provides a means to produce these molecules in heterologous hosts. This offers a means to obtain previously inaccessible structures and build a systematic understanding of the features underlying activity. To take a first step towards this goal, in this chapter a suite of structurally related analogues of oxidised β-amyrin are generated by heterologous expression in *N. benthamiana* and purified in milligram quantities. The bioactive properties of these compounds are then compared for anti-proliferative and anti-inflammatory activity using a range of human cancer cell lines.

5.3 - Results and discussion

5.3.1 - Obtaining a series of oxidised β-amyrin derivatives

In Chapter 4, co-expression of β-amyrin synthase in combination with different β-amyrin oxidases (P450s) in the *N. benthamiana* transient expression system was used to generate a series of β-amyrin derivatives that differed in the type and position of oxidation. This suite of P450s allowed modification of the β-amyrin scaffold at six different positions (**Figure 5-2**, centre). The products included oleanolic acid and other oxygenated forms of β-amyrin. The triterpene biosynthetic toolkit established in the Osbourn lab also contains enzymes enabling production of other structurally diverse triterpenes including protopanaxadiol, a bioactive triterpene produced by *Panax* (ginseng) species (**Figure 5-2**, right). This enzyme toolkit provides an invaluable resource for generating structural analogues, enabling for the first time a systematic investigation of structure-activity relationships within the triterpenes.

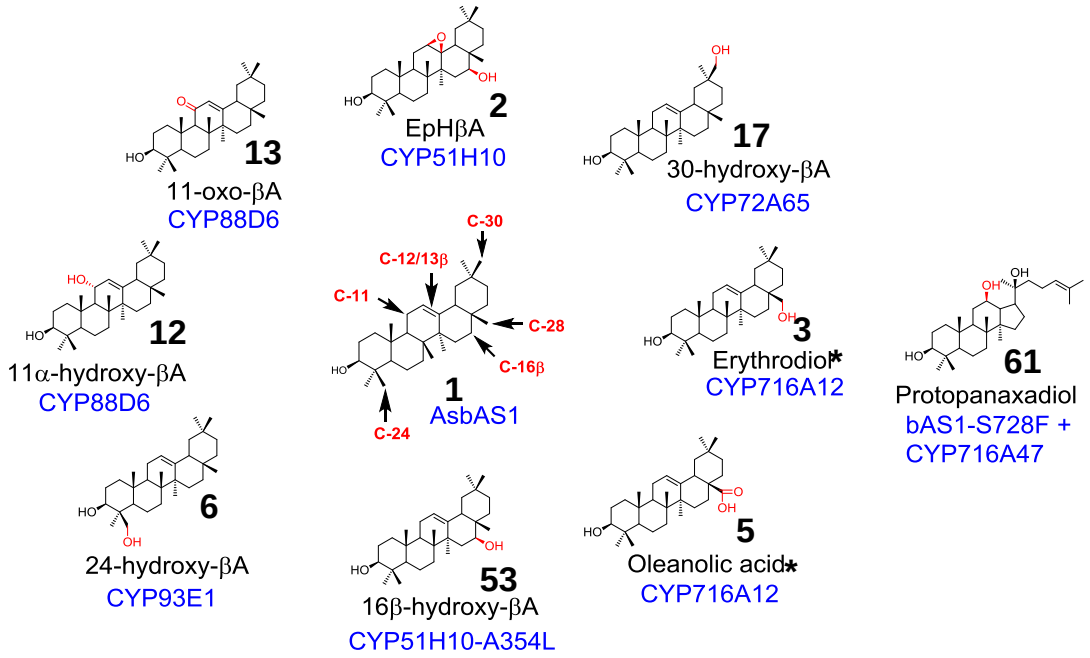


Figure 5-2: Structures of oxidised forms of β-amyrin investigated in this study. The β-amyrin scaffold is shown in the centre with carbons that were shown to be oxidised by P450s in Chapter 4. The non-β-amyrin-based ginsenoside triterpene dammarenediol-II (61) was also included in this study (shown on the right. The numbering of the compounds is as in Chapter 4. ***** These compounds were purchased commercially but are known to be products of CYP716A12 [97] (section 4.3.1.1). **Abbreviations:** βA – β-amyrin.

Based on the activity of the P450s characterised in Chapter 4, a set of eight oxidised forms of β-amyrin (**1**) were selected for investigation (**Figure 5-2**). These compounds included the product of the CYP51H10 enzyme, 12,13β-epoxy,16β-hydroxy-β-amyrin (EpHβA (**2**)); the CYP716A12 products, the C-28 alcohol (erythrodiol (**3**)) and carboxylic acid (oleanolic acid (**5**)); the C-24-hydroxylated (**6**) and C-30-hydroxylated (**17**) forms of β- amyrin, produced by CYP93E1 and CYP72A65, respectively; and two C-11-oxidised forms of β-amyrin, 11-α-hydroxy-β-amyrin (**12**) and 11-oxo-β-amyrin (**13**), which are produced by CYP88D6. Finally, the A354L mutant form of CYP51H10 was selected to allow production of the C-16β-hydroxylated form of β-amyrin (**53**) (**Figure 5-2**). A full list of the compounds and constructs used to generate them is given in **Table 5-1**.

#	Compound	Constructs	# Plants (DW)	Yield (total)	Yield (per g DW)
1	β-Amyrin	Purchased	-	-	-
2	EpHβA	*tHMGR/AsbAS1/ CYP51H10*	70 (20.4 g)	79 mg	3.87 mg
3	Erythrodiol	Purchased	-	-	-
5	Oleanolic acid	Purchased	-	-	-
6	24-Hydroxy-β-amyrin	*tHMGR/AsbAS1/ CYP93E1*	17 (3 g)	4.7 mg	1.56 mg
12	11α-Hydroxy-β-amyrin	*tHMGR/AsbAS1/ CYP88D6*	23 (10.4 g)	7 mg	0.67 mg
13	11-Oxo-β-amyrin	*tHMGR/AsbAS1/ CYP88D6*	23 (10.4 g)	10.4 mg	1 mg
17	30-Hydroxy-β-amyrin	*tHMGR/AsbAS1/ CYP72A65*	30 (19 g)	6.5 mg	0.34 mg
37	11-Oxo, 24-hydroxy-β-amyrin	*tHMGR/AsbAS1/ CYP88D6/ CYP93E1*	112 (18.5 g)	41 mg	2.2 mg
45	11-Oxo-oleanolic acid	*tHMGR/AsbAS1/ CYP716A12/ CYP88D6*	141 (70 g)	46 mg	0.66 mg
53	16β-Hydroxy-β-amyrin	*tHMGR/AsbAS1/ CYP51H10-A354L*	45 (17.9 g)	6.6 mg	0.37 mg
61	Protopanaxadiol	*tHMGR/AsbAS1*-S728F/CYP716A47	90 (31 g)	18 mg	0.58 mg
63	Glycyrrhetinic acid	Purchased	-	-	-

Table 5-1: Triterpenes tested in the present study. Compounds are numbered as in Chapter 4. Where compounds were purified from *N. benthamiana,* the combination of *A. tumefaciens* strains (OSCs and P450s) are indicated under 'Constructs' (compounds 1, 3, 5 and 63 were purchased from commercial suppliers). The number of plants and purified yield of compound are provided here for quick reference. Further detail on purification methods of individual compounds is provided in the methods sections. Supplier information is given in section 2.1.6.1. Abbreviations: DW, Dry weight; EpHβA, 12,13β-epoxy,16β-hydroxy-β-amyrin.

With the exception of erythrodiol (**3**) and oleanolic acid (**5**), none of these oxidised forms of β-amyrin are available to purchase commercially, therefore these compounds were accessed by production in *N. benthamiana*. In Chapter 3, strategies to enhance the production capacity of *N. benthamiana* were explored. This demonstrated that *tHMGR* (a truncated form of the MVA pathway rate-limiting enzyme 3-hydroxy, 3-methylglutaryl-CoA reductase) could significantly increase yields of both simple and oxidised triterpenes in transient expression experiments in *N. benthamiana.* Using this strategy, over 70mg of pure EpHβA (**2**) was obtained by co-expression of tHMGR with the oat β-amyrin synthase (*AsbAS1*) and the CYP51H10 oxidase in *N. benthamiana* (section 3.2.5). This strategy was therefore employed to obtain sufficient quantities of the other oxidised derivatives for use in bioassays. Separate *A. tumefaciens* strains harbouring *tHMGR, AsbAS1* and individual P450 expression constructs were co-infiltrated into *N. benthamiana* leaves in order to generate the relevant compound (**Table 5-1**). This enabled purification of β-amyrin derivatives **6**, **12**, **13, 17** and **53** (**Figure 5-2**). Full details concerning infiltration and purification of individual products are given in **method section 5.5.1**. Verification of the products was performed by GC-MS analysis of the purified compounds **6**, **12**, **13, 17** compared to previously published spectra [98-100]. The structure of **53** has been verified by NMR analysis (Prof Robert Minto, personal communication). GC-MS chromatograms of the purified compounds versus the crude *N. benthamiana* extracts are given in **Supplementary Figure 22-Supplementary Figure 26**.

Finally, the tetracyclic triterpene dammarenediol-II (DM-II) was produced by expression of a mutant enzyme AsbAS1-S728F [137] (also see section 4.3.5). Conversion of this product to the ginsenoside triterpene protopanaxadiol **61** (**Figure 5-2** right) was then achieved by coexpression of this enzyme with *CYP716A47* [103] (section 4.3.5). Protopanaxadiol was purified from *N. benthamiana* leaves expressing *tHMGR/AsbAS1-S728F/CYP716A47* by Dr Michael Stephenson and Dr Ramesha Thimmappa, Osbourn lab. This compound was also included in the bioassays, giving a total of ten test compounds (**Figure 5-2**).

5.3.2 – Investigating anti-proliferative activity

A common means of assessing the activity of a test compound is to measure its ability to inhibit the growth of cells when incubated across a range of concentrations. The concentration at which a test compound inhibits cell growth by 50 percent relative to controls (IC_{50}) can subsequently be calculated. For structure-activity studies, comparison of IC_{50} values between compounds provides a simple means to

ascertain the relative potency of individual modifications. Therefore the anti-proliferative activities of the ten compounds shown in **Figure 5-2** were tested on a range of human cancer cell lines. The cell proliferation assay utilises MTS (5-[3-(carboxymethoxy)phenyl]-3-(4,5-dimethyl-2-thiazolyl)-5-[(phenylamino)-carbonyl]-2H-tetrazolium inner salt), a tetrazolium reagent that undergoes a colour change upon reduction to a formazan product. This reduction depends on NADH (or NADPH), which is produced by actively respiring cells. MTS reduction is facilitated by use of phenazine methyl sulphate (PMS) as an intermediate electron acceptor, since MTS has poor cellular permeability [200] (**Figure 5-3**). The formation of the formazan product can be measured by absorption at 490nm and provides a simple readout for ascertaining cell proliferation.

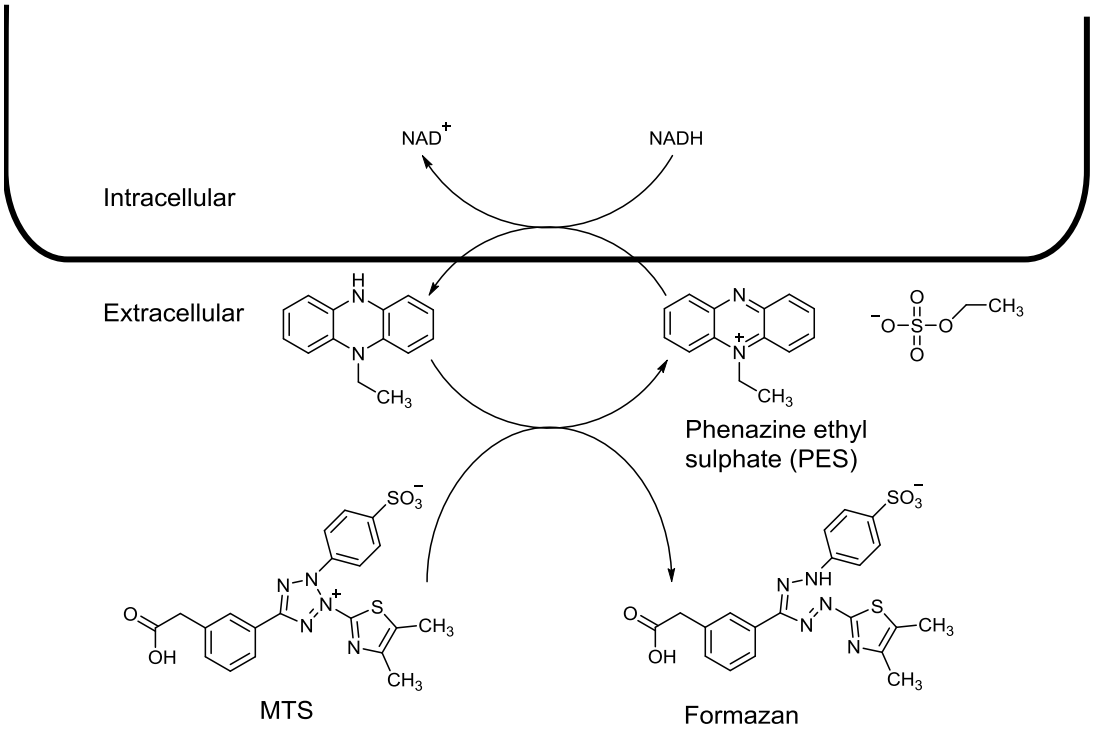


Figure 5-3: Schematic of the MTS assay, as used to detect actively respiring cells. The reduction of MTS by NADH (*via* PES) creates a formazan product that absorbs strongly at 490nm. Therefore the relative abundance of living cells can be quantified by measuring absorbance of the MTS reagent at this wavelength.

The anti-proliferative effects of each of the ten compounds were evaluated using four different human cancer cell lines. These included MCF-7 (breast adenocarcinoma); HL60 (promyelocytic leukaemia); MIA PaCa-2 (pancreatic cancer); and M202 (melanoma). Due to the lipophilic nature of the simple

triterpenes, compounds were dissolved in dimethyl sulphoxide (DMSO) to make the relevant stock and serial dilutions. A series of eight two-fold dilutions of each test compound were made with the aim of determining the IC_{50} values (**methods 5.5.2**). The synthetic triterpene bardoxolone methyl was included as a positive control, with the vehicle (DMSO) as a control. Cell lines were incubated for 72 hours in the presence of the test compounds at various concentrations prior to assessment of viability using the MTS assay (**methods 5.5.4**).

Analysis of the results revealed clear differences in the relative activities of the various triterpenes tested. Six of the test triterpenes, including β-amyrin, appeared to be completely inactive across the full range of concentrations tested. However the remaining four showed clear dose-dependent anti-proliferative activity for the four cell lines tested. These compounds included the ginsenoside protopanaxadiol (**61**); the β-amyrin C-28 alcohol erythrodiol (**3**); the C-28 carboxylic acid (oleanolic acid (**5**); and 11-oxo-β-amyrin (**13**), which features a C-11 ketone. Some precipitation was evident after 72 hours at the highest concentrations tested (generally 100-200μM) owing to the poor water solubility of the triterpenes. However this was observed for all of the test compounds and is unlikely to account for the differences in observed activity between the different compounds. A summary of the IC_{50} values for these compounds is provided in **Table 5-2**.

	M202	**MCF-7**	**MIA PaCa-2**	**HL60**
Erythrodiol (**3**)	19.1 ± 4.9	15.9 ± 1.8	17.4 ± 3.1	23.7 ± 4.9
Oleanolic acid (**5**)	44.6 ± 6.7	45.4 ± 3.8	50.5 ± 3.8	26.5 ± 6.7
11-oxo-β-amyrin (**13**)	86.8 ± 37.3	64.8 ± 8.4	59.7 ± 24.0	26.3 ± 1.1
Protopanaxadiol (**61**)	57.1 ± 34.8	48.5 ± 1.9	43.7 ± 13.7	27.9 ± 3.9
Bardoxolone methyl	4.5	2.4	4.8	0.7

Table 5-2: Antiproliferative effects of oxidised triterpenes. IC_{50} values (μM) are shown for the four cell lines. The data are means of at least three biological and technical replicates in each of the cell lines (± SD), with the exception of bardoxolone methyl, for which only a single biological replicate (three technical replicates) was included.

Erythrodiol (**3**) showed the strongest anti-proliferative activity for all four cell lines, with IC_{50} values in the region of 15-25 μM. Oleanolic acid (**5**) and protopanaxadiol (**61**) had similar IC_{50} values for each of the cell lines. 11-Oxo-β-amyrin (**13**) overall had the weakest anti-proliferative activity, with the exception of the HL60 cell line, where it was comparable with oleanolic acid. When comparing between cell lines, the four active compounds had similar IC_{50} values across the three adherent cell lines, M202, MCF-7 and MIA PaCa-2. Values were generally lower for the HL60 cells for all compounds except erythrodiol for which the HL60 IC_{50} value was

highest. Finally, as expected, the IC_{50} values for these test triterpenes were several fold higher than the bardoxolone methyl control which was found to compare well with previous results in the O'Connell lab. Bardoxolone is reported to have a growth inhibitory range between 0.1µM and 1µM in most cell lines with up to 5µM required for inducing apoptosis [187]. Therefore these observed values are slightly higher than might be expected, but still within in a similar range.

The IC_{50} values for the test triterpene compounds appear to compare well with previous studies where data was available. Oleanolic acid has previously been reported to have IC_{50} values of 28µM and 30µM after 72 hours incubation for MCF-7 and HL60 cells respectively [201]. Erythrodiol has been reported to have an IC_{50} value of 48.8µM after 72 hours in a human colon cancer cell line HT-29 [202] and 18.3µM after 48 hours in a bladder cancer cell line (NTUB1) [203]. Furthermore, a direct comparison between oleanolic acid and erythrodiol found that erythrodiol was generally more potent across a range of concentrations (0-100µM) between one and five days in MCF-7 cells [204].

There are relatively few studies examining the bioactivity of 11-oxo-β-amyrin. However this compound has been isolated from the tropical forest tree *Duabanga grandiflora* and found to have IC_{50} values between 35-66µM in six cancer cell lines including one leukaemia (MOLT-3) and five carcinoma lines (HuCCA-1, A549, HeLa, HepG2 and MDA-MB-231) [205].

Finally, for protopanaxadiol, IC_{50} values of 68.4 µM have been reported in MCF-7 lines [206]. Similar ranges (20-78 µM) were also observed across a range of cancers from various tissues [206].

5.3.3 – Utilising combinatorial biosynthesis to further investigate structure-function relationships

Amongst the β-amyrin derivatives that showed activity in the anti-proliferation assays, the presence of the C-11 ketone and C-28 alcohol/acid functionalities on the β-amyrin scaffold appear to be important for activity. The effects of combining both of these modifications on a single scaffold was therefore investigated. The pairwise combination of *CYP88D6* (C-11 oxidase) and *CYP716A12* (C-28 oxidase) results in the production of 11-oxo-oleanolic acid (**45**) (section 4.3.2.3), which features both the C-11 ketone and the C-28 carboxylic acid (**Figure 5-4**). This compound is not commercially available, but had been purified from *N. benthamiana* as part of work for Chapter 4. A positional isomer of 11-oxo-oleanolic acid can be purchased in the form of glycyrrhetinic acid (**63**). Glycyrrhetinic acid is one of the

active constituents of liquorice [47] and features the carboxylic acid at the C-30 position instead (**Figure 5-4**). Hence glycyrrhetinic acid was also tested in parallel (**Table 5-1**).

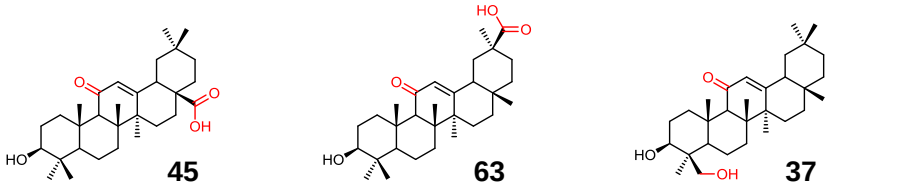


Figure 5-4: Additional products tested for antiproliferative activity. These compounds are the result of the action of two P450s (blue) on β-amyrin. The numbering of the compounds corresponds to the numbering scheme used in Chapter 4. Compounds **45** and **37** were produced in *N. benthamiana* in Chapter 4. ***** compound **63** was commercially purchased, but is known to be synthesised by the consecutive action of CYP88D6 and CYP72A154 [100]. **Abbreviations:** βA – β-amyrin. Species abbreviations are as listed in **Figure 5-1**.

Both 11-oxo-oleanolic acid and glycyrrhetinic acid showed activity across each of the four cell lines. However, in most cases the IC_{50} values for these compounds were slightly lower than any of the four active compounds tested initially (**Table 5-3**). These tests were conducted at a similar time to the previous tests and are therefore expected to be directly comparable. Between the two, 11-oxo-oleanolic acid was less active than glycyrrhetinic acid in the four cell lines (**Table 5-3**). As above, IC_{50} values were generally lower in HL60 cells versus the other three lines. This could be evidence of selectivity towards these cells, however given that this trend was consistent for most compounds; it may simply reflect a higher sensitivity of these cells. This is consistent with previous observations in the O'Connell lab.

In Chapter 4, combinatorial biosynthesis in *N. benthamiana* also afforded the isolation of a second 11-oxo-β-amyrin derivative in the form of 11-oxo-24-hydroxy-β-amyrin (**37**) (See **Figure 5-4** and **Table 5-1**). This compound was therefore also tested against the HL60 cell lines as before. No antiproliferative activity was detected for this compound over the 72 incubation period (**Table 5-3**). Together, this suggests that while the presence of the ketone at C-11 can influence antiproliferative activity over β-amyrin alone, this is apparently dramatically influenced by the type and position of other modifications to the scaffold.

	M202	MCF-7	MIA PaCa-2	HL60
11-oxo-oleanolic acid (**45**)	85.4 ± 13.3	154.4 ± 41.6	103.2 ± 16.3	30.6 ± 4.2
18 β-glycyrrhetinic acid (**63**)	59.2 ± 10.3	99.6 ± 21.3	75.0 ± 18.8	25.7 ± 2.9
11-oxo,24-hydroxy-β-amyrin (**37**)	NT	NT	NT	>200

Table 5-3: Antiproliferative activity of oxidised triterpenes. IC_{50} values (μM) ± standard deviation for additional test compounds (**Figure 5-4**). Results are average of at least three biological and technical replicates in each of the cell lines. NT = Not tested.

Glycyrrhetinic acid (**63**) has previously been reported to have an IC_{50} of 38μM and 25μM in HL60s after over 24 and 48 hours, incubation respectively [207] which compares well to the current values. It has also been reported to have an IC_{50} of 32.4μM in MCF-7 cells after 48 hours, which is lower than observed here [208]. However other studies have reported an IC_{50} activity of 84.7μM in MCF-7 cells using a sulphorodamine B cytotoxicity assay after 96 hours in MCF-7 cells (and similar values in 14 other lines) [209]. Although this assay differs from the formazan-based assay as used here, these techniques are generally comparable when used as an end-point assay [210].

There appear to be few studies that have investigated the anti-proliferative activity of 11-oxo-oleanolic acid (**45**). Assefa et al noted that both 11-oxo-oleanolic acid (**45**) and oleanolic acid (**5**) had comparable IC_{50} values towards human melanoma SK-MEL cells (average 131μM and 112μM, respectively) [211]. In the present study, the activity of 11-oxo-oleanolic acid (**45**) was comparable to oleanolic acid only for the HL60 line. In the other cell lines, 11-oxo-oleanolic acid (**45**) tended to be much weaker than oleanolic acid (**5**) in all cell lines including the M202 melanoma cell line.

5.3.4 – Investigating the anti-inflammatory effects of oxidised β-amyrin compounds.

The above investigation showed that simple modifications to the β-amyrin scaffold could have profound influences on the anti-proliferative activity of the molecules. A number of naturally occurring and synthetic β-amyrin derivatives have been demonstrated to have anti-inflammatory activity. Therefore this prompted an investigation of the potential anti-inflammatory properties of these compounds.

Inflammation is triggered in response both to tissue damage and to the presence of foreign particles or organisms, including ‘pathogen-associated molecular patterns’

(PAMPS) such as the bacterial endotoxin lipopolysaccharide (LPS) [190]. Monocytes serve a key role in detecting pathogens and coordinating the inflammatory response. LPS triggers a pro-inflammatory intracellular signalling cascade that culminates in release of pro-inflammatory cytokines, such as tumour necrosis factor alpha (TNFα). These cytokines are responsible for local recruitment and activation of macrophages and neutrophils to combat the immediate threat, as well as exerting systemic effects [190]. Thus one method of gauging the anti-inflammatory potential of a test compound is to assess its ability to attenuate cytokine production in response to a pro-inflammatory stimulus. To this end, an enzyme-linked immunosorbent assay (ELISA) was used to measure the production of TNFα by THP-1 cells (**methods 5.5.5.1**). THP-1 cells are a human monocytic cell line that serve as a useful *in vitro* model system for this purpose as they have comparable behaviour to primary monocytes and rapidly respond to LPS stimulation [212].

The compounds that had been tested for anti-proliferative activity in sections **5.3.2** and **5.3.3** were then evaluated using the TNFα ELISA assay in order to ascertain their ability to suppress TNFα secretion in LPS-stimulated THP-1 cells. The cells were pre-incubated in the presence of 100 μM of each compound (or 100-200 nM bardoxolone methyl) for 30 minutes prior to treatment with LPS, and then incubated for a further three hours. The supernatants were frozen and analysed for TNFα by ELISA (**methods 5.5.5**).

The compounds that were found to be inactive in the anti-proliferative assays (**1**,**2**,**6**,**12**,**17**,**53**) similarly showed little or no anti-inflammatory activity, with levels of TNFα production comparable to vehicle-only treated controls (**Supplementary Figure 27**). Amongst the compounds that had previously shown anti-proliferative activity, 11-oxo-β-amyrin (**13**), protopanaxadiol (**61**) and glycyrrhetinic acid (**63**) dramatically alleviated TNFα production at 100μM. 11-oxo-oleanolic acid also produced a consistent reduction in TNFα relative to the vehicle-only control, but was less effective than the other compounds. In contrast, erythrodiol did not appear to attenuate TNFα production (**Figure 5-5**). None of the compounds tested here appeared to be pro-inflammatory at 100μM, since TNFα was not detected in non-LPS-treated samples (data not shown).

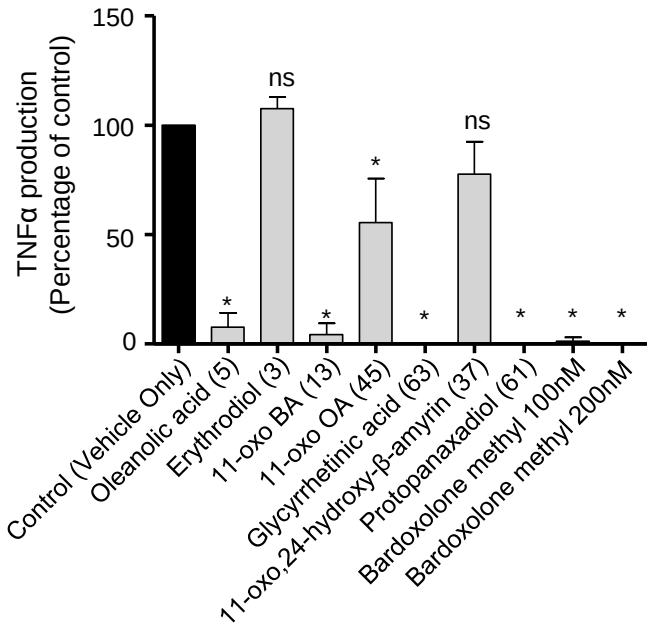


Figure 5-5: TNFα production in LPS-stimulated THP-1 cells incubated with triterpenes. Cells were pre-incubated with 100 µM of each of the test compounds prior to LPS stimulation. Bardoxolone methyl was used at concentrations of 100 or 200 nM as indicated. TNFα production was quantified by ELISA after 3 hours and is shown relative to DMSO (vehicle)-treated control cells. Values are means of three biological replicates except oleanolic acid (**5**) and 11-oxo,24-hydroxy-β-amyrin (**37**), which due to time constraints show three technical replicates. Error bars show standard deviation. * denotes significant difference compared to the control group ($p<0.05$); ns, not significant.

Work is currently ongoing to establish the dose-dependent anti-inflammatory response of the active triterpenes shown here. This will allow the relative potency of the anti-inflammatory compounds to be tested. Initial tests using protopanaxadiol (**61**) and glycyrrhetinic acid (**63**) have suggested that the concentration needed to alleviate the TNFα production by 50% is 47.6µM and 55.5µM respectively **(Figure 5-6**). As demonstrated here, bardoxolone methyl is clearly inhibitory at 200nM, which demonstrates as expected that these simple triterpenes are significantly less potent than bardoxolone-methyl. However this analysis provides a useful benchmark with which to compare further compounds.

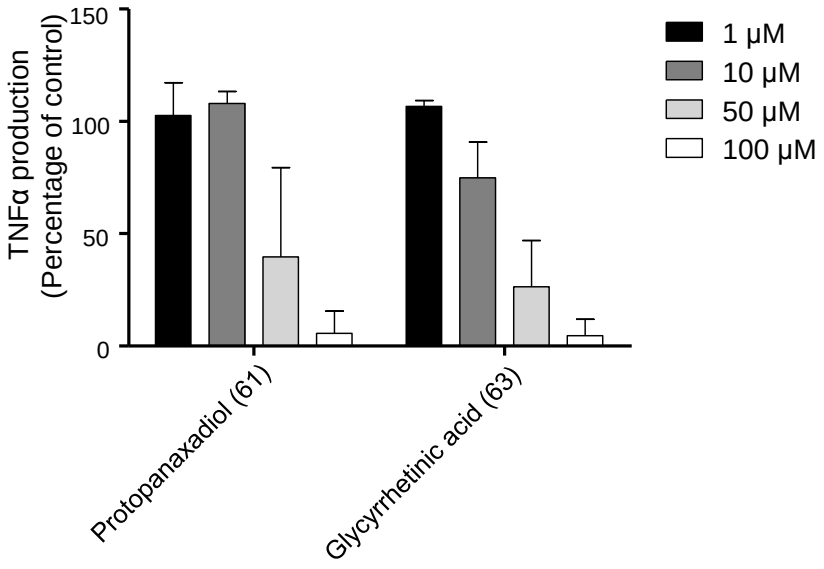


Figure 5-6: Dose-dependent TNFα production following incubation with different triterpene concentrations**.** THP-1 cells following LPS treatment. Cells were preincubated with concentrations of Protopanaxadiol and glycyrrhetinic acid from 1-100μM as indicated. TNFα production was quantified after 3 hours and is shown relative to the DMSO (vehicle)-treated control cells. These data were used to calculate IC_{50} values.

5.3.5 – Results summary: insights into structural features underlying activity

When considering anti-proliferative activity, the results of the present study indicate that modifications at the C-28 position are important for activity, since both alcohol and acid functionalities at this position result in anti-proliferative activity. Modifications to the C-11 position also emerge as potentially important; while 11-oxo-β-amyrin (**13**) showed anti-proliferative activity, 11α-hydroxy-β-amyrin (**12**) (**Figure 5-2**) was inactive, suggesting that hydroxylation at this position is less favourable than the ketone group. Triterpenes with hydroxyl groups at the C-16β (**53**), C-24 (**6**) or C-30 (**17**) positions of β-amyrin were also inactive. Combinatorial products containing the C-11 ketone with either a C-30 (glycyrrhetinic acid) or C-28 carboxylic acid group were also anti-proliferative, although their activity appeared to be generally lower than the parent compounds in most cell lines.

The general importance of the ketone moiety at the C-11 position in pentacyclic triterpenes has not yet been resolved. Goncalvez et al tested the cytotoxicity of various derivatives of asiatic acid (an α-amyrin scaffold containing a C-28 acid) towards HeLa cells [213]. The cytotoxicity of certain compounds was increased upon incorporation of this functionality, but for others, a decrease in activity was observed compared to the parent scaffold [213]. In another study, a series of

oleanolic acid-derived triterpenoids isolated from the Southeast Asian tree *Styrax tonkinensis* were tested in HL60s. Of these, oleanolic acid (**5**) was found to be the most potent antiproliferative compound (IC_{50}~8.9µM). The presence of an additional C-6β hydroxyl group led to a moderate drop in activity (3-fold) while further presence of the C-11 ketone decreased antiproliferative activity more than 30-fold compared to oleanolic acid [214]. The C-11 ketone is present in glycyrrhetinic acid and has been suggested to be important for inducing oxidative stress in mitochondria resulting in membrane permeability transition and apoptosis [215]. Interestingly, 11-oxo-oleanolic acid (**45**) has been found to induce apoptosis, while oleanolic acid (**5**) did not [211]. Equally however, the select removal of this group from glycyrrhetinic acid (**63**) had little effect on the overall antiproliferative activity across 13 cell lines compared to the parent compound and resulted in comparable levels of apoptosis [216]. Hence ascertaining the cellular or molecular mechanisms underlying the antiproliferative activity of these compounds will be an important future task to further qualify the mechanisms behind structural modifications.

From an investigation of the anti-inflammatory activity, similar to the anti-proliferative assay, it was found that a β-amyrin scaffold featuring a C-11 ketone appears to be important. This was evident by the fact that three molecules tested here with this feature (11-oxo-β-amyrin (**13**), 11-oxo-oleanolic acid (**45**) and glycyrrhetinic acid (**63**)) all show some degree of anti-inflammatory activity. Glycyrrhetinic acid (**63**) is a well-known anti-inflammatory agent and a synthetic derivative has been marketed as carbenoxolone for the treatment of peptic ulcers. The biological activity of **63** is attributed to its ability to inhibit the activity of the 11β-hydroxy-steroid dehydrogenase 2 (11βHSD2) [47]. This enzyme is responsible for the inactivation of the steroidal hormone cortisol by conversion to cortisone, hence leading to higher levels of cortisol (**Figure 5-7**). Although this mechanism may be less relevant for *in vitro* cell cultures, glycyrrhetinic acid is also capable of binding directly to the mineralocorticoid receptor [47]. Accordingly, glycyrrhetinic acid was shown to inhibit LPS-induced production of TNFα in mouse RAW264.7 macrophages via the glucocorticoid receptor between 1-5µM [217]. Glucocorticoid signalling plays important physiological roles in reducing inflammation [218]. Hence it may be that the anti-inflammatory effects attributed to **13** and **45** likewise reflect an interaction with this signalling mechanism.

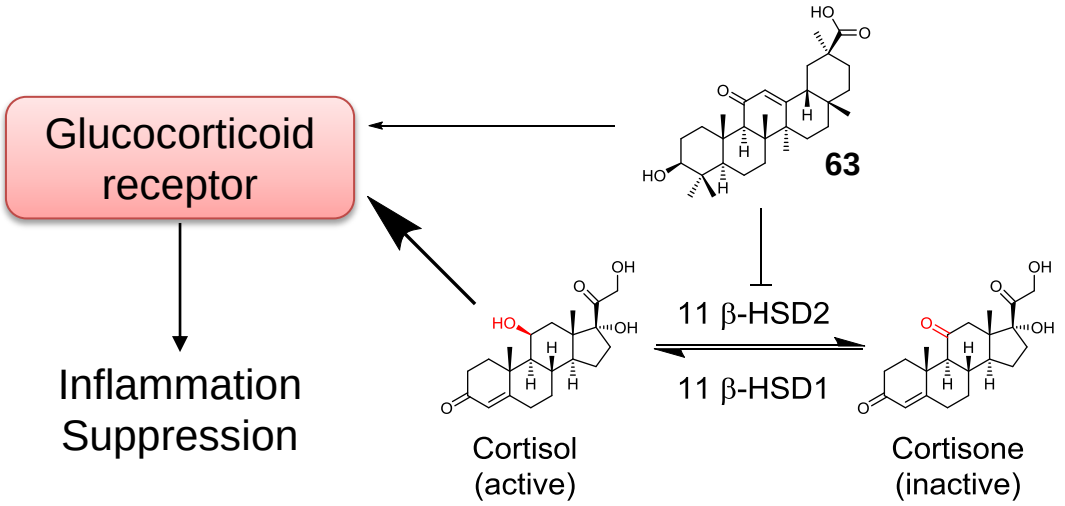


Figure 5-7: Mechanisms of glycyrrhetinic acid anti-inflammatory activity**.** Glycyrrhetinic acid (**63**) can inhibit the 11β-hydroxy steroid dehydrogenase 2 enzyme (11β-HSD) which is responsible for inactivation of cortisol. In addition, like cortisol it can also bind directly to the glucocorticoid receptor.

Other mechanisms are also plausible for the anti-inflammatory activity of these compounds. For example, the α,β-unsaturated ketone (enone) present in the C-ring might allow conjugate addition to thiols in cysteine residues within proteins (**Figure 5-8**). Several proteins use this as a redox-scavenging mechanism and amongst these is the Kelch-like ECH-associated protein 1 (Keap1). Keap1 is a negative regulator of the Nrf2 transcription factor, which is responsible for activating a number of 'phase 2 response' genes with antioxidant and detoxification functions [219]. Moreover, activation of Nrf2 also represses the activity of pro-inflammatory targets such as NFκB [219]. The Keap1-Nrf2 pathway has been heavily implicated in the function of bardoxolone-methyl with a critical role for enone functionalities [57, 196]. In future, the relevance of the conjugate addition mechanism could be investigated by incubation with glutathione [220].

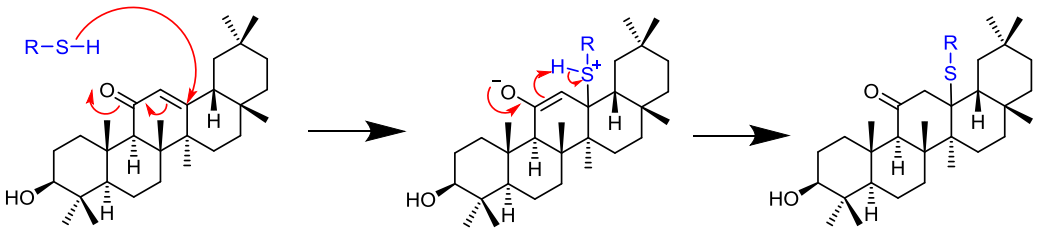


Figure 5-8: Conjugate addition of C-ring enone triterpenes. Derivatives of β-amyrin with an α,β-unsaturated ketone at C-11 may react with thiols present in cysteines as may be found in proteins or glutathione (indicated as R-group). Interaction with such groups on the surface of redox-sensitive proteins (such as Keap-1) may underlie the anti-inflammatory activity of these molecules by activating cellular detoxification mechanisms.

5.4 – Conclusions and perspectives

Understanding triterpene structure-activity relationships has been confounded by the huge structural diversity and complexity of triterpenoids isolated from natural sources. An alternative approach is the systematic modification of simple triterpenes using enzymes, allowing the importance of individual modifications to be tested directly.

To demonstrate this latter approach, here *N. benthamiana* was used to express triterpene oxidases in order to generate a series of structural analogues of oxidised forms of β-amyrin. These were screened for anti-proliferative and anti-inflammatory activity to gain insight into the structural features that are important for these processes. Two modifications emerged as potentially important – the presence of a ketone at the C-11 position, and a carboxylic acid at C-28. Forms of β-amyrin containing these modifications inhibited proliferation of a range of cells and showed the ability to alleviate TNFα production in THP-1 cells.

Combinatorial biosynthesis also allowed the effects of combining these modifications to be examined. Derivatives of 11-oxo-β-amyrin (**13**) containing carboxylic acids at C-28 (**45**), C-30 (**63**) or an alcohol at C-24 (**37**) were also examined. Derivatives of 11-oxo-β-amyrin with C-28 (**45**) or C-30 (**63**) carboxylic acids were also active in both proliferation and anti-inflammatory assays. However surprisingly incorporation of a C-24 hydroxyl group (**37**) appeared to completely abolish the antiproliferative activity observed for 11-oxo-β-amyrin, and similarly attenuated the anti-inflammatory response.

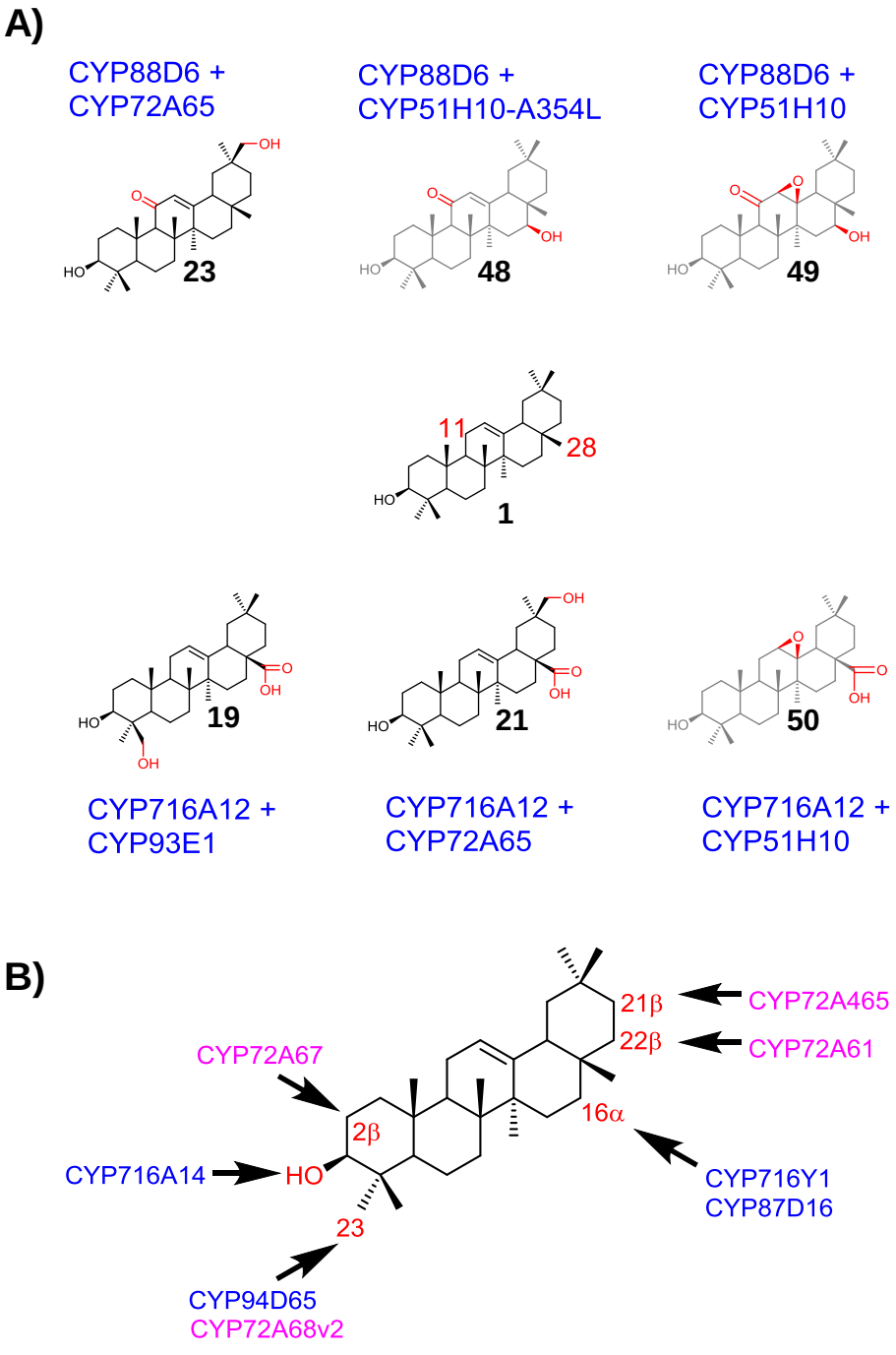


Figure 5-9: Possible future test molecules based on the results of this study. **A)** Compounds identified through combinatorial biosynthesis (Chapter 4) which are predicted to contain either the C-11 oxo or C-28 carboxylic acid functionalities. The P450s which resulted in these compounds are shown in blue. Numbering of compounds is as indicated in Chapter 4. **B)** Positions around the β-amyrin for which other oxidases (indicated in blue) are known. Enzymes indicated in magenta are those which do act directly on β-amyrin but require an oxidised form. CYP72A475 and CYP94D65 are detailed in Chapter 6. Further information on the other enzymes is detailed in **Table 1-1**.

It will be particularly interesting to investigate the activity of other products containing the C-11 or C-28 functionalities. Several such products emerged from the combinatorial biosynthesis work described in Chapter 4 (**Figure 5-9A).** Likewise, continued investigation into the effects of oxidation at other positions around the β-amyrin scaffold is an obvious next step. Within the enzymes investigated in Chapter 4, it was shown that oleanolic acid can be further oxidised with hydroxyls at the C-2β and C23 positions (**Supplementary Figure 18-19**). Similarly, in the following chapter, an investigation into avenacin biosynthesis in *A. strigosa* highlights additional P450s with putative C-21, C-23 and C-30 oxidase activity. Other characterised enzymes which were not tested here should also allow other positions to be investigated (**Figure 5-9B**). As more β-amyrin oxidases continue to be characterised from other species, extensive targeted oxidation of this backbone will be possible.

Furthermore, a number of β-amyrin oxidases have previously been shown to be active on other triterpene backbones [95, 97, 108]. Hence this opens the opportunity to extend this approach to other scaffolds. Indeed here protopanaxadiol was shown to be one of the most active molecules in both the anti-proliferative and anti-inflammatory assays. Therefore the importance of the type of triterpene scaffold should be further investigated.

N. benthamiana has proven to be an excellent system for the generation of triterpene analogues. As the molecular pathways underlying triterpene biosynthesis in the plant kingdom continue to be uncovered, an exciting array of previously-inaccessible triterpenoids can be realised in this system. This approach holds the potential to greatly improve our understanding of structure-function relationships underlying triterpene bioactivity and to direct biosynthesis for production of drugs and other useful compounds.

5.5 – Materials and Methods for Chapter 5

5.5.1 – Production and purifications of compounds from *N. benthamiana*.

For all compounds, plants were infiltrated by hand (**section 2.1.3.3**) with strains of *A. tumefaciens* carrying *tHMGR*, *AsbAS1* and p450s as detailed in **section 2.1.3.2**. The *A. tumefaciens* strains carried the relevant genes within the pEAQ-*HT*-DEST1 expression vector as detailed **in section 4.5.1** with the exception of strains carrying *CYP88D6* and *CYP93E1*. Further information on these strains is given in **section 5.5.1.1.** The infiltrated solution had a total OD_{600} of 0.6, giving an OD_{600} of 0.2 per strain. After 5 days, infiltrated leaves were harvested, processed and crude extracts were made as per **section 2.1.4**. The extracts were dissolved in toluene and loaded onto the specified column and run according to programs indicated in the relevant sections below. Where further techniques were performed such as decolourisation and/or recrystallization, these are detailed in **sections 2.1.4.4** and **2.1.4.4.**

5.5.1.1 – Golden Gate vectors carrying *CYP88D6* and *CYP93E1*

The strains carrying the *CYP88D6* and *CYP93E1* vectors were created prior to and later replaced by those detailed in **section 4.5.1.1**. These constructs were created using a Golden Gate cloning strategy by Dr Aymeric Leveau according to Weber et al [135]. Briefly, the *CYP88D6* and *CYP93E1* genes were synthesised by the Engineering Nitrogen Symbiosis for Africa (ENSA) project (https://www.ensa.ac.uk/) with silent mutations introduced into the coding sequence to remove type II restriction enzyme sites [135]. Vectors were based on the Golden Gate level 2 pAGM4723 backbone [135], and were constructed from Golden Gate level 1 vectors containing cassettes for a) the P450 flanked by CPMV-*HT* 5' and 3' UTRs under control of the 35S promoter and terminator and b) the P19 R43W mutant gene (for regenerating transgenic plants if necessary [221]) also under the control of the 35S promoter and terminator. A schematic of the T-DNA for these vectors is given in **Figure 5-10**. Vector sequences have been deposited with ENSA (https://www.ensa.ac.uk/) under the numbers EC81392 (CYP88D6) and EC81537 (CYP93E1).

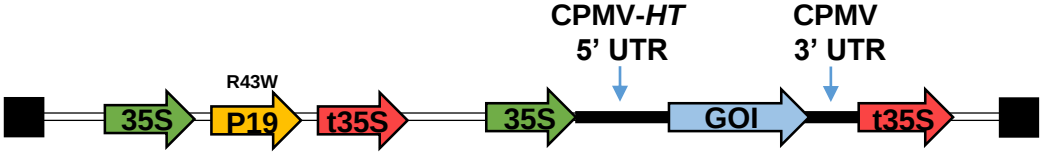


Figure 5-10: Schematic of the T-DNA portion of pAGM4723 containing either *CYP88D6* or *CYP93E1* (represented by the blue GOI arrow).

5.5.1.2 – 30-hydroxy-β-amyrin

30 plants expressing *tHMGR/AsbAS1*/*CYP72A65* yielded 19g dry leaf material. Flash chromatography of the extract was performed using a SNAP KP-Sil 25g column (Biotage) according to program P-9 (**Table 2-4**). Fractions containing 30-hydroxy-β-amyrin (as identified by TLC and GC-MS) were pooled and dried by rotary evaporation. This was resuspended in toluene and subjected to further flash chromatography using a SNAP ultra 10g column (Biotage) (program P-10) (**Table 2-4**). This yielded 8.9mg of an orange solid. The orange impurity was removed by trituration in methanol (room temperature) to give a total of 6.5mg of a white amorphous solid.

5.5.1.3 – 11-hydroxy- and 11-oxo- β-amyrin

23 plants expressing *tHMGR/AsbAS1*/*CYP88D6* yielded 10.4g dry leaf material. Flash chromatography of the leaf extracts was performed using a SNAP KP-Sil 25g column (Biotage) as according to program P-9 (**Table 2-4**). Fractions were analysed by GC-MS and TLC. Those containing 11-oxo-β-amyrin were pooled and likewise for 11-hydroxy-β-amyrin. The two semi-purified extracts each underwent further flash chromatography with a SNAP ultra 10g column (Biotage) according to program P-11 (**Table 2-4**) (performed twice for each). The fractions containing the respective metabolites were isolated and dried. This gave 7mg of off-white crystals of 11-hydroxy-β-amyrin and approximately 20mg of a yellow solid containing 11-oxo-β-amyrin. The 11-oxo-β-amyrin was further purified by recrystallization in hot ethanol to yield 10.4mg of white crystals with a further 9mg of yellow solid remaining.

5.5.1.4 – 24-hydroxy-β-amyrin

17 plants expressing *tHMGR/AsbAS1*/*CYP93E1* yielded 3g dry leaf material. Flash chromatography of the extract was performed using a SNAP ultra 10g column (Biotage) according to program P-9 (**Table 2-4**). This was repeated twice and the fractions containing 24-hydroxy-β-amyrin (monitored by TLC/GC-MS), were pooled and dried, then dissolved in toluene. A second flash column was performed (SNAP ultra 10g column (Biotage)) by program P-11 (**Table 2-4**). The purest fractions containing 24-hydroxy-β-amyrin were pooled and dried to give 4.7mg of a white crystalline solid.

5.5.1.5 – 16β-hydroxy-β-amyrin

45 plants expressing *tHMGR/AsbAS1/CYP51H10-A354L* yielded 17.9g dry leaf material. Flash chromatography of the extract was performed using a SNAP ultra 50g column (Biotage) according to program P-9 (**Table 2-4**). Fractions were

monitored by TLC and GC-MS and those containing the 16β-hydroxy-β-amyrin product were pooled and dried. These were dissolved in toluene and loaded onto a SNAP Ultra 10g cartridge (Biotage). Separation was achieved with a gradient according to program P-12 (**Table 2-4**). This gave an orange solid of which approximately 20mg was decolourised using activated charcoal (Sigma) and then recrystallized using hot ethanol to give 14mg of white crystals.

5.5.1.6 – 12,13β-epoxy,16β-hydroxy-β-amyrin

Please see **section 3.4.8**.

5.5.1.7 – Protopanaxadiol

Protopanaxadiol (PPD) was kindly supplied by Dr Michael Stephenson and Dr Ramesha Thimmappa. PPD was produced in *N. benthamiana* by the coexpression of *tHMGR, AsbAS1, AsbAS1-S728F* and *CYP716A47*. Approximately 90 plants were infiltrated by vacuum infiltration which gave 31g of leaf material after harvesting and freeze drying. The extract was saponified and extracted with ethyl acetate before drying. The extract was adsorbed onto silica gel and loaded onto a SNAP ultra 10g column (Biotage) and run manually under isocratic 25:75 ethyl acetate:hexane conditions to remove impurities. The mobile phase was then increased to 100% ethyl acetate over 3 column volumes and further run at 100% ethyl acetate to elute the product. Extracts were analysed by GC-MS and TLC to isolate fractions containing PPD. These were pooled and adsorbed onto silica gel and loaded onto a second SNAP ultra 10g column (Biotage). A gradient was performed according to program P-13 (**Table 2-4**). Recovered fractions were analysed by TLC to isolate those containing PPD. These were pooled and decolourised with activated charcoal followed by recrystallization with hot ethanol as in **section 2.1.4.4**. A single round of recrystallization resulted in 18mg of pure PPD with 119mg remaining of the semi-pure extract.

5.5.2 – Preparation of compounds for assays

Stock solutions of each of the compounds were made in sterile DMSO. Where possible, these were made to 20mM, though some were used at 10mM due to poor solubility. Stock solutions were vortexed vigorously and heated at 60°C for 10 mins to ensure solubility prior to the dilutions. A series of eight different stocks were made for each compound by two-fold dilutions in DMSO to give a stock concentrations ranging from 0.156mM to 20mM or 0.078mM to 10mM depending on solubility of the compounds in DMSO. These stock solutions were added as detailed below in the relevant assays.

Compound Name	Formula	Mass	Stock	Supplier
Oleanolic acid	$C_{30}H_{48}O_3$	456.71	20mM	Extrasynthese
Erythrodiol	$C_{30}H_{50}O_2$	442.73	20mM	Extrasynthese
Protopanaxadiol	$C_{30}H_{52}O_3$	460.74	20mM	-
11-oxo-β-amyrin	$C_{30}H_{48}O_2$	440.71	20mM	-
11α-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.73	20mM	-
30-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.73	20mM	-
24-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.73	10mM	-
16β-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.73	10mM	-
EpHβA	$C_{30}H_{50}O_3$	458.73	10mM	-
β-amyrin	$C_{30}H_{50}O$	426.73	10mM	Extrasynthese
11-oxo-oleanolic acid	$C_{30}H_{46}O_4$	470.69	20mM	-
Glycyrrhetinic acid	$C_{30}H_{46}O_4$	470.69	20mM	Extrasynthese
11-oxo,24-hydroxy-β-amyrin	$C_{30}H_{48}O_3$	456.74	20mM	-

Table 5-4: Compounds used for the MTS assays and ELISAs. Suppliers of compounds are detailed. Where suppliers are not given, the compound was purified from N. benthamiana as detailed in 5.5.1, except EpHβA (**section 3.4.8**), 11-oxo,24-hydroxy-β-amyrin (section 4.5.6) and 11-oxo-oleanolic acid (**section 4.5.7**).

5.5.3 – Culture of human cell lines

All cell lines were obtained from the European Collection of Authenticated Cell Cultures (ECACC). THP-1, HL-60 and M202 cells were cultured in RPMI 1640 media (Hyclone), while MCF-7 and MIA PaCa-2 cells were cultured in DMEM media (Hyclone). All media was supplemented with 10% heat-inactivated bovine foetal calf serum (FCS - Hyclone), L-glutamine (2mM) and antibiotics (penicillin (100U/mL) and streptomycin (100µg/mL)) (Gibco). All cell cultures were maintained at 37°C in a humidified atmosphere with 5% CO_2 and passaged every 3.5 days.

5.5.3.1– Cell density measurements

Cell density was measured using a Malassez haemocytometer. A 50µL solution of 0.4% trypan blue was mixed with the cells in an equal ratio. 15µL of this was added to the haemocytometer for counting under a light microscope.

For the adherent Mia PaCa-2, M202 and MCF-7 lines, cells first underwent trypsinisation. The media was removed and cells were washed in prewarmed Dulbecco's phosphate-buffered saline (PBS, Gibco) for 10 seconds. Cells were then treated with a 0.25% trypsin/EDTA solution (Sigma) and incubated at 37°C for 2-5 minutes before addition of 10mL fresh media. Following measurements of cell density as above, M202 and MCF-7 cells were diluted to $8x10^4$/mL in RPMI 1640 and DMEM media, respectively. MIA PaCa-2 cells were diluted to $5x10^4$/mL in

DMEM media. All media contained Fetal Bovine Serum (FBS - Hyclone) plus glutamine and antibiotics as above. Cells were aliquoted into 96 well plates using 100µL of cells in each well. Cells were further incubated overnight to allow for adherence.

For suspension cell lines THP-1 and HL-60, cells were measured as above and adjusted to a final density of $3x10^5$ cells/mL on the day of the assay by dilution in RPMI 1640 media as detailed in **section 5.5.3**. Cells were aliquoted into 96 well plates at 100µL per well to give a total of $3x10^4$ cells per well.

5.5.4 – MTS antiproliferation assay

MTS assays were performed in 96-well plates under sterile conditions. The outermost wells contained 250µL water in order to maintain humidity and minimise media evaporation. For each plate, two compounds were assayed in triplicate across eight concentrations. 1µL of the appropriate stock concentration was added per 100µL well to give a 100-fold dilution from the original stock. Untreated and vehicle (DMSO)-treated controls were also included for reference. Finally, three wells containing media only were included in order to subtract media absorbance during measurement.

After addition of the compounds, cells were incubated for 72 hours at 37°C (5% CO_2). Following this period, the MTS reagent (Cell Titer 96 Aqueous non-radioactive assay, Promega) was prepared and 10µL was added to all wells (excluding outermost wells). Cells were returned to the incubator for 2.5-4 hours depending on the individual cell line. Finally, absorbance at 492nm was measured and the averaged media-only reading was subtracted from all wells.

Data analysis was performed using GraphPad Prism software (version 5). IC_{50} values were generated by using non-linear regression (log of treatment concentration versus response, variable slope, least squares method).

5.5.5 – Anti-inflammatory assay

THP-1 cells were grown as detailed in **section 5.5.3**. Cell density was measured and the cells were pelleted by centrifugation (1200 rpm, 7 min). Cells were resuspended in an appropriate volume of medium to give a final density of $1x10^6$ cells/mL. Cells were aliquoted into 24 well plates at 500µL per well to give $5x10^5$ cells per well. 2.5µL of the test compounds in DMSO were added to the wells (1 in 200 dilution, giving 100µM final concentration in most cases) and incubated for 30 mins at 37°C (5% CO_2). After this time, to stimulate TNFα production, 1µL LPS

(0.5mg/mL stock in RPMI media) was added to the appropriate wells (1µg/mL LPS final). Cells were incubated as above for a further three hours. Finally supernatants were harvested by transferring cells to microcentrifuge tubes and centrifuging at 2000rpm for 5 mins. Supernatants were transferred to fresh microcentrifuge tubes on ice and stored at -80°C until needed.

5.5.5.1 – ELISA setup

Solutions of the following reagents were prepared for the ELISA assay:

Phosphate Buffered Saline (PBS) solution

Per litre (pH 7.0):

8g NaCl

1.17g Na_2HPO_4

0.2g KH_2PO_4

0.2g KCl

Coating buffer

Per litre (pH 6.5):

16.1g NaH_2PO_4

11.8g Na_2HPO_4

Wash buffer

1L PBS (as above)

50µL Tween-20

Assay diluent

90% Phosphate buffered saline solution

10% Foetal bovine serum

pH to 7.0

Stop solution

2N H_2SO_4

5.5.5.2 – TNFα ELISA

Determination of TNFα production was carried out in 96 well plates using a Human TNF ELISA kit (BD biosciences). Plates were coated in 100µL of human TNF

capture antibody diluted in coating buffer, sealed and stored overnight at 4°C. Cells were aspirated three times with wash buffer, blocked with >200µL assay diluent and incubated at room temperature for 1 hour. Cells were washed again three times in wash buffer before addition of samples. 100µL of supernatants from THP-1 cells were added (**section 5.5.5**) to the plate. Supernatants derived from LPS-treated cells were first diluted in assay diluent four-fold in media in order to bring these samples to within the range of the standard curve. A TNFα standard (BD Biosciences) was included across a two-fold serial dilution range from 500pg/mL to 7.8pg/mL (diluted in PBS). This was used to generate the standard curve for each assay for determination of TNFα production. A pre-prepared 100pg/mL TNF standard was also included across all plates to ensure consistency of the assay between replicates. Plates were incubated at room temperature for 2 hours and then washed 5 times in wash buffer. 100µL of the detection antibody was added before sealing and incubating the plate at RT for 1 hour. After this period, the plate was washed again seven times with wash buffer, leaving the wash buffer in the wells for at least 30 seconds each time. 100µL of substrate solution (TMB substrate reagent kit, BD Biosciences) was added to each well and left in the dark for 30 minutes before addition of 50µL stop solution. Plates were measured by absorption at both 450nm and 570nm and the value for 570nm was subtracted from the value for 450nm.

Data analysis was performed using GraphPad Prism software (version 5). A linear regression analysis was performed on the TNFα standard and this was used to convert the absorbance values of test samples to pg/mL. The values were divided by the value for the DMSO + LPS-treated control to show relative TNFα production. These were averaged from three biological replicates. For statistical analysis, a one way ANOVA was performed with post-hoc analysis performed by Tukey's multiple comparison test.

Chapter 6:

Identification of P450s required for avenacin biosynthesis in oat

Acknowledgements:

The author designed and performed all RT-PCR and qPCR expression profiling experiments as well as GC-MS analysis of *N. benthamiana* leaves. Dr Aymeric Leveau performed cloning of oat P450s and assisted with identification of candidate enzymes. Dr Sam Mugford performed LC-MS analysis of oat mutants. Dr Tim Langdon and Prof Bin Han are gratefully acknowledged for providing genomic sequence data from *A. atlantica* and *A. strigosa,* respectively.

6.1 – Introduction

6.1.1 – Saponins in *Avena* species

The saponins are a structurally diverse group of triterpene and steroidal glycosides that serve as a major class of defensive agents in plants [33, 34, 39]. Saponins are distributed throughout the angiosperms, but are rarer in monocotyledonous species [37, 72] and largely absent from *Graminaceous* species, including most cultivated cereals [222]. Oats (*Avena spp*) are a notable exception due to their ability to produce antimicrobial triterpene saponins (avenacins). The avenacins confer defence against soil-borne pathogens such as *Gaeumannomyces graminis* var. tritici, the causative agent of 'take-all' disease. Indeed, oats are frequently used as a 'break crop' to manage take-all and other diseases. An understanding of the biosynthesis of avenacins in oat offers the potential to engineer production of avenacins and other structurally related antimicrobial triterpenes into other important crops such as wheat for disease control.

6.1.2 – Structure and biosynthesis of avenacins

Four structurally related avenacins are present in oat, named A-1, B-1, A-2 and B-2 (**Figure 6-1A**). There are three major structural features common to all four avenacins, including a core oxidised β-amyrin skeleton, a C-3 branched trisaccharide and a C-21 acyl group (**Figure 6-1A**). Other modifications to the β-amyrin scaffold that are present in all four avenacins include several oxidations – namely, a β-hydroxyl group at C-16, an aldehyde at C-30, and a β-epoxide bridging C-12 and C-13. The four avenacins also all have the same trisaccharide moiety attached at the C-3 position, which consists of an α-L-arabinose linked at the 2 and 4 positions to two β-D-glucose molecules [182]. Avenacins A-1 and A-2 differ from B-1 and B-2 in that they contain a hydroxyl moiety at the C-23 position. A second distinction is the type of acyl ester at the C-21 position, which may be either an *N*-methyl-anthraniloyl group (A1 and B1) or a benzoyl group (A2 and B2) (**Figure 6-1A**). Avenacin A-1 and B-1 are strongly UV fluorescent due to the presence of the *N*-methyl anthranilate group, while avenacins A-2 and B-2 are not (**Figure 6-1B**).

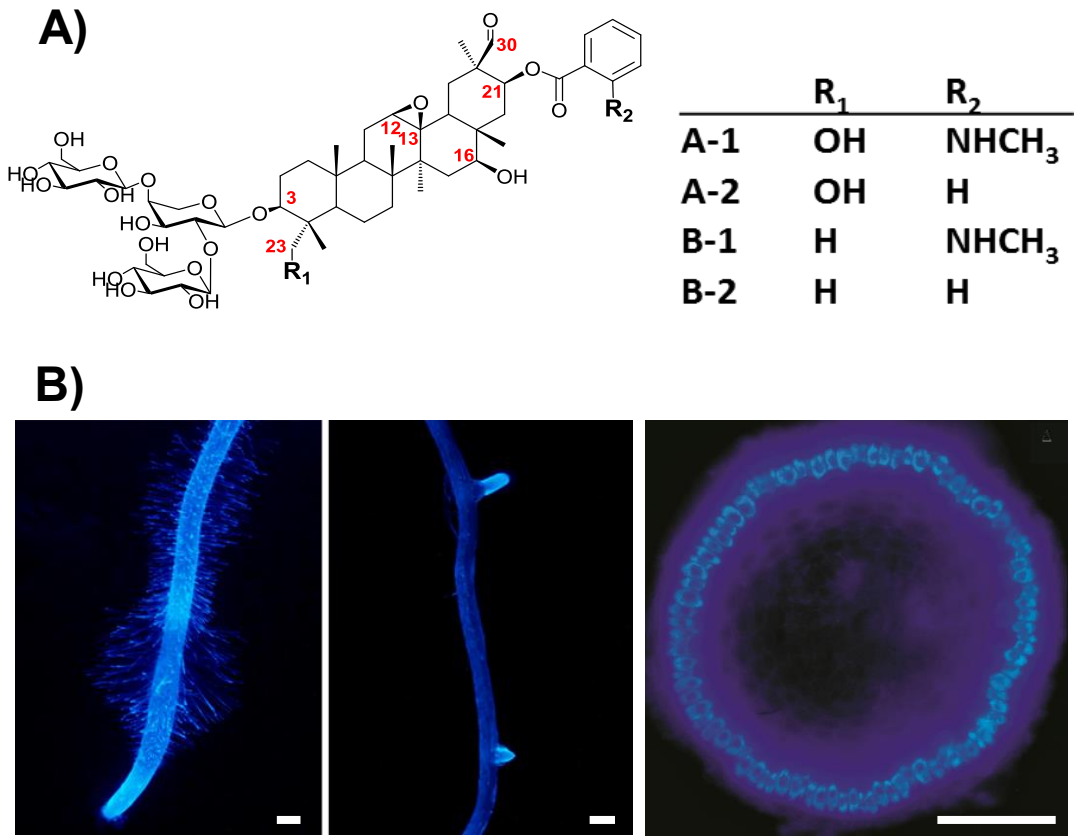


Figure 6-1: Avenacins in oat **A)** Structures of the oat avenacins. The four structural variants are distinguished by modifications at two positions, including the presence or absence of hydroxylation at C-23 and the type of acyl group present at C-21; *N*-methyl-anthranilate or benzoate **B)** Oat roots exhibit bright blue fluorescence under UV light due to the presence of the major avenacin, A-1. This fluorescence is restricted to the root tips (left) and lateral roots (middle) . Scale bars = 200µm. A root cross section shows that the fluorescence associated with avenacin A-1 is located in the epidermal cells (right). Scale bar = 50µm. Adapted from [29] and [223].

Avenacin A-1 is the major avenacin produced in oats [224]. The biosynthesis of avenacin A-1 depends on two converging pathways, one that synthesises the fully formed triterpene glycoside lacking the acyl group [des-acyl avenacin (DAA)], and the other that generates the *N*-methyl anthraniloyl-*O*-glucose acyl donor for addition at the C-21 position (**Figure 6-2A**). The first of these pathways is carried out between the endoplasmic reticulum (ER) and the cytosol. The second involves methylation and glycosylation of anthranilate to form the *N*-methyl anthraniloyl-*O*-glucose acyl donor. These two pathways converge in the vacuole, where the *N*-methyl anthranilate is conjugated to DAA at the C-21β position [225, 226]. At the start of this project five genes required for avenacin biosynthesis had been cloned and the encoded enzymes characterised. These include the β-amyrin synthase *AsbAS1* (*Sad1*), which forms the triterpene scaffold, and the *AsCYP51H10* β-amyrin

oxidase (*Sad2*), responsible for initial oxidation of β-amyrin (**Figure 6-2A**) [85, 86, 225-228]. Two additional genes include *AsMT1* (*Sad9*) and *AsUGT74H5* (*Sad10*) which encode a methyltransferase and glycosyltransferase, respectively, that make *N*-methyl-anthraniloyl glucoside from anthranilate [225, 228] (**Figure 6-2A**). Finally, *AsSCPL1* (*Sad7*) encodes the serine carboxypeptidase-like acyltransferase that conjugates the acyl group to the triterpene scaffold [226] (**Figure 6-2A**). The enzymes responsible for oxidation at the C-21, C-23 and C-30 positions and for addition of the sugar chain at the C-3 position were unknown.

6.1.3 – Approaches for discovering new *Sad* gene candidates.

Elucidation of the avenacin pathway thus far has been facilitated by the establishment of a collection of avenacin-deficient mutants of *A. strigosa* following sodium azide mutagenesis, using a screen for loss of root fluorescence [38]. This work resulted in the identification of 8 independent *saponin deficient* (*Sad*) loci [38, 227]. Four of these loci (*Sad1, Sad2, Sad7* and *Sad9)* have been cloned and characterized and shown to encode AsbAS1, AsCYP51H10, AsSCPL1 and AsMT1, as described above [86, 225-228]. The remaining loci defined by genetic analysis had not yet been uncharacterized at the start of this project.

Intriguingly, the loci for avenacin biosynthesis are genetically linked and form a biosynthetic gene cluster [94]. This cluster is the result of the independent recruitment of a number of unrelated genes from other parts of the genome during oat evolution and is not present in other related cereal species [92, 229, 230]. The first pathway gene to be cloned was *AsbAS1.* Establishment of a bacterial artificial chromosome (BAC) library for *A. strigosa* then facilitated identification of the four other characterized genes within the avenacin cluster by assembly and sequencing of a BAC contig [86, 225, 226, 228]. These genes lie within a 200 kb region with no other obvious intervening genes [225] (**Figure 6-2B**). Further loci are known to be genetically linked to the cluster, but have not yet been characterised. Extension of the sequences flanking the cloned avenacin biosynthetic genes is therefore expected to reveal other candidate pathway genes.

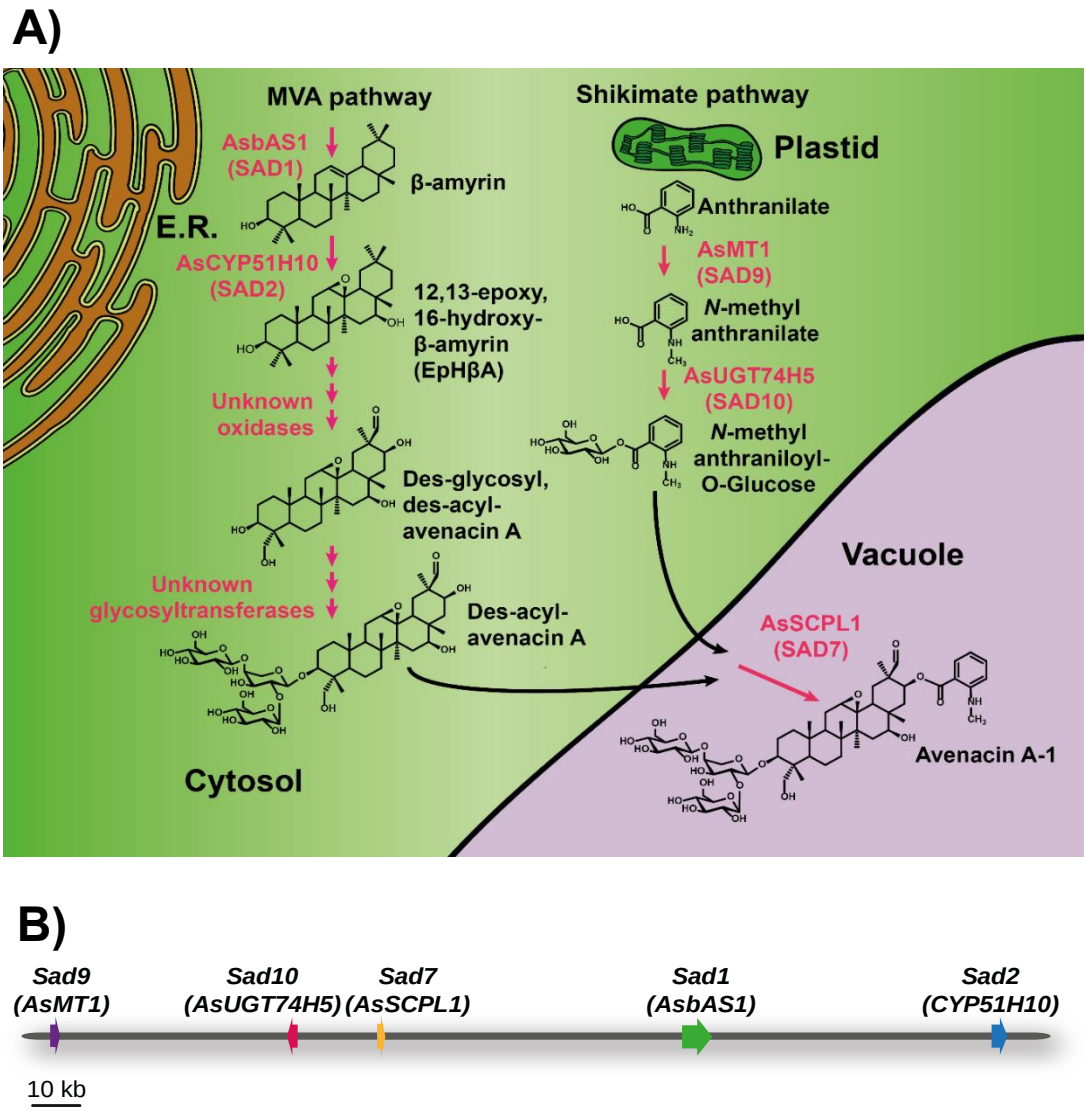


Figure 6-2: Biosynthesis of avenacins in oat **A)** Proposed pathway for biosynthesis of avenacin A-1 in *A. strigosa.* Synthesis and oxidation of β-amyrin is believed to occur on the cytosolic face of the endoplasmic reticulum (ER). The oxidised scaffold is expected to be glycosylated prior to transport to the vacuole, where it is then conjugated to the shikimate pathway-derived *N*-methyl anthranilate group by the acyltransferase AsSCPL1. Enzymatic steps are indicated by red arrows, and transport by black arrows. **B**) The avenacin gene cluster. The five genes that have so far been shown to be involved in avenacin biosynthesis lie within a 200kb region of the oat genome. Additional biosynthetic loci defined by mutagenesis are also known to be linked to the cluster.

6.1.4 – Aims

Acylation of the avenacins is an important modification required for antifungal activity. Conjugation of the acyl group requires functionalisation of the triterpene scaffold at the C-21β position, likely by P450-mediated hydroxylation. At present however, no C-21β oxidase has been identified in oat or any other species. Additionally, the enzymes responsible for oxidation of the triterpene scaffold at the C-23 and C-30 positions are unknown. Oxidation at these and other positions may also be important for antifungal activity [85, 224]. In this chapter, three *A. strigosa* P450s are implicated in avenacin biosynthesis. These findings represent an important advance in understanding biosynthesis of antimicrobial triterpenes in oat and will facilitate metabolic engineering of other crop plants for enhanced disease resistance. Furthermore, these newly characterized triterpene oxidases offer an important addition to the suite of enzymes available for generating triterpenes in other species. This will be important for investigating structure-function relationships and creating novel antifungal saponins.

6.2 – Results and discussion

6.2.1 – Identification of new avenacin P450 candidates

All currently known β-amyrin oxidases are members of the cytochrome P450 superfamily [171]. Thus a P450 was expected to be the most likely candidate for a C-21 hydroxylase in oat. The Osbourn lab has generated an oat root tip 454 transcriptomic database [29] within which all of the known avenacin biosynthetic genes are represented. Thus this database was expected to contain the additional P450s required for avenacin biosynthesis. However the database contains a total of 168 P450 sequences (Dr Jenni Rant, personal communication) which makes screening such a large number of candidates difficult.

Recently, next-generation sequencing-based mapping of recombinant inbred lines (RILs) derived from crosses of *Avena atlantica* and *A.strigosa* was performed by Dr Tim Langdon (Aberystwyth University, Wales). This has allowed mapping of *A. atlantica* genomic sequence data to 'bins' which correspond to relatively large genomic segments. In collaboration with Dr Langdon, these bins were mined for known *Sad* genes. This gave a series of genomic contigs that could potentially harbour physically linked genes participating in avenacin biosynthesis.

Identification of oat P450s which are both linked to known *Sad* loci and expressed in root tips allows the number of candidates to be considerably narrowed. Therefore, the *A. atlantica* genomic sequence contigs were used as search queries against the *A. strigosa* root tip 454 transcriptomic database [29] (**methods 6.4.1**). Eight P450 sequences were identified in the *A. strigosa* root tip database using this approach, including the previously characterised *CYP51H10 (Sad2*) gene. The *A. strigosa* P450 sequences were submitted to the P450 nomenclature committee for assignment [82] and are listed in **Table 6-1**.

Name	Clan	454 number	Length (bp)
CYP51H10/Sad2	CYP51	17168	1473
CYP71E22	CYP71	23466/23468	1518
CYP72A475	CYP72	18622/00941/00942	1608
CYP72A476	CYP72	09269	1548
CYP88A75	CYP85	13408/06752	1482
CYP89N1	CYP71	18440	1485
CYP706C45	CYP71	19837	1764
CYP711A54	CYP711	06108/19967	1605

Table 6-1: P450s identified in this study. The corresponding 454 contig numbers and predicted lengths of the protein coding regions are given.

Avenacin biosynthesis is confined to the epidermal cells of oat root tips (**Figure 6-1B**), and in accordance with this the five previously characterized *Sad* genes are expressed specifically in the root tips and not in other tissues [29, 86, 225-228]. Therefore, RT-PCR was used to investigate the expression profiles of the new candidate genes in root and shoot tissues of *A. strigosa* seedlings (**methods 6.4.8**). The *AsbAS1* (*Sad1*) and *CYP51H10* (*Sad2*) genes were included for comparison, along with the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*).

The majority of the candidate genes could be successfully amplified by PCR in cDNA preparations from *A. strigosa* root tips, in accordance with their representation in the 454 database. However of the seven candidates, only two (*CYP72A475* and *CYP72A476*) showed a *Sad* gene-like expression pattern, with strongest expression in the root tips and little or no expression in the elongation zone or shoots (**Figure 6-3**). Three of the other candidates (*CYP71E22, CYP88A75 and CYP711A54*) were expressed at very low levels in the root tips. *CYP706C45, CYP71E22, CYP89N1* and *CYP88A75* also showed expression in the elongation zone and/or shoot (**Figure 6-3**). Thus, based on gene expression profiling, *CYP72A475* and *CYP72A476* are the most promising candidates for missing steps in avenacin biosynthesis.

6.2.2 – Expression of P450 candidates in *N. benthamiana*

As previously mentioned, the *A. strigosa CYP51H10* (*Sad2*) gene encodes the second enzyme in the avenacin biosynthetic pathway and converts β-amyrin to 12,13β-epoxy, 16β-hydroxy-β-amyrin (EpHβA) [85]. To establish whether any of the seven new oat P450s were able to modify β-amyrin or EpHβA, the predicted full-length coding sequences of the candidate genes were amplified from oat root tip cDNA for transient expression in *N. benthamiana* (performed by Dr Aymeric Leveau). This was successful for five out of seven candidates; however it was not possible to obtain full length products for *CYP88A75* or *CYP71E22*, which is broadly in accordance with their poor amplification from root tip cDNA during the previous RT-PCR analysis (**Figure 6-3**). The five full length coding sequences were cloned into the pEAQ-*HT*-DEST1 vector and used to transform *A. tumefaciens* (**methods 6.4.4**). Each candidate was then coexpressed either with *AsbAS1* alone, or with *AsbAS1* and *CYP51H10*. Additionally, *tHMGR* was included in all combinations to enhance the levels of the substrate and allow better identification of novel metabolites (Chapter 3).

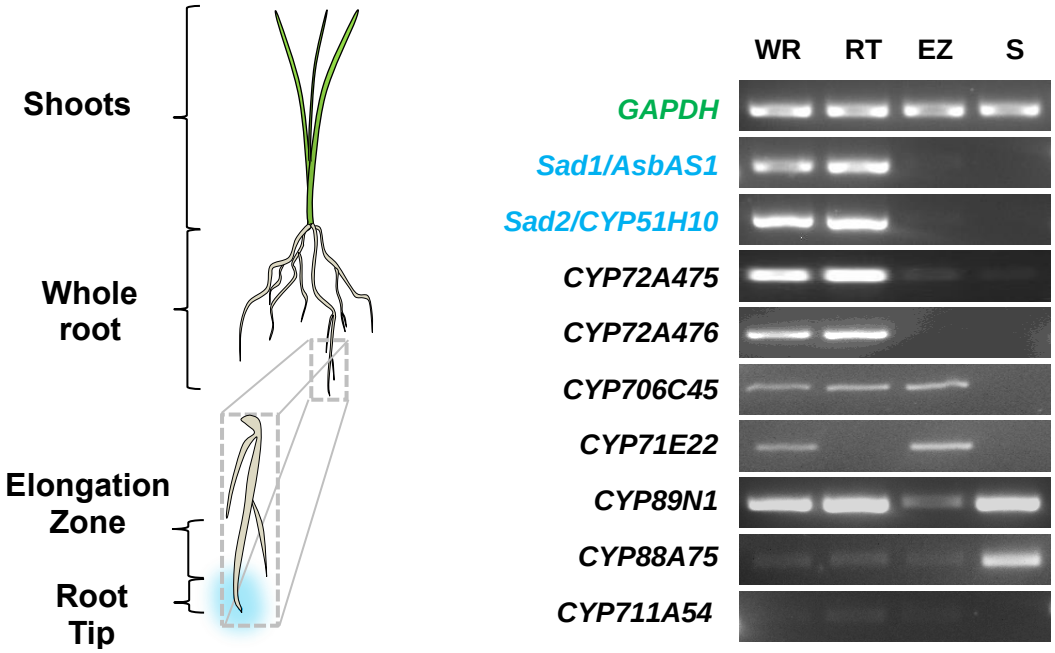


Figure 6-3: Analysis of the expression profiles of candidate P450 genes. (Left) Avenacin production is restricted to the root tips (blue glow) and lateral roots (not shown). (Right) RT-PCR expression profiling of candidate genes. The profiles of the previously characterized *AsbAS1* (*Sad1*) and *AsCYP51H10* (*Sad2*) genes were included as positive controls (indicated in blue), along with that of the housekeeping gene GAPDH (shown in green). RNA was extracted from 4 day old oat seedlings. Abbreviations: RT = root tip, WR = whole root, EZ = elongation zone, S = shoot.

Five days after infiltration with *A. tumefaciens, N. benthamiana* leaves were harvested and extracts analysed by GC-MS. Of the five candidates tested, none showed any activity on β-amyrin when coexpressed with *tHMGR/AsbAS1* (data not shown). Similarly, *CYP706C45, CYP89N1* or CYP711A54 showed no activity on EpHβA when coexpressed with *tHMGR/AsbAS1/CYP51H10* (data not shown). This could be due to problems with expression of these enzymes in *N. benthamiana.* However, new products could be detected upon coexpression of either *CYP72A475* (15.5 mins) or *CYP72A476* (15.0 mins) with *tHMGR*l*AsbAS1/CYP51H10,* suggesting that these enzymes had activity on EpHβA (**Figure 6-4A**). Furthermore, the new peak in the *tHMGR/AsbAS1/CYP51H10/CYP72A475* sample was accompanied by a large decrease in the EpHβA peak compared to controls, suggesting that much of the EpHβA had been converted (**Figure 6-4A**). In contrast, EpHβA remained the major product in the *tHMGR/AsbAs1/ CYP51H10/CYP72A476* extract (**Figure 6-4A**). Analysis of the electron-impact (EI) mass spectra revealed that each of the new peaks had a nominal parent mass of *m/z* 690 corresponding to a hydroxylated form of EpHβA (**Figure 6-4A**). The difference in retention times indicated that each of the new products represents different hydroxylated forms of the EpHβA product.

The avenacins are oxidised at up to five positions across the β-amyrin scaffold including a C-ring epoxide (C-12-C-13) and a D-ring hydroxyl (C-16), both of which are present in the EpHβA substrate. The remaining positions include E-ring aldehyde (C-30) and hydroxyl (C-21) groups, and an optional A-ring C-23 hydroxyl (present only in avenacins A-1 and A-2) (**Section 6.1.2**/**Figure 6-1**). Thus the C-21, C-23 and C-30 are the obvious candidate positions for oxidation. To investigate which modifications were likely to be present in the new products, the mass spectra were further analysed. The EI fragmentation of β-amyrin and related products is described in detail in section 4.1.2. Briefly, EI fragmentation of β-amyrin proceeds via a retro-Diels-Alder (rDA) mechanism, giving a characteristic ion at *m/z* 218 (representing the C*DE rings) (**Supplementary Figure 17**) [173, 174]. An alternative fragmentation route results in a minor ion at *m/z* 279 and corresponds to the ABC* rings (with silylated C-3 hydroxyl) [162, 174]. The EpHβA mass spectrum also shows a prominent *m/z* 279 ion (ABC* rings) while the *m/z* 218 ion in β-amyrin (C*DE rings) shifts to *m/z* 322 due to the gain of the C-ring epoxide and the (silylated) C-16 hydroxyl group (**Figure 6-4A** and **Supplementary Figure 17**)**.** Changes to these ions can thus inform the location of new modifications on the scaffold.

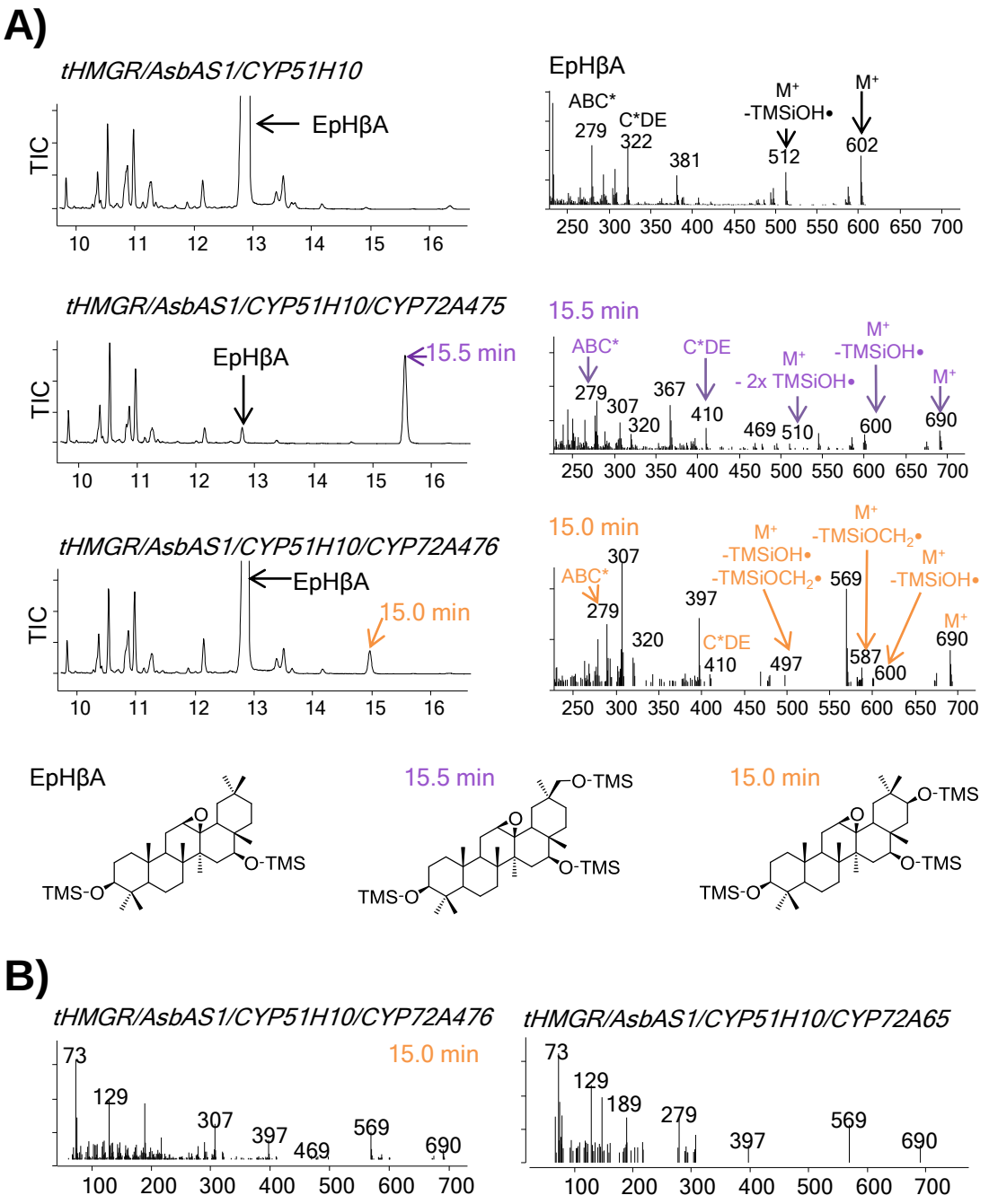


Figure 6-4: Analysis of *N. benthamiana* leaf extracts by GC-MS. **A)** Total ion chromatograms (TIC) of extracts from leaves expressing *tHMGR/AsbAS1/ CYP51H10* versus those expressing *tHMGR/AsbAS1/CYP51H10/ CYP72A475* and *tHMGR/AsbAS1/CYP51H10/CYP72A476.* The EI mass spectra for the novel peaks in each sample are given to the right. Below are shown the putative structures of the new products. **B)** The novel product in *tHMGR/AsbAS1/CYP51H10/CYP72A476* samples has a matching mass spectrum to a peak in *tHMGR/AsbAS1/CYP51H10/CYP72A65* samples (Chapter 4, compound 36, **Figure 4-8**). This suggests the product may be a C-30 hydroxylated form of EpHβA.

Analysis of the novel peaks showed that both contain the *m/z* 279 ion but have lost the *m/z* 322 ion, together suggesting that neither enzyme oxidises the A-ring-localised C-23 position (**Figure 6-4A**). A distinction between the remaining possible oxidations (C-21 and C-30) can be made due to the fact that hydroxyl groups located on methyl groups (as for C-30) are usually lost as $TMSiOCH_2$• (loss of *m/z* 103) whereas ring hydroxyls are lost as TMSiOH• (loss of *m/z* 90) [173]. Accordingly, the fragmentation of the peak at 15 min in the *tHMGR/AsbAS1/CYP51H10/CYP72A476* sample showed ions indicating the presence of a methyl-localised hydroxyl group (**Figure 6-4A**). These ions were not present in the *tHMGR/AsbAs1/CYP51H10/CYP72A475* sample, which instead showed losses consistent with ring-localised hydroxyl groups (**Figure 6-4A**). Together these observations suggest that the *CYP72A475* may be a C-21 oxidase and the *CYP72A476* is a C-30 oxidase, catalysing production of 12,13-epoxy,16,21-dihydroxy-β-amyrin and 12,13-epoxy,16,30-dihydroxy-β-amyrin, respectively.

In work described in Chapter 4, combinatorial biosynthesis of β-amyrin oxidases in *N. benthamiana* allowed for production of novel oxidised forms of EpHβA. In particular, coexpression of *tHMGR, AsbAS1* and *CYP51H10* with the *M. truncatula* C-30 oxidase *CYP72A65* resulted in a novel peak putatively identified as 12,13-epoxy, 16,30-dihydroxy-β-amyrin (**section 4.3.2**). Comparison of the mass spectra of this product with the novel product from *tHMGR/AsbAS1/CYP51H10/CYP72A476* revealed that the two mass spectra are very similar (**Figure 6-4B**). This offers further evidence that *CYP72A476* is a C-30 oxidase.

The apparent specificity of *CYP72A475* and *CYP72A476* for EpHβA over β-amyrin is evidence for their involvement in avenacin biosynthesis. However while *CYP72A475* demonstrates almost total conversion of the EpHβA substrate, *CYP72A476* converted only minor amounts. This could indicate that EpHβA is not the preferred substrate for the enzyme. Further coexpression of both P450s together with *tHMGR, AsbAS1* and *CYP51H10* was conducted, but new products were not detected (data not shown). This may suggest that *CYP72A476* is poorly active in *N. benthamiana* or that none of the molecules synthesised here are good substrates for this enzyme. Other CYP72A enzymes (*CYP72A224* and *CYP72A1*) have been shown to utilise glycosylated substrates [122, 231]. Identification of other enzymes in the avenacin biosynthetic pathway may allow the order of modifications to be better understood.

It is interesting to note that CYP72A476 apparently catalyses oxidation to a C-30 alcohol, but not the aldehyde present in the four avenacins. P450-mediated formation of aldehydes occurs via the 2-step oxidation of methyl groups via an alcohol intermediate [232]. Sequential oxidation at a single position is often catalysed by a single P450, but it is equally possible that other unknown enzymes are required [233]. It is also possible that the aldehyde constituent is formed but subsequently reduced back to an alcohol by endogenous reductases in *N. benthamiana.* Similar activity has been noted previously; production of aldehyde-containing precursors of artemisinin are converted to their respective alcohols when expressed transiently in *N. benthamiana* [234] or when fed exogenously to the closely related *Nicotiana tabacum* [235]. As such, it may be worth testing this enzyme in other systems such as yeast to see if additional metabolites can be detected.

6.2.3 – CYP72A475 is synonymous with the *Sad6* locus

The above results suggested that *CYP72A475* may be a C-21β oxidase. Absence of the C-21β hydroxyl group would prevent acylation at this position and cause a loss-of-fluorescence phenotype as used in the original mutagenesis screen [38]. Two of the avenacin-deficient mutants within the Osbourn group collection (#825 and #1243) have recently been profiled by LC-MS by Dr Sam Mugford (John Innes Centre) and shown to accumulate a compound that is not present in the wild type S75 *A. strigosa* line (**Figure 6-5**). Genetic complementation analyses indicates that mutant lines #825 and #1243 are allelic, and are herein referred to as *sad6* mutants (Prof A. Osbourn, personal communication).

The compound accumulating in the *sad6* mutants has a mass spectrum similar to that of des-acyl avenacin B, which accumulates in *sad7* mutants (**Figure 6-6, left**) [226]. This suggests that this compound lacks the acyl group at C-21 and is missing an oxygen atom at one position. Further examination of the mass spectra revealed the presence of a group of four ions at *m/z* 435, 453, 471 and 489 (**Figure 6-6, right**). These have consistent mass differences of 18, likely representing the aglycone (*m/z* 489) and fragmentation products that have undergone sequential loss of hydroxyl groups as water (at C-3, C-16, C-21 and C-23, highlighted in red in **Figure 6-6**). This pattern can be seen in the product from the *sad6* mutants as well as des-acyl avenacin B (*sad7*) (**Figure 6-6, right**). Des-acyl avenacin A by contrast contains five ions at *m/z* 433, 451, 469, 487 and 505 due to the presence of the additional C-23 hydroxyl group (**Figure 6-6, right**). Therefore the product in *sad6* mutant root extracts is expected to contain four hydroxyl groups. These should

include the C-3 and C-16 hydroxyl groups as the *sad6* mutation affects a step downstream of EpHβA production. The C-23 is also likely to be present, owing to the difference in retention time between these products and des-acyl avenacin B (**Figure 6-5**). Therefore absence of the C-21 hydroxyl group is the most likely interpretation. This new compound is hereby named des-acyl, des-21-hydroxy-avenacin A (**Figure 6-5**).

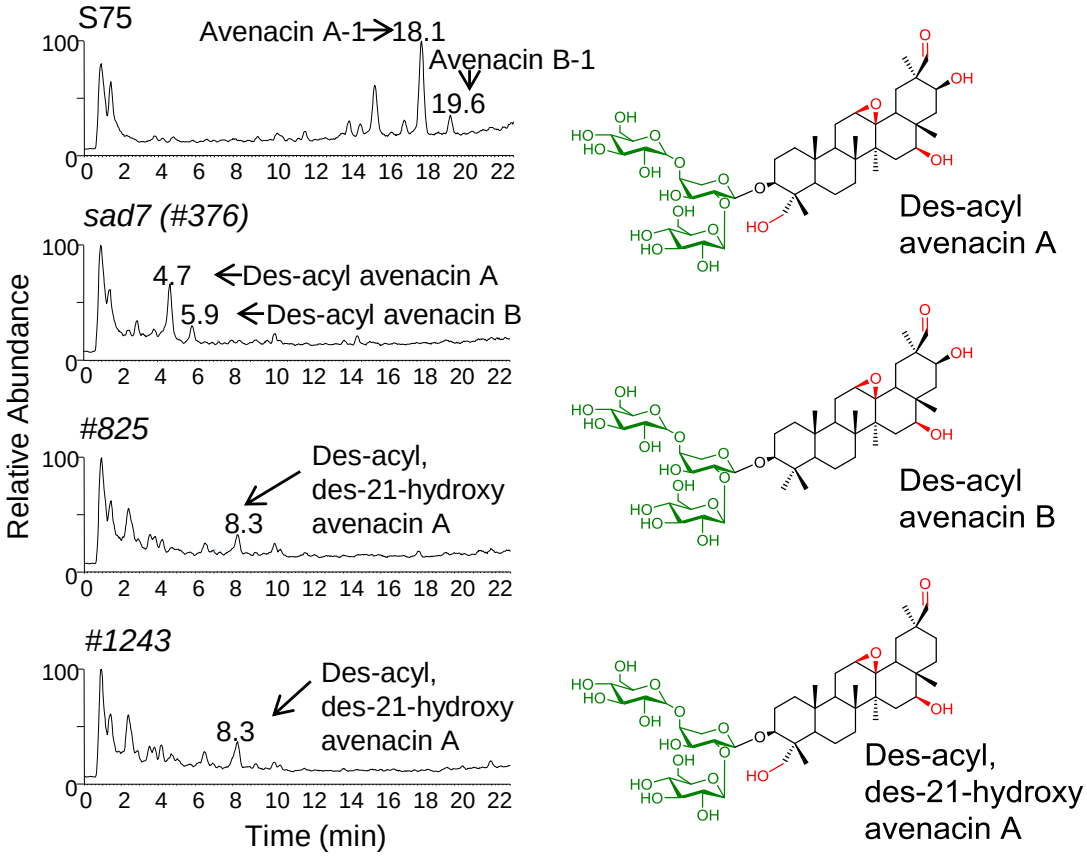


Figure 6-5: LC-MS analysis of oat roots. Extracts of the wild type *A. strigosa* line S75, a *sad7 mutant*, and the two *sad6* (#825 and #1243) mutants are shown. The mutant numbers are as in Papadopoulou et al [38]. The numbers above the peaks are the retention times (minutes). The compound detected at 8.3 mins in mutants #825 and #1243 is likely to be des-acyl, des-21-hydroxy avenacin A based on the mass fragmentation pattern. Supporting mass spectra are given for these assignments in **Figure 6-6**. Data kindly provided by Dr Sam Mugford.

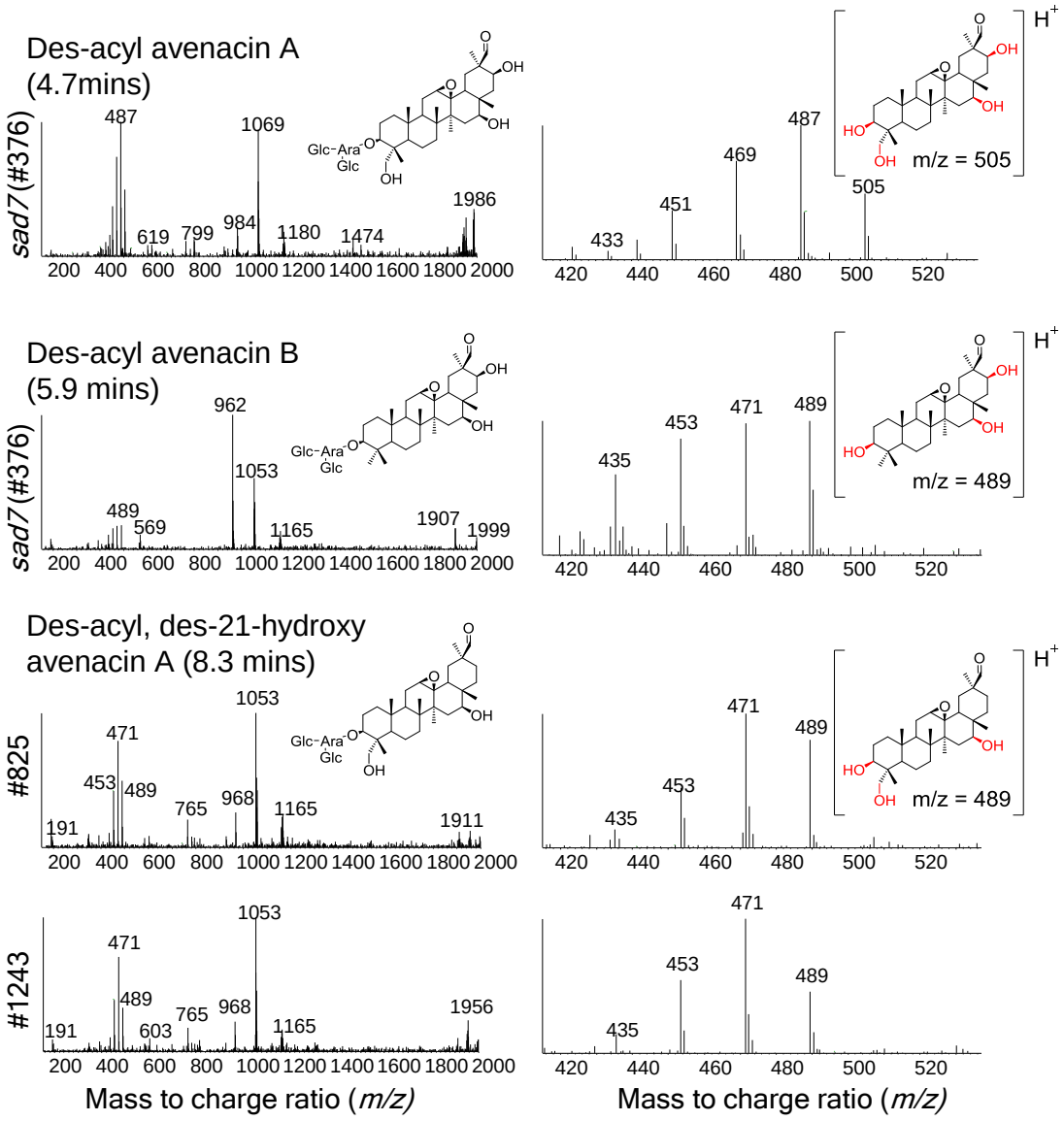


Figure 6-6: ESI mass spectra for compounds in root extracts of *A. strigosa sad* mutants. **Left**: *sad6* mutants (#825 and #1243) accumulate a compound with similar mass spectrum to des-acyl avenacin B, (which accumulates in *sad7* mutant roots), but different retention times (8.3 minutes vs 5.9 minutes respectively). **Right**: In-chamber fragmentation of the compounds can be observed, which results in loss of the trisaccharide at C-3. The aglycone can undergo further fragmentation with characteristic losses of *m/z* 18 (loss of the hydroxyl groups as water). The compound in #825 and #1243 plants is therefore expected to have 3 hydroxyl groups. The lack of the acyl group implicates the C-21 hydroxyl group as being absent in the compound at 8.3 minutes. Data kindly provided by Dr Sam Mugford.

To genotype the two mutant lines, attempts were made to clone *CYP72A475* from cDNA derived from root tissues of #825 and #1243 seedlings (**methods 6.4.5.1**). A PCR product of the correct size was amplified from the #1243 cDNA template, but not the #825 template. The product from the #1243 line was cloned and sequenced. This revealed a guanine-to-adenine substitution at position 1068 in mutant that is predicted to result in conversion of Trp356 to a premature stop codon (TGG → TGA) (**Figure 6-7A**). The same mutation has also been identified from genomic sequencing of this line (Dr Aymeric Leveau, personal communication).

To eliminate the possibility that sample preparation errors were responsible for the failure to amplify a product in #825, qPCR expression analysis of *AsbAS1* and *CYP72A475* was performed in wild type and the two mutant lines (**methods 6.4.9**). This revealed that *CYP72A475* transcript levels were markedly reduced in #1243 mutants compared to the wild type, which may be due to nonsense-mediated RNA decay [236] (**Figure 6-7B**). Again, no amplification product was detected for mutant #825, despite clear amplification of the *AsbAS1* transcript. This is further evidence that the *CYP72A475* is disrupted in these two mutant lines. Further recent analysis suggests that mutant #825 has undergone a deletion event in the region of the *CYP72A475* gene (Dr Aymeric Leveau, personal communication).

Taken together, these results suggest that the allelic *sad6* mutants #825 and #1243 both fail to hydroxylate the β-amyrin backbone at C-21, and that this is due to absence of a functional *CYP72A475* gene. Oxidation of the C-21β position is a pre-requisite for addition of the acyl group, which in turn is important for anti-fungal activity [225]. Accordingly, lines #825 and #1243 both show increased susceptibility to infection by the take-all fungus *Gaeumannomyces graminis* var. *tritici,* highlighting the importance of this modification for disease resistance [38].

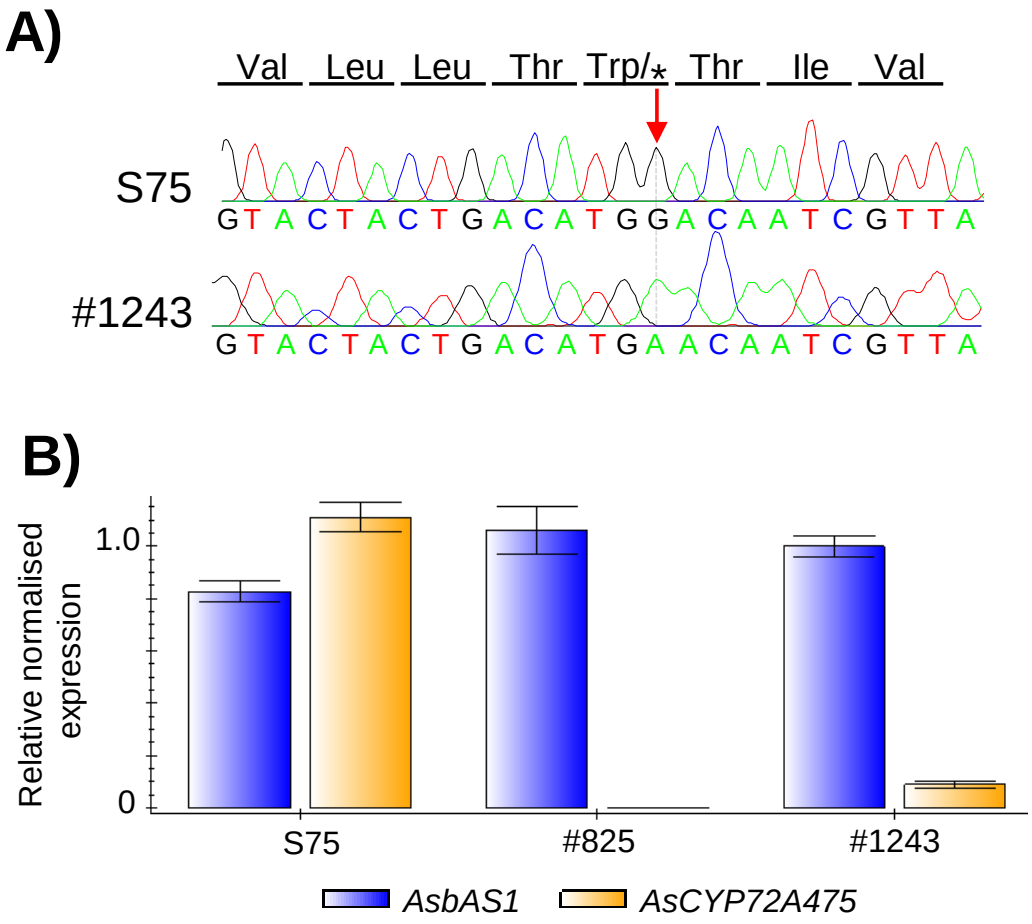


Figure 6-7: Analysis of *CYP72A475* in oat mutants #825 and #1243 **A)** Sequencing of the *CYP72A475* transcript in the wild type (S75) and in mutant line #1243. The #1243 line carries a nonsense point mutation at position +1068 (G→A). Sequence data for the wild type (S75) transcript was kindly provided by Dr Aymeric Leveau **B)** qPCR analysis of transcript levels of *AsbAS1* and *CYP72A475* in the wild type (S75) and the two *sad6* mutants #825 and #1243. The *CYP72A475* transcript is abolished in #825 and dramatically reduced in #1243, while the *AsbAS1* transcript can be readily detected in all lines. Expression levels are shown relative to #1243 *AsbAS1* transcript and normalised to transcript levels for the oat elongation factor1-α (EF1-α) using the ΔΔCq method.

6.2.4 – A CYP94D family member lies next to *CYP72A475*

Genome sequencing of *A. strigosa* S75 is currently underway in collaboration between the Osbourn lab and the group of Professor Bin Han, (China). Recent progress in this area has resulted in sequence data providing 89.4% coverage of the genome. This provided the opportunity to further explore the sequence surrounding the P450s characterised in this study.

BLAST searches of this database were used to identify *A. strigosa* genomic scaffolds containing *CYP72A475* and *CYP72A476* (**methods 6.4.10**). In the *A. atlantica* genomic sequence data, the *CYP72A475* and *CYP72A476* were located

on a single scaffold 5947bp apart. However they were not represented on a single scaffold in the *A. strigosa* genome (v0.6) because of a gap in the assembly (**Figure 6-8A**). Nevertheless, the two scaffolds that were retrieved extended the regions flanking the two P450s and were therefore searched for the presence of additional genes using GENSCAN [237]. This revealed two additional neighbouring genes. The first was identified as the *UGT74H5/Sad10* homologue, *UGT74H7*. *UGT74H7* sits 17.6kb upstream of *CYP72A476* (**Figure 6-8A**) and is also known to be expressed in root tips [228]. The second was a predicted P450 gene (assigned as CYP94D65) that was located 73.4 kb from the *CYP72A475* gene (**Figure 6-8A**). *CYP94D65* transcripts were also represented within the 454 transcriptomic database (contig 20449), indicating that this gene is likely to be expressed in *A. strigosa* root tips. qPCR profiling revealed that the expression profile of the *CYP94D65* was similar to those of other *Sad* genes although expression levels in whole roots was somewhat higher (**Figure 6-8B**). The similarity in expression pattern suggested that this gene may also have a role in avenacin biosynthesis.

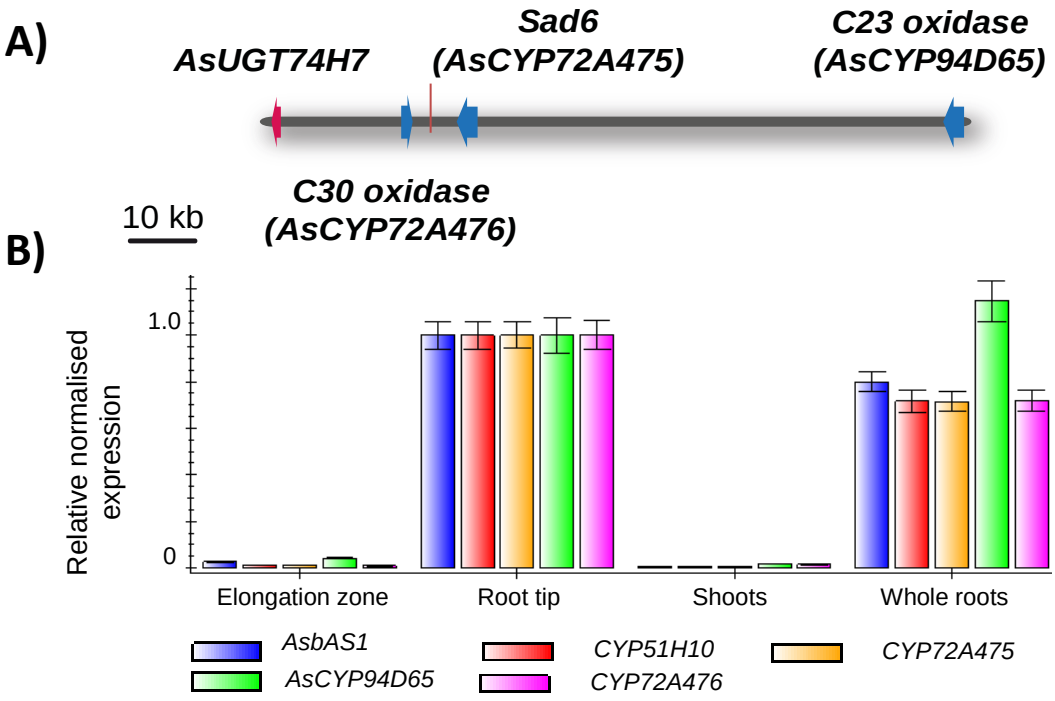


Figure 6-8: Clustering and coexpression of novel avenacin biosynthetic genes **A)** Schematic of the three oat P450s characterised in this study by analysis of genome sequence data. The red line between *CYP72A475* and *CYP72A476* represents the adjoining ends of the two *A. strigosa* scaffolds which are expected to be linked based on the sequence data from *A. atlantica*. These genes are expected to be located adjacent to the other known *Sad* genes based on genetic analysis (indicated in **Figure 6-2B**), but the intervening region has not yet been resolved. **B)** qPCR analysis of *CYP72A476, CYP72A475* and *CYP94D65* expression in wild-type (S75) *A. strigosa* seedlings. *AsbAS1* and *CYP51H10* are included as positive controls. Expression levels are shown relative to root tips and normalised using the ΔΔCq method to the oat elongation factor1-α (EF1-α).

6.2.5 – Product profiling of *CYP94D65*

To determine a possible role in avenacin biosynthesis, the *CYP94D65* gene was cloned from root cDNA into the pEAQ-*HT*-DEST1 vector and transformed into *A. tumefaciens* for transient expression in *N. benthamiana* (**methods 6.4.4**). As before, *CYP94D65* was coexpressed with *tHMGR/AsbAS1* and *tHMGR/AsbAS1/ CYP51H10*. In contrast to *CYP72A475* and *CYP72A476* which were inactive on β-amyrin, GC-MS analysis of the *tHMGR/AsbAS1/CYP94D65* sample revealed a new peak at 11.7 min. This was not present in a *tHMGR/CYP94D65* control (**Figure 6-9**). Analysis of the mass spectra revealed a putative parent mass ion of *m/z* 586 consistent with a hydroxylated β-amyrin. Furthermore, the most abundant ion was found to be *m/z* 218, which is also present in β-amyrin and represents the unmodified C*DE rings following rDA fragmentation [173]. Together, these observations indicate a derivative of β-amyrin hydroxylated within the ABC* rings (see **section 6.2.2**). Given that avenacins A-1 and A-2 contain a hydroxyl group at C-23, it is possible that CYP94D65 is a β-amyrin C-23 oxidase. This is also supported by the fact that *A. strigosa sad2* mutants (lacking functional CYP51H10) have been reported to accumulate a mixture of β-amyrin and 23-hydroxy-β-amyrin and indicates that the C-23 hydroxylase can utilise β-amyrin as a substrate [238]. Comparison of *sad2* mutant root extracts with the *tHMGR/AsbAS1/CYP94D65* sample showed that a peak within this sample had a mass spectrum identical to the putative 23-hydroxy-β-amyrin which could not be found in wild type controls (**Supplementary Figure 28**). This further supports the idea that CYP94D65 is a β-amyrin C-23 hydroxylase.

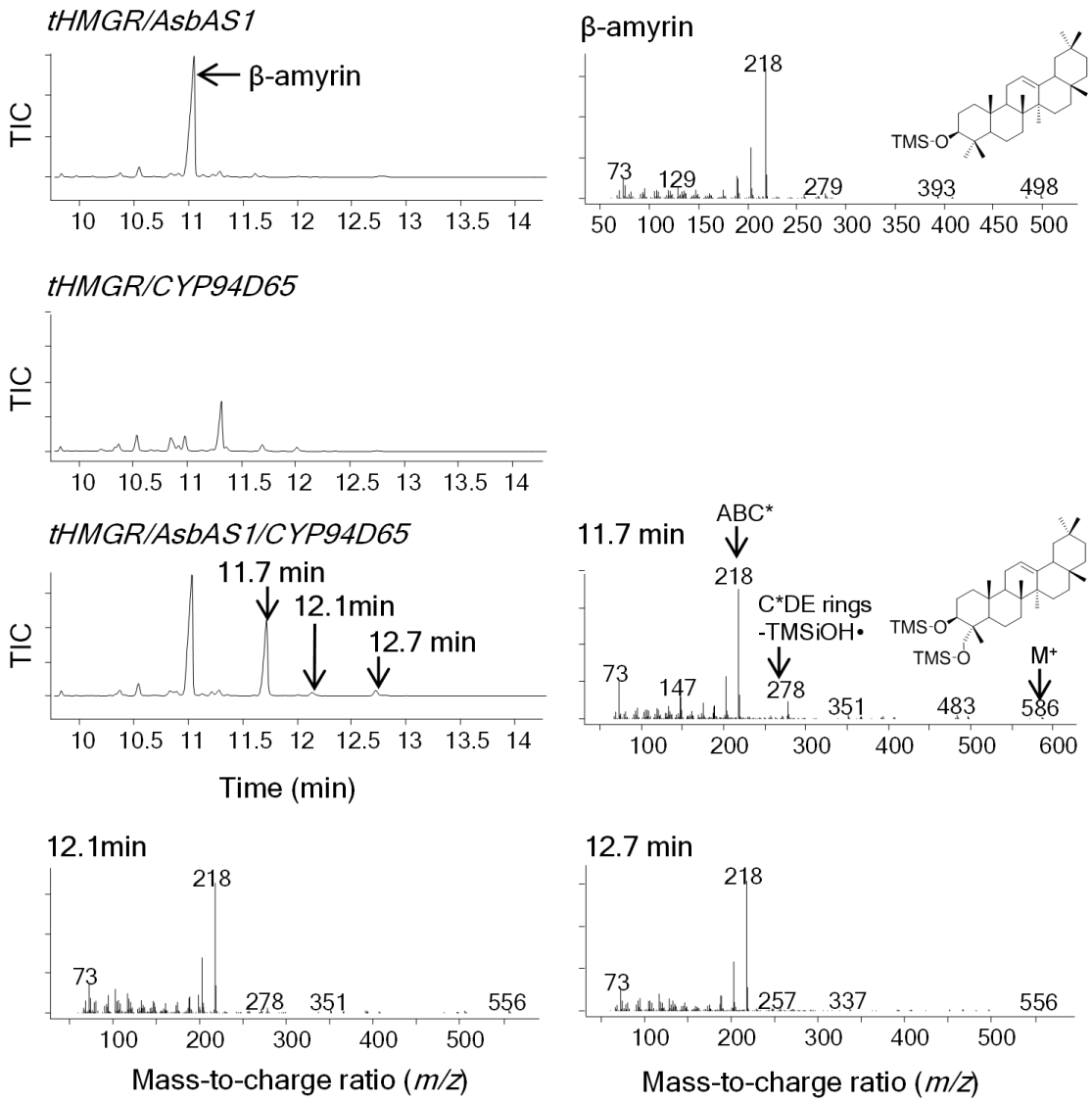


Figure 6-9: Activity of *CYP94D65* towards β-amyrin. GC-MS analysis of *N. benthamiana* leaves expressing *tHMGR/AsbAS1/CYP94D65* are shown above. The total ion chromatograms (TIC) are displayed on the left, with EI mass spectra of labelled peaks to the right and below relevant chromatograms. The peak at 11.7 mins closely matches a peak in *sad2* mutants (**Supplementary Figure 28**) which is believed to be 23-hydroxy-β-amyrin. Minor peaks at 12.1 and 12.7 minutes could also be observed which had putative parent ions at *m/z* 556. As described previously (section 4.3.2.2), these are likely artefacts of the extraction method resulting from transesterification of the product at 11.7 min.

For the sample containing *tHMGR/AsbAS1/CYP51H10*/*CYP94D65*, several new peaks could be seen relative to the control. The first of these coeluted with the EpHβA product at 12.9 mins, but had a clearly distinct mass spectrum, including a parent mass at *m/z* 690 suggesting a hydroxylated form of EpHβA (**Figure 6-10**). This spectrum was clearly distinct from either of the previous products of *CYP72A475* and *CYP72A476* and shared an ion at *m/z* 322 with EpHβA which (as described in Section 6.2.2) is suggested to be the C*DE rings. This product was therefore assigned putatively as 12,13β-epoxy, 16β, 23-dihydroxy-β-amyrin. The identity of the other peaks was less apparent. Two had likely parent masses of *m/z* 660 (13.4 and 14.5 min), while a third had a mass of *m/z* 704 (14.3 min) (**Figure 6-10**). The products with *m/z* 660 are likely to be the C-23-acetylated form of the EpHβA. The peak with *m/z* 704 is 14 mass units above the putative 23-hydroxy-EpHβA and suggests that an additional carbonyl group may be present. However the overall resemblance of the spectrum to that of 23-hydroxy-EpHβA indicates that the additional group was likely to be in proximity to the 23-hydroxyl group. One possibility is that this compound is a C-23 carboxylic acid (**Figure 6-10**), however avenacins carrying this functional group do not appear to have been reported previously. Isolation of this product will be necessary to confirm its identity.

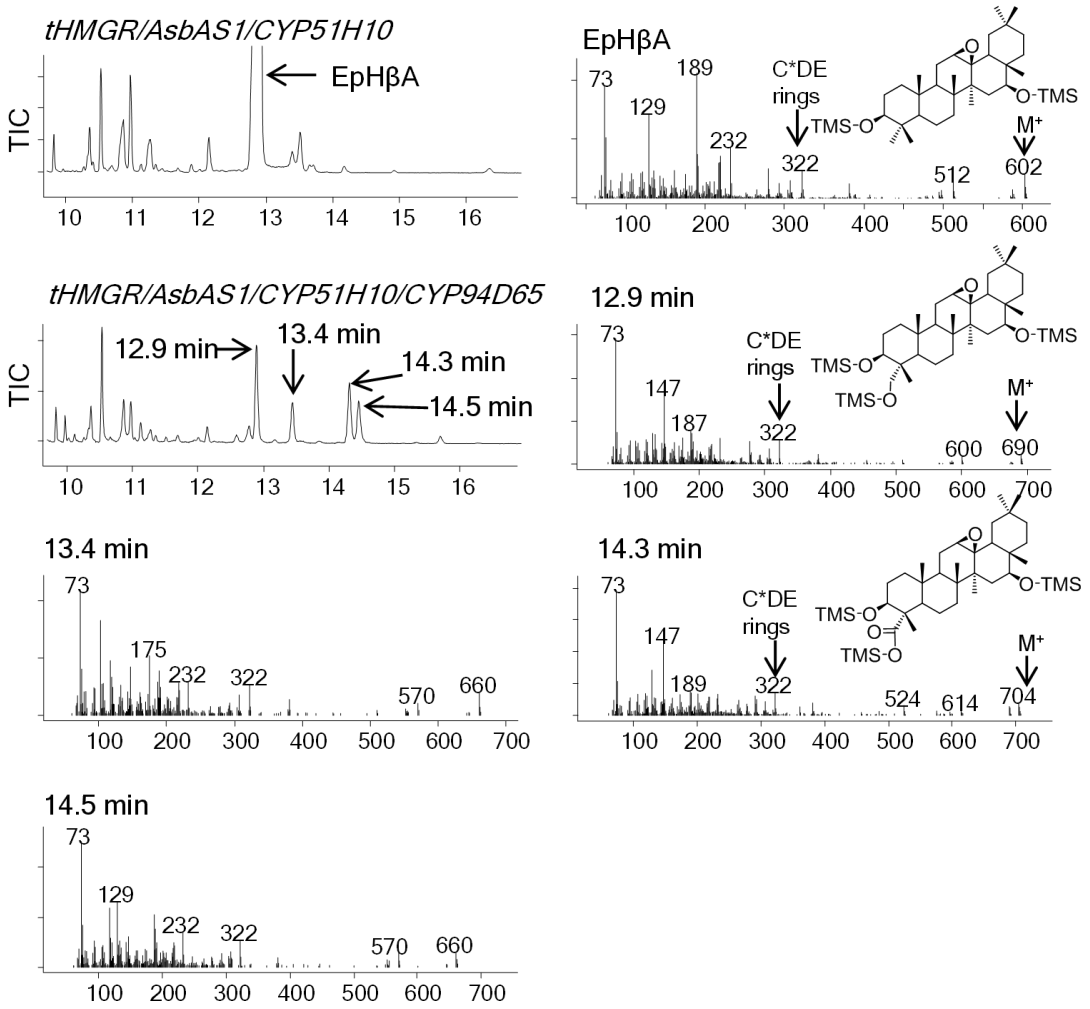


Figure 6-10: Formation of putative C-23 oxidised EpHβA in *N. benthamiana.* GC-MS total ion chromatograms (TIC) from leaf extract expressing *tHMGR/AsbAS1/CYP51H10/ CYP94D65* are shown on the left side, with the mass spectra for indicated peaks to the right or below chromatograms. Predicted structures are given inset above the mass spectra.

Finally, given that both CYP72A475 and CYP94D65 showed substantial oxidative capacity towards the EpHβA product, the compatibility of these enzymes was investigated. Analysis of extracts of *N. benthamiana* leaves coexpressing *tHMGR/AsbAS1/CYP51H10/CYP72A475/CYP94D65* showed a peak at 15.6 min which coeluted with a similar peak in the *tHMGR/AsbAS1/CYP51H10*/*CYP72A475* control but had a unique mass spectrum (**Figure 6-11**). This product had a putative parent ion at *m/z* 778, which is in accordance with EpHβA plus 2 derivatised hydroxyl groups. Based on these results, it is suggested that that the new product may be 12,13-epoxy, 16,21,23-trihydroxy-β-amyrin.

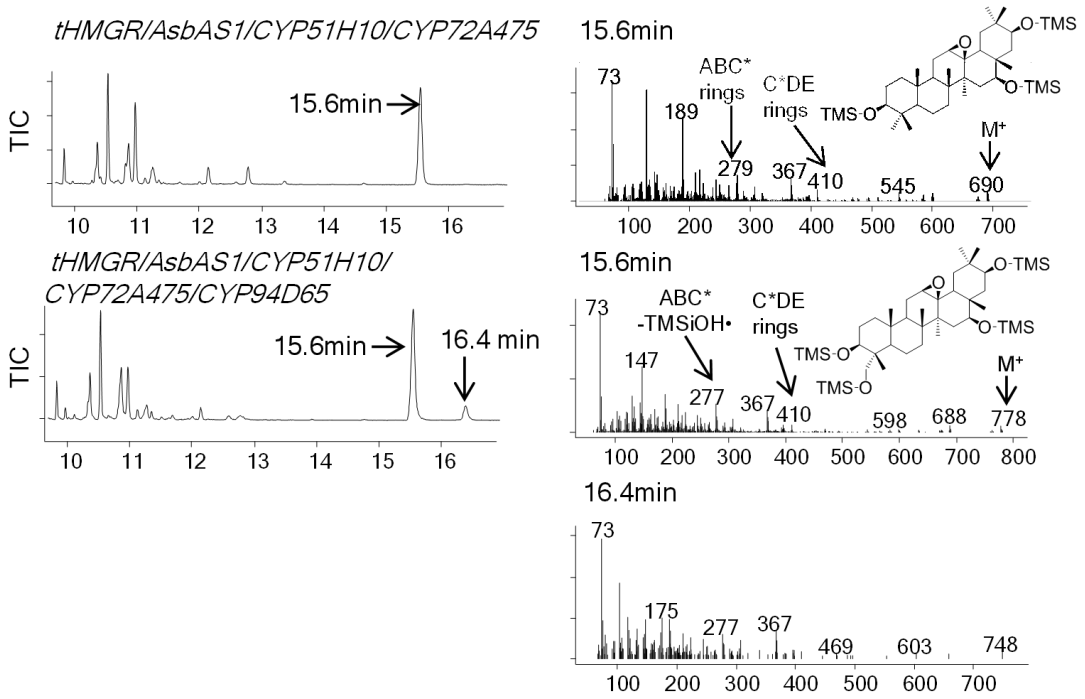


Figure 6-11: Formation of novel products through transient expression of *CYP94D65* with *CYP51H10* and *CYP72A475* in *N. benthamiana.* The total ion chromatograms are shown on the left side, with the mass spectra for indicated peaks to the right or below chromatograms. The major peak in the *tHMGR/AsbAS1/CYP51H10/ CYP72A475/CYP94D65* sample coeluted with a peak in the control (15.6 mins). Predicted structures are inset above the mass spectra.

In the last five years, members of the CYP94B and CYP94C families have been established as key players in jasmonate homeostasis [239]. Jasmonates are important plant hormones mediating responses to various environmental stresses such as wounding and pathogen infection. The C-12 oxidation of jasmonates by CYP94B3, CYP94C-1 and (to a lesser extent) CYP94B1 in *A. thaliana* has been shown to be an important mechanism for attenuating the jasmonate signalling response [233, 240-242]. It is likely that at least some of the same subfamily members are relevant to jasmonate inactivation in monocots [243]. The CYP94D family sits in a separate clade to the CYP94A, CYP94B and CYP94C families and its members are functionally uncharacterised [244]. The discovery here of CYP94D65 as an enzyme capable of oxidising β-amyrin suggests that (at least in oat), the CYP94D family may function in specialised metabolism. A fifth subfamily, CYP94E, is closely related to the CYP94D subfamily and is present exclusively in monocots [244]. Future characterisation of further monocot and dicot CYP94D members will be of interest to understand the functional evolution of this group.

6.3 – Conclusions and perspectives

6.3.1 – P450s involved in avenacin biosynthesis.

With the exception of CYP51H10, all of the currently known triterpene-oxidising cytochrome P450s are from dicot species (see **Table 1-1**). Here, three new triterpene-oxidising P450s were identified from oat and implicated in avenacin biosynthesis. *CYP72A475* was found to be synonymous with the *Sad6* gene (Section 6.2.3). It is expected that this P450 catalyses oxidation of the C-21β position of the triterpene scaffold allowing for acylation. At the time of writing, no other C-21β oxidase has yet been described. A second closely-related P450, CYP72A476, is expected to be a C-30 triterpene oxidase (Section 6.2.2). No *A. strigosa* mutants have yet been identified as deficient in this function, although they may be present amongst the uncharacterised *sad* mutants. Identification of such mutants may help to shed light on the importance of this functionality for antifungal activity.

The third P450, CYP94D65 is likely to be the avenacin C-23 oxidase. One other P450 has previously been characterised as an oleanolic acid C-23 oxidase (CYP72A68v2). The CYP72 and CYP94 families are part of the CYP72 and CYP86 P450 clans, respectively, and thus these enzymes have likely evolved their C-23-oxidising activity independently (**Figure 6-12**). Interestingly, CYP94D65 appears to be the first triterpene oxidase identified within the CYP86 clan (**Figure 6-12**). This therefore raises questions about whether members of this clan may have wider significance in triterpene biosynthesis in monocots.

Although the genome of *A. strigosa* has not yet been finalised, the three P450s investigated in this study are thought to be located across a 100kb region with no obvious intervening genes (**Figure 6-8A**). The *CYP72A475/Sad6* gene is also known to be closely linked to the cluster [94]. The discovery of putative C-23 and C-30 hydroxylases within the vicinity suggests the cluster encompasses at least two additional genes. Therefore at least eight unique genes are likely to be present within the avenacin cluster.

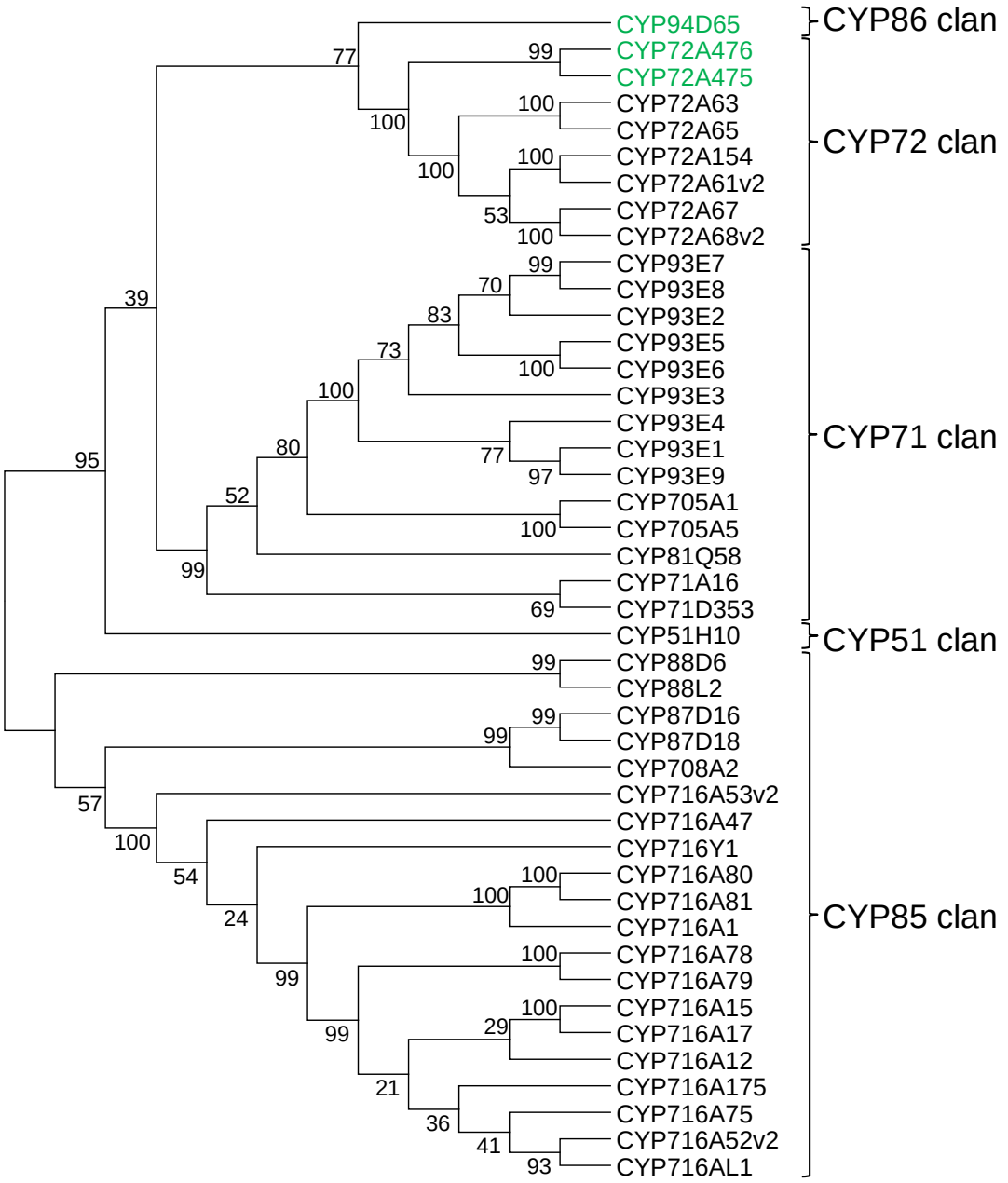


Figure 6-12: Phylogenetic tree of cytochrome P450s implicated in triterpenoid biosynthesis. The three oat P450s characterised in this study are highlighted in green. CYP72A475 and CYP72A476 cluster with other CYP72A subfamily members as expected. CYP94D65 occupies a separate clade to the other characterised P450s including another C-23 oxidase (CYP72A68v2) and the two are found within different P450 clans. The tree was constructed using the Neighbour-joining method with 10,000 bootstrap replicates (bootstrap support shown as percentages). Gaps were eliminated from the data set.

These P450s offer an exciting opportunity to understand the intricate structure-function relationships underlying the bioactivity of these molecules. A combinatorial approach may be used to generate avenacin analogues lacking one or more functional groups (**Figure 6-13**). This may also be facilitated by the recently-generated mutant variants of CYP51H10 capable of producing either the C-12-13 epoxide or C-16-hydroxyl group (see section 4.3.3). Furthermore, as demonstrated in section 4.3.2, P450s from other species may be used to extend the modifications beyond those found naturally in the avenacins. Such studies will be particularly informative for engineering antifungal saponin biosynthesis in vulnerable crop species.

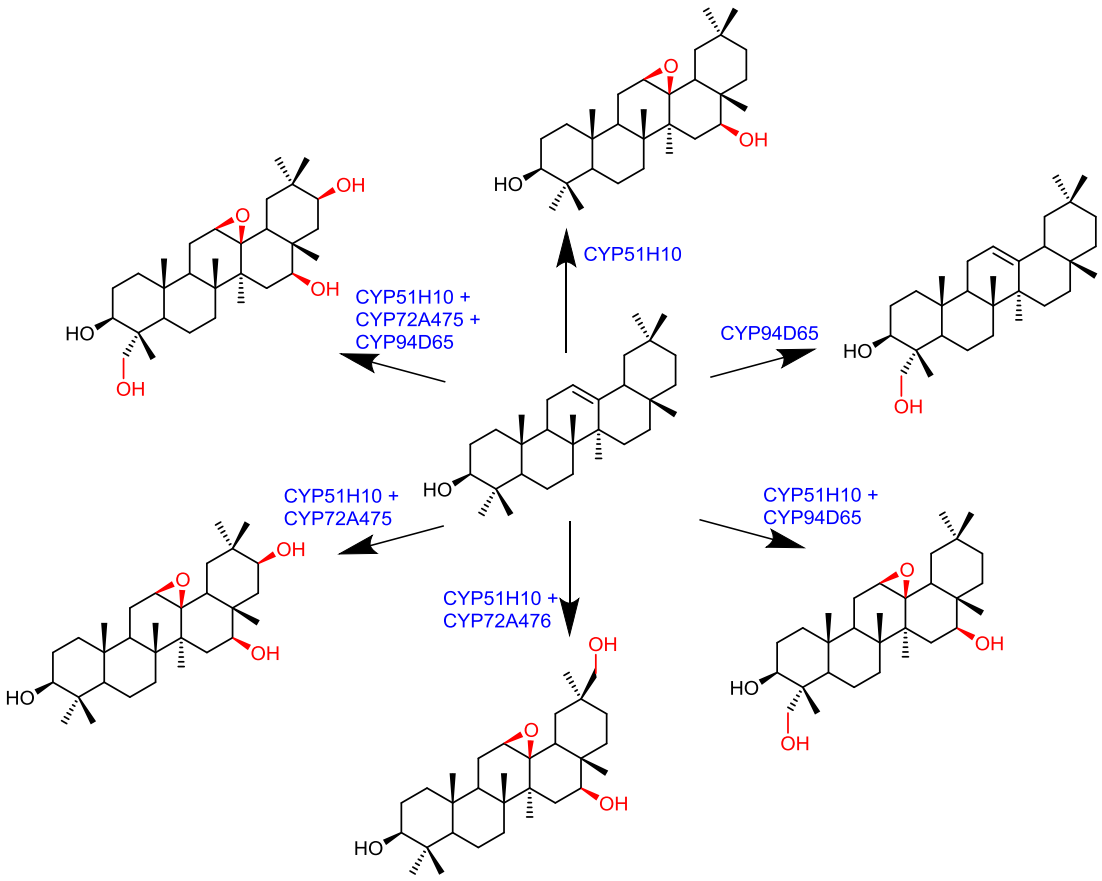


Figure 6-13: Combinations of oat P450s expressed in this study and the predicted products. Clockwise from top: EpHβA; 23-hydroxy-β-amyrin; 12,13β-epoxy,16β,23-dihydroxy-β-amyrin; 12,13β-epoxy,16β,30-dihydroxy-β-amyrin; 12,13β-epoxy,16β,21β-dihydroxy-β-amyrin; 12,13β-epoxy,16β,21β,23-trihydroxy-β-amyrin

6.4 – Materials and Methods for Chapter 6

6.4.1 – Identification of P450 candidates

Genomic sequence data from *A. atlantica* were kindly provided by Dr Tim Langdon. The entire genomic scaffold sequences (approximately 5-50kb each) were used as queries to search (by BLASTn) against the *A. strigosa* 454 root tip transcriptomic database. This allowed identification of orthologous *A. strigosa* genes which were expressed in the root tips. The *A. strigosa* contigs with the lowest e-values were selected (typically e = zero). These were in turn used as BLAST queries against the Genbank database to ascertain the relevant enzyme families. Non-P450 enzymes were subsequently eliminated from the list of candidates. Sequences from the oat 454 database were translated in all 6 reading frames to identify the expected coding sequence (http://web.expasy.org/translate/).

6.4.2 – Growth and harvesting of *A. strigosa* seedlings

Seeds of *A. strigosa* were dehusked by hand and sterilised using a 5% sodium hypochlorite solution for 5 minutes. The following steps were performed in a sterile laminar flow hood. The seeds were washed multiple times using sterile water and air dried on sterile filter paper. Seeds were transferred to distilled water agar medium in 10cm petri dishes with the radicle facing towards the bottom of the plate. Plates were stored at 4°C in the dark for 24 hours before transfer to a growth cabinet (22°C constant, 16 hour lighting). Seedlings were incubated for 4-7 days depending on rate of growth.

For harvesting different tissues for expression profiling of the *A. strigosa* S75 accession (qPCR or RT-PCR), a set of the seedlings were divided into whole roots and shoots, with another set having their roots further divided into the terminal 2-3mm ‘tips’ and the elongation zone (~5mm) immediately above the tips. These steps were performed on a glass plate placed on dry ice and the harvested tissues were transferred immediately to microcentrifuge tubes placed in dry ice. Harvested samples were stored at -80°C until needed.

6.4.3 – RNA extraction from oat seedlings and cDNA synthesis

For extraction of RNA from oat tissues, tissues were taken from the -80°C freezer and transferred to liquid nitrogen. Using a micropestle, the tissues were ground to a fine powder. RNA extraction was performed using a RNeasy Plant Mini kit (Qiagen) according to the manufacturer’s instructions with minor changes noted here. Briefly, following transfer of the samples to the RNeasy Mini Spin Column, an on-column

DNase digestion was performed (Promega RQ1 DNase). The columns were washed with 350µl buffer RW1 and centrifuged. 80µl DNase (prepared according to manufacturer's instructions) was transferred to the column membrane and incubated for 1 hour at room temperature. Columns were washed a second time with 350µl buffer RW1 and the rest of the protocol was followed as normal. RNA concentration was checked using a Thermo Nanodrop 2000 and the integrity of RNA was confirmed by running 5µl of the samples on a 1% agarose gel (plus ethidium bromide) run at 100V for 15 minutes. cDNA synthesis was performed using the Superscript-II First Strand Synthesis System (Invitrogen) with the supplied OligodT primers. For all samples, 1µg total RNA was used per synthesis reaction to give approximately equal template concentrations for downstream applications.

6.4.4 – Cloning of candidate *A. strigosa* P450s

Cloning was performed by Dr Aymeric Leveau from *A. strigosa* S75 root tip cDNA, using primers as detailed in **Table 6-2**. Amplified products were cloned into pDONR207 for sequencing before transfer to pEAQ-*HT*-DEST1. The constructs were transformed into *A. tumefaciens* LBA4404 as detailed in section 2.1.2.3

6.4.5 – Cloning Primers

Name	Adapter/ 5'sequence	Gene-specific sequence
CYP71E22-GW-F	attB1F-3	ATGGTGATCCTGCACCCCTCGC
CYP71E22-GW-R	attB2R-3	CTAGGCGCGCGGCGGCACGTATAG
CYP72A475	attB1Fs-2	ATGGCACTGCTTGTAGAGTTGTTC
CYP72A475	attB2Rs-2	TCATATAGCCTTGAGCTTAATATG
CYP72A476-GW-F	attB1F-3	ATGAGTAGTTTTGAGATGGAATTC
CYP72A476-GW-R	attB2R-3	TCMGATGGCCTTGAGAGTAATCTGAGC
CYP88A75-GW-F	attB1F-3	ATGTGGTGGCCGTGGTGGGCGG
CYP88A75-GW-R	attB2R-3	TCAGTGTCCACCGGAGCCGGAGG
CYP89N1-GW-F	attB1F-3	ATGGAGGGGCTCACCATCCTCTTG
CYP89N1-GW-R	attB2R-3	TCMCTGCTTCATGTCCCTTGGGATGATG
CYP94D65	attB1Fs-2	ATGGAGCCGGCGCCCTTGAGCTC
CYP94D65	attB2Rs-2	TTAGTACTGATTTCTAGGTGC
CYP706C45-GW-F	attB1F-3	ATGCCACATACGTTAACGTTTTCC
CYP706C45-GW-R	attB2R-3	TTACTCAGACGAATAATAGAGTTC
CYP711A54-GW-F	attB1F-3	ATGGGCCGCCTGGCAGAAGCACTG
CYP711A54-GW-R	attB2R-3	TCATATGTTCCGCTCGATGACTTGGAG

Table 6-2: Primers used to clone candidate P450s from *A. strigosa*. Gene-specific sequences are given on the right, with 5' adapters for Gateway cloning indicated in the middle column and detailed in **Table 2-1**

6.4.5.1 – Cloning and sequencing *CYP72A465* from #1243

Seedlings from lines #825 and #1243 were grown as in **section 6.4.2.** Roots were harvested and used to synthesise cDNA (**section 6.4.3**). A PCR was performed

using Phusion polymerase (NEB) for cDNA from each line. A strong band of the correct size was formed in the #1243 cDNA template. Two very weak bands were formed in the #825 sample (~1.5 and 2.5kb). Gel purification of the bands was performed (QIAquick gel extraction kit, Qiagen) and cloned into pDONR207 before transforming *E. coli* DH5α. Colonies were recovered on all plates and two from each were sent for sequencing with attL1-F/2-R primers (**Table 2-1**). The #1243 mutant full-length *CYP72A465* transcript was found to contain a G1068A point mutation. None of the clones from #825 contained the *CYP72A465* gene.

6.4.6 – Infiltration, extraction and analysis of *N. benthamiana* leaves

Infiltration was performed by Dr Aymeric Leveau and James Reed according to the protocol in section 2.1.3.3. Extraction of the leaves was performed according to **section 4.5.5.** For GC-MS analysis, ethyl acetate leaf extracts of *N. benthamiana* plants were dried under nitrogen before adding 50µl Tri-Sil Z reagent (Sigma). Samples were heated at 70°C for 30 minutes before loading onto the GC-MS. Samples were analysed according to **section 2.1.5.2.**

6.4.7 – LC-MS analysis of oat mutants

LC-MS analysis of oat mutants was performed by Dr Sam Mugford as previously described [226].

6.4.8 – RT-PCR profiling

Primers for RT-PCR were designed using Primer3 [245]. The following parameters were specified: 18-23nt, T_m = 57-62°C, GC % = 30-70% and product size = 270bp. For most candidates, primers were designed towards the 3' end of the cDNA. For GAPDH, primers were designed to span predicted intron sequences to identify the presence of contaminating genomic DNA by the product size (gDNA = 427bp, cDNA = 246bp). 25µL PCR was carried out using GoTaq Green polymerase (Promega) with primers listed in (**Table 6-4**) and thermal cycle detailed in **Table 6-3**. cDNA templates were from different tissues of 4-day old S75 seedlings (detailed in **section 6.4.2**). PCR products were analysed using a 2% agarose gel.

	Step	Temp (°C)	Time
30 cycles	Denaturation	98	30 sec
	Denaturation	98	10 sec
	Annealing	55	10 sec
	Extension	72	60 sec
	Extension	72	3 min

Table 6-3: RT-PCR thermal cycle parameters

	Name	Sequence
RT-PCR expression profiling	GAPDH-RT-PCR-F	GGTGGTCATTTCAGCCCCTA
	GAPDH-RT-PCR-R	CTCCCACCTCTCCAGTCCTT
	AsbAS1-RT-PCR-F	AACTACGCCAACTACCGCAA
	AsbAS1-RT-PCR-R	ACGATTAAACCGCGGGACTT
	CYP51H10-RT-PCR-F	ACACACTTACAAGGACCCCG
	CYP51H10-RT-PCR-R	CGCACACAAAGATACATCCGT
	CYP72A475-RT-PCR-F	CCGAGGCCTCAAAGAATCCA
	CYP72A475-RT-PCR-R	TGACCCATTATTACACTGTTGCA
	CYP72A476-RT-PCR-F	TTCCAAGGCGCATATCTCCC
	CYP72A476-RT-PCR-R	TCAGATCCCGATTCATCCGC
	CYP706C45-RT-PCR-F	CTCGAGTTCAACCCGGAGAG
	CYP706C45-RT-PCR-R	AATTCAAGTTCATGCCGCCT
	CYP71E22-RT-PCR-F	GGTCAATGGCGCTGATGAAC
	CYP71E22-RT-PCR-R	TTTATTTGGGAACGACGGGC
	CYP89N1-RT-PCR-F	CACACTGCATGCCGAGTACT
	CYP89N1-RT-PCR-R	TGAATTAAGAGGACCGGGCC
	CYP88A75-RT-PCR-F	GACATGATGGACCGGCTGAT
	CYP88A75-RT-PCR-R	AAAGTGTCTCGAGTCGCCTG
	CYP711A54-RT-PCR-F	GCGATACGGGCCAATCTACA
	CYP711A54-RT-PCR-R	CGGCCCTGTTGATGTAAGGT

Table 6-4: Primers used for RT-PCR expression profiling of P450s *A. strigosa.*

6.4.9 – qPCR analysis

Primers for *EF1-α*, *AsbAS1* and *CYP51H10* were the same as described previously [29]. For *CYP72A475*, *CYP72A476* and *CYP94D65*, primers were designed using primer 3 [245] to give similar amplicons and annealing temperatures. One primer in each pair was designed to span a predicted exon-exon boundary by comparison of the cDNA sequence with the *A. strigosa* genomic sequence. Primers for *CYP72A475* were designed to anneal upstream of the mutation identified in line #1243. Primer sequences are listed in **Table 6-5.**

For expression profiling of *AsbAS1, CYP51H10, CYP72A475, CYP72A476* and *CYP94D5* in wild-type plants, cDNA templates were synthesised from root and shoot tissues as described in **sections 6.4.2** and **6.4.3**. For *CYP72A475* expression profiling in S75, and #825 and #1245 mutants, cDNA templates were synthesised from whole roots from 5-7 day-old seedlings.

Samples were set up in triplicate in white 96 well plates (4titude) using the DYNAMO Flash SYBR Green kit (Thermo) according to the manufacturer's instructions. qPCR profiling was conducted using a BIO-RAD CFX96 Touch Real-Time PCR Detection system. All targets were measured using the ΔΔCq method relative to the *A. strigosa* housekeeping gene *EF1-α* (elongation factor 1-α).

Name	Sequence
EF1-α-qPCR-F	TCCCCATCTCTGGATTTGAG
EF1-α-qPCR-R	TCTCTTGGGCTCGTTGATCT
AsbAS1-qPCR-F	TGCGGAATTCACAAAGAACA
AsbAS1-qPCR-R	GCTTGGCTTCTGTCGGAATA
CYP51H10-qPCR-F	TCGACAGGAAGTGGAGGAGT
CYP51H10-qPCR-R	ATCTCGGACCTCACTTCCAA
CYP72A475-qPCR-F	GGAAGCTGCTAGGTAATGG
CYP72A475-qPCR-R	CCGACAACTTGAGCTTGA
CYP72A476-qPCR-F	GCATTCAACGAGATGCTC
CYP72A476-qPCR-R	AAATTGTTCTTGATGGCCTG
CYP94D65-qPCR-F	GTCATTACCAGGTTCGTGT
CYP94D65-qPCR-R	GATGATGTTGTCGGCGTA

Table 6-5: Primers used for q-PCR expression profiling of *A. strigosa* P450s.

6.4.10 – BLAST searches of the *A. strigosa* genome

Searches were performed by Dr Aymeric Leveau. Briefly, genomic sequence scaffolds from version 0.6 of the *A. strigosa* genome were kindly supplied by Prof Bin Han (National Centre for Gene Research, China). A database of the sequences was compiled using CLC (Qiagen) and BLAST searches were conducted using the *A. strigosa* 454 coding sequences for *CYP72A475* and *CYP72A476* to identify scaffolds containing the genomic sequences.

6.4.11 – Extraction and GC-MS analysis of oat roots

For growing seedlings, ten seeds for each of three *A. strigosa* lines were used (S75/*sad1* line #109/*sad2* line #1027) [38]. Seeds were sterilised and grown as in **section 6.4.2** and harvested after 5 days. The whole roots were cut away from the rest of the seedling using a razor blade and stored at -80°C. The roots were

lyophilised overnight and weighed the following day. The roots were then saponified and extracted as in **section 3.4.7**. GC-MS analysis was performed according to **section 2.1.5.1.**

Chapter 7

General Discussion

7.1 – Introduction

The abundance and diversity of the triterpenes is testament to their importance throughout the plant kingdom. Naturally, these molecules have also found a wide range of applications throughout human history with uses ranging from traditional medicines to cleaning agents. Triterpenes still have huge potential for development in the commercial, agrochemical and pharmaceutical sectors, however much of this potential remains unrealised due to the difficulties in sourcing and selective modification of these compounds. Recent understanding of the biosynthetic pathways of these molecules in a range of plant species now provides for the first time an opportunity to use heterologous host production platforms to access and exploit the full potential of the triterpenes. In the present work, transient expression in *N. benthamiana* was developed as a powerful and rapid strategy for functional analysis of triterpene biosynthetic enzymes and for triterpene engineering. The value of this expression platform for combinatorial triterpene biosynthesis is demonstrated. This system is also used to generate suites of analogues of known and new-to-nature triterpenes in milligram quantities and the biological activities of these derivatives is investigated.

7.2 – *N. benthamiana* as a platform for triterpene production

At present, yeast has been the major organism of choice for functional characterisation of triterpene biosynthetic enzymes. Efforts have also been made to optimise production in this host [108, 147, 148]. Nevertheless, *N. benthamiana* was considered to have promise because it is a plant-based system and might have certain advantages over yeast (e.g. availability of precursors, coenzymes and appropriate subcellular compartmentalisation mechanisms). Prior to the start of this project, there were few reported cases of triterpenes being successfully produced in *N. benthamiana* [85, 93]. Therefore a major aim of the present PhD thesis was to determine the suitability of *N. benthamiana* as a production platform for a range of different triterpenes. To this end, a triterpene 'toolkit' was established, consisting of a collection of genes encoding characterised triterpene biosynthetic enzymes. These genes were each cloned into expression constructs and introduced into *A. tumefaciens* strains. Each *A. tumefaciens* strain carried a single gene encoding an OSC or P450, allowing the strains to be used in a modular fashion, enabling rapid assembly and validation of new combinations.

Agroinfiltration of the OSC-carrying strains in *N. benthamiana* confirmed that a diverse range of triterpene scaffolds could be produced in this system (Chapter 4).

Further experiments demonstrated that several triterpenes including β-amyrin and dammarenediol-II could be further oxidised *in planta* by coexpression of previously characterised P450s with the relevant scaffold-generating OSC. In each case, the major product observed upon GC-MS analysis conformed to the expected product. This is particularly encouraging as perhaps one of the factors presently limiting more widespread use of *N. benthamiana* is the frequency with which unintended modifications to the produced compound are observed [140]. For example, certain diterpenes have been reported to be modified by endogenous oxidases [128]. Although unknown peaks were found in some of the chromatograms in the present study, in many cases these could be resolved as extraction or analytical artefacts (section 4.3). Glycosylation has also been reported upon production of simple and oxidised monoterpenes and sesquiterpenes [122, 127, 132, 234]. While production of triterpene monoglucosides in *N. benthamiana* has also been reported to elicit further glycosylation, this did not appear to occur for triterpene aglycones [105]. As GC-MS was the major analytical tool employed throughout this study, it is possible that some conversion of simple triterpenes occurred in *N. benthamiana* which were not detected. Nevertheless, the abundance of the triterpene products as assessed by GC-MS suggested that substantial quantities are present in the expected form. Ultimately in the present study it was possible to use this system to produce and purify a number of oxidised triterpenes in milligram quantities. These results indicate that *N. benthamiana* is a very good system for the production and oxidation of a diverse range of triterpenes.

7.3 – Further enhancing triterpene production in *N. benthamiana*

Optimising triterpene production in *N. benthamiana* is desirable for increasing the levels of product, facilitating purification, structural analysis and use for downstream applications. Metabolic engineering of the host species is one way in which the production capacity of the host system can be improved. However metabolic engineering in *N. benthamiana,* (and indeed plants in general) is in its infancy compared to microbial systems such as yeast, for which an array of molecular tools are available [246]. Nevertheless, there is clear promise for metabolic engineering of *N. benthamiana* by transient expression. Strategies investigated include alternate cellular compartmentalisation [127, 132, 133] and use of rate-limiting enzymes [127, 128, 132]. However such strategies had not been demonstrated for triterpenes. Therefore, Chapter 3 investigated strategies for improving triterpene production in *N. benthamiana.* To this end, potential rate-limiting steps within the native triterpene-producing MVA pathway were considered. Coexpression of a truncated,

feedback-insensitive form of 3-hydroxy,3-methylglutaryl-CoA reductase (tHMGR) increased the yields of β-amyrin by approximately 4-fold when compared to plants expressing β-amyrin synthase (AsbAS1) alone (section 3.2.2). Inclusion of tHMGR also enhanced production of oxidised forms of β-amyrin, including 12,13β-epoxy,16β-hydroxy-β-amyrin (EpHβA), when AsbAS1 was co-expressed with P450s. To test how the use of tHMGR would impact isolated yields, this product was scaled up and purified, resulting in a total of 79 mg of pure EpHβA, equivalent to 4mg/g dw (section 3.2.5). This is almost four-fold higher than previous reports [85].

That such a yield should be obtained with relatively little engineering is very encouraging for future development. Thus further investigation of other candidate rate-limiting enzymes within the plant cell should be considered for future work. Although tHMGR gave the largest increase in β-amyrin, squalene synthase (SQS) was also found to improve β-amyrin production. Further increases in β-amyrin were consequently observed when both *tHMGR* and *SQS* were expressed together with *AsbAS1* (section 3.2.3). This demonstrates that additional yield increases may be obtained by co-expressing addition upstream precursor pathway genes. In a recent example, overexpression of every gene in the MVA pathway in *A. thaliana* illustrated that several of the genes resulted in enhanced sterol accumulation [169]. Thus, one strategy to improve yields might simply be the overexpression of every gene in the MVA pathway, a strategy that has been successfully used to improve sesquiterpene production in yeast [247]. Rather than coinfiltrating several *A. tumefaciens* strains carrying single MVA genes, a more efficient method would be to integrate multiple genes into a single vector. Golden Gate cloning [135] is one such approach which facilitates integration of multiple genes into a single binary vector for expression in *N. benthamiana.* This system is currently being utilised in the Osbourn lab. Alternatively, several genes might be fused into a single open reading frame separated by sequence encoding a 'self-cleaving' peptide. A viral 2A peptide has been successfully applied for transient expression in *N. benthamiana* to fuse an amorphadiene synthase, FPS and tHMGR enzymes [132].

In conjunction with overexpressing rate-limiting enzymes, other strategies may also have merit for further boosting triterpene yields. For example, the loss of metabolic intermediates might also be prevented by using transient expression to silence endogenous *N. benthamiana* enzymes which compete with triterpene biosynthesis (such as cycloartenol synthase). This has been shown to be a viable strategy for improving sesquiterpene content in *N. benthamiana* transient expression studies

[131]. Spatial control of key enzymes in the pathway might also be used to channel metabolism towards triterpene biosynthesis. This could be achieved through fusion of key enzymes into a single polypeptide [248] or alternatively through tethering multiple enzymes to protein scaffolds [249].

The relative abundance of individual proteins might be 'tuned' to overcome metabolic bottlenecks and optimise product yield. The pEAQ vectors used in the present study have significant potential for such an approach. For example, control over individual proteins might be achieved through use of both wild type and *HT* CPMV 5' UTRs. In this way the stoichiometry of Bluetongue virus proteins could be optimised allowing for the enhanced formation of virus-like particles [250]. Modifications to the 3' UTR also have similar potential for altering protein levels [251]. It should be noted that significant variation in relative expression of individual proteins may be observed between plant cells when the respective T-DNA encoding these proteins are derived from separate strains of *A. tumefaciens* [252]. Therefore, efforts to explore the above approaches would be best served through integration of several genes into a common vector as described above.

7.4 – Expanding triterpene diversity

Triterpenes are amongst the most chemically diverse terpenes and the overwhelming majority of these products are oxygenated [73]. Oxidation provides the functional handles for further modification by tailoring enzymes such as glycosyltransferases and hence serves a key role in generating triterpene structural diversity. As described above, a core aim of this PhD was to assemble a collection of known oxidases for engineering triterpenes in *N. benthamiana*. In addition, this collection was also expanded through functional characterisation of putative C-21, C-23 and C-30 P450s from oat with a role in avenacin biosynthesis (Chapter 6). Avenacins are triterpene glycosides (saponins) with strong antifungal activity, thus these enzymes represent important advances in our understanding of this pathway. The ability to produce triterpene saponins is very rare amongst cereals. Thus understanding avenacin biosynthesis provides an opportunity for engineering avenacin production into crop species which are susceptible to fungal pathogens such as take-all disease, a major pathogen of wheat.

In addition to recreating known triterpenoids in heterologous hosts, an important application for the collection of triterpene biosynthetic enzymes is engineering the production of new triterpenes through combination of enzymes from different species. To demonstrate this principle, a core set of the β-amyrin oxidases from the

triterpene toolkit were systematically combined in Chapter 4. This approach showed that almost all of the pairwise combinations resulted in the formation of new products. Two such hybrid compounds, 11-oxo,24-hydroxy-β-amyrin and 11-oxo-oleanolic acid were subsequently purified and the identity of the compounds was verified by NMR. Work is currently ongoing to verify the identity of the other compounds, but the expected products do not appear to have been previously isolated. This clearly demonstrates the utility of this approach for accessing novel structural diversity.

7.4.1 – Engineering triterpene glycosides

Triterpene glycosides (saponins) found within the plant kingdom frequently feature an array of diverse sugars including glucose, arabinose, rhamnose, xylose and galactose [72]. Most of the necessary sugar donors required for their synthesis are not present in microbial systems. Therefore an important test of the potential of *N. benthamiana* as a production host will be the capacity to produce a range of diverse triterpene glycosides. In the present work, it was demonstrated that high levels of the ginsenoside scaffolds protopanaxadiol and protopanaxatriol could be produced through coexpression of a mutant form of *AsbAS1* with P450s from *Panax ginseng*. These scaffolds are present in over 90% of ginsenosides [44] and are normally present as glycosides. Recent progress in elucidating UDP-dependent glycosyltransferases (UGTs) from *P. ginseng* [253] has allowed for the reproduction of a number of key ginsenosides in *N. benthamiana.* (Dr Thomas Louveau and Charlie Owen, Osbourn lab). Furthermore, a combinatorial approach using UDP-dependent glycosyltransferases (UGTs) in the Osbourn lab has revealed that a range of triterpene glycosides can be produced in *N. benthamiana* upon coexpression of triterpene-modifying UGTs from various plant species. This includes the ability to glycosylate the scaffold with non-glucose sugars. Furthermore, glycosides of several P450 combinatorial products described within this thesis have also been produced (Dr Thomas Louveau, personal communication). This further demonstrates that *N. benthamiana* may be an excellent system for generating structurally diverse triterpenoids and saponins.

An obvious question to consider is the extent to which combinatorial biosynthesis can be applied for a multi-step pathway. For example, while some of the pairwise P450 combinations used herein were highly active (such as CYP88D6 and CYP93E1), others gave relatively poor accumulation of new products (such as CYP51H10 and CYP716A12). Such bottlenecks will be detrimental to further elaboration of the scaffold. Protein engineering approaches might therefore be

applied to enhance the activity of individual P450s [110]. Alternatively, functionally conserved orthologues of individual enzymes have been shown to have differences in their catalytic efficiency towards a common substrate [99]. By extension, similar differences might be observed in efficiency towards novel substrates. In the present thesis, two C-30 oxidases (CYP72A63 and CYP72A65) were investigated using the combinatorial biosynthesis approach. Of these, CYP72A65 was found to be a better candidate for use in combinatorial biosynthesis (Chapter 4). Therefore this stresses the importance of fully exploiting the available sequence data to not only to identify enzymes with novel functionalities, but also functionally 'redundant' enzymes which may be better suited for individual applications.

7.4.2 – Expanding the enzyme collection further

The number of characterised genes in triterpene biosynthesis has increased significantly in recent years. However there are still relatively few species for which multi-step triterpene biosynthesis pathways have been defined. Rapid decreases in sequencing costs mean that the available plant sequence data is likely to increase hugely over the coming years. Data from projects such as the 1000 plants initiative (1KP - www.onekp.com) offer access to transcriptomic data from a diverse variety of plant species [254]. These may be mined for new OSCs or common triterpene oxidase families. Additionally, the growing number of sequenced plant genomes provides an opportunity to use computational methods to search for clustered triterpene biosynthetic genes [255].

Ultimately, as the number of available triterpene biosynthetic genes increases, the opportunity for combinatorial biosynthesis will increase exponentially. Hence high-throughput combinatorial screening approaches should be considered. A recent approach developed by Dr T. Rademacher (Fraunhofer IME, Aachen, Germany) allows generation of a 3-D cell 'pack' from plant cell cultures which can be transformed by *A. tumefaciens*. This approach is suited to a 96-well format and might be applied to high-throughput screening of triterpene combinatorial libraries while maintaining the benefits of the plant-based expression and modular *A. tumefaciens* toolkit approach.

7.5 – Understanding structural features behind biological activity

The number of possible chemical structures (chemical space) is vast; small organic molecules of 30 atoms or less have been estimated to number more than 10^{60} compounds [256]. However within this vast space, it is now recognised that a much smaller subset of these molecules occupy 'biologically relevant' chemical space

[257]. Natural products inherently occupy this area and consequently have been hugely important for the discovery and development of therapeutics [186]. Triterpenes, like most natural products, have a high number of chiral centres and display structural rigidity, properties which may be important mediators of biological activity, such as their ability to bind protein targets [258]. Equally, their similarity to other biological products such as sterols may enhance their bioavailability, such as through interaction with cellular transporters. Therefore triterpenes can be considered excellent bioactive leads from which explore further structural variants.

Much of the vast structural variation present within the triterpenes has not yet been accessed. The current understanding of triterpene structure-activity relationships is very limited, and has been restricted to synthetic derivatives of a small number of readily available scaffolds. The use of P450s to selectively functionalise the β-amyrin scaffold provided a unique opportunity for systematic investigation into the types and positions of oxidation which contribute to activity. As a proof-of-concept a series of oxidised forms of β-amyrin was generated and these analogues investigated for their antiproliferative and anti-inflammatory activity in human cells (Chapter 5). Clear differences could be observed between the different oxidised compounds, suggesting that the type and position of the modification plays an important role in determining their activity. This was attributed to the presence of a ketone at C-11 and modifications to the C-28 position. Further combinatorial biosynthesis allowed structural variants of 11-oxo-β-amyrin to be assessed, and the resulting compounds again showed clear differences in their activity. Hence there is significant value in extending this approach to other positions around the β-amyrin scaffold to gain further insight into structure-activity relationships. Other triterpene scaffolds also have great potential for exploring bioactivity. In the present work, the dammarenediol-II-based triterpene protopanaxadiol was also active in both the antiproliferative and anti-inflammatory assays. The cucurbitacins are another class of structurally diverse cytotoxic triterpenes which have considerable potential as cancer therapeutics [91]. The biosynthesis of these molecules is beginning to be elucidated in species such as cucumber [89]. Equally, as described above (section 7.4.2), genome mining efforts will undoubtedly lead to novel scaffold-generating OSCs and associated tailoring enzymes, providing access to a rich diversity of previously-uncharacterised triterpenes which can be tested for activity.

Moving forward in our understanding of triterpene structure-function relationships will require further qualification of mechanisms by which these compounds exert their effect. This ‘target deconvolution’ step can be approached at the level of

functional genetics to identify individual protein targets, or alternatively, considered in the context of their cellular response such as through transcriptomic profiling [259]. Identification of the relevant cellular targets will allow for hypothesis-driven testing of the features contributing to bioactivity.

Bridging the gap between discovery of a lead molecule and its deployment as an effective therapeutic agent often requires the use of synthetic chemistry. Synthetic modifications of natural products are frequently deployed to improve the therapeutic profile of the molecule, increasing potency, reducing toxicity and increasing solubility or chemical stability [260]. Presently most chemical syntheses require access to relatively pure starting material. In future it could be possible to move towards both biosynthesis of triterpenes and subsequent chemical modification *in planta.* Such 'one-pot' reactions would reduce both the time and cost required to purify compounds prior to synthetic modification. At the same time, this approach could itself facilitate purification of products which prove hard to separate from endogenous *N. benthamiana* metabolites by altering polarity or adding chromo- or fluorophores which aid detection.

Ultimately, biological components may be seen as another tool for synthetic chemistry. As described above, the ability of enzymes such as P450s to carry out regio- and stereoselective scaffold oxidation can provide the necessary functional handles allowing modification of otherwise synthetically intractable C-H bonds. Equally, other classes of enzyme may also find use in the synthetic chemical toolkit. For example, chemical moieties such as acetyl groups are frequently used as protecting agents to prevent unwanted chemical modification of specific functional groups. Hence regiospecific acylation of triterpenes could be achieved by use of acyltransferases, further reducing the number of downstream synthetic steps needed. Although relatively few triterpene acyltransferases are currently known, there is significant evidence to suggest that triterpene acyltransferases may be found clustered with OSCs [89, 225]. Therefore the number of characterised acyltransferases can be expected to increase significantly over the coming years.

The therapeutic potential of plants has been known since ancient times [1]. Even today plants continue to be a source of important therapeutics, highlighted by the recent 2015 Nobel Prize in Physiology or Medicine being awarded to Prof Youyou Tu for the discovery of artemisinin. Triterpenes show clear potential for application in a range of areas from cancer and inflammation [48, 50, 57] to vaccine adjuvants [53, 56] and obesity treatment [49]. Nevertheless the difficulty in isolating many of

these compounds coupled with the need for accessing new, biologically relevant, chemical diversity underscores the need for development of biological platforms to engineer their production. The rapidly accumulating sequence data from across the plant kingdom coupled with advances in gene synthesis and assembly methods present for the first time exciting opportunities to assemble and engineer production of diverse new triterpenes, enabling the full potential of these very important molecules to be exploited.

References

1. Atanasov, A.G., Waltenberger, B., Pferschy-Wenzig, E.M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E.H., Rollinger, J.M., Schuster, D., Breuss, J.M., Bochkov, V., Mihovilovic, M.D., Kopp, B., Bauer, R., Dirsch, V.M., and Stuppner, H., (2015) Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnol Adv*. 33, 1582-614.
2. Miralpeix, B., Rischer, H., Hakkinen, S., Ritala, A., Seppanen-Laakso, T., Oksman-Caldentey, K.-M., Capell, T., and Christou, P., (2013) Metabolic Engineering of Plant Secondary Products: Which Way Forward? *Curr Pharm Des*. 19, 5622-39.
3. Rischer, H., Hakkinen, S.T., Ritala, A., Seppanen-Laakso, T., Miralpeix, B., Capell, T., Christou, P., and Oksman-Caldentey, K.M., (2013) Plant cells as pharmaceutical factories. *Curr Pharm Des*. 19, 5640-60.
4. Kolewe, M.E., Gaurav, V., and Roberts, S.C., (2008) Pharmaceutically active natural product synthesis and supply via plant cell culture technology. *Mol Pharm*. 5, 243-56.
5. Brown, S., Clastre, M., Courdavault, V., and O'Connor, S.E., (2015) De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc Natl Acad Sci U S A*. 112, 3205-10.
6. Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S., and Newman, J.D., (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*. 496, 528-32.
7. Galanie, S., Thodey, K., Trenchard, I.J., Filsinger Interrante, M., and Smolke, C.D., (2015) Complete biosynthesis of opioids in yeast. *Science*. 349, 1095-100.
8. Renault, H., Bassard, J.E., Hamberger, B., and Werck-Reichhart, D., (2014) Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Curr Opin Plant Biol*. 19, 27-34.
9. Bouvier, F., Rahier, A., and Camara, B., (2005) Biogenesis, molecular regulation and function of plant isoprenoids. *Prog Lipid Res*. 44, 357-429.
10. Crowell, D.N. and Huizinga, D.H., (2009) Protein isoprenylation: the fat of the matter. *Trends Plant Sci*. 14, 163-70.
11. Vranova, E., Coman, D., and Gruissem, W., (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu Rev Plant Biol*. 64, 665-700.
12. Chadwick, M., Trewin, H., Gawthrop, F., and Wagstaff, C., (2013) Sesquiterpenoids lactones: benefits to plants and people. *Int J Mol Sci*. 14, 12780-05.
13. Tallamy, D.W., Stull, J., Ehresman, N.P., Gorski, P.M., and Mason, C.E., (1997) Cucurbitacins as Feeding and Oviposition Deterrents to Insects. *Environ Entomol*. 26, 678-83.
14. Roy, A. and Saraf, S., (2006) Limonoids: overview of significant bioactive triterpenes distributed in plants kingdom. *Biol Pharm Bull*. 29, 191-201.
15. Dinan, L., (2001) Phytoecdysteroids: biological aspects. *Phytochemistry*. 57, 325-39.
16. Kato-Noguchi, H. and Peters, R.J., (2013) The role of momilactones in rice allelopathy. *J Chem Ecol*. 39, 175-85.
17. Das, A., Lee, S.-H., Hyun, T.K., Kim, S.-W., and Kim, J.-Y., (2012) Plant volatiles as method of communication. *Plant Biotechnol Rep*. 7, 9-26.

18. Unsicker, S.B., Kunert, G., and Gershenzon, J., (2009) Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. *Curr Opin Plant Biol*. 12, 479-85.
19. Dudareva, N., Klempien, A., Muhlemann, J.K., and Kaplan, I., (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol*. 198, 16-32.
20. Schwab, W., Fuchs, C., and Huang, F.-C., (2013) Transformation of terpenes into fine chemicals. *Eur J Lipid Sci Technol*. 115, 3-8.
21. van Beilen, J.B. and Poirier, Y., (2007) Establishment of new crops for the production of natural rubber. *Trends Biotechnol*. 25, 522-29.
22. Beller, H.R., Lee, T.S., and Katz, L., (2015) Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids. *Nat Prod Rep*. 32, 1508-26.
23. Rabinovitch-Deere, C.A., Oliver, J.W., Rodriguez, G.M., and Atsumi, S., (2013) Synthetic biology and metabolic engineering approaches to produce biofuels. *Chem Rev*. 113, 4611-32.
24. Paddon, C.J. and Keasling, J.D., (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat Rev Microbiol*. 12, 355-67.
25. Brocks, J.J., Logan, G.A., Buick, R., and Summons, R.E., (1999) Archean molecular fossils and the early rise of eukaryotes. *Science*. 285, 1033-36.
26. Buschhaus, C. and Jetter, R., (2011) Composition differences between epicuticular and intracuticular wax substructures: how do plants seal their epidermal surfaces? *J Exp Bot*. 62, 841-53.
27. Buschhaus, C. and Jetter, R., (2012) Composition and physiological function of the wax layers coating Arabidopsis leaves: β-amyrin negatively affects the intracuticular water barrier. *Plant Physiol*. 160, 1120-29.
28. Delis, C., Krokida, A., Georgiou, S., Pena-Rodriguez, L.M., Kavroulakis, N., Ioannou, E., Roussis, V., Osbourn, A.E., and Papadopoulou, K.K., (2011) Role of lupeol synthase in *Lotus japonicus* nodule formation. *New Phytol*. 189, 335-46.
29. Kemen, A.C., Honkanen, S., Melton, R.E., Findlay, K.C., Mugford, S.T., Hayashi, K., Haralampidis, K., Rosser, S.J., and Osbourn, A., (2014) Investigation of triterpene synthesis and regulation in oats reveals a role for β-amyrin in determining root epidermal cell patterning. *Proc Natl Acad Sci U S A*. 111, 8679-84.
30. Field, B., Fiston-Lavier, A.S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A.E., (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci U S A*. 108, 16116-21.
31. Go, Y.S., Lee, S.B., Kim, H.J., Kim, J., Park, H.Y., Kim, J.K., Shibata, K., Yokota, T., Ohyama, K., Muranaka, T., Arseniyadis, S., and Suh, M.C., (2012) Identification of marneral synthase, which is critical for growth and development in Arabidopsis. *Plant J*. 72, 791-804.
32. Field, B. and Osbourn, A.E., (2008) Metabolic diversification--independent assembly of operon-like gene clusters in different plants. *Science*. 320, 543-47.
33. Faizal, A. and Geelen, D., (2013) Saponins and their role in biological processes in plants. *Phytochem Rev*. 12, 877-93.
34. Sparg, S.G., Light, M.E., and van Staden, J., (2004) Biological activities and distribution of plant saponins. *J Ethnopharmacol*. 94, 219-43.
35. Augustin, J.M., Kuzina, V., Andersen, S.B., and Bak, S., (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry*. 72, 435-57.
36. Osbourn, A., Goss, R.J., and Field, R.A., (2011) The saponins: polar isoprenoids with important and diverse biological activities. *Nat Prod Rep*. 28, 1261-68.

37. Hostettmann, K. and Marston, A., *Saponins* (1995). Cambridge University Press, Cambridge, UK.
38. Papadopoulou, K., Melton, R.E., Leggett, M., Daniels, M.J., and Osbourn, A.E., (1999) Compromised disease resistance in saponin-deficient plants. *Proc Natl Acad Sci U S A*. 96, 12923-28.
39. Moses, T., Papadopoulou, K.K., and Osbourn, A., (2014) Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit Rev Biochem Mol Biol*. 49, 439-62.
40. Podolak, I., Galanty, A., and Sobolewska, D., (2010) Saponins as cytotoxic agents: a review. *Phytochem Rev*. 9, 425-74.
41. Armah, C.N., Mackie, A.R., Roy, C., Price, K., Osbourn, A.E., Bowyer, P., and Ladha, S., (1999) The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. *Biophys J*. 76, 281-90.
42. Morrissey, J.P., Wubben, J.P., and Osbourn, A.E., (2000) Stagonospora avenae secretes multiple enzymes that hydrolyze oat leaf saponins. *Mol Plant Microbe Interact*. 13, 1041-52.
43. Osbourn, A.E., Clarke, B.R., Dow, J.M., and Daniels, M.J., (1991) Partial characterization of avenacinase from *Gaeumannomyces graminis* var. *avenae. Physiol Mol Plant P*. 38, 301-312.
44. Christensen, L.P., (2009) Ginsenosides chemistry, biosynthesis, analysis, and potential health effects. *Adv Food Nutr Res*. 55, 1-99.
45. Leung, K.W. and Wong, A.S., (2010) Pharmacology of ginsenosides: a literature review. *Chin Med*. 5, 20.
46. Qi, L.W., Wang, C.Z., and Yuan, C.S., (2011) Ginsenosides from American ginseng: chemical and pharmacological diversity. *Phytochemistry*. 72, 689-99.
47. Omar, H.R., Komarova, I., El-Ghonemi, M., Fathy, A., Rashad, R., Abdelmalak, H.D., Yerramadha, M.R., Ali, Y., Helal, E., and Camporesi, E.M., (2012) Licorice abuse: time to send a warning message. *Ther Adv Endocrinol Metab*. 3, 125-38.
48. Yadav, V.R., Prasad, S., Sung, B., Kannappan, R., and Aggarwal, B.B., (2010) Targeting inflammatory pathways by triterpenoids for prevention and treatment of cancer. *Toxins*. 2, 2428-66.
49. Liu, J., Lee, J., Salazar Hernandez, M.A., Mazitschek, R., and Ozcan, U., (2015) Treatment of obesity with celastrol. *Cell*. 161, 999-1011.
50. Kamble, S.M., Goyal, S.N., and Patil, C.R., (2014) Multifunctional pentacyclic triterpenoids as adjuvants in cancer chemotherapy: a review. *RSC Adv*. 4, 33370-82.
51. Oda, K., Matsuda, H., Murakami, T., Katayama, S., Ohgitani, T., and Yoshikawa, M., (2003) Relationship between adjuvant activity and amphipathic structure of soyasaponins. *Vaccine*. 21, 2145-51.
52. Sun, H.X., Qin, F., and Ye, Y.P., (2005) Relationship between haemolytic and adjuvant activity and structure of protopanaxadiol-type saponins from the roots of *Panax notoginseng. Vaccine*. 23, 5533-42.
53. Sun, H.X., Xie, Y., and Ye, Y.P., (2009) Advances in saponin-based adjuvants. *Vaccine*. 27, 1787-96.
54. Adams, M.M., Damani, P., Perl, N.R., Won, A., Hong, F., Livingston, P.O., Ragupathi, G., and Gin, D.Y., (2010) Design and synthesis of potent *Quillaja* saponin vaccine adjuvants. *J Am Chem Soc*. 132, 1939-45.
55. Chea, E.K., Fernandez-Tejada, A., Damani, P., Adams, M.M., Gardner, J.R., Livingston, P.O., Ragupathi, G., and Gin, D.Y., (2012) Synthesis and preclinical evaluation of QS-21 variants leading to simplified vaccine adjuvants and mechanistic probes. *J Am Chem Soc*. 134, 13448-57.

56. Fernandez-Tejada, A., Chea, E.K., George, C., Pillarsetty, N., Gardner, J.R., Livingston, P.O., Ragupathi, G., Lewis, J.S., Tan, D.S., and Gin, D.Y., (2014) Development of a minimal saponin vaccine adjuvant based on QS-21. *Nat Chem*. 6, 635-43.
57. Sporn, M.B., Liby, K.T., Yore, M.M., Fu, L., Lopchuk, J.M., and Gribble, G.W., (2011) New synthetic triterpenoids: potent agents for prevention and treatment of tissue injury caused by inflammatory and oxidative stress. *J Nat Prod*. 74, 537-45.
58. Suh, N., Wang, Y., Honda, T., Gribble, G.W., Dmitrovsky, E., Hickey, W.F., Maue, R.A., Place, A.E., Porter, D.M., Spinella, M.J., Williams, C.R., Wu, G., Dannenberg, A.J., Flanders, K.C., Letterio, J.J., Mangelsdorf, D.J., Nathan, C.F., Nguyen, L., Porter, W.W., Ren, R.F., Roberts, A.B., Roche, N.S., Subbaramaiah, K., and Sporn, M.B., (1999) A novel synthetic oleanane triterpenoid, 2-cyano-3,12-dioxoolean-1,9-dien-28-oic acid, with potent differentiating, antiproliferative, and anti-inflammatory activity. *Cancer Res*. 59, 336-41.
59. Lombard, J. and Moreira, D., (2011) Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol*. 28, 87-99.
60. Hemmerlin, A., Harwood, J.L., and Bach, T.J., (2012) A raison d'etre for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog Lipid Res*. 51, 95-148.
61. Degenhardt, J., Kollner, T.G., and Gershenzon, J., (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*. 70, 1621-37.
62. Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., and Osbourn, A., (2014) Triterpene biosynthesis in plants. *Annu Rev Plant Biol*. 65, 225-57.
63. Poulter, C.D., (2002) Biosynthesis of non-head-to-tail terpenes. Formation of 1'-1 and 1'-3 linkages. *Acc Chem Res*. 23, 70-77.
64. Siedenburg, G. and Jendrossek, D., (2011) Squalene-hopene cyclases. *Appl Environ Microbiol*. 77, 3905-15.
65. Racolta, S., Juhl, P.B., Sirim, D., and Pleiss, J., (2012) The triterpene cyclase protein family: a systematic analysis. *Proteins*. 80, 2009-19.
66. Abe, I., (2007) Enzymatic synthesis of cyclic triterpenes. *Nat Prod Rep*. 24, 1311-31.
67. Domingo, V., Arteaga, J.F., Quilez del Moral, J.F., and Barrero, A.F., (2009) Unusually cyclized triterpenes: occurrence, biosynthesis and chemical synthesis. *Nat Prod Rep*. 26, 115-34.
68. Xu, R., Fazio, G.C., and Matsuda, S.P., (2004) On the origins of triterpenoid skeletal diversity. *Phytochemistry*. 65, 261-91.
69. Xiong, Q., Wilson, W.K., and Matsuda, S.P., (2006) An *Arabidopsis* oxidosqualene cyclase catalyzes iridal skeleton formation by Grob fragmentation. *Angew Chem Int Ed Engl*. 45, 1285-88.
70. Phillips, D.R., Rasbery, J.M., Bartel, B., and Matsuda, S.P., (2006) Biosynthetic diversity in plant triterpene cyclization. *Curr Opin Plant Biol*. 9, 305-14.
71. Lodeiro, S., Xiong, Q., Wilson, W.K., Kolesnikova, M.D., Onak, C.S., and Matsuda, S.P., (2007) An oxidosqualene cyclase makes numerous products by diverse mechanisms: a challenge to prevailing concepts of triterpene biosynthesis. *J Am Chem Soc*. 129, 11213-22.
72. Vincken, J.P., Heng, L., de Groot, A., and Gruppen, H., (2007) Saponins, classification and occurrence in the plant kingdom. *Phytochemistry*. 68, 275-97.
73. Hamberger, B. and Bak, S., (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Phil Trans R Soc B*. 368, 20120426.

74. Castillo, D.A., Kolesnikova, M.D., and Matsuda, S.P., (2013) An effective strategy for exploring unknown metabolic pathways by genome mining. *J Am Chem Soc*. 135, 5885-94.
75. Sohrabi, R., Huh, J.H., Badieyan, S., Rakotondraibe, L.H., Kliebenstein, D.J., Sobrado, P., and Tholl, D., (2015) In planta variation of volatile biosynthesis: an alternative biosynthetic route to the formation of the pathogen-induced volatile homoterpene DMNT via triterpene degradation in Arabidopsis roots. *Plant Cell*. 27, 874-90.
76. Moses, T., Pollier, J., Faizal, A., Apers, S., Pieters, L., Thevelein, J.M., Geelen, D., and Goossens, A., (2014) Unravelling the Triterpenoid Saponin Biosynthesis of the African Shrub *Maesa lanceolata. Mol Plant*. 8, 122-35.
77. Munro, A.W., Girvan, H.M., Mason, A.E., Dunford, A.J., and McLean, K.J., (2013) What makes a P450 tick? *Trends Biochem Sci*. 38, 140-50.
78. Jensen, K. and Moller, B.L., (2010) Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry*. 71, 132-41.
79. Guengerich, F.P. and Munro, A.W., (2013) Unusual cytochrome p450 enzymes and reactions. *J Biol Chem*. 288, 17065-73.
80. Mizutani, M. and Sato, F., (2011) Unusual P450 reactions in plant secondary metabolism. *Arch Biochem Biophys*. 507, 194-203.
81. Nelson, D. and Werck-Reichhart, D., (2011) A P450-centric view of plant evolution. *Plant J*. 66, 194-211.
82. Nelson, D.R., (2006) Cytochrome P450 nomenclature, 2004. *Methods Mol Biol*. 320, 1-10.
83. Nutzmann, H.W., Huang, A., and Osbourn, A., (2016) Plant metabolic clusters - from genetics to genomics. *New Phytol*. 211, 771-89.
84. Lepesheva, G.I. and Waterman, M.R., (2007) Sterol 14α-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. *Biochim Biophys Acta*. 1770, 467-77.
85. Geisler, K., Hughes, R.K., Sainsbury, F., Lomonossoff, G.P., Rejzek, M., Fairhurst, S., Olsen, C.E., Motawia, M.S., Melton, R.E., Hemmings, A.M., Bak, S., and Osbourn, A., (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc Natl Acad Sci U S A*. 110, E3360-67.
86. Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A., Melton, R., and Osbourn, A., (2006) A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. *Proc Natl Acad Sci U S A*. 103, 18848-53.
87. Han, J.Y., Kim, M.J., Ban, Y.W., Hwang, H.S., and Choi, Y.E., (2013) The involvement of β-amyrin 28-oxidase (CYP716A52v2) in oleanane-type ginsenoside biosynthesis in *Panax ginseng. Plant Cell Physiol*. 54, 2034-46.
88. Han, J.Y., Hwang, H.S., Choi, S.W., Kim, H.J., and Choi, Y.E., (2012) Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in Panax ginseng. *Plant Cell Physiol*. 53, 1535-45.
89. Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., Zeng, J., Zhou, Q., Wang, S., Gu, W., Liu, M., Ren, J., Gu, X., Zhang, S., Wang, Y., Yasukawa, K., Bouwmeester, H.J., Qi, X., Zhang, Z., Lucas, W.J., and Huang, S., (2014) Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science*. 346, 1084-88.
90. Zhang, J., Dai, L., Yang, J., Liu, C., Men, Y., Zeng, Y., Cai, Y., Zhu, Y., and Sun, Y., (2016) Oxidation of Cucurbitadienol Catalyzed by CYP87D18 in the Biosynthesis of Mogrosides from *Siraitia grosvenorii. Plant Cell Physiol*. 57, 1000-07.

91. Chen, J.C., Chiu, M.H., Nie, R.L., Cordell, G.A., and Qiu, S.X., (2005) Cucurbitacins and cucurbitane glycosides: structures and biological activities. *Nat Prod Rep*. 22, 386-99.
92. Boutanaev, A.M., Moses, T., Zi, J., Nelson, D.R., Mugford, S.T., Peters, R.J., and Osbourn, A., (2015) Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci U S A*. 112, E81-88.
93. Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Pena-Rodriguez, L.M., Katsarou, D., Field, B., Osbourn, A.E., and Papadopoulou, K.K., (2013) A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol*. 200, 675-90.
94. Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A., (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc Natl Acad Sci U S A*. 101, 8233-38.
95. Moses, T., Pollier, J., Shen, Q., Soetaert, S., Reed, J., Erffelinck, M.L., Van Nieuwerburgh, F.C., Vanden Bossche, R., Osbourn, A., Thevelein, J.M., Deforce, D., Tang, K., and Goossens, A., (2015) OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of *Artemisia annua*. *Plant Cell*. 27, 286-301.
96. Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T., and Ebizuka, Y., (2006) Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. *FEBS J*. 273, 948-59.
97. Fukushima, E.O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., Saito, K., and Muranaka, T., (2011) CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol*. 52, 2050-61.
98. Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., Akashi, T., Aoki, T., Saito, K., and Muranaka, T., (2008) Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proc Natl Acad Sci U S A*. 105, 14204-09.
99. Moses, T., Thevelein, J.M., Goossens, A., and Pollier, J., (2014) Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. *Phytochemistry*. 108, 47-56.
100. Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T., Sudo, H., Fukushima, E.O., Akashi, T., Aoki, T., Saito, K., and Muranaka, T., (2011) Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell*. 23, 4112-23.
101. Fukushima, E.O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., and Muranaka, T., (2013) Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol*. 54, 740-9.
102. Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., Scotti, C., and Calderini, O., (2015) CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. *Mol Plant*. 8, 1493-506.
103. Han, J.Y., Kim, H.J., Kwon, Y.S., and Choi, Y.E., (2011) The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in Panax ginseng. *Plant Cell Physiol*. 52, 2062-73.
104. Fiallos-Jurado, J., Pollier, J., Moses, T., Arendt, P., Barriga-Medina, N., Morillo, E., Arahana, V., de Lourdes Torres, M., Goossens, A., and Leon-Reyes, A., (2016) Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in *Chenopodium quinoa* leaves. *Plant Sci*. 250, 188-97.
105. Khakimov, B., Kuzina, V., Erthmann, P.O., Fukushima, E.O., Augustin, J.M., Olsen, C.E., Scholtalbers, J., Volpin, H., Andersen, S.B., Hauser, T.P., Muranaka, T., and Bak, S., (2015) Identification and genome organization of saponin pathway genes from a

wild crucifer, and their use for transient production of saponins in *Nicotiana benthamiana. Plant J*. 84, 478-90.

106. Andre, C.M., Legay, S., Deleruelle, A., Nieuwenhuizen, N., Punter, M., Brendolise, C., Cooney, J.M., Lateur, M., Hausman, J.F., Larondelle, Y., and Laing, W.A., (2016) Multifunctional oxidosqualene cyclases and cytochrome P450 involved in the biosynthesis of apple fruit triterpenic acids. *New Phytol*. 211, 1279-94.
107. Huang, L., Li, J., Ye, H., Li, C., Wang, H., Liu, B., and Zhang, Y., (2012) Molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in *Catharanthus roseus. Planta*. 236, 1571-81.
108. Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreno, M.A., Martins, J.C., Thevelein, J.M., and Goossens, A., (2014) Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16α hydroxylase from *Bupleurum falcatum. Proc Natl Acad Sci U S A*. 111, 1634-39.
109. Yasumoto, S., Fukushima, E.O., Seki, H., and Muranaka, T., (2016) Novel triterpene oxidizing activity of *Arabidopsis thaliana* CYP716A subfamily enzymes. *FEBS Lett*. 590, 533-40.
110. Ignea, C., Ioannou, E., Georgantea, P., Loupassaki, S., Trikka, F.A., Kanellis, A.K., Makris, A.M., Roussis, V., and Kampranis, S.C., (2015) Reconstructing the chemical diversity of labdane-type diterpene biosynthesis in yeast. *Metab Eng*. 28, 91-103.
111. Jia, M., Potter, K.C., and Peters, R.J., (2016) Extreme promiscuity of a bacterial and a plant diterpene synthase enables combinatorial biosynthesis. *Metab Eng*. 37, 24-34.
112. Andersen-Ranberg, J., Kongstad, K.T., Nielsen, M.T., Jensen, N.B., Pateraki, I., Bach, S.S., Hamberger, B., Zerbe, P., Staerk, D., Bohlmann, J., Moller, B.L., and Hamberger, B., (2016) Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angew Chem Int Ed Engl*. 55, 2142-46.
113. Mafu, S., Jia, M., Zi, J., Morrone, D., Wu, Y., Xu, M., Hillwig, M.L., and Peters, R.J., (2016) Probing the promiscuity of ent-kaurene oxidases via combinatorial biosynthesis. *Proc Natl Acad Sci U S A*. 113, 2526-31.
114. Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R., Tikunov, Y., Bovy, A., Chikate, Y., Singh, P., Rogachev, I., Beekwilder, J., Giri, A.P., and Aharoni, A., (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*. 341, 175-9.
115. Potter, K., Criswell, J., Zi, J., Stubbs, A., and Peters, R.J., (2014) Novel product chemistry from mechanistic analysis of ent-copalyl diphosphate synthases from plant hormone biosynthesis. *Angew Chem Int Ed Engl*. 53, 7198-202.
116. Laursen, T., Borch, J., Knudsen, C., Bavishi, K., Torta, F., Martens, H.J., Silvestro, D., Hatzakis, N.S., Wenk, M.R., Dafforn, T.R., Olsen, C.E., Motawia, M.S., Hamberger, B., Moller, B.L., and Bassard, J.E., (2016) Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science*. 354, 890-93.
117. Singleton, C., Howard, T.P., and Smirnoff, N., (2014) Synthetic metabolons for metabolic engineering. *J Exp Bot*. 65, 1947-54.
118. Osbourn, A.E., O'Maille, P.E., Rosser, S.J., and Lindsey, K., (2012) Synthetic biology. *New Phytologist*. 196, 671-77.
119. Nielsen, J. and Keasling, J.D., (2016) Engineering Cellular Metabolism. *Cell*. 164, 1185-97.
120. Farhi, M., Marhevka, E., Ben-Ari, J., Algamas-Dimantov, A., Liang, Z., Zeevi, V., Edelbaum, O., Spitzer-Rimon, B., Abeliovich, H., Schwartz, B., Tzfira, T., and

Vainstein, A., (2011) Generation of the potent anti-malarial drug artemisinin in tobacco. *Nat Biotechnol*. 29, 1072-74.

121. Fuentes, P., Zhou, F., Erban, A., Karcher, D., Kopka, J., and Bock, R., (2016) A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *eLife*. 5, e13664.
122. 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., and Werck-Reichhart, D., (2014) The seco-iridoid pathway from *Catharanthus roseus. Nat Commun*. 5, 3606.
123. Pfalz, M., Mikkelsen, M.D., Bednarek, P., Olsen, C.E., Halkier, B.A., and Kroymann, J., (2011) Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in *Arabidopsis* indole glucosinolate modification. *Plant Cell*. 23, 716-29.
124. Geu-Flores, F., Nielsen, M.T., Nafisi, M., Moldrup, M.E., Olsen, C.E., Motawia, M.S., and Halkier, B.A., (2009) Glucosinolate engineering identifies a gamma-glutamyl peptidase. *Nat Chem Biol*. 5, 575-77.
125. Nielsen, A.Z., Ziersen, B., Jensen, K., Lassen, L.M., Olsen, C.E., Moller, B.L., and Jensen, P.E., (2013) Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth Biol*. 2, 308-15.
126. Fischer, M.J., Meyer, S., Claudel, P., Perrin, M., Ginglinger, J.F., Gertz, C., Masson, J.E., Werck-Reinhardt, D., Hugueney, P., and Karst, F., (2013) Specificity of *Ocimum basilicum* geraniol synthase modified by its expression in different heterologous systems. *J Biotechnol*. 163, 24-9.
127. Dong, L., Jongedijk, E., Bouwmeester, H., and Van Der Krol, A., (2016) Monoterpene biosynthesis potential of plant subcellular compartments. *New Phytol*. 209, 679-90.
128. Bruckner, K. and Tissier, A., (2013) High-level diterpene production by transient expression in *Nicotiana benthamiana. Plant Methods*. 9, 46.
129. Gnanasekaran, T., Vavitsas, K., Andersen-Ranberg, J., Nielsen, A.Z., Olsen, C.E., Hamberger, B., and Jensen, P.E., (2015) Heterologous expression of the isopimaric acid pathway in *Nicotiana benthamiana* and the effect of N-terminal modifications of the involved cytochrome P450 enzyme. *J Biol Eng*. 9, 24.
130. Zerbe, P., Chiang, A., Dullat, H., O'Neil-Johnson, M., Starks, C., Hamberger, B., and Bohlmann, J., (2014) Diterpene synthases of the biosynthetic system of medicinally active diterpenoids in *Marrubium vulgare. Plant J*. 79, 914-27.
131. Cankar, K., Jongedijk, E., Klompmaker, M., Majdic, T., Mumm, R., Bouwmeester, H., Bosch, D., and Beekwilder, J., (2015) (+)-Valencene production in *Nicotiana benthamiana* is increased by down-regulation of competing pathways. *Biotechnol J*. 10, 180-9.
132. van Herpen, T.W., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H.J., and Beekwilder, J., (2010) *Nicotiana benthamiana* as a production platform for artemisinin precursors. *Plos One*. 5, e14222.
133. Liu, Q., Majdi, M., Cankar, K., Goedbloed, M., Charnikhova, T., Verstappen, F.W., de Vos, R.C., Beekwilder, J., van der Krol, S., and Bouwmeester, H.J., (2011) Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana. Plos One*. 6, e23255.
134. Jin, J., Kim, M.J., Dhandapani, S., Tjhang, J.G., Yin, J.L., Wong, L., Sarojam, R., Chua, N.H., and Jang, I.C., (2015) The floral transcriptome of ylang ylang (*Cananga odorata* var. *fruticosa*) uncovers biosynthetic pathways for volatile organic compounds and a multifunctional and novel sesquiterpene synthase. *J Exp Bot*. 66, 3959-75.

135. Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S., (2011) A modular cloning system for standardized assembly of multigene constructs. *Plos One*. 6, e16765.
136. Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D., Patron, N.J., and Marillonnet, S., (2014) A golden gate modular cloning toolbox for plants. *ACS Synth Biol*. 3, 839-43.
137. Salmon, M., Thimmappa, R.B., Minto, R.E., Melton, R.E., Hughes, R.K., O'Maille, P.E., Hemmings, A.M., and Osbourn, A., (2016) A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. *Proc Natl Acad Sci U S A*. 113, E4407-14.
138. Huang, A.X., Xiong, Z., and Corey, E.J., (1999) An Exceptionally Short and Simple Enantioselective Total Synthesis of Pentacyclic Triterpenes of the β-Amyrin Family. *J Am Chem Soc*. 121, 9999-10003.
139. Surendra, K. and Corey, E.J., (2009) A short enantioselective total synthesis of the fundamental pentacyclic triterpene lupeol. *J Am Chem Soc*. 131, 13928-29.
140. Arendt, P., Pollier, J., Callewaert, N., and Goossens, A., (2016) Synthetic biology for production of natural and new-to-nature terpenoids in photosynthetic organisms. *Plant J*. 87, 16-37.
141. O'Connor, S.E., (2015) Engineering of Secondary Metabolism. *Annu Rev Genet*. 49, 71-94.
142. Vasilev, N., Schmitz, C., Dong, L., Ritala, A., Imseng, N., Häkkinen, S.T., van der Krol, S., Eibl, R., Oksman-Caldentey, K.-M., Bouwmeester, H., Fischer, R., and Schillberg, S., (2014) Comparison of plant-based expression platforms for the heterologous production of geraniol. *Plant Cell Tiss Organ Cult*. 117, 373-80.
143. Hemmerlin, A., (2013) Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant Sci*. 203-204, 41-54.
144. Burg, J.S. and Espenshade, P.J., (2011) Regulation of HMG-CoA reductase in mammals and yeast. *Prog Lipid Res*. 50, 403-10.
145. Leivar, P., Antolin-Llovera, M., Ferrero, S., Closa, M., Arro, M., Ferrer, A., Boronat, A., and Campos, N., (2011) Multilevel control of Arabidopsis 3-hydroxy-3-methylglutaryl coenzyme A reductase by protein phosphatase 2A. *Plant Cell*. 23, 1494-511.
146. Pollier, J., Moses, T., Gonzalez-Guzman, M., De Geyter, N., Lippens, S., Vanden Bossche, R., Marhavy, P., Kremer, A., Morreel, K., Guerin, C.J., Tava, A., Oleszek, W., Thevelein, J.M., Campos, N., Goormachtig, S., and Goossens, A., (2013) The protein quality control system manages plant defence compound synthesis. *Nature*. 504, 148-52.
147. Kirby, J., Romanini, D.W., Paradise, E.M., and Keasling, J.D., (2008) Engineering triterpene production in *Saccharomyces cerevisiae*-β-amyrin synthase from *Artemisia annua. FEBS J*. 275, 1852-9.
148. Dai, Z., Liu, Y., Zhang, X., Shi, M., Wang, B., Wang, D., Huang, L., and Zhang, X., (2013) Metabolic engineering of *Saccharomyces cerevisiae* for production of ginsenosides. *Metab Eng*. 20, 146-56.
149. Holmberg, N., Harker, M., Wallace, A.D., Clayton, J.C., Gibbard, C.L., and Safford, R., (2003) Co-expression of N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase and C24-sterol methyltransferase type 1 in transgenic tobacco enhances carbon flux towards end-product sterols. *Plant J*. 36, 12-20.
150. Harker, M., Holmberg, N., Clayton, J.C., Gibbard, C.L., Wallace, A.D., Rawlins, S., Hellyer, S.A., Lanot, A., and Safford, R., (2003) Enhancement of seed phytosterol levels by expression of an N-terminal truncated *Hevea brasiliensis* (rubber tree) 3-hydroxy-3-methylglutaryl-CoA reductase. *Plant Biotechnol J*. 1, 113-21.

151. Kim, O.T., Kim, S.H., Ohyama, K., Muranaka, T., Choi, Y.E., Lee, H.Y., Kim, M.Y., and Hwang, B., (2010) Upregulation of phytosterol and triterpene biosynthesis in *Centella asiatica* hairy roots overexpressed ginseng farnesyl diphosphate synthase. *Plant Cell Rep*. 29, 403-11.
152. Lee, M.H., Jeong, J.H., Seo, J.W., Shin, C.G., Kim, Y.S., In, J.G., Yang, D.C., Yi, J.S., and Choi, Y.E., (2004) Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene. *Plant Cell Physiol*. 45, 976-84.
153. Seo, J.W., Jeong, J.H., Shin, C.G., Lo, S.C., Han, S.S., Yu, K.W., Harada, E., Han, J.Y., and Choi, Y.E., (2005) Overexpression of squalene synthase in *Eleutherococcus senticosus* increases phytosterol and triterpene accumulation. *Phytochemistry*. 66, 869-77.
154. Zelcer, N., Sharpe, L.J., Loregger, A., Kristiana, I., Cook, E.C., Phan, L., Stevenson, J., and Brown, A.J., (2014) The E3 ubiquitin ligase MARCH6 degrades squalene monooxygenase and affects 3-hydroxy-3-methyl-glutaryl coenzyme A reductase and the cholesterol synthesis pathway. *Mol Cell Biol*. 34, 1262-70.
155. Han, J.Y., In, J.G., Kwon, Y.S., and Choi, Y.E., (2010) Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng. Phytochemistry*. 71, 36-46.
156. Veen, M., Stahl, U., and Lang, C., (2003) Combined overexpression of genes of the ergosterol biosynthetic pathway leads to accumulation of sterols in *Saccharomyces cerevisiae. FEMS Yeast Res*. 4, 87-95.
157. Vranova, E., Coman, D., and Gruissem, W., (2012) Structure and dynamics of the isoprenoid pathway network. *Mol Plant*. 5, 318-33.
158. Wu, S., Schalk, M., Clark, A., Miles, R.B., Coates, R., and Chappell, J., (2006) Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nat Biotechnol*. 24, 1441-47.
159. Wu, S., Jiang, Z., Kempinski, C., Eric Nybo, S., Husodo, S., Williams, R., and Chappell, J., (2012) Engineering triterpene metabolism in tobacco. *Planta*. 236, 867-77.
160. Sainsbury, F. and Lomonossoff, G.P., (2008) Extremely high-level and rapid transient protein production in plants without the use of viral replication. *Plant Physiol*. 148, 1212-18.
161. Sainsbury, F., Thuenemann, E.C., and Lomonossoff, G.P., (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J*. 7, 682-93.
162. Assimopoulou, A.N. and Papageorgiou, V.P., (2005) GC-MS analysis of penta- and tetra-cyclic triterpenes from resins of *Pistacia* species. Part I. *Pistacia lentiscus* var. Chia. *Biomed Chromatogr*. 19, 285-311.
163. Shan, H., Segura, M.J., Wilson, W.K., Lodeiro, S., and Matsuda, S.P., (2005) Enzymatic cyclization of dioxidosqualene to heterocyclic triterpenes. *J Am Chem Soc*. 127, 18008-09.
164. Rasbery, J.M., Shan, H., LeClair, R.J., Norman, M., Matsuda, S.P., and Bartel, B., (2007) *Arabidopsis thaliana* squalene epoxidase 1 is essential for root and seed development. *J Biol Chem*. 282, 17002-13.
165. Bick, J.A. and Lange, B.M., (2003) Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch Biochem Biophys*. 415, 146-54.
166. Hasunuma, T., Takeno, S., Hayashi, S., Sendai, M., Bamba, T., Yoshimura, S., Tomizawa, K., Fukusaki, E., and Miyake, C., (2008) Overexpression of 1-Deoxy-D-xylulose-5-phosphate reductoisomerase gene in chloroplast contributes to increment of isoprenoid production. *J Biosci Bioeng*. 105, 518-26.

167. Henry, L.K., Gutensohn, M., Thomas, S.T., Noel, J.P., and Dudareva, N., (2015) Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. *Proc Natl Acad Sci U S A*. 112, 10050-55.
168. Liao, P., Wang, H., Hemmerlin, A., Nagegowda, D.A., Bach, T.J., Wang, M., and Chye, M.L., (2014) Past achievements, current status and future perspectives of studies on 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) in the mevalonate (MVA) pathway. *Plant Cell Rep*. 33, 1005-22.
169. Lange, I., Poirier, B.C., Herron, B.K., and Lange, B.M., (2015) Comprehensive Assessment of Transcriptional Regulation Facilitates Metabolic Engineering of Isoprenoid Accumulation in Arabidopsis. *Plant Physiol*. 169, 1595-606.
170. Zgoła-Grześkowiak, A. and Grześkowiak, T., (2011) Dispersive liquid-liquid microextraction. *Trends Anal Chem*. 30, 1382-99.
171. Seki, H., Tamura, K., and Muranaka, T., (2015) P450s and UGTs: Key Players in the Structural Diversity of Triterpenoid Saponins. *Plant Cell Physiol*. 56, 1463-71.
172. Assimopoulou, A.N. and Papageorgiou, V.P., (2005) GC-MS analysis of penta- and tetra-cyclic triterpenes from resins of *Pistacia* species. Part II. *Pistacia terebinthus* var. Chia. *Biomed Chromatogr*. 19, 586-605.
173. Burnouf-Radosevich, M., Delfel, N.E., and England, R., (1985) Gas chromatography-mass spectrometry of oleanane- and ursane-type triterpenes—application to *Chenopodium quinoa* triterpenes. *Phytochemistry*. 24, 2063-66.
174. Budzikiewicz, H., Wilson, J.M., and Djerassi, C., (1963) Mass Spectrometry in Structural and Stereochemical Problems. XXXII.1Pentacyclic Triterpenes. *J Am Chem Soc*. 85, 3688-3699.
175. Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., Graham, N., Odoardi, M., Piano, E., Arcioni, S., May, S., Scotti, C., and Calderini, O., (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell*. 23, 3070-81.
176. Segura, M.J., Meyer, M.M., and Matsuda, S.P., (2000) *Arabidopsis thaliana* LUP1 converts oxidosqualene to multiple triterpene alcohols and a triterpene diol. *Org Lett*. 2, 2257-59.
177. Zhang, Q., Lu, Z., Li, X., Zheng, Y., Yao, D., Gu, Y., Huo, C., and Cong, B., (2015) Triterpenoids and Steroids from the Leaves of *Forsythia suspensa. Chem Nat Compd*. 51, 178-80.
178. Ikuta, A., Kamiya, K., Satake, T., and Saiki, Y., (1995) Triterpenoids from callus tissue cultures of *Paeonia* species. *Phytochemistry*. 38, 1203-07.
179. Al-Jaber, H.I., Abrouni, K.K., Al-Qudah, M.A., and Abu Zarga, M.H., (2012) New terpenes from *Salvia palaestina* Benth. and *Salvia syriaca* L. growing wild in Jordan. *J Asian Nat Prod Res*. 14, 618-25.
180. Kitagawa, I., (2002) Licorice root. A natural sweetener and an important ingredient in Chinese medicine. *Pure Appl Chem*. 74, 1189-98.
181. Kitagawa, I., Hori, K., Sakagami, M., Zhou, J.L., and Yoshikawa, M., (1993) Saponin and sapogenol. XLVIII. On the constituents of the roots of *Glycyrrhiza uralensis* Fischer from northeastern China. (2). Licorice-saponins D3, E2, F3, G2, H2, J2, and K2. *Chem Pharm Bull*. 41, 1337-45.
182. Crombie, L., Crombie, W.M.L., and Whiting, D.A., (1984) Structures of the four avenacins, oat root resistance factors to 'take-all' disease. *J. Chem. Soc., Chem. Commun.* 4, 246-48.
183. Itkin, M., Davidovich-Rikanati, R., Cohen, S., Portnoy, V., Doron-Faigenboim, A., Oren, E., Freilich, S., Tzuri, G., Baranes, N., Shen, S., Petreikov, M., Sertchook, R., Ben-Dor, S., Gottlieb, H., Hernandez, A., Nelson, D.R., Paris, H.S., Tadmor, Y., Burger, Y., Lewinsohn, E., Katzir, N., and Schaffer, A., (2016) The biosynthetic

pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii. Proc Natl Acad Sci U S A*. 113, E7619-28.

184. Qi, L.W., Wang, C.Z., and Yuan, C.S., (2011) Isolation and analysis of ginseng: advances and challenges. *Nat Prod Rep*. 28, 467-95.
185. Qiu, W.-W., Li, J., Li, H., Zou, H., Gao, L.-X., Liu, T., Yang, F., Li, J.-Y., and Tang, J., (2012) Synthesis and Biological Evaluation of Oleanolic Acid Derivatives as Novel Inhibitors of Protein Tyrosine Phosphatase 1B. *Heterocycles*. 85, 1117.
186. Newman, D.J. and Cragg, G.M., (2016) Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod*. 79, 629-61.
187. Liby, K.T. and Sporn, M.B., (2012) Synthetic oleanane triterpenoids: multifunctional drugs with a broad range of applications for prevention and treatment of chronic disease. *Pharmacol Rev*. 64, 972-1003.
188. Jager, S., Trojan, H., Kopp, T., Laszczyk, M.N., and Scheffler, A., (2009) Pentacyclic triterpene distribution in various plants - rich sources for a new group of multi-potent plant extracts. *Molecules*. 14, 2016-31.
189. Laszczyk, M.N., (2009) Pentacyclic triterpenes of the lupane, oleanane and ursane group as tools in cancer therapy. *Planta Med*. 75, 1549-60.
190. Ashley, N.T., Weil, Z.M., and Nelson, R.J., (2012) Inflammation: Mechanisms, Costs, and Natural Variation. *Ann Rev Ecol Evol Syst*. 43, 385-406.
191. Medzhitov, R., (2008) Origin and physiological roles of inflammation. *Nature*. 454, 428-35.
192. Elinav, E., Nowarski, R., Thaiss, C.A., Hu, B., Jin, C., and Flavell, R.A., (2013) Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer*. 13, 759-71.
193. Grivennikov, S.I., Greten, F.R., and Karin, M., (2010) Immunity, inflammation, and cancer. *Cell*. 140, 883-99.
194. de Santana Souza, M.T., Almeida, J.R.G.d.S., de Souza Araujo, A.A., Duarte, M.C., Gelain, D.P., Moreira, J.C.F., dos Santos, M.R.V., and Quintans-Júnior, L.J., (2014) Structure-Activity Relationship of Terpenes with Anti-Inflammatory Profile - A Systematic Review. *Basic Clin Pharmacol Toxicol*. 115, 244-56.
195. Sun, H.F., WS; Wang, WZ; Chun H, (2006) Structure-activity relationships of oleanane- and ursane- type triterpenoids. *Botanical Studies*. 47, 339-68.
196. Dinkova-Kostova, A.T., Liby, K.T., Stephenson, K.K., Holtzclaw, W.D., Gao, X., Suh, N., Williams, C., Risingsong, R., Honda, T., Gribble, G.W., Sporn, M.B., and Talalay, P., (2005) Extremely potent triterpenoid inducers of the phase 2 response: correlations of protection against oxidant and inflammatory stress. *Proc Natl Acad Sci U S A*. 102, 4584-89.
197. Ahmad, R., Raina, D., Meyer, C., Kharbanda, S., and Kufe, D., (2006) Triterpenoid CDDO-Me blocks the NF-kB pathway by direct inhibition of IKKβ on Cys-179. *J Biol Chem*. 281, 35764-69.
198. Chin, M.P., Reisman, S.A., Bakris, G.L., O'Grady, M., Linde, P.G., McCullough, P.A., Packham, D., Vaziri, N.D., Ward, K.W., Warnock, D.G., and Meyer, C.J., (2014) Mechanisms contributing to adverse cardiovascular events in patients with type 2 diabetes mellitus and stage 4 chronic kidney disease treated with bardoxolone methyl. *Am J Nephrol*. 39, 499-508.
199. Chin, M.P., Wrolstad, D., Bakris, G.L., Chertow, G.M., de Zeeuw, D., Goldsberry, A., Linde, P.G., McCullough, P.A., McMurray, J.J., Wittes, J., and Meyer, C.J., (2014) Risk factors for heart failure in patients with type 2 diabetes mellitus and stage 4 chronic kidney disease treated with bardoxolone methyl. *J Card Fail*. 20, 953-58.
200. Berridge, M.V., Herst, P.M., and Tan, A.S., (2005) Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. *Biotechnol Annu Rev*. 11, 127-52.

201. Shan, J.Z., Xuan, Y.Y., Ruan, S.Q., and Sun, M., (2011) Proliferation-inhibiting and apoptosis-inducing effects of ursolic acid and oleanolic acid on multi-drug resistance cancer cells in vitro. *Chin J Integr Med*. 17, 607-11.
202. Juan, M.E., Wenzel, U., Daniel, H., and Planas, J.M., (2008) Erythrodiol, a natural triterpenoid from olives, has antiproliferative and apoptotic activity in HT-29 human adenocarcinoma cells. *Mol Nutr Food Res*. 52, 595-99.
203. Chen, H.L., Lin, K.W., Huang, A.M., Tu, H.Y., Wei, B.L., Hour, T.C., Yen, M.H., Pu, Y.S., and Lin, C.N., (2010) Terpenoids induce cell cycle arrest and apoptosis from the stems of *Celastrus kusanoi* associated with reactive oxygen species. *J Agric Food Chem*. 58, 3808-12.
204. Allouche, Y., Warleta, F., Campos, M., Sánchez-Quesada, C., Uceda, M., Beltrán, G., and Gaforio, J.J., (2011) Antioxidant, Antiproliferative, and Pro-apoptotic Capacities of Pentacyclic Triterpenes Found in the Skin of Olives on MCF-7 Human Breast Cancer Cells and Their Effects on DNA Damage. *J Agric Food Chem*. 59, 121-30.
205. Kaweetripob, W., Mahidol, C., Prachyawarakorn, V., Prawat, H., and Ruchirawat, S., (2012) 5-formylfurfuryl esters from *Duabanga grandiflora*. *Phytochemistry*. 76, 78-82.
206. Wang, W., Zhao, Y., Rayburn, E.R., Hill, D.L., Wang, H., and Zhang, R., (2007) In vitro anti-cancer activity and structure-activity relationships of natural products isolated from fruits of *Panax ginseng*. *Cancer Chemother Pharmacol*. 59, 589-601.
207. Pirzadeh, S., Fakhari, S., Jalili, A., Mirzai, S., Ghaderi, B., and Haghshenas, V., (2014) Glycyrrhetinic Acid Induces Apoptosis in Leukemic HL60 Cells Through Upregulating of CD95/ CD178. *Int J Mol Cell Med*. 3, 272-78.
208. Sharma, G., Kar, S., Palit, S., and Das, P.K., (2012) 18β-glycyrrhetinic acid induces apoptosis through modulation of Akt/FOXO3a/Bim pathway in human breast cancer MCF-7 cells. *J Cell Physiol*. 227, 1923-31.
209. Schwarz, S. and Csuk, R., (2010) Synthesis and antitumour activity of glycyrrhetinic acid derivatives. *Bioorg Med Chem*. 18, 7458-74.
210. Keepers, Y.P., Pizao, P.E., Peters, G.J., van Ark-Otte, J., Winograd, B., and Pinedo, H.M., (1991) Comparison of the sulforhodamine B protein and tetrazolium (MTT) assays for in vitro chemosensitivity testing. *Eur J Cancer*. 27, 897-900.
211. Assefa, H., Nimrod, A., Walker, L., and Sindelar, R., (1999) Synthesis and evaluation of potential complement inhibitory semisynthetic analogs of oleanolic acid. *Bioorg Med Chem Lett*. 9, 1889-94.
212. Chanput, W., Mes, J.J., and Wichers, H.J., (2014) THP-1 cell line: an in vitro cell model for immune modulation approach. *Int Immunopharmacol*. 23, 37-45.
213. Gonçalves, B.M.F., Salvador, J.A.R., Marín, S., and Cascante, M., (2016) Synthesis and biological evaluation of novel asiatic acid derivatives with anticancer activity. *RSC Adv*. 6, 3967-85.
214. Wang, F., Hua, H., Pei, Y., Chen, D., and Jing, Y., (2006) Triterpenoids from the resin of *Styrax tonkinensis* and their antiproliferative and differentiation effects in human leukemia HL-60 cells. *J Nat Prod*. 69, 807-10.
215. Fiore, C., Salvi, M., Palermo, M., Sinigaglia, G., Armanini, D., and Toninello, A., (2004) On the mechanism of mitochondrial permeability transition induction by glycyrrhetinic acid. *Biochim Biophys Acta*. 1658, 195-201.
216. Csuk, R., Schwarz, S., Kluge, R., and Strohl, D., (2012) Does one keto group matter? Structure-activity relationships of glycyrrhetinic acid derivatives modified at position C-11. *Arch Pharm Chem Life Sci*. 345, 28-32.
217. Kao, T.C., Shyu, M.H., and Yen, G.C., (2010) Glycyrrhizic acid and 18β-glycyrrhetinic acid inhibit inflammation via PI3K/Akt/GSK3β signaling and glucocorticoid receptor activation. *J Agric Food Chem*. 58, 8623-29.

218. Rhen, T. and Cidlowski, J.A., (2005) Antiinflammatory action of glucocorticoids--new mechanisms for old drugs. *N Engl J Med*. 353, 1711-23.
219. Kim, J., Cha, Y.N., and Surh, Y.J., (2010) A protective role of nuclear factor-erythroid 2-related factor-2 (Nrf2) in inflammatory disorders. *Mutat Res*. 690, 12-23.
220. Logashenko, E.B., Salomatina, O.V., Markov, A.V., Korchagina, D.V., Salakhutdinov, N.F., Tolstikov, G.A., Vlassov, V.V., and Zenkova, M.A., (2011) Synthesis and pro-apoptotic activity of novel glycyrrhetinic acid derivatives. *ChemBioChem*. 12, 784-94.
221. Saxena, P., Hsieh, Y.C., Alvarado, V.Y., Sainsbury, F., Saunders, K., Lomonossoff, G.P., and Scholthof, H.B., (2011) Improved foreign gene expression in plants using a virus-encoded suppressor of RNA silencing modified to be developmentally harmless. *Plant Biotechnol J*. 9, 703-12.
222. Osbourn, A.E., (2003) Saponins in cereals. *Phytochemistry*. 62, 1-4.
223. Wegel, E., Koumproglou, R., Shaw, P., and Osbourn, A., (2009) Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. *Plant Cell*. 21, 3926-36.
224. Mary, W., Crombie, L., and Crombie, L., (1986) Distribution of avenacins A-1, A-2, B-1 and B-2 in oat roots: Their fungicidal activity towards 'take-all' fungus. *Phytochemistry*. 25, 2069-73.
225. Mugford, S.T., Louveau, T., Melton, R., Qi, X., Bakht, S., Hill, L., Tsurushima, T., Honkanen, S., Rosser, S.J., Lomonossoff, G.P., and Osbourn, A., (2013) Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell*. 25, 1078-92.
226. Mugford, S.T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R.K., Papadopoulou, K., Melton, R., Philo, M., Sainsbury, F., Lomonossoff, G.P., Roy, A.D., Goss, R.J., and Osbourn, A., (2009) A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell*. 21, 2473-84.
227. Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., and Osbourn, A., (2001) A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc Natl Acad Sci U S A*. 98, 13431-26.
228. Owatworakit, A., Townsend, B., Louveau, T., Jenner, H., Rejzek, M., Hughes, R.K., Saalbach, G., Qi, X., Bakht, S., Roy, A.D., Mugford, S.T., Goss, R.J., Field, R.A., and Osbourn, A., (2013) Glycosyltransferases from oat (*Avena*) implicated in the acylation of avenacins. *J Biol Chem*. 288, 3696-704.
229. Inagaki, Y.S., Etherington, G., Geisler, K., Field, B., Dokarry, M., Ikeda, K., Mutsukado, Y., Dicks, J., and Osbourn, A., (2011) Investigation of the potential for triterpene synthesis in rice through genome mining and metabolic engineering. *New Phytol*. 191, 432-48.
230. Chu, H.Y., Wegel, E., and Osbourn, A., (2011) From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant J*. 66, 66-79.
231. Irmler, S., Schröder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J., (2008) Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J*. 24, 797-804.
232. Guengerich, F.P., (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol*. 14, 611-50.
233. Heitz, T., Widemann, E., Lugan, R., Miesch, L., Ullmann, P., Desaubry, L., Holder, E., Grausem, B., Kandel, S., Miesch, M., Werck-Reichhart, D., and Pinot, F., (2012) Cytochromes P450 CYP94C1 and CYP94B3 catalyze two successive oxidation steps

of plant hormone Jasmonoyl-isoleucine for catabolic turnover. *J Biol Chem*. 287, 6296-306.

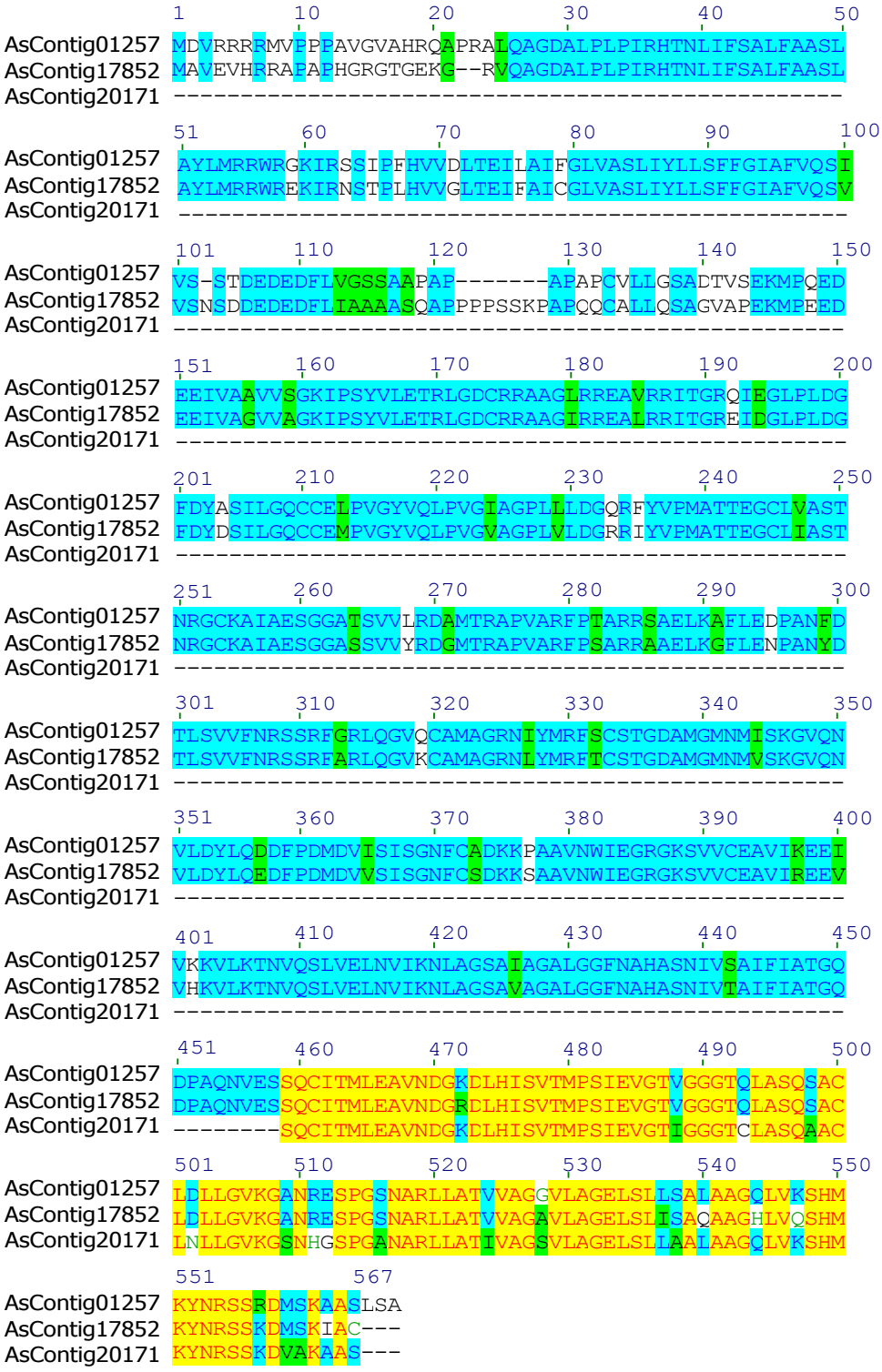
234. Ting, H.M., Wang, B., Ryden, A.M., Woittiez, L., van Herpen, T., Verstappen, F.W., Ruyter-Spira, C., Beekwilder, J., Bouwmeester, H.J., and van der Krol, A., (2013) The metabolite chemotype of *Nicotiana benthamiana* transiently expressing artemisinin biosynthetic pathway genes is a function of CYP71AV1 type and relative gene dosage. *New Phytol*. 199, 352-66.
235. Gruszka, D.S., I.; Maluszynski, M., (2012) Sodium Azide as a Mutagen. *Plant Mutation Breeding and Biotechnology*.
236. Chiba, Y. and Green, P.J., (2009) mRNA Degradation Machinery in Plants. *J Plant Biol*. 52, 114-24.
237. Burge, C. and Karlin, S., (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 268, 78-94.
238. Qin, B., Eagles, J., Mellon, F.A., Mylona, P., Pena-Rodriguez, L., and Osbourn, A.E., (2010) High throughput screening of mutants of oat that are defective in triterpene synthesis. *Phytochemistry*. 71, 1245-52.
239. Koo, A.J. and Howe, G.A., (2012) Catabolism and deactivation of the lipid-derived hormone jasmonoyl-isoleucine. *Front Plant Sci*. 3, 19.
240. Kitaoka, N., Matsubara, T., Sato, M., Takahashi, K., Wakuta, S., Kawaide, H., Matsui, H., Nabeta, K., and Matsuura, H., (2011) *Arabidopsis* CYP94B3 encodes jasmonyl-L-isoleucine 12-hydroxylase, a key enzyme in the oxidative catabolism of jasmonate. *Plant Cell Physiol*. 52, 1757-65.
241. Koo, A.J., Cooke, T.F., and Howe, G.A., (2011) Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. *Proc Natl Acad Sci U S A*. 108, 9298-303.
242. Koo, A.J., Thireault, C., Zemelis, S., Poudel, A.N., Zhang, T., Kitaoka, N., Brandizzi, F., Matsuura, H., and Howe, G.A., (2014) Endoplasmic reticulum-associated inactivation of the hormone jasmonoyl-L-isoleucine by multiple members of the cytochrome P450 94 family in *Arabidopsis. J Biol Chem*. 289, 29728-38.
243. Kurotani, K., Hayashi, K., Hatanaka, S., Toda, Y., Ogawa, D., Ichikawa, H., Ishimaru, Y., Tashita, R., Suzuki, T., Ueda, M., Hattori, T., and Takeda, S., (2015) Elevated levels of CYP94 family gene expression alleviate the jasmonate response and enhance salt tolerance in rice. *Plant Cell Physiol*. 56, 779-89.
244. Widemann, E., Grausem, B., Renault, H., Pineau, E., Heinrich, C., Lugan, R., Ullmann, P., Miesch, L., Aubert, Y., Miesch, M., Heitz, T., and Pinot, F., (2015) Sequential oxidation of Jasmonoyl-Phenylalanine and Jasmonoyl-Isoleucine by multiple cytochrome P450 of the CYP94 family through newly identified aldehyde intermediates. *Phytochemistry*. 117, 388-99.
245. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G., (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res*. 40, e115.
246. Jensen, M.K. and Keasling, J.D., (2015) Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS Yeast Res*. 15, 1-10.
247. Westfall, P.J., Pitera, D.J., Lenihan, J.R., Eng, D., Woolard, F.X., Regentin, R., Horning, T., Tsuruta, H., Melis, D.J., Owens, A., Fickes, S., Diola, D., Benjamin, K.R., Keasling, J.D., Leavell, M.D., McPhee, D.J., Renninger, N.S., Newman, J.D., and Paddon, C.J., (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci U S A*. 109, E111-18.
248. Albertsen, L., Chen, Y., Bach, L.S., Rattleff, S., Maury, J., Brix, S., Nielsen, J., and Mortensen, U.H., (2011) Diversion of flux toward sesquiterpene production in

Saccharomyces cerevisiae by fusion of host and heterologous enzymes. *Appl Environ Microbiol*. 77, 1033-40.

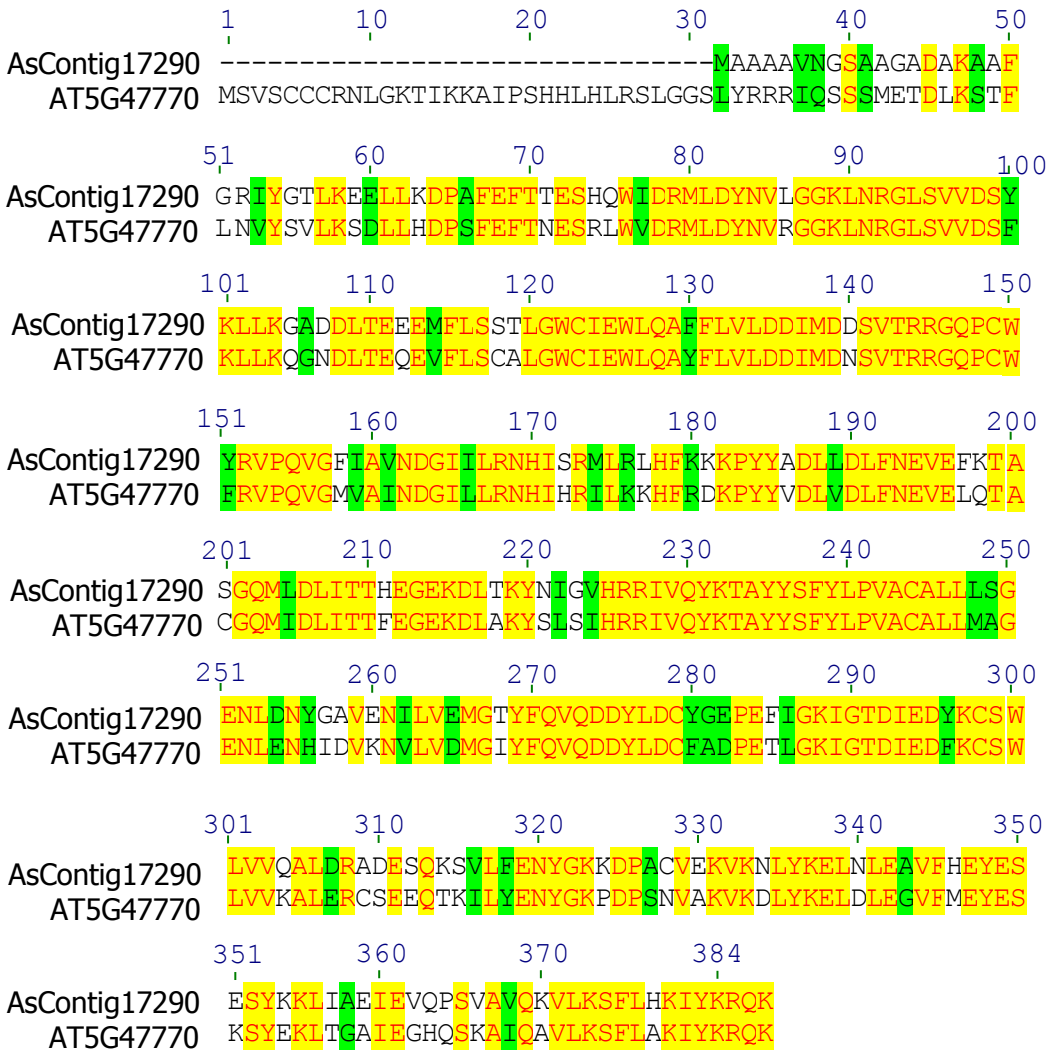
249. Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L., and Keasling, J.D., (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol*. 27, 753-9.
250. Thuenemann, E.C., Meyers, A.E., Verwey, J., Rybicki, E.P., and Lomonossoff, G.P., (2013) A method for rapid production of heteromultimeric protein complexes in plants: assembly of protective bluetongue virus-like particles. *Plant Biotechnol J*. 11, 839-46.
251. Meshcheriakova, Y.A., Saxena, P., and Lomonossoff, G.P., (2014) Fine-tuning levels of heterologous gene expression in plants by orthogonal variation of the untranslated regions of a nonreplicating transient expression system. *Plant Biotechnol J*. 12, 718-27.
252. Montague, N.P., Thuenemann, E.C., Saxena, P., Saunders, K., Lenzi, P., and Lomonossoff, G.P., (2011) Recent advances of cowpea mosaic virus-based particle technology. *Hum Vaccin*. 7, 383-90.
253. Wang, P., Wei, Y., Fan, Y., Liu, Q., Wei, W., Yang, C., Zhang, L., Zhao, G., Yue, J., Yan, X., and Zhou, Z., (2015) Production of bioactive ginsenosides Rh2 and Rg3 by metabolically engineered yeasts. *Metab Eng*. 29, 97-105.
254. Matasci, N., Hung, L.H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., Burleigh, J.G., Gitzendanner, M.A., Wafula, E., Der, J.P., dePamphilis, C.W., Roure, B., Philippe, H., Ruhfel, B.R., Miles, N.W., Graham, S.W., Mathews, S., Surek, B., Melkonian, M., Soltis, D.E., Soltis, P.S., Rothfels, C., Pokorny, L., Shaw, J.A., DeGironimo, L., Stevenson, D.W., Villarreal, J.C., Chen, T., Kutchan, T.M., Rolf, M., Baucom, R.S., Deyholos, M.K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J., and Wong, G.K., (2014) Data access for the 1,000 Plants (1KP) project. *Gigascience*. 3, 17.
255. Medema, M.H. and Osbourn, A., (2016) Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. *Nat Prod Rep*. 33, 951-62.
256. Bohacek, R.S., McMartin, C., and Guida, W.C., (1996) The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev*. 16, 3-50.
257. Harvey, A.L., Edrada-Ebel, R., and Quinn, R.J., (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov*. 14, 111-29.
258. Feher, M. and Schmidt, J.M., (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci*. 43, 218-27.
259. Schirle, M. and Jenkins, J.L., (2016) Identifying compound efficacy targets in phenotypic drug discovery. *Drug Discov Today*. 21, 82-9.
260. Szychowski, J., Truchon, J.F., and Bennani, Y.L., (2014) Natural products in medicine: transformational outcome of synthetic chemistry. *J Med Chem*. 57, 9292-308.
261. Wahlberg, I., Karlsson, K., Enzell, C.R., Svensson, S., Koskikallio, J., and Kachi, S., (1971) A Mass Spectral Fragmentation Reaction Characteristic of 11-Oxo-α-amyrin and 11-Oxo-β-amyrin Derivatives. *Acta Chem Scand*. 25, 3192-3193.
262. Brendolise, C., Yauk, Y.K., Eberhard, E.D., Wang, M., Chagne, D., Andre, C., Greenwood, D.R., and Beuning, L.L., (2011) An unusual plant triterpene synthase with predominant α-amyrin-producing activity identified by characterizing oxidosqualene cyclases from *Malus* x *domestica*. *FEBS J*. 278, 2485-99.

263. Sawai, S., Shindo, T., Sato, S., Kaneko, T., Tabata, S., Ayabe, S.-i., and Aoki, T., (2006) Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus. Plant Science*. 170, 247-57.
264. Fazio, G.C., Xu, R., and Matsuda, S.P., (2004) Genome mining to identify new plant triterpenoids. *J Am Chem Soc*. 126, 5678-79.

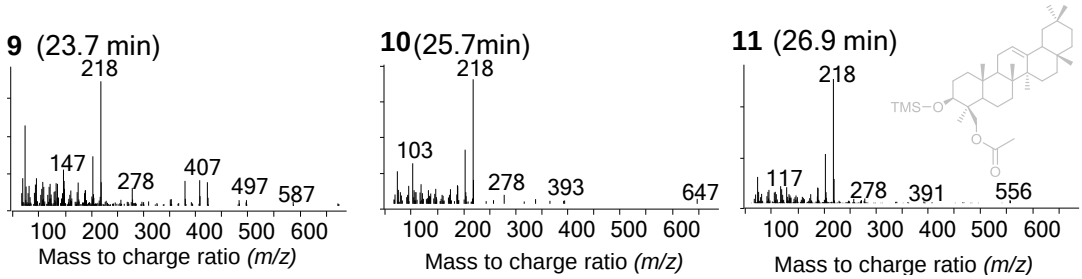
Supplementary Figures



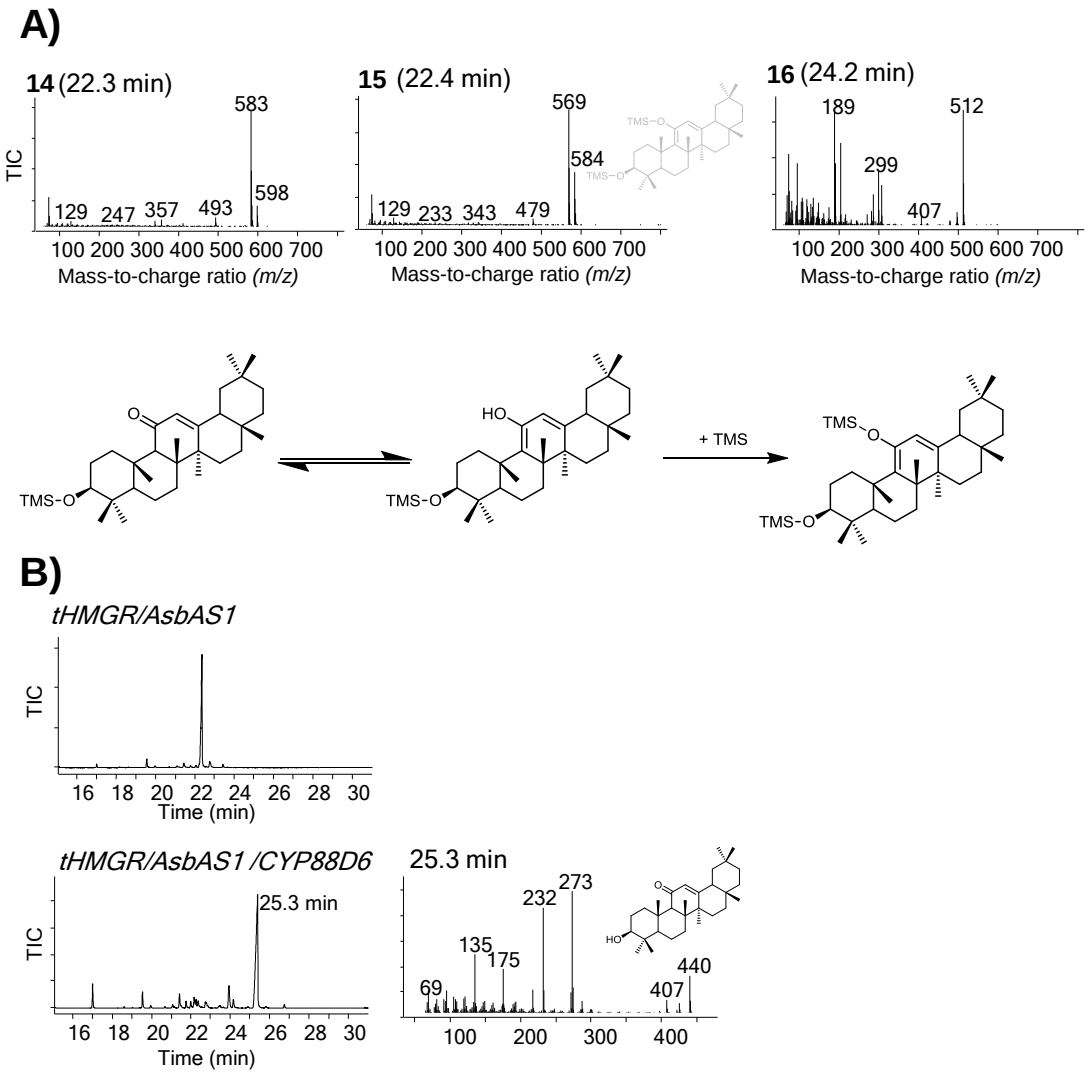
Supplementary Figure 1: Amino acid sequence alignment of the three *A. strigosa* isoforms of HMGR as identified in the root tip 454 transcriptomic database. Sequence conservation is indicated by colour. Yellow indicates absolute conservation of residues between the three sequences while blue indicates conservation in two sequences only. Only a partial sequence was available for AsContig20171. Green indicates different but chemically similar amino acids.



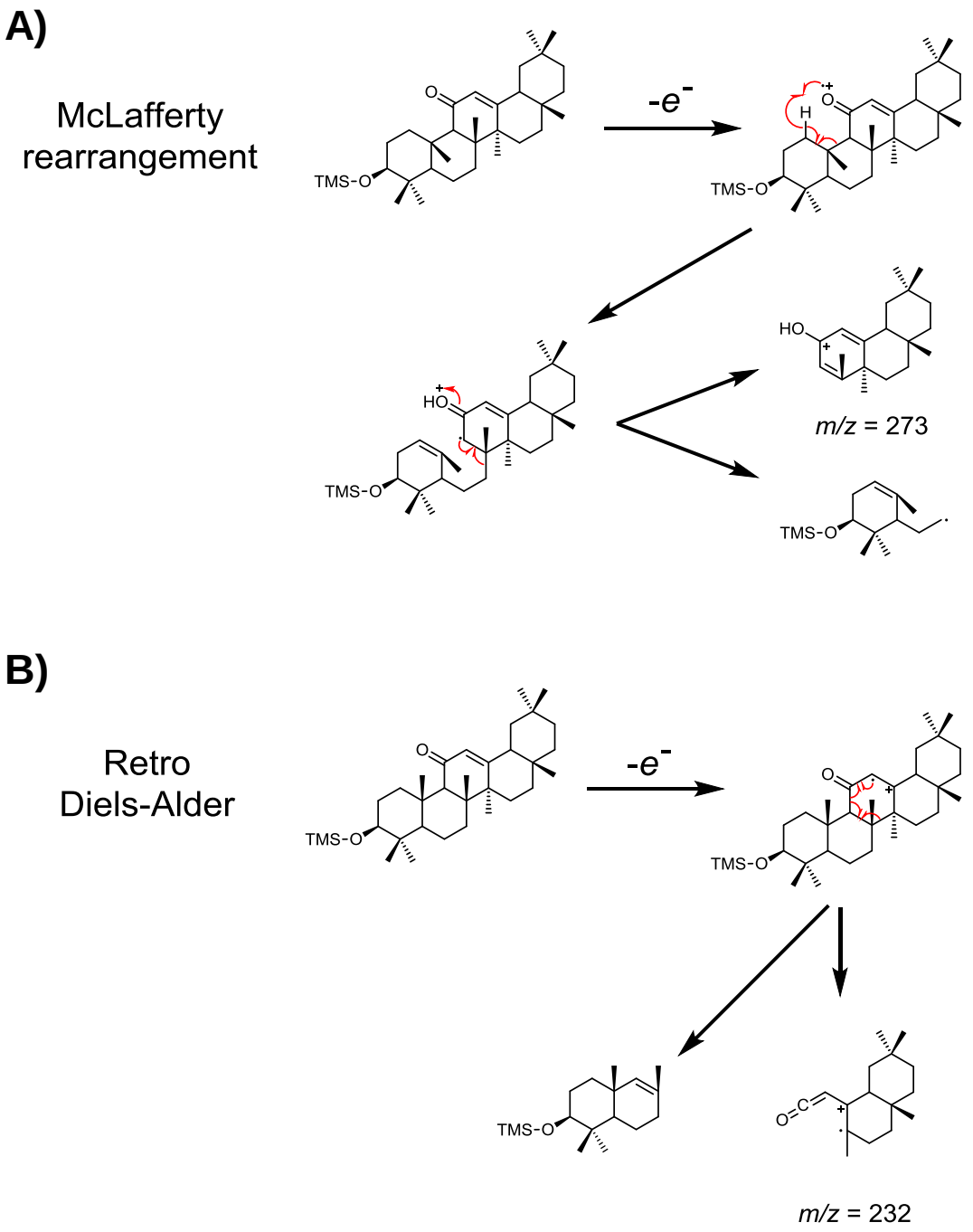
Supplementary Figure 2: Amino acid sequence alignment between the *A. thaliana* and *A. strigosa* FPS proteins. The mitochondrially-targeted *A. thaliana* FPS (AT5G47770) and the putative *A. strigosa* FPS (AsContig17290) used in this study were aligned using vector NTI. Yellow highlighting indicates absolute sequence conservation between the two proteins, while green indicates different but chemically similar residues.



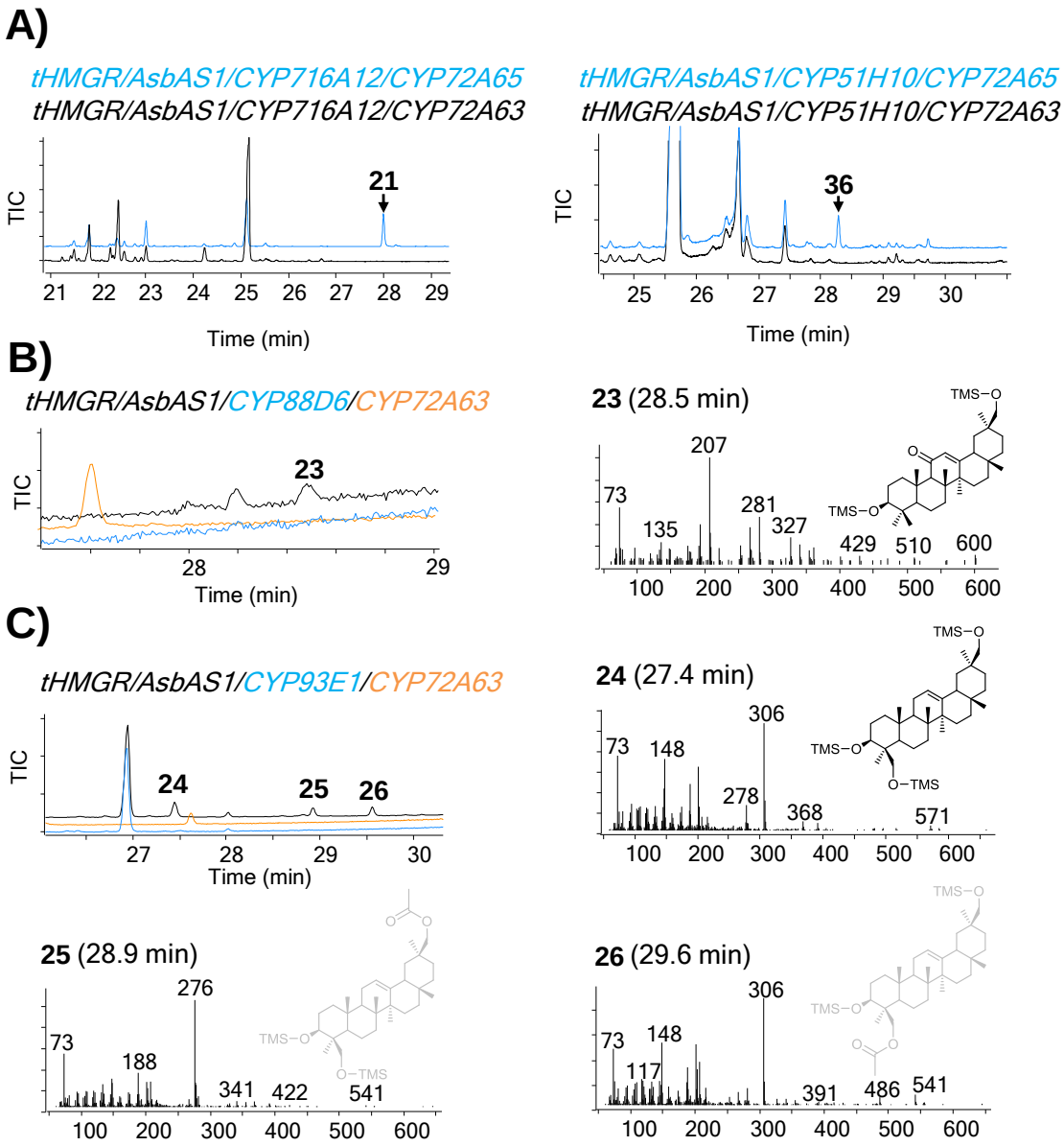
Supplementary Figure 3: Mass spectra for products not shown in **Figure 4-4** from leaves expressing *tHMGR/AsbAS1/CYP93E1.* Peak **11** is expected to be the C-24 acetylated form of 24-hydroxy-β-amyrin.



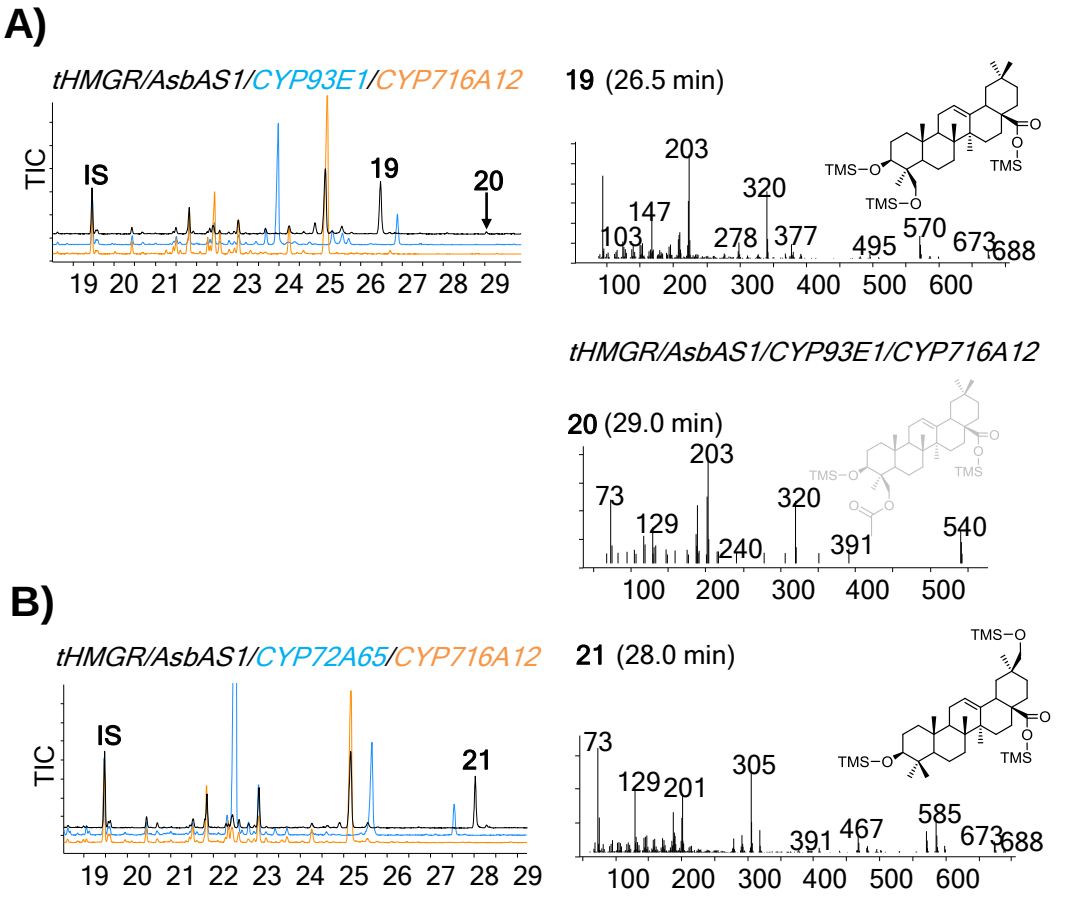
Supplementary Figure 4: Additional products observed upon GC-MS analysis of 11-oxo-β-amyrin. **A)** Mass spectra for products not shown in **Figure 4-5**. The product at 22.4 minutes is likely to be a tautomerised and silylated form of 11-oxo-β-amyrin resulting from the basic environment of the pyridine solvent present in the derivitising reagent. (see scheme below) **B)** GC-MS analysis of underivitised *N. benthamiana* leaf extracts expressing *tHMGR/AsbAS1/CYP88D6.* In contrast to the derivitised samples, one major peak was dominant in the sample. This had the expected mass of non-silylated 11-oxo-β-amyrin (MW = 440).



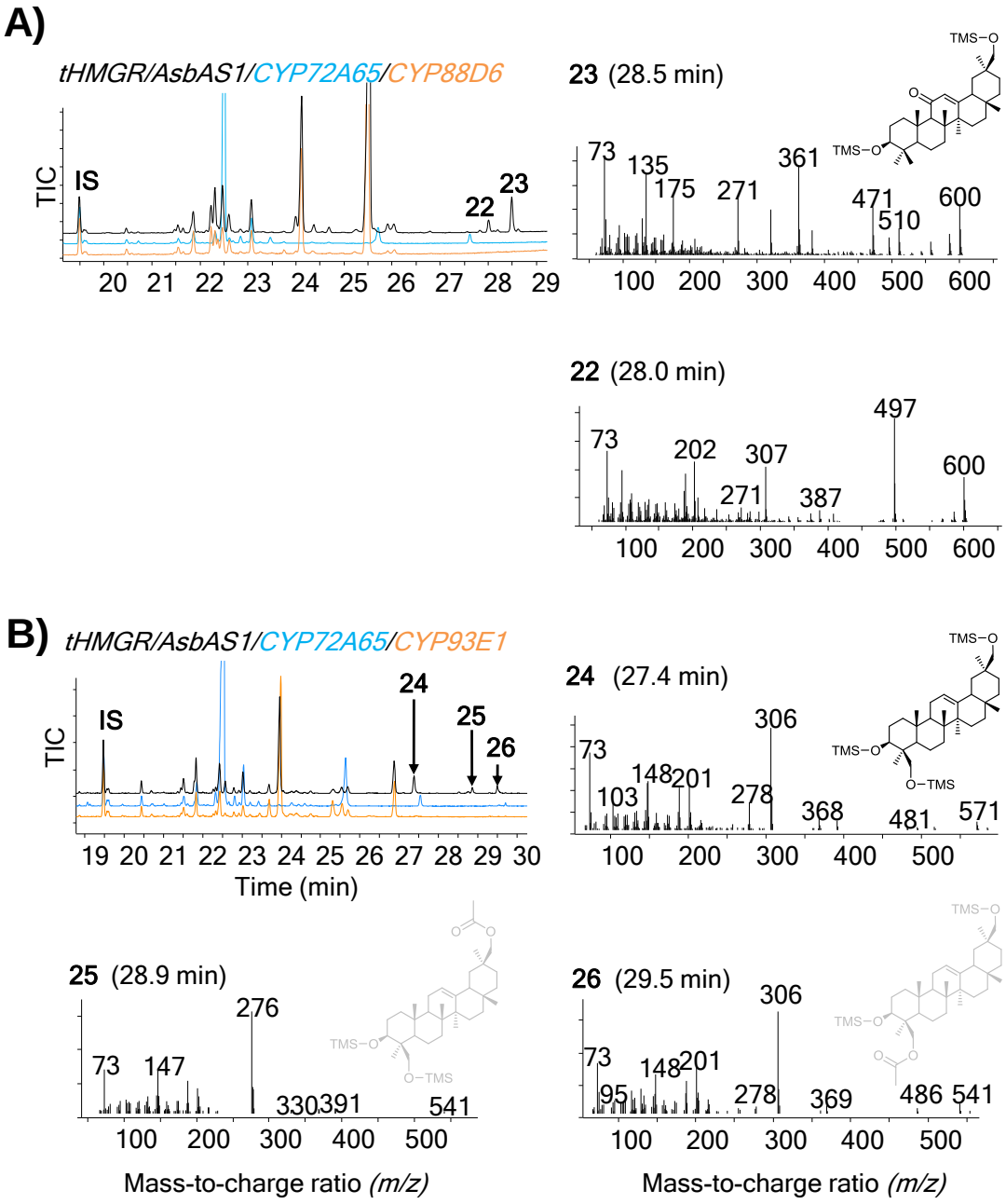
Supplementary Figure 5: EI fragmentation of 11-oxo-β-amyrin. A) The presence of a C-11 carbonyl group in β-amyrin derivatives allows γ- (McLafferty) rearrangement during fragmentation. For 11-oxo-β-amyrin this gives rise to a characteristic ion at *m/z* 273. This mass shifts accordingly with the presence of functionalities in the C,D or E rings. B) 11-oxo-β-amyrin can also undergo retro Diels-Alder fragmentation as with β-amyrin.



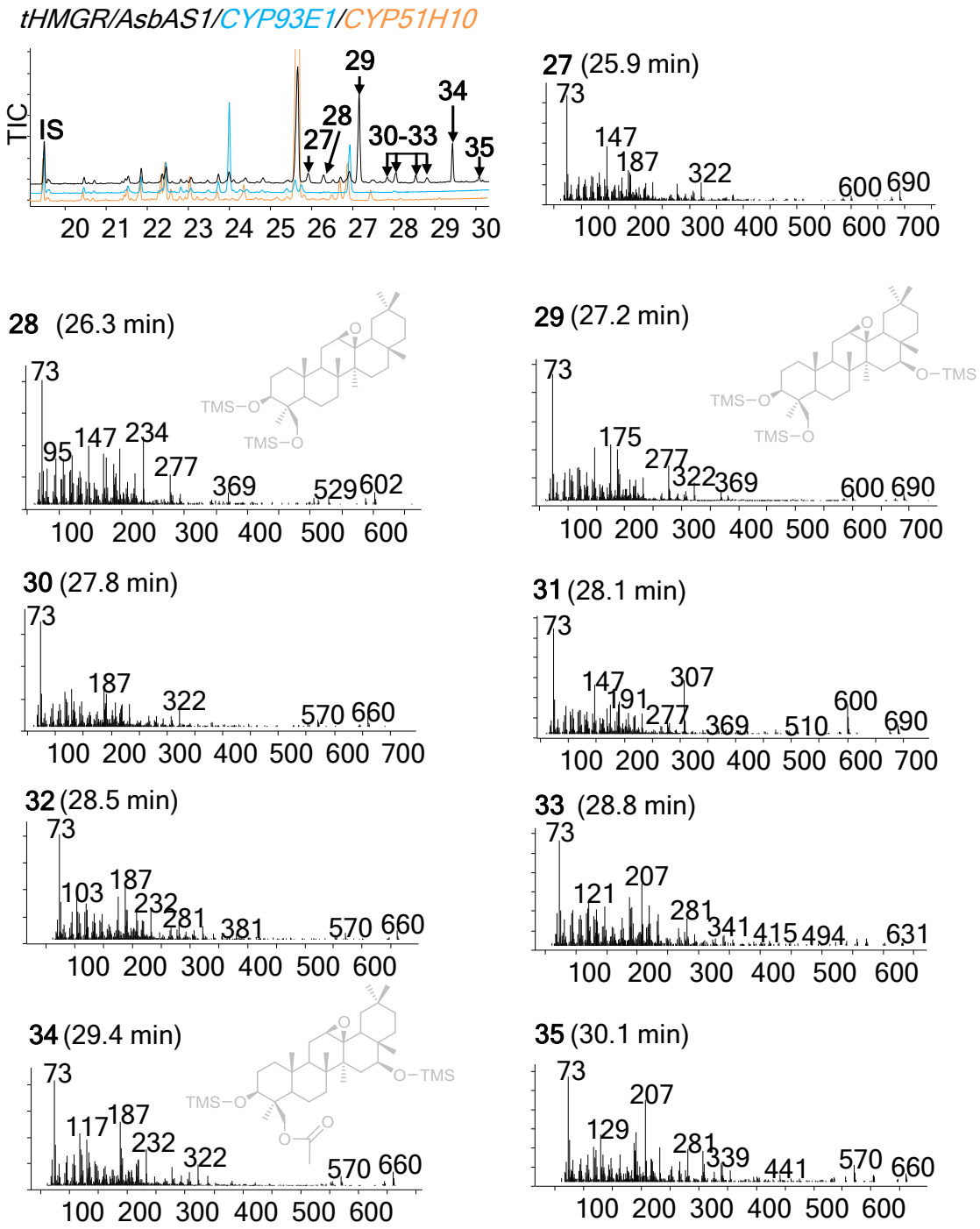
Supplementary Figure 6: Combinatorial biosynthesis with CYP72A63. These combinations resulted in much poorer accumulation of the products versus the equivalent combination expressing CYP72A65. **A)** Combinations of CYP72A63 with either CYP716A12 (**left**) or CYP51H10 (**right**) failed to reveal any novel products. Black chromatograms represent the combinatorial samples containing CYP72A63, and blue chromatograms represent the equivalent sample with CYP72A65. The combinatorial products found in the CYP72A65 samples are labelled. **B)** Combinations of *tHMGR/AsbAS1/CYP88D6/CYP72A63* and **C)** *tHMGR/AsbAS1/CYP93E1/CYP72A63* resulted in trace amounts of products which matched those shown in **Supplementary Figure 8**. Mass spectra from the CYP72A63 sample are shown.



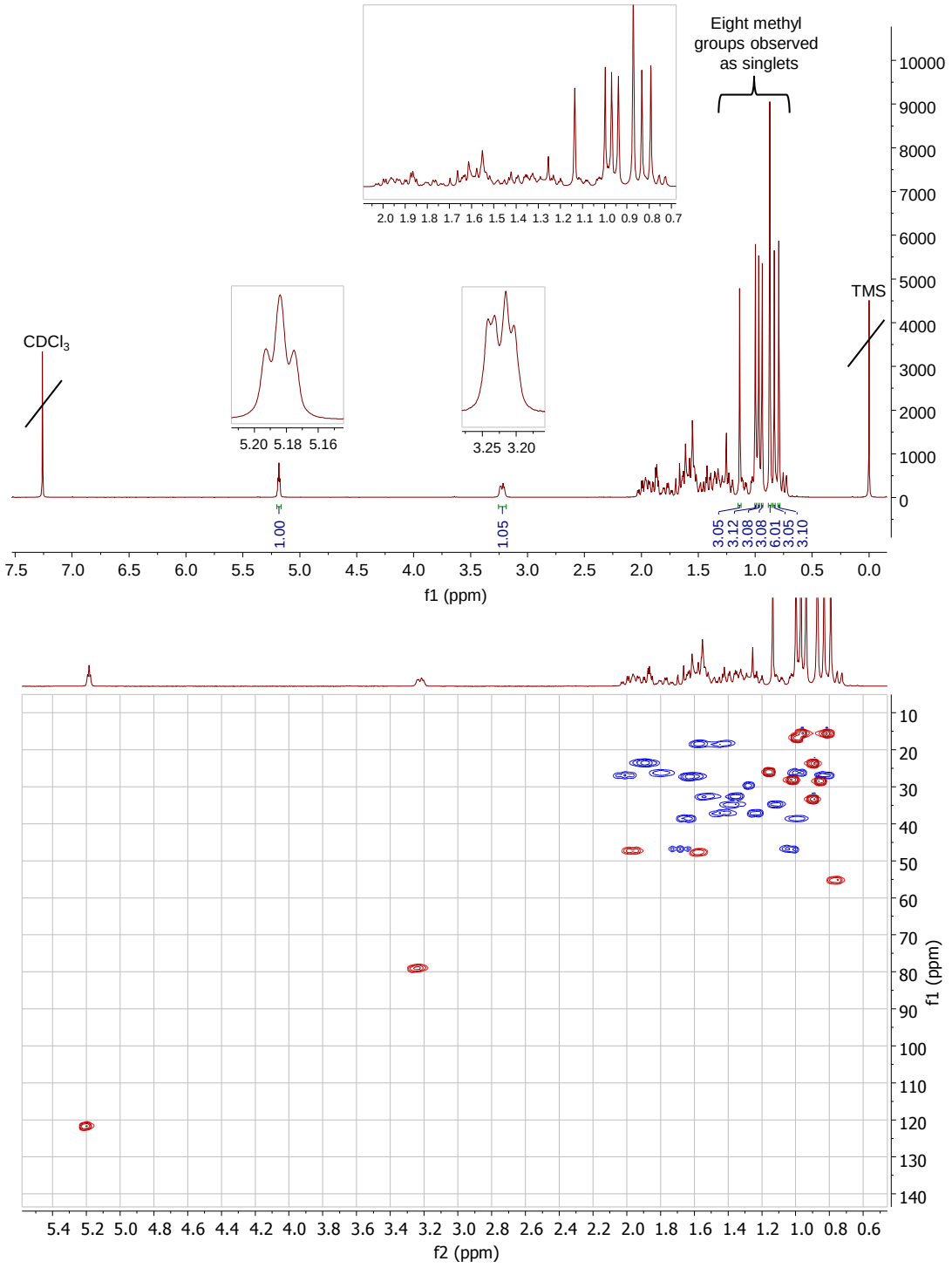
Supplementary Figure 7: Combinatorial biosynthesis of compounds which could be verified against previously-published data. Total ion chromatograms (TIC) of *N. benthamiana* leaves expressing *tHMGR*, *AsbAS1* and **A)** *CYP93E1* and *CYP716A12*, **B)** *CYP72A65* and *CYP716A12*, Chromatograms are shown on the left side, with the black chromatograms representing the combinatorial sample, and single-P450 controls in orange and blue. In each case the EI mass spectrum for the major combinatorial product is given on the right and below. IS – Internal standard (Coprostanol).



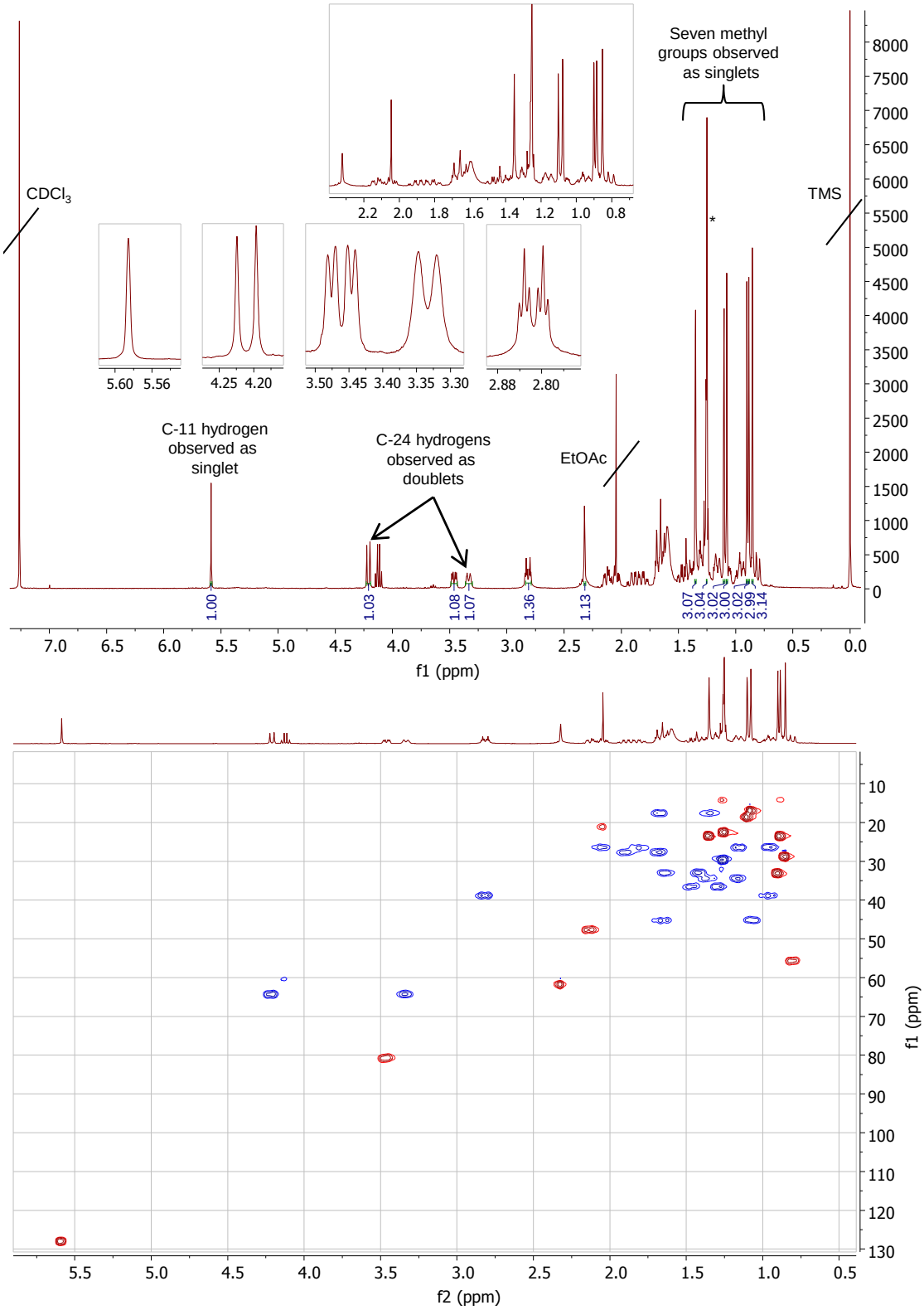
Supplementary Figure 8: Combinatorial biosynthesis of compounds which could be verified against previously-published data. Total ion chromatograms (TIC) of *N. benthamiana* leaves expressing *tHMGR*, *AsbAS1* and **A)** *CYP72A65* and *CYP88D6* and **B)** *CYP72A65* and *CYP93E1*. Chromatograms are shown on the left side, with the black chromatograms representing the combinatorial sample, and single-P450 controls in orange and blue. In each case the EI mass spectrum for the major combinatorial product is given on the right and below. IS – Internal standard (Coprostanol).



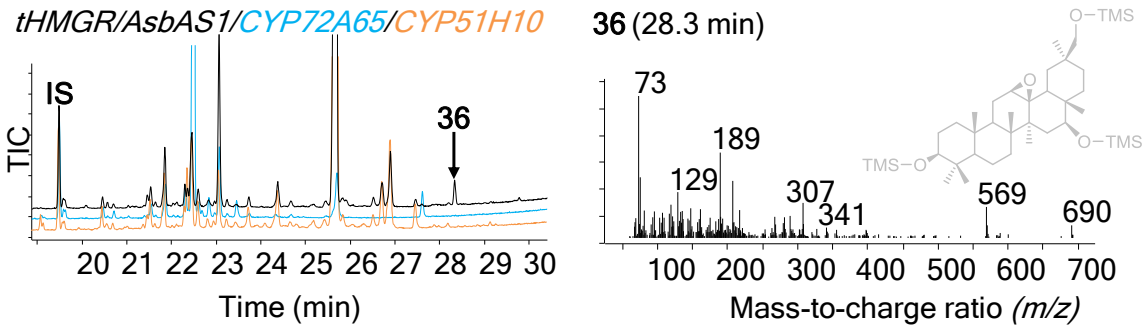
Supplementary Figure 9: GC-MS analysis of *N. benthamiana* leaf extracts expressing *tHMGR/AsbAS1/CYP51H10/CYP93E1.* The total ion chromatogram (TIC) is shown on the upper left. Black chromatograms represent the combinatorial sample, with single-P450 controls in orange and blue. The EI mass spectra for predicted combinatorial products is given to the right and below the chromatogram. IS – Internal standard (Coprostanol).



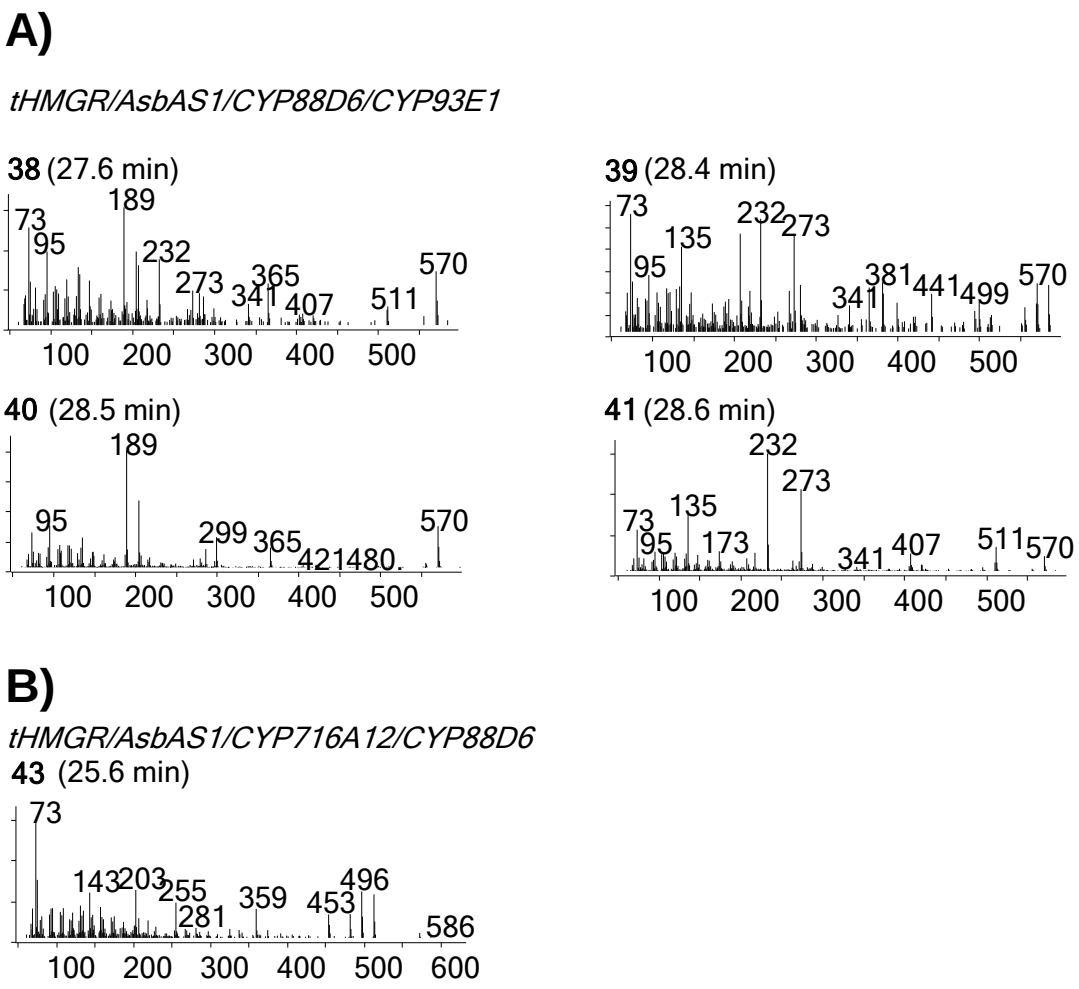
Supplementary Figure 10: NMR spectra for β-amyrin. Upper panel: Proton (^{1}H) NMR of β-amyrin in $CDCl_3$. Lower panel: DEPT-edited HSQC.



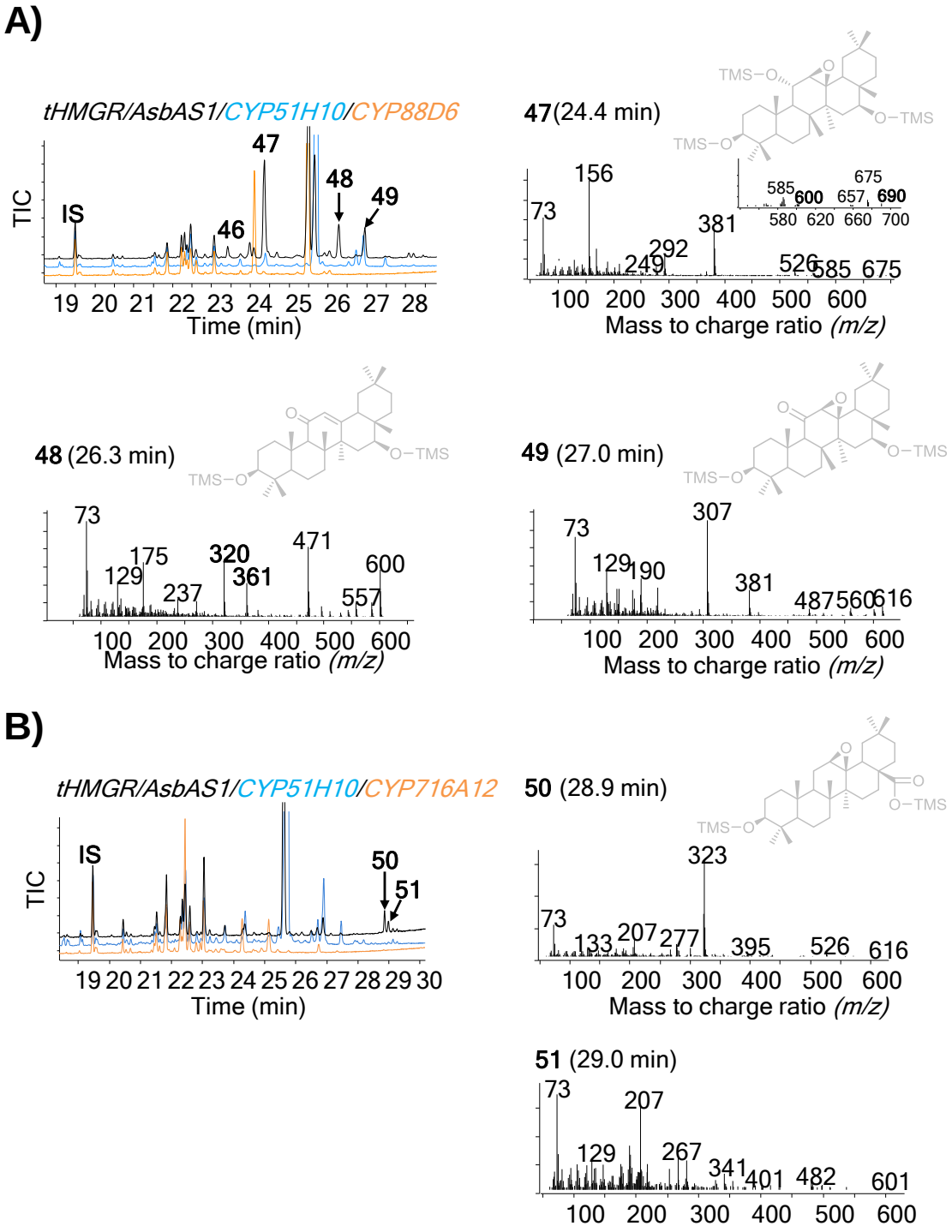
Supplementary Figure 11: NMR spectra for 11-oxo, 24-hydroxy-β-amyrin (**37**). Upper panel: Proton (^{1}H) NMR of β-amyrin in $CDCl_3$. Important resonances described in the main text are annotated. The singlet marked * overlapped with a contaminating impurity. Lower panel: DEPT-edited HSQC. A full assignment of this compound (in deuterated pyridine) is provided in **Table 4-5.**



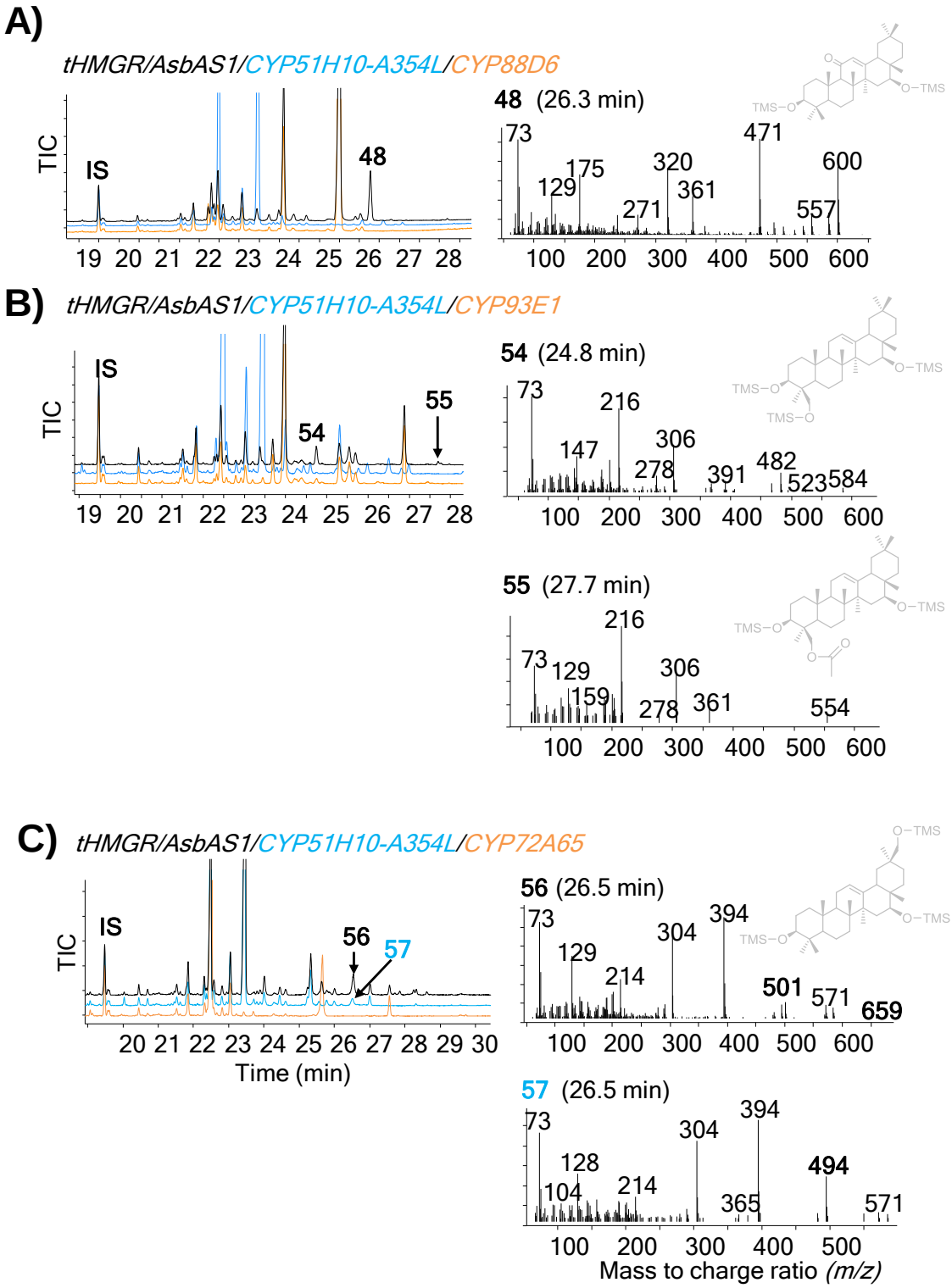
Supplementary Figure 12: GC-MS analysis of *N. benthamiana* leaf extracts expressing *tHMGR/AsbAS1* with *CYP72A65* and *CYP51H10.* The total ion chromatograms (TIC) is shown on the left side with the black chromatogram representing the combinatorial sample and single-P450 controls in orange and blue. The EI mass spectra for peak 36 is given to the right of the chromatogram. IS – Internal standard (Coprostanol).



Supplementary Figure 13: Mass spectra for peaks not shown in the main text **A)** Peaks 38-41 not shown in **Figure 4-9A. B)** Peak 43 not shown in **Figure 4-11A**

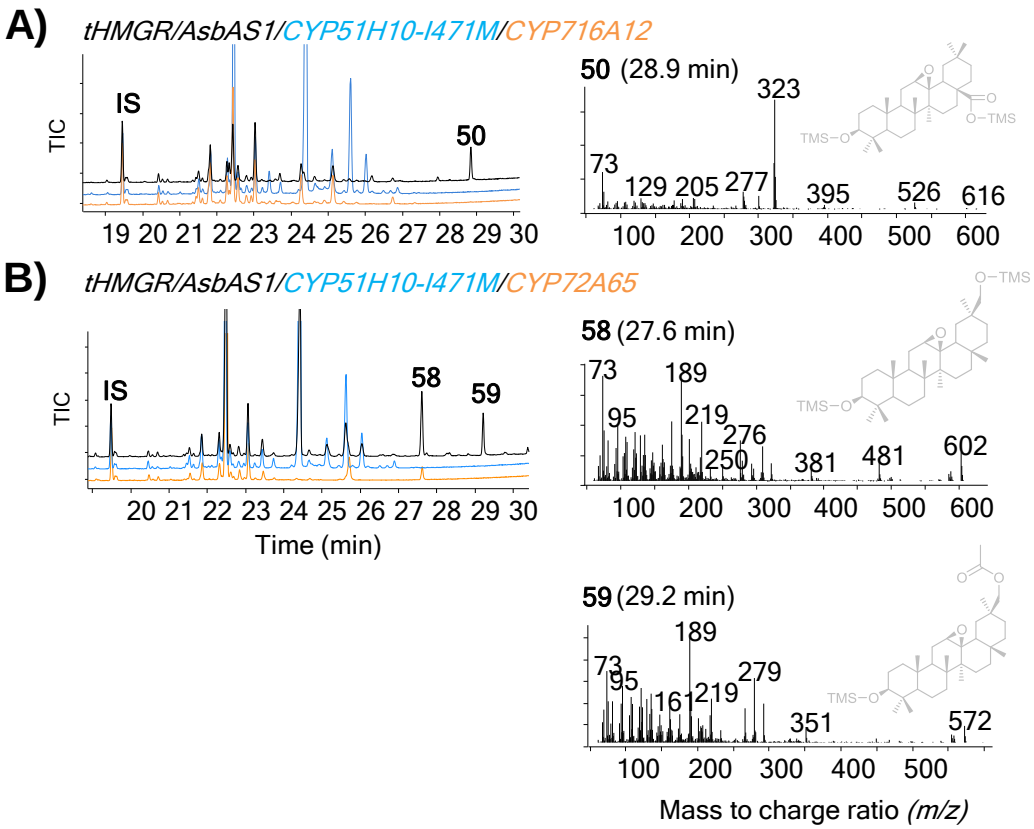


Supplementary Figure 14: GC-MS analysis of *N. benthamiana* leaves expressing *tHMGR* and *AsbAS1* with **A)** *CYP51H10* and *CYP88D6* and **B)** *CYP51H10* and *CYP716A12*. Total ion chromatograms (TIC) are given in the upper left of each section with the combinatorial sample in black and control samples in orange and blue. Mass spectral data for the major products to the right and/or below. Putative product structures are given in grey.

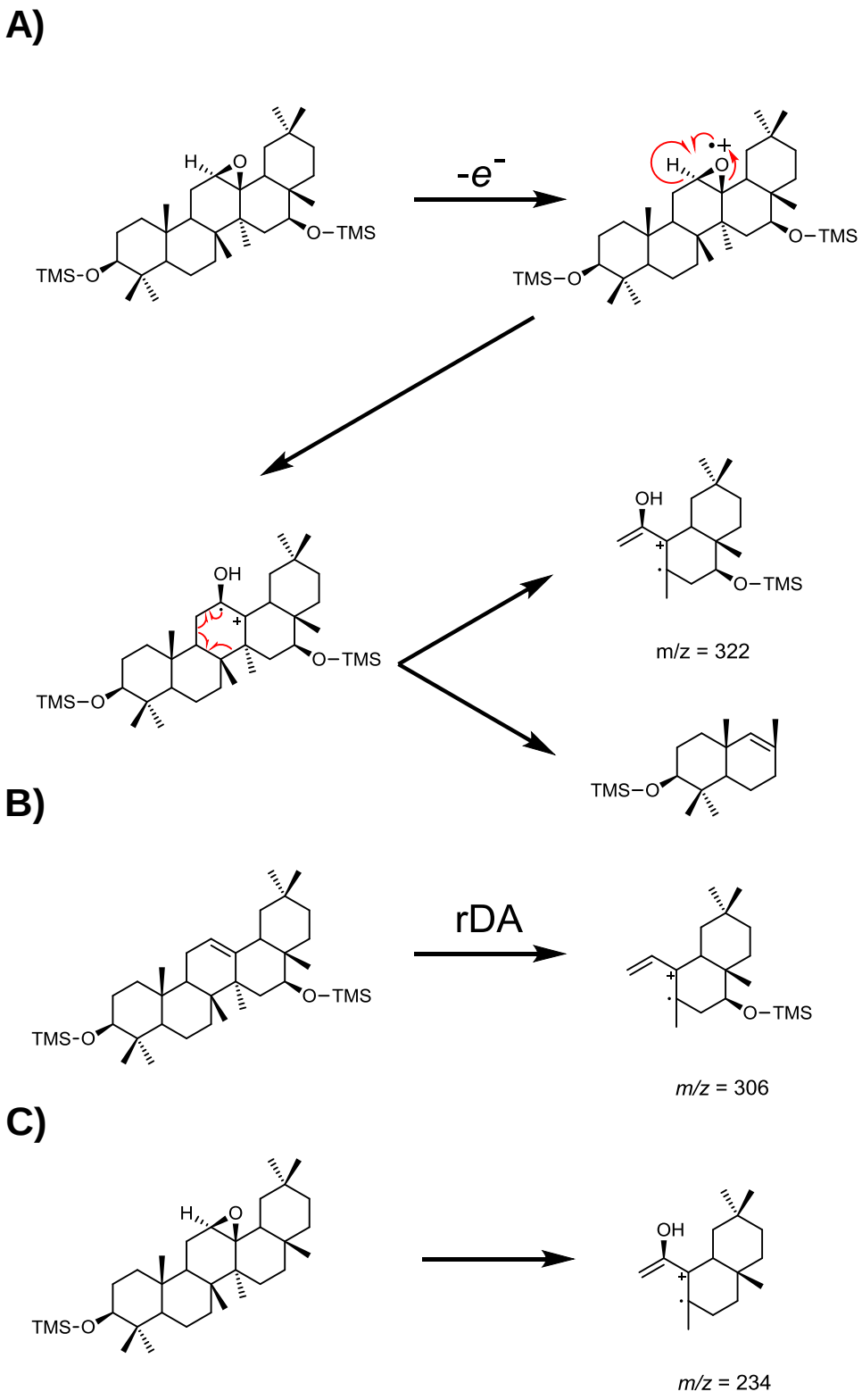


Supplementary Figure 15: Combinatorial biosynthesis utilising CYP51H10-A354L. *N. benthamiana* leaves expressing **A)** *tHMGR/AsbAS1/CYP51H10-A354L/CYP88D6* **B)** *tHMGR/AsbAS1/ CYP51H10-A354L/CYP93E1* **C)** *tHMGR/AsbAS1/CYP51H10-A354L/CYP72A65.* GC-MS total ion chromatograms (TIC) are given on the left with the combinatorial sample (black) overlaid with single-P450 controls (orange and blue). Mass spectra for the major products and predicted structures are given on the right and below. For C) *tHMGR/AsbAS1/CYP51H10/CYP72A65* leaves accumulate a product (**56**) which has a similar retention time and mass spectrum to a peak observed in the control sample (**57**). The

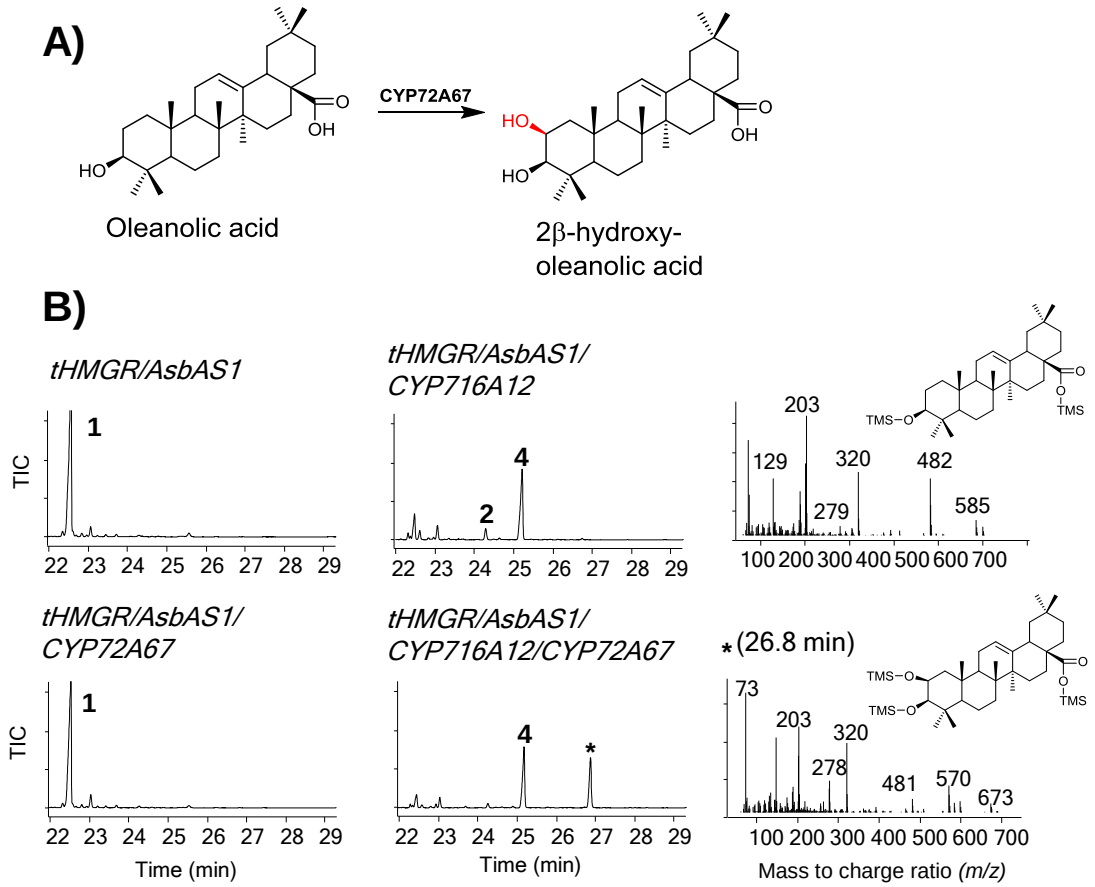
presence of unique ions (bold) in each sample suggest that these are two different compounds.



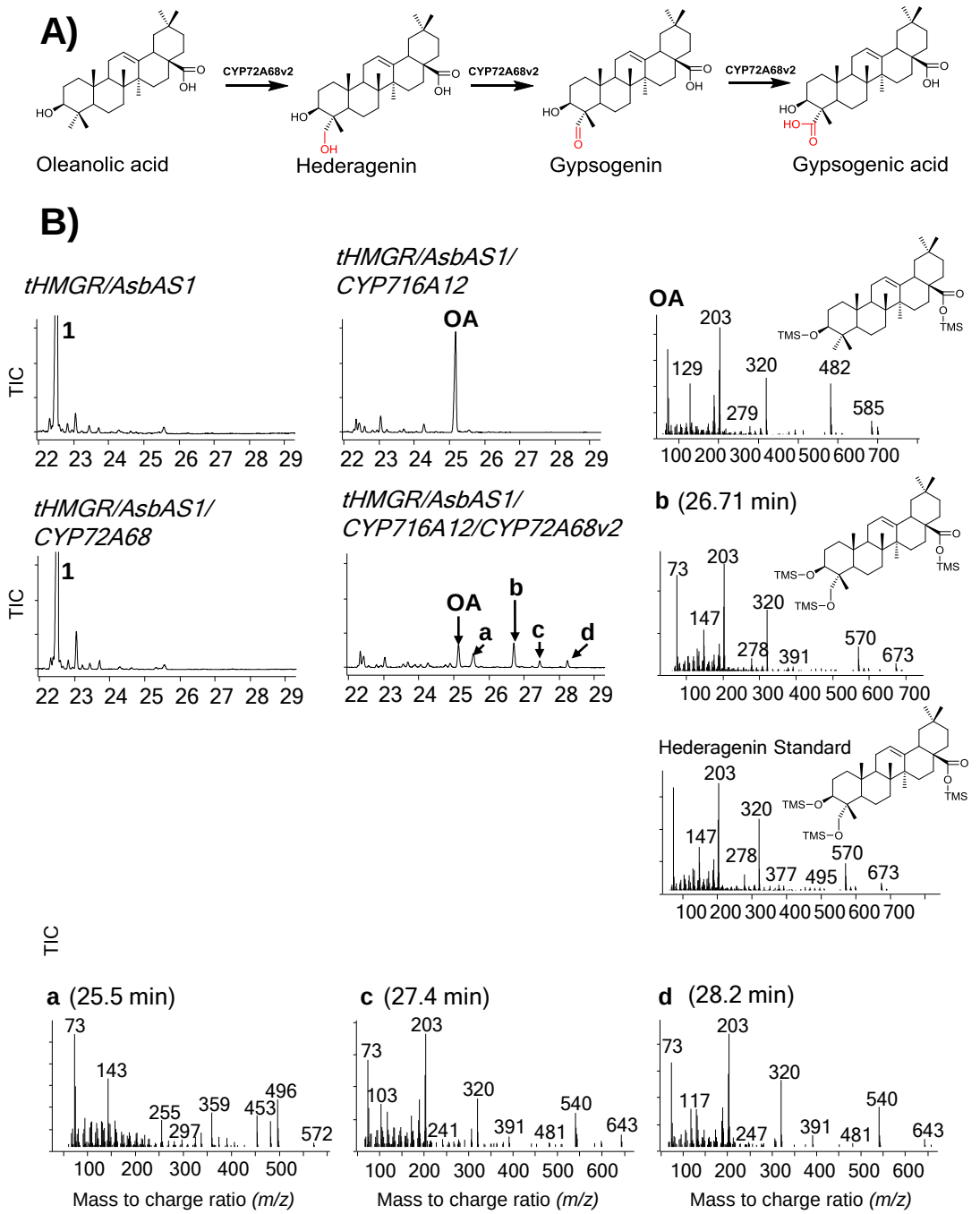
Supplementary Figure 16 Combinatorial biosynthesis utilising CYP51H10-I471M. GC-MS analysis of *N. benthamiana* leaves expressing **A)** *tHMGR/AsbAS1/CYP51H10-I471M/CYP716A12* and **B)** *tHMGR/AsbAS1/CYP51H10-I471M/CYP72A65.* Total ion chromatograms (TIC) are displayed on the left side with combinatorial sample chromatograms given in black and controls in blue and orange. Mass spectra and putative structures for the major products given to the right and below.



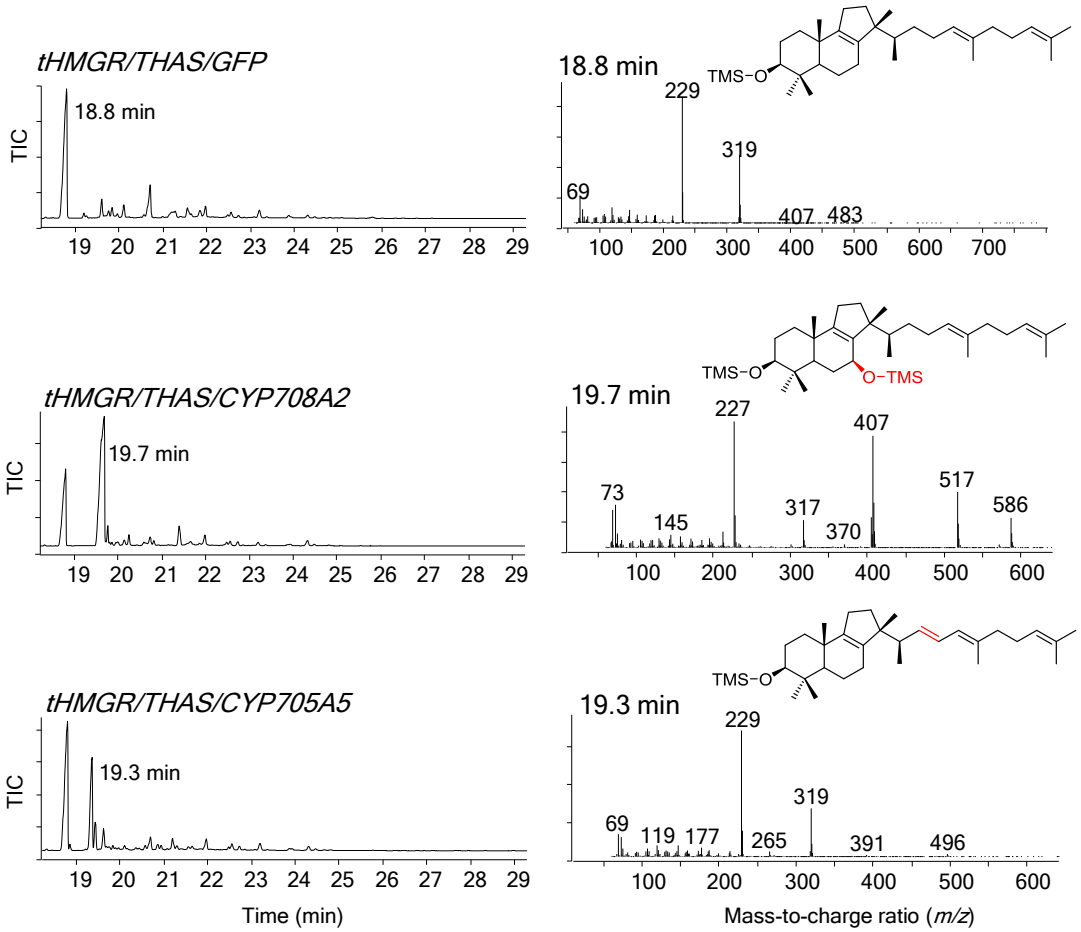
Supplementary Figure 17: EI fragmentation of EpHβA and CYP51H10 mutants. **A)** Proposed fragmentation for the wild-type CYP51H10 product (12,13β-epoxy,16β-hydroxy-β-amyrin) leading to the formation of the ion 322. **B)** retro Diels-Alder (rDA) fragmentation of 16β-hydroxy-β-amyrin, the product of CYP51H10-A354L. **C)** Proposed fragmentation of the CYP51H10-I471M product 12,13β-epoxy-β-amyrin. The mechanism is likely to be similar to that in A), leading to the ion at *m/z* 234.



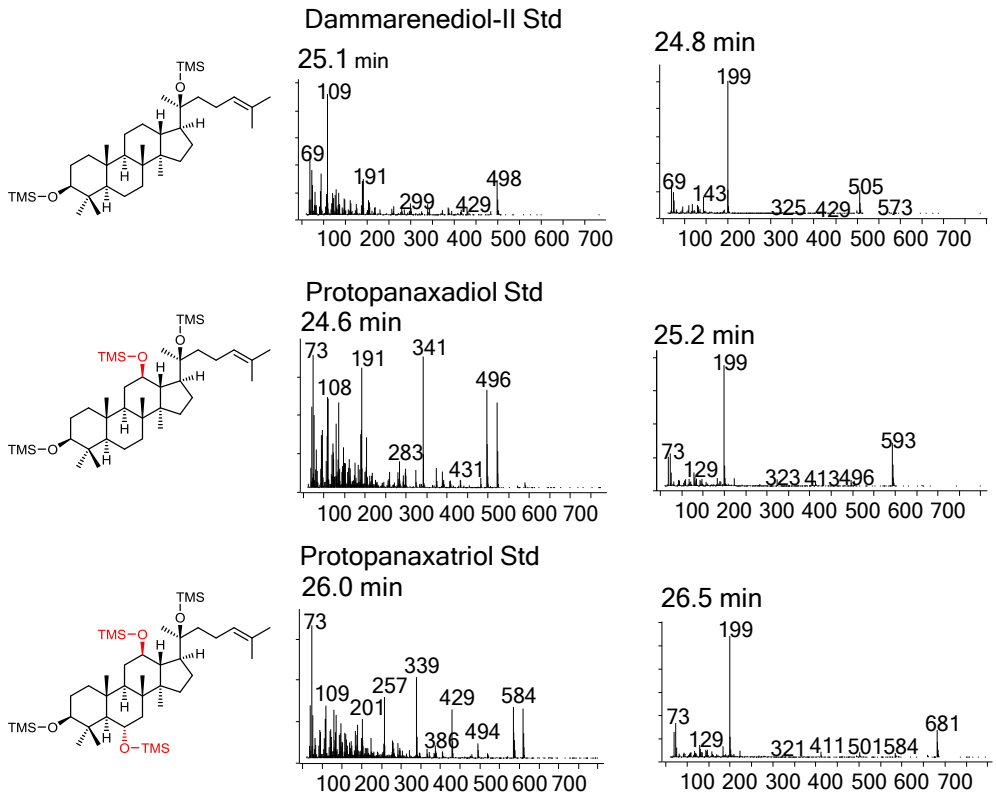
Supplementary Figure 18: Production of 2β-hydroxy-oleanolic acid in *N. benthamiana* by coexpression of *tHMGR/AsbAS1/CYP716A12/CYP72A67*. **A)** Schematic for C-2 oxidation of oleanolic acid to 2β-hydroxy-oleanolic acid. **B)** GC-MS analysis of leaf extracts from coexpression of *tHMGR/AsbAS1/CYP716A12/CYP72A67* in *N. benthamiana.*



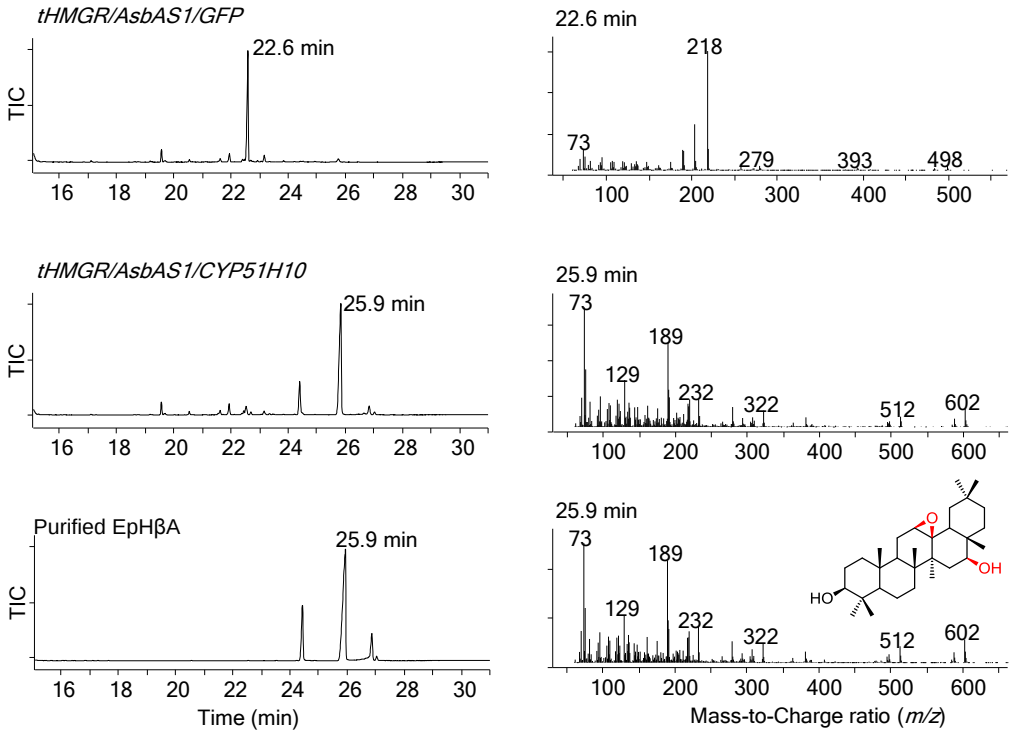
Supplementary Figure 19: Production of hederagenin in *N. benthamiana* by coexpression of *tHMGR, AsbAS1, CYP716A12* and CYP72A68v2. **A)** Schematic for the step C-23 oxidation of oleanolic acid to gypsogenic acid. **B)** GC-MS data from *N. benthamiana* leaves expressing *tHMGR/AsbAS1/ CYP716A12/CYP72A68v2* plus control extracts. Gypsogenin and gypsogenic acid were not detected using the present method.



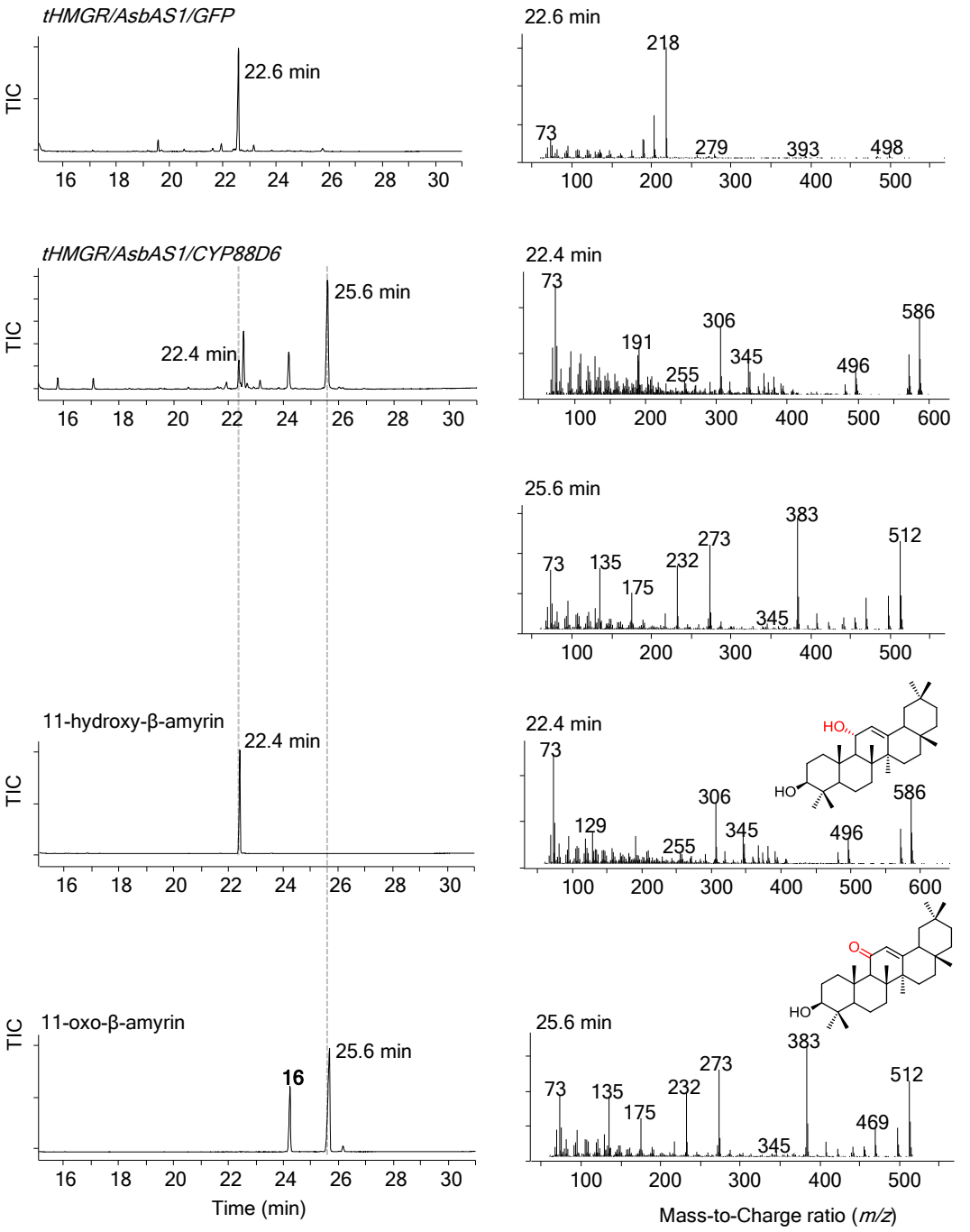
Supplementary Figure 20: Production and oxidation of thalianol in *N. benthamiana.* GC-MS total ion chromatograms (TIC) are shown on the left from analysis of extracts of leaves expressing *tHMGR* with the *Arabidopsis thaliana* thalianol synthase (THAS). Oxidation of this scaffold was further achieved by coexpression of the thalianol oxidases *CYP708A2* (middle) and *CYP705A5* (bottom). Details of these genes are given in **Supplementary Table 5**.



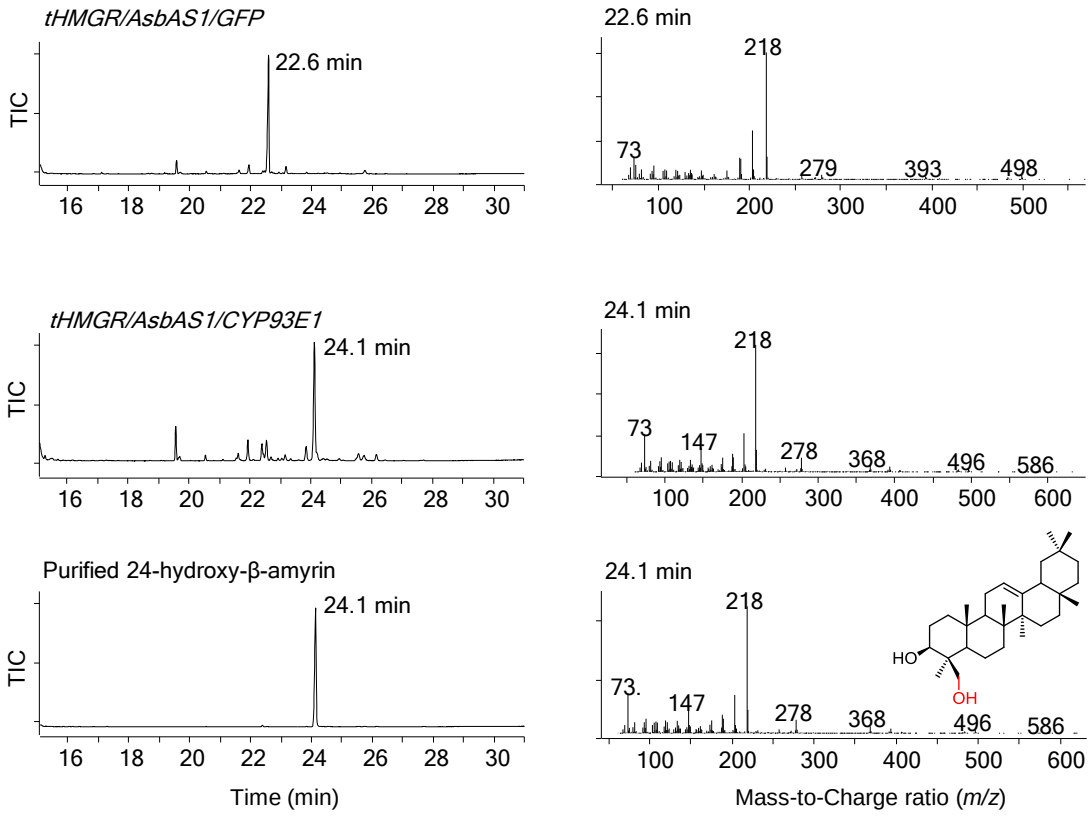
Supplementary Figure 21: Mass spectra for peaks detailed in **Figure 4-15**. Each of the standards for Dammarenediol-II, Protopanaxadiol and Protopanaxatriol result in two peaks by GC-MS.



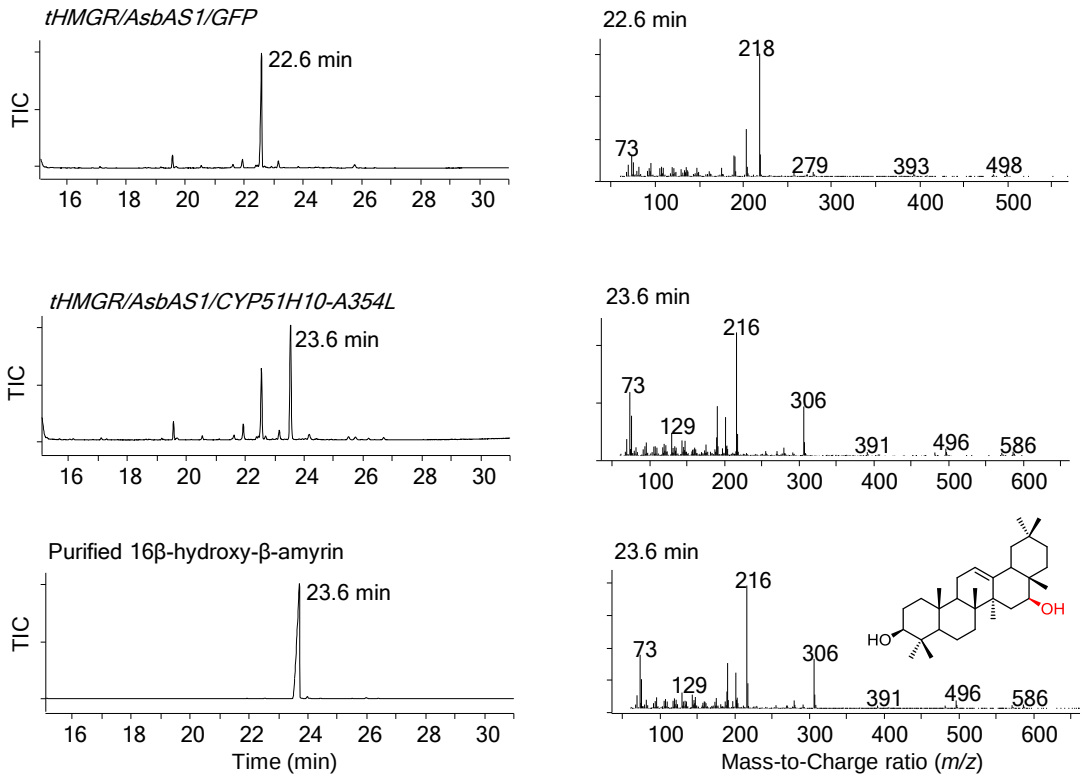
Supplementary Figure 22: GC-MS analysis of the purified EpHβA product. Total ion chromatograms (TIC) for *N. benthamiana* leaves expressing *tHMGR/AsbAS1/CYP51H10* (middle left). Control *tHMGR/AsbAS1/GFP* leaves are shown at the top. GC-MS analysis of the purified EpHβA product is shown to the bottom left. EpHβA can degrade under the conditions of the GC-MS to form multiple peaks. The mass spectrum for the major peaks in the samples is shown to the right of the chromatogram.



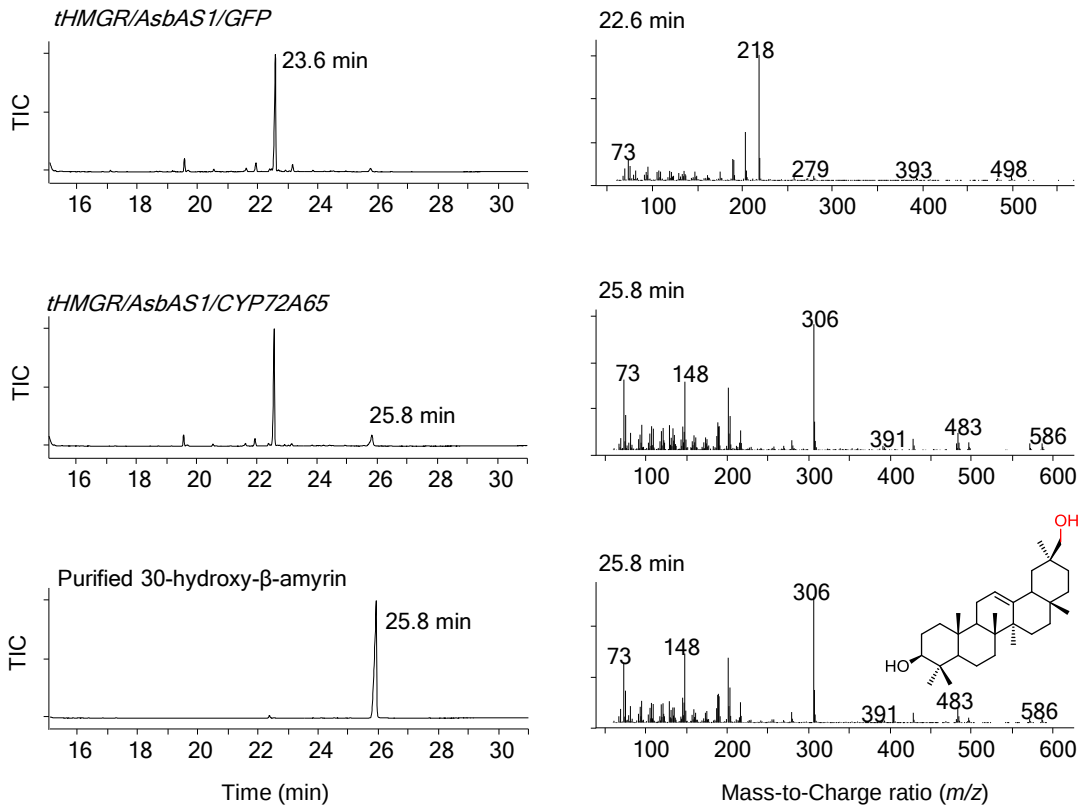
Supplementary Figure 23: GC-MS analysis of the purified 11-α-hydroxy-β-amyrin and 11-oxo-β-amyrin. Total ion chromatograms (TIC) for control *N. benthamiana* leaves expressing *tHMGR/AsbAS1* are shown at the top left. Leaves expressing *tHMGR/AsbAS1/CYP88D6* are shown below. GC-MS analyses of the purified products are shown to the bottom left. Product **16** in the 11-oxo-β-amyrin sample is believed to be an artefact of the derivatisation process. The mass spectrum for the indicated peaks in the samples is shown to the right of the chromatograms.



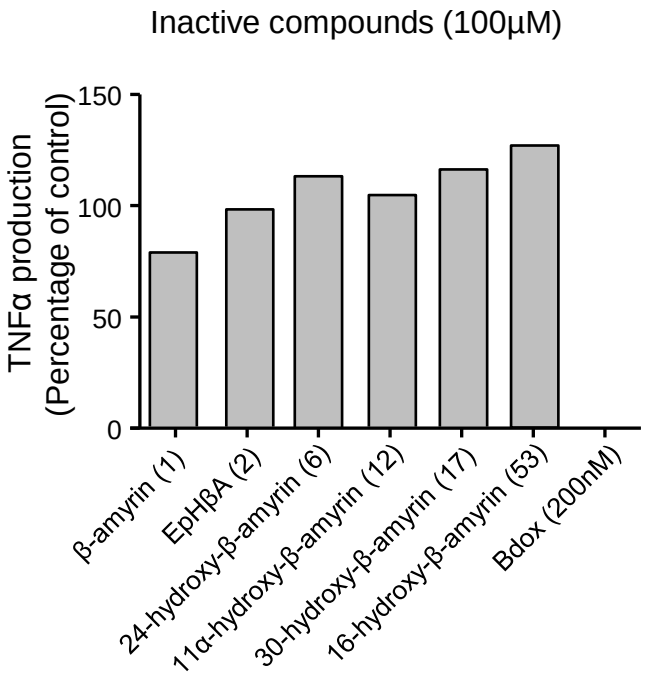
Supplementary Figure 24: GC-MS of the purified 24-hydroxy-β-amyrin. Total ion chromatograms (TIC) for control *N. benthamiana* leaves expressing *tHMGR/AsbAS1/GFP* are shown at the top left. Leaves expressing *tHMGR/AsbAS1/CYP93E1* are shown middle-left. The TIC chromatogram for the purified 24-hydroxy-β-amyrin is shown to the bottom left. Mass spectra are shown on the right of the chromatograms.



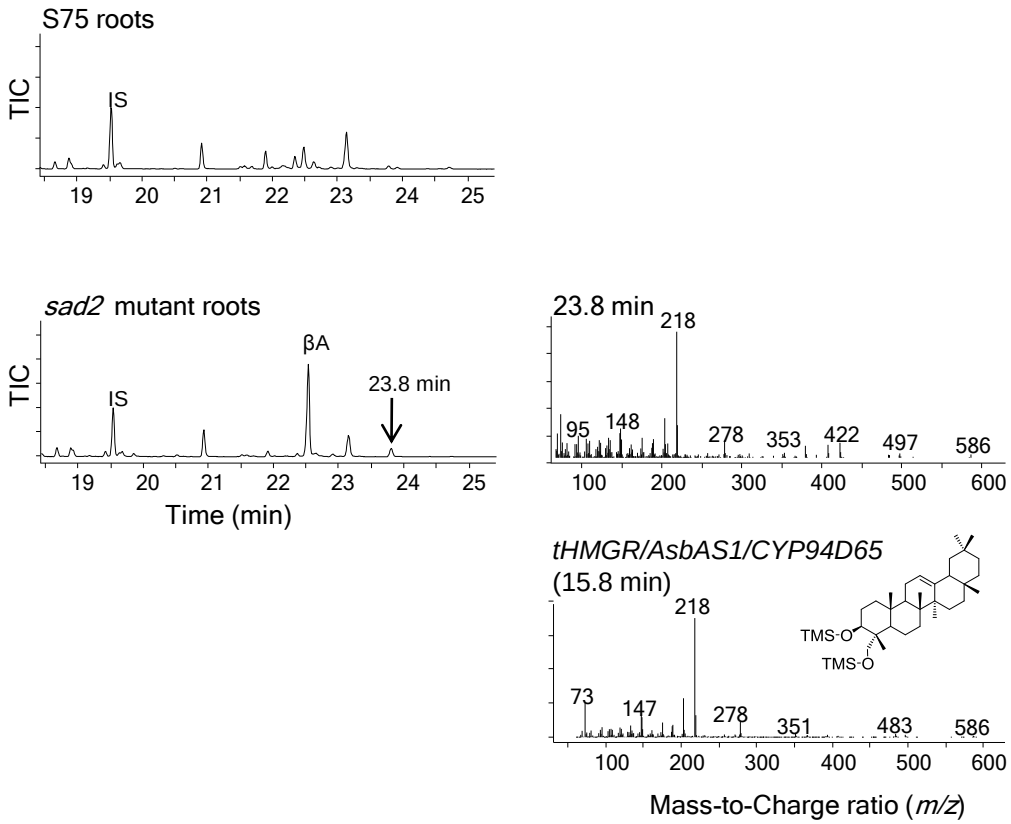
Supplementary Figure 25: GC-MS of the purified 16β-hydroxy-β-amyrin. Total ion chromatograms (TIC) of control *N. benthamiana* leaves expressing *tHMGR/AsbAS1/GFP* are shown at the top left. Leaves expressing *tHMGR/AsbAS1/CYP51H10-A354L* are shown middle-left. The TIC chromatogram for the purified 16β-hydroxy-β-amyrin is shown to the bottom left. Mass spectra are shown on the right of the chromatograms.



Supplementary Figure 26: GC-MS of the purified 30-hydroxy-β-amyrin. Total ion chromatograms (TIC) of control *N. benthamiana* leaves expressing *tHMGR/AsbAS1/GFP* are shown at the top left. Leaves expressing *tHMGR/AsbAS1/CYP72A65* are shown middle-left. The TIC chromatogram for the purified 30-hydroxy-β-amyrin is shown to the bottom left. Mass spectra are shown on the right of the chromatograms.



Supplementary Figure 27: ELISA measurements of TNFα in LPS-stimulated THP-1 cells treated with inactive compounds. The compounds indicated were inactive in the MTS assay. Cells were preincubated with 100µM of the test compounds indicated prior to LPS stimulation. Bardoxolone-methyl (Bdox) was used at 200nM as indicated. TNFα production was quantified after 3 hours and is shown relative to the DMSO (vehicle)-treated control cells. Only a single test was conducted.



Supplementary Figure 28: GC-MS comparison of oat roots. The *A. strigosa sad2* mutants accumulate β-amyrin (βA) as previously reported [238]. A second peak can also be seen at 23.8 minutes with a mass spectrum consistent with the product found in *N. benthamiana* leaves expressing *tHMGR/AsbAS1/CYP94D65*. The two chromatograms cannot be compared directly due to differences in the GC program.

Supplementary Tables

Enzyme name	EC number	*A. thaliana* homologue	*A. thaliana* short name	Oat 454 Contig	e value (TBLASTN)
HMGR	1.1.1.34	AT1G76490	AtHMGR1	01257	6.90E-208
				17852	1.30E-206
				20171	4.70E-40
		AT2G17370	AtHMGR2	01257	1.90E-189
				17852	4.30E-185
				20171	2.40E-35
FPS	2.5.1.10	AT5G47770	AtFPS1	17290	5.30E-138
		AT4G17190	AtFPS2	17290	8.00E-135
SQS	2.5.1.21	AT4G34640	AtSQS1	17273	8.30E-150
		AT4G34650	AtSQS2	17273	1.30E-137
SQE	1.14.99.7	AT1G58440	AtSQE1	01780	2.80E-212
		AT2G22830	AtSQE2	01780	1.60E-200
		AT4G37760	AtSQE3	01780	3.30E-209
		AT5G24140	AtSQE4	01780	4.30E-120
		AT5G24150	AtSQE5	01780	5.20E-131
		AT5G24160	AtSQE6	01780	4.80E-128
		AT5G24155	AtSQE7	01780	9.20E-17

Supplementary Table 1: Searching an oat root tip transcriptomic database for MVA pathway genes. Protein sequences for the known *A. thaliana* genes were used as search queries in a translated *A. strigosa* 454 database (TBLASTN cutoff value e-10). Oat 454 contig numbers and e-values are given in the right-most columns. A comparison of the three forms of oat HMGR are given in **Supplementary Figure 1.** The *AtFPS1* gene (AT5G47770) encodes both long- and short protein variants which localise to either the mitochondria or cytosol respectively depending on the presence or absence of 41 amino acids at the N-terminus of the protein[41]. Analysis of the oat FPS contig 17290 did not identify any obvious targeting sequences. Furthermore, amino acid sequence alignment with the mitochondrial-targeted form of A. thaliana AtFPS1 revealed that the oat FPS is shorter and does not align within the N-terminal mitochondrial peptide (**Supplementary Figure 2**).

Number:	1	3	5	6	8	13	17	52	53	2
Compound	β-amyrin	Erythrodiol	Oleanolic acid	24-hydroxy-β-amyrin	Olean12-ene, 3β,24-oic-acid	11-oxo-β-amyrin	30-hydroxy-β-amyrin	12,13β-epoxy-β-amyrin	16β-hydroxy-β-amyrin	EpHβA
Fragment loss None (M+)	498	586	600	586	600	512	586	514	586	602
CH_3	483	571	585	571	585	497	571	499	571	587
TMSiOH	408	496	510	496	510	422	496	424	496	512
$TMSiOCH_2$	-		-	483	-	-	483	-	-	-
TMSiOOCH	-	-	482	-		-	-	-	-	-
ABC* rings	218	306	320	218	218	232	306	234	306	322
C*DE rings	279	279	279	368	382	-	279	279	279	279
ABC rings (McLafferty)	-	-	-	-	-	273	-	-	-	-
ABC* + TMSiOH	-	216	-	-	-	-	-	-	216	232
ABC* + $TMSiOCH_2$	-	-	-	-	-	-	203	-	-	-
ABC* + TMSiCOOH	-	-	203	-	-	-	-	-	-	-
C*DE + TMSiOH	189	189	189	278	292	189	189	189	189	189

Supplementary Table 2: Summary of ions observed by from EI fragmentation of oxidised forms of β-amyrin. The column on the left describes the likely fragment which is lost to give the observed ion. Where ABC* rings is indicated, this generally refers to retro-Diels Alder fragmentation, with the exception of molecules containing the C-12,13 epoxide as shown in **Supplementary Figure 17**. Loss of the ABC rings by McLafferty rearrangement is also observed only in compounds containing the C-11 carbonyl group (11-oxo-β-amyrin) [100, 174, 261] (A scheme for this rearrangement is shown in **Supplementary Figure 5**).

	Combination	CYP716A12 + CYP93E1	CYP716A12 + CYP72A65	CYP88D6 + CYP72A65	CYP93E1 + CYP72A65
Number:		19	21	23	24
	Compound	4-hydroxy-oleanolic acid	Queretaroic acid	11-oxo, 30-hydroxy-β-amyrin	24,30-dihydroxy-β-amyrin
Fragment loss					
None (M+)		688	688	600	674
CH3		673	673	585	659
TMSiOH		598	598	510	584
$TMSiOCH_2$		585	585	497	571
TMSiOOCH		570	570	-	-
ABC* rings		320	408	320	306
C*DE rings		368	279	-	368
AB rings (McLafferty)		-	-	361	-
ABC* + TMSiOH		-	318	230	216
ABC* + $TMSiOCH_2$		-	305	217	203
ABC* + TMSiCOOH		202	291	-	-
C*DE + TMSiOH		278	189	189	278

Supplementary Table 3: Summary of ions observed by EI fragmentation of previously described combinatorial compounds. The fragment loss is as described in **Supplementary Table 2**. Ions labelled in red were not observed in the present study likely due to low abundance of the compound. Ions labelled in grey were minor constituents of the mass spectra and could not be clearly defined as a characteristic fragment.

	Combination	CYP51H10 + CYP93E1	CYP51H10 + CYP93E1	CYP51H10 + CYP72A65	CYP88D6 + CYP93E1	CYP88D6 + CYP716A12	CYP51H10 + CYP88D6	A354L + CYP93E1	I471M + CYP72A65
Number:		28	29	36	37	45	47	48	59
	Compound	12,13β-epoxy, 24-hydroxy-β-amyrin	12,13β-epoxy, 16β,24-dihydroxy-β-amyrin	12,13β-epoxy, 16β,30-dihydroxy-β-amyrin	11-oxo,24-hydroxy-β-amyrin	11-oxo-oleanolic acid	11-oxo, 16-hydroxy-β-amyrin	16β,24-dihydroxy-β-amyrin	12,13β-epoxy, 30-hydroxy-β-amyrin
Fragment loss									
None (M+)		602	690	690	600	614	600	674	602
CH_3		587	675	675	585	599	585	659	587
TMSiOH		512	600	600	510	524	510	584	512
$TMSiOCH_2$		499	587	587	497	-	-	571	499
TMSiOOCH		-	-	-	-	496	-	-	-
ABC* rings		234	322	410	232	334	320	306	322
C*DE rings		367	368	279	-	-	-	367	279
AB rings (McLafferty)		-	-	-	273	375	361	-	-
ABC* + TMSiOH		-	232	320	-	-	230	216	232
ABC* + $TMSiOCH_2$		-	-	307	-	-	-	-	219
ABC* + TMSiCOOH		-	-	-	-	216	-	-	-
C*DE + TMSiOH		277	277	189	277	289	189	278	189

Supplementary Table 4: Ions observed upon EI fragmentation of combinatorial products for which reference spectra were not available. The fragment loss scheme is as described in **Supplementary Table 2**. With the exception of 11-oxo-oleanolic acid and 11-oxo,24-hydroxy-β-amyrin, product identities are putatively inferred from GC-MS data. Not shown are the putative 11-oxo,12,13-epoxy,16β-hydroxy-β-amyrin, 11,16β-dihydroxy,12,13-epoxy-β-amyrin and 12,13β-epoxy-oleanolic acid as their fragmentations differed significantly from the usual fragmentation pattern.

Name	Substrate	Product	Accession	Ref
CYP72A67	Oleanolic acid	2β-hydroxy-Oleanolic acid	AB558149	[102]
CYP72A68v2	Oleanolic acid	Gypsogenic acid	AB558150	[101]
CYP716A47	Dammarenediol-II	Protopanaxadiol	JN604536	[103]
CYP716A53v2	Protopanaxadiol	Protopanaxatriol	JX036031	[88]
CYP708A2	Thalianol	7β-hydroxy-thalianol	NM_180822	[32, 74]
CYP705A5	Thalianol	Desaturated thalianol	NM_124173	[32]

Name	Substrate	Product	Accession	Ref
MdOSC1	Oxidosqualene	α-amyrin	FJ032006	[262]
LjOSC3	Oxidosqualene	Lupeol	AB181245	[263]
CsCPQ	Oxidosqualene	Cucurbitadienol	NP_ 001292630	[89]
AsbAS1-S728F	Oxidosqualene	Dammarenediol-II	ABG88962 (w/t)	[137]
THAS	Oxidosqualene	Thalianol	NM_124175	[264]

Supplementary Table 5: Additional P450s and OSCs expressed as part of the triterpene toolkit.

Appendix 1 – Publications

J. Reed contributed to the following publications during this PhD:

Moses, T., Pollier, J., Shen, Q., Soetaert, S., Reed, J., Erffelinck, M.L., Van Nieuwerburgh, F.C., Vanden Bossche, R., Osbourn, A., Thevelein, J.M., Deforce, D., Tang, K., and Goossens, A., (2015) OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of Artemisia annua. *Plant Cell*. 27, 286-301.

Zhou, Y,. Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., Lin, T., Liu Z., Zeng K., Zhong, Y., Zhang, S., Hu, Q., Liu, M., Zhang, H., Reed, J., Moses, T., Liu, X., Huang, P., Qing, Z., Liu, X., Tu, P., Kuang, H., Zhang, Z., Osbourn, A., Ro, DK., Shang, Y., Huang, S. Convergence and divergence of bitterness biosynthesis and regulation in cucumber, melon and watermelon. *Nature Plants*. 2, 16183 (2016).