Directing Biosynthesis of Bioactive Triterpenes for Pharmaceutical Applications

A thesis submitted to the University of East Anglia in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Rebecca Emily Casson

December 2022

Abstract

Triterpenes are a large, structurally diverse class of plant specialised metabolites, with a history of use as components of traditional medicine, and potential as a source of therapeutics for the treatment of cancer and inflammatory diseases. The development of triterpene-based therapeutics would be greatly aided by a systematic structure-activity relationship investigation, requiring the procurement of a suite of structurally related molecules for investigation. However, the low abundance of triterpenes in nature and their recalcitrance to chemical synthesis presents a challenge. Recent understanding of triterpene biosynthesis in heterologous hosts provides a promising new platform to produce specific triterpenes for evaluation, overcoming these problems.

In this thesis, the triterpene toolkit developed in the Osbourn lab is used to produce a range of triterpenes in *N. benthamiana*. Novel putative triterpenes are generated through combination of toolkit enzymes (Chapter 3). Expansion of the toolkit allows for the identification of new functionalities on the β-amyrin scaffold (Chapter 4). Large-scale production allows for structural determination of a suite of triterpenes with a range of regio- and stereo-chemistries (Chapter 5). These are evaluated for anti-proliferative and anti-inflammatory effects (Chapter 6). A detailed structure-activity relationship investigation is carried out, and methods for improving the throughput of triterpene biological testing are considered (Chapter 7).

Key findings include the production of 64 triterpenes through combinatorial biosynthesis, including 31 novel structures, the expansion of the triterpene toolkit to include bacterial and human cytochromes P450, the generation and purification of a suite of 20 β-amyrin derivatives (14 of which were fully structurally verified by NMR), the discovery of three new positions on the β-amyrin scaffold responsible for anti-proliferative activity, and three triterpenes with anti-inflammatory activity. A structure-activity relationship investigation also predicts novel bioactive triterpenes. This work will provide the basis for the design of triterpenes optimised for improved biological activity.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to thank my supervisors, Prof. Anne Osbourn and Prof. Maria O'Connell, for giving me the opportunity to take on this project and truly make it my own over the course of this study, and also for their continued support both personal and academic along the course of this journey.

I would additionally like to thank the members of the Osbourn and O'Connell labs, past and present, for a supportive and welcoming research environment. Particular thanks must go to: Dr. James Reed, for his constant help, advice, and support; Dr. Michael Stephenson, for his advice on NMR spectroscopy; Dr. Hannah Hodgson, for her assistance with LC-MS and for her particular support during the write-up process; Dr. Charlotte Owen, for practical assistance throughout the lab and for providing a supportive and listening ear; and Jenny Jo, for her support and to bringing light and life into the lab throughout the PhD. In the O'Connell lab, thanks go to Emily Hobson, Maya Valmiki, and Adenike Odedele for their assistance with biological assays.

The work contained in this thesis would not be possible without the technology platforms at JIC and the work they do in supporting our science. I would like to thank horticultural services, the NMR platform, and metabolite services for their tireless work supporting this science. Special thanks go to Dr. Lionel Hill and Dr. Paul Brett of JIC metabolomics for their help and advice in all things GC- and LC-MS related.

I also have to thank my friends and family for their ongoing support. I owe many thanks to my parents, David and Jill, for their unwavering support of me throughout the years of my education, and for continuing to support and encourage me at home throughout this project. I also thank my twin Tom for his encouragement and for generally brightening my life, and my friends Raven and Amy for listening to me ramble endlessly about my PhD for the last four years, and for endeavouring to support me through the highs and lows of this project.

Finally, I thank Rosie, my cat, who contributed nothing practical to this project and yet remains a source of joy.

List of Abbreviations

ANOVA- Analysis of variance

APCI- Atmospheric pressure chemical ionisation

BSA- Bovine serum albumin

CAD- Charged aerosol detector

CBP- CREB-binding protein

CNC- Cap 'n' Collar

COSY- Correlation spectroscopy

CSL- Cellulose synthase-like

CYP/CYP450- Cytochrome P450

DAD- Diode array detector

DEPT- Distortionless enhancement by polarisation transfer

DMF- Dimethyl fumarate

DMAPP- Dimethyl allyl diphosphate

DMSO- Dimethyl sulphoxide

ECACC- European Collection of Authenticated Cell Cultures

ECL- Electrochemiluminescence

EDTA- Ethylenediaminetetraacetic acid

EI- Electron impact ionisation

EIC- Extracted ion chromatogram

ESI- Electrospray ionisation

ELISA- Enzyme-linked immunosorbent assay

ELSD- Evaporative light scattering detector

FBS- Foetal bovine serum

GC-MS- Gas chromatography – mass spectrometry

GFP- Green fluorescent protein

GPP- Geranyl diphosphate

GT1- Family 1 glycosyltransferases

HMBC- Heteronuclear multiple bond coherence

HMGR- 3-hydroxy,3-methylglutaryl-CoA reductase

HPLC- High pressure liquid chromatography

HSQC- Heteronuclear single quantum coherence

IC_{50}- Half-maximal inhibitory concentration

IκB – Inhibitor of κB

IKK- IκB kinase

IL-6– Interleukin 6

iNOS- Inducible nitric oxide synthase

IPP- Isopentenyl diphosphate

KEAP1- Kelch-like ECH associated protein 1

LC-MS- Liquid chromatography – mass spectrometry

LPS- Lipopolysaccharide

MeCN- Acetonitrile

MeOH- Methanol

MES- 2-(N-morpholino)ethanesulphonic acid

MEP- Methylerythritol 4-phosphate

MOPS- 3-(N-morpholino)propanesulphonic acid

MTS- 5-[3-(carboxymethoxy)phenyl]-3-(4,5-dimethyl-2-thiazolyl)-5-[(phenylamino)-carbonyl]-2H-tetrazolium inner salt

MVA- Mevalonate

m/z- mass/charge ratio

NAD(P)H- Nicotinamide adenine dinucleotide (phosphate)

NCI- National cancer institute

Neh- Nrf2-ECH homology

NfκB- Nuclear factor kappa-light-chain-enhancer of activated B cells

Nrf2- Nuclear factor erythroid 2-related factor 2

NMR- Nuclear magnetic resonance

NPNPD- National cancer institute programme for natural products discovery

NOS/RNS- Reactive nitrogen species

OS- 2,3-oxidosqualene

OSC- Oxidosqualene cyclase

PBS- Phospho-buffered saline

PCR- Polymerase chain reaction

PDA- Photodiode array

PEG- Polyethylene glycol

PES- Phenazine ethyl sulphate

ROESY- Rotational frame overhauser enhancement spectroscopy

ROS- Reactive oxygen species

RPMI- Roswell Park memorial institute

SAD1– *Avena strigosa* β-amyrin synthase

SCPL- Serine carboxypeptidase-like

SDS- Sodium dodecyl sulphate

SPE- Solid-phase extraction

SQE- Squalene epoxidase

TAE- Tris-acetate-EDTA

TBS- Tris-buffered saline

TBST- Tris-buffered saline Tween

tHMGR– Truncated HMGR

TG- Transglucosidases

TIC– Total ion chromatogram

TLC- Thin layer chromatography

TMS- Trimethylsilyl

TNFα– Tumour necrosis factor alpha

TTK- Triterpene toolkit

UDP- Uracil diphosphate

UV- Ultraviolet

1 General Introduction

1.1 Plant Specialised Metabolism

Plants live a sedentary lifestyle and are unable to escape from threats in their environment by moving away from them. As a result, plants have evolved the ability to produce a diverse range of chemicals that protect them against biotic and abiotic stresses. These specialised metabolites thus have important ecological roles. They are also of great value to humans as an important source of therapeutics, including painkillers, anti-cancer, and anti-parasitic compounds [1-3]. However, the exploitation of plant compounds as medicines is often hampered by the difficulty in cultivating plants that produce important natural products at commercial scale. Many plant species are slow-growing and produce only small amounts of the desired bioactive products. This can lead to high prices for natural products used in the clinic such as taxol or artemisinin [4].

As a result, there is ongoing research into other ways of generating plant or plant-inspired natural products. Production methods such as total chemical synthesis and plant cell culture have been developed but can be difficult to deploy at scale [5]. Another strategy is the production of valuable plant molecules in heterologous hosts. The revolution in genome sequencing over the last decade, coupled with the discovery and elucidation of a growing number of plant biosynthetic enzymes and pathways, has led to step-changes in our understanding of plant natural product biosynthesis. At the same time, innovations in metabolic engineering and heterologous expression platforms have led to the production of numerous natural products in significant yields [6,7]. Additionally, the amount of diversity currently accessible in terms of available synthetic enzymes and characterised pathways has made the possibility of engineering novel compounds in heterologous systems an achievable reality [8].

The current study focuses on the triterpenes, a therapeutically and commercially important class of plant specialised metabolites. This chapter introduces the triterpenes and the current understanding of their biosynthesis, before going on to discuss methods for producing triterpenes in heterologous hosts, the biological activities and mechanisms of action of these compounds, and evaluation of

triterpene structure-activity relationships for the engineering of more potent novel molecules in future.

1.2 Introduction to Triterpenes

Triterpenes are 30-carbon molecules derived from 5-carbon isoprene units. The term "triterpene" technically encompasses two classes of molecules. Sterols serve essential functions within eukaryotic organisms and can be considered part of primary metabolism, although they can also be modified into sterol-derived molecules that have more specialised roles. The second class comprises non-steroidal molecules and contains many structurally diverse and specialised compounds. In this thesis, the term "triterpenes" will usually be used to refer to this second group of biomolecules.

Triterpenes and sterols have a variety of functions *in planta*. Sterols are essential for normal plant growth and development. They are important components of membranes and are also precursors for steroidal phytohormones [9]. They also serve as precursors for a diverse range of steroid-derived specialized metabolites. The non-steroidal triterpenes have diverse roles. They are found in the waxy cuticles of plants [10], where they may have a structural role [11]. They have been implicated in growth and development, for example in oats, where elevated accumulation of the triterpene scaffold β-amyrin results in short super-hairy roots [12]. Many triterpenes have been implicated in plant defence. This is particularly the case for triterpene glycosides (also known as saponins), which are reported to have antifungal, insecticidal, and molluscicidal effects [13-16].

Triterpenes are also exploited by humans for medicinal purposes. Many traditional medicines contain triterpenes and saponins as active constituents. For example, glycyrrhizin (**Figure 1-1**) is a triterpene saponin found in liquorice and is used as a sweetener. Its aglycone, glycyrrhetinic acid, interacts with glucocorticoid receptors and is associated with the anti-inflammatory activity of liquorice [17]. The boswellic acids, especially 11-oxo-β-boswellic acid and 3-O-Acetyl-11-oxo-β-boswellic acid (**Figure 1-1**), are major components of frankincense and are also known to exhibit anti-inflammatory effects [18,19].

Aside from these "traditional" uses of triterpenes, there is interest in the development of triterpenes as novel therapeutic agents. For example, the triterpene saponin QS-21 (**Figure 1-1**), extracted from the bark of the tree *Quillaja saponaria*, is an important immunostimulant that is used as an adjuvant in vaccines for shingles, malaria, and COVID-19 [20-22]. QS-21 is a complex molecule with multiple functional groups, including eight diverse sugar groups and a C-18 acyl chain [23]. Another triterpene that shows significant pharmaceutical promise is celastrol (**Figure 1-1**), a pentacyclic triterpene extracted from the thunder god vine (*Tripterygium wilfordii*). This nortriterpene quinone methide has a variety of biological effects, most notably anti-inflammatory activity [24].

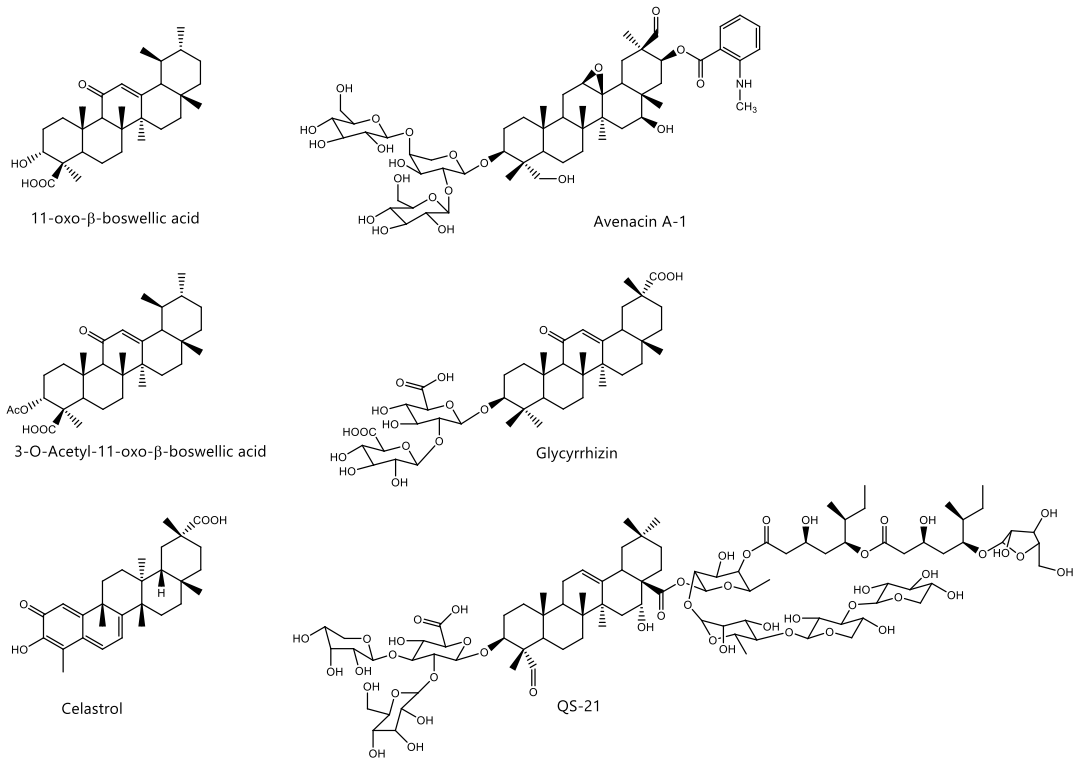


Figure 1-1: Examples of naturally occurring triterpenes with important biological activities.

The bioactive properties of naturally occurring triterpenes have generated significant interest in the chemical synthesis of triterpenes and in the semi-synthetic derivatisation of simple triterpenes extracted from nature. Full chemical synthesis of triterpenes is difficult due to their structural complexity. For example, the first reported total synthesis of the simple pentacyclic triterpene β-amyrin gave a 1.3 % yield over 17 steps, later improved to a 20.3 % yield over 10 steps [25,26]. Synthesis

of more complex triterpenes, such as erythrodiol, oleanolic acid, or QS-21 has also been reported [25,27], but these methods still represent a greater investment of time and resources than simply extracting the triterpenes in question from natural sources. Greater success has been achieved through the creation of semi-synthetic derivatives, with a number of glycosylated, aminated, acylated, and polyethylene glycol (PEG)-ylated derivatives of common triterpenes having been synthesised [28]. However, semi-synthetic methods are limited due to their reliance on already existing chemical "handles" for modification. One notable success story of semi-synthetic triterpene derivatisation is the development of bardoxolone methyl, a semi-synthetic triterpene derivative with anti-inflammatory effects, which is discussed in further detail in **Section 1.7.2** [29,30].

1.3 Triterpene Biosynthesis in Plants

Triterpenes are biosynthesised in plants from the mevalonate (MVA) pathway, via the linear precursor 2,3-oxidosqualene (OS). Triterpene scaffolds are formed by cyclization of OS, with subsequent functionalisation and tailoring by downstream biosynthetic pathway enzymes.

1.3.1 Triterpenes are Produced by the MVA Pathway in the Cytosol

Two distinct pathways exist for the biosynthesis of terpenoids: the mevalonate (MVA) pathway, found in most eukaryotes, archaea and some bacteria [31], and the non-mevalonate or Methylerythritol 4-phosphate (MEP) pathway, found in most bacteria [32]. Plants are unusual in possessing both pathways, utilising the MVA pathway in the cytosol and the MEP pathway in the plastid (**Figure 1-2**) [33]. Despite the differences in biochemistry between these pathways, both lead to the production of the same five-carbon building blocks for terpene biosynthesis, the isomers isopentenyl diphosphate (IPP) and dimethyl allyl diphosphate (DMAPP) [34]. Condensation of IPP and DMAPP yields the 10-carbon compound geranyl diphosphate (GPP), with the addition of additional IPP units leading to longer carbon chains (**Figure 1-2**) [35].

These linear substrates are then transformed into cyclic derivatives, with the resulting products dependent on the length of the linear chain. In the case of the

triterpenes, two molecules of farnesyl diphosphate, a C-15 carbon chain, combine in head-to-head condensation to form the C-30 linear molecule squalene, leading to the loss of the diphosphate groups [36]. While bacteria may cyclise squalene directly through the action of squalene-hopene cyclases [37], in plants squalene is oxidised prior to cyclisation by squalene epoxidase (SQE) to form 2,3,-oxidosqualene (**Figure 1-2**) [38]. The formation of this epoxide leads to the characteristic C-3 oxygen functionality found in the sterols and triterpenes.

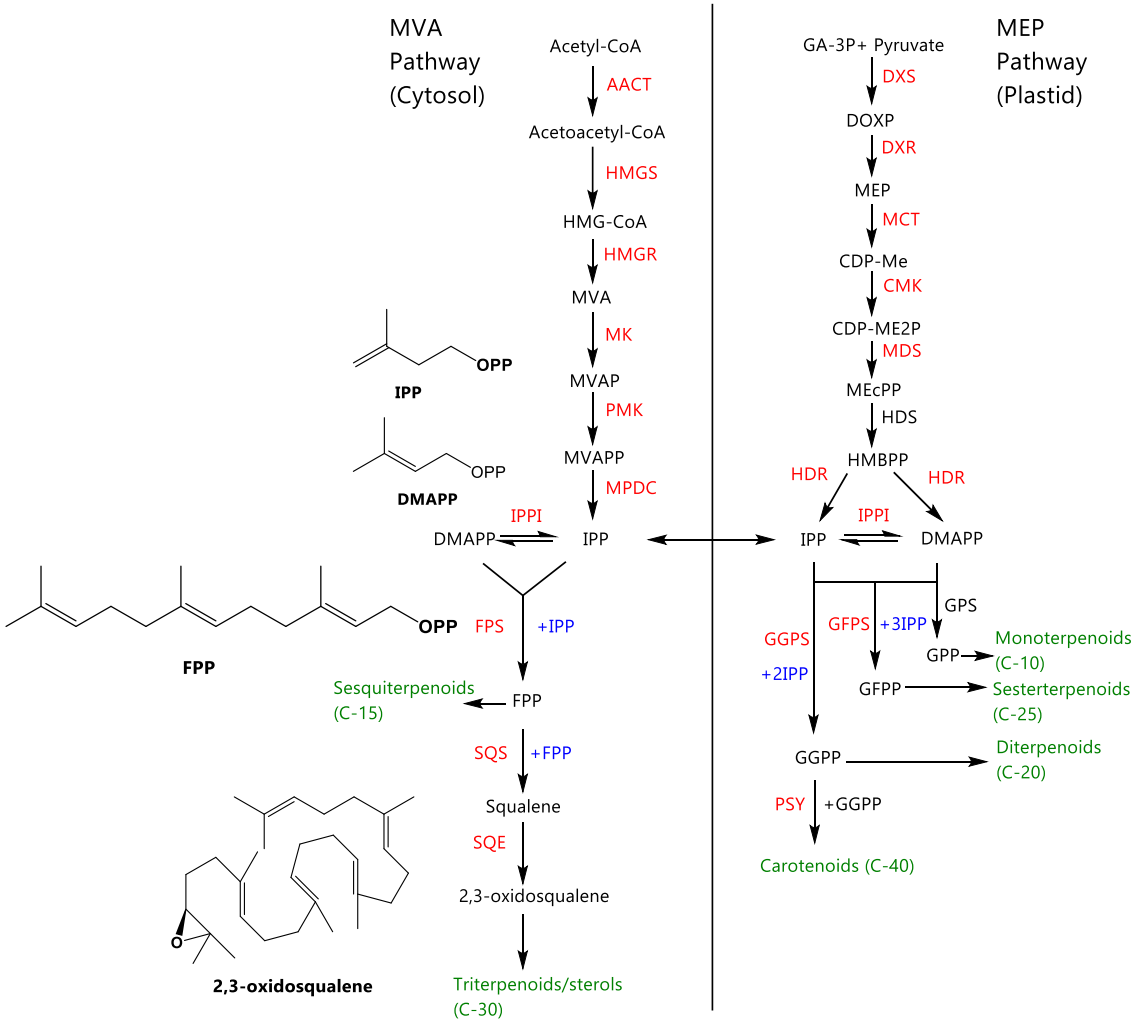


Figure 1-2: MVA and MEP pathways. Enzymes are shown in red, additional molecules for condensation in blue, and terpenoid products in green. **Abbreviations**: AACT, Acetyl-CoA C-acetyltransferase; CoA, Coenzyme A; HMGS, HMG-CoA synthase; HMG, 3 hydroxy 3 methylglutaryl-CoA; HMGR, HMG-CoA reductase; MVA, mevalonate; MK, MVA kinase; MVAP, melavonate-5-diphosphate; PMK, Phospho-MVA kinase; MVAPP, mevalonate-5-diphosphate; MPDC, diphospho-MVA kinase; GA-3P, D-glyceraldehyde-3-phosphate; DXS, DXP synthase; DXP, 1-deoxy-D-xylulose 5-phosphate; DXR, DXP reductoisomerase; MEP, 2-C-Methyl-D-erythitol; MCT, MEP cytidylyltransferase; CDP-ME, 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol; CMK, CDP-ME kinase; CDP-ME2P, 2-phospho-4-(cytidine 5′-

diphospho)-2-C-methyl-D-erythritol); MDS, MEcPP synthase; MEcPP, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; HDS, HMBPP synthase; HMBPP, 4-hydroxy-3-methylbut-2-enyl-diphosphate; HDR, HMBPP reductase; IPP, isopentenyl diphosphate; DMAPP, dimethyl allyl diphosphate; IPPI, IPP isomerase; FPS, FPP synthase; FPP, farnesyl diphosphate; GGPS, GGPP synthase; GGPP, geranylgeranyl diphosphate; GFPS, GFPP synthase; GFPP, geranylfarnesyl diphosphate; GPS, GPP synthase; GPP, geranyl diphosphate; SQS, squalene synthase; SQE, squalene epoxidase; PSY, phytoene synthase. Adapted from [33] and [34].

1.3.2 Oxidosqualene Cyclisation Leads to Triterpene Scaffold Diversity

2,3-oxidosqualene produced by the MVA pathway is cyclised into triterpene scaffolds by oxidosqualene cyclase (OSC) enzymes. Cyclisation is initiated by protonation of the epoxide of 2,3-oxidosqualene, which leads to the formation of a carbocation [39]. The double bonds of the substrate are folded into proximity of the cation through conformational change in the active site of the OSC (**Figure 1-4**). This leads to a cascade of electrophilic attacks, resulting in ring formation (**Figure 1-4**). The initial cyclised intermediates may undergo ring expansions as well as methyl or hydride shifts to yield the final products [40,41]. The degree of cyclisation produced through this process can vary widely with the formation of anywhere between one and five rings (mono- to pentacyclic triterpenes) [42]. Termination of the reaction often occurs through deprotonation, resulting in a characteristic double bond [42]. In some cases, water capture occurs instead, leading to the formation of triterpene diols [43].

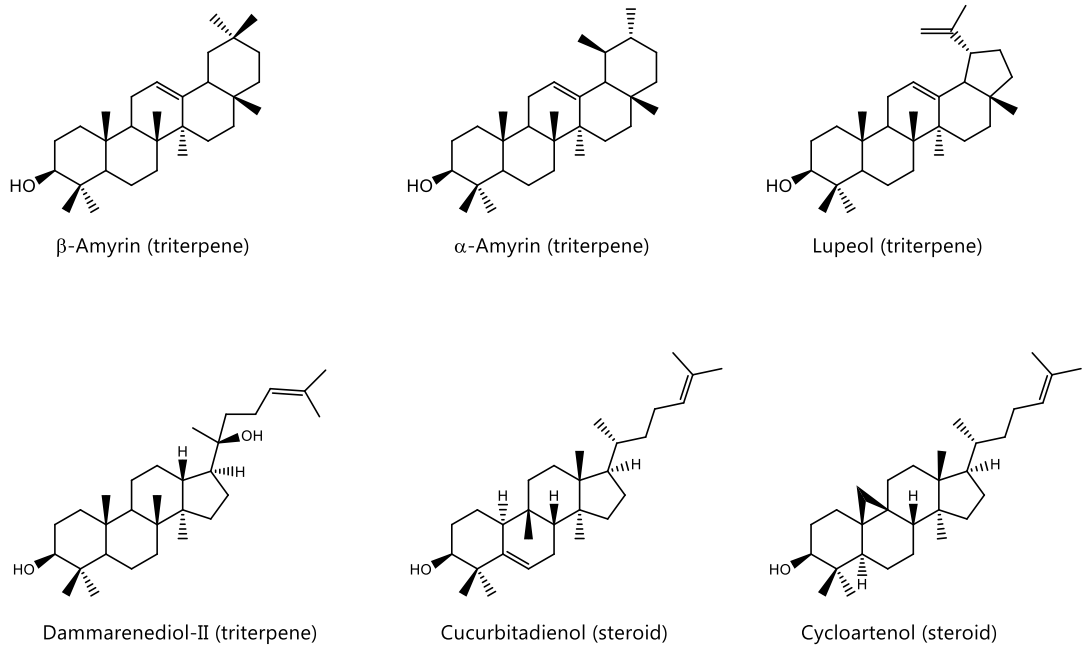


Figure 1-3: Examples of triterpene and steroid scaffolds produced by OSCs .

The differing mechanisms of OSC-catalysed reactions give rise to diverse triterpene scaffolds, with over 100 scaffold types reported [39]. Traditionally, the majority of triterpene scaffolds have been found to originate from one of two cyclisation routes, differing depending on the configuration of a 6,6,6,5 tetracyclic cationic intermediate. Sterols and certain tetracyclic triterpene scaffolds (such as curcurbitadienol) originate through the protosteryl cation, formed through cyclisation of 2,3-oxidosqualene with the A, B, and C rings in the chair-boat-chair configuration (**Figure 1-4**) [44]. In contrast, many pentacyclic triterpenes (such as β-amyrin or lupeol) originate through the dammerenyl cation, formed through cyclisation in which the A, B, and C rings of 2,3-oxidosqualene are in the chair-chair-chair configuration (**Figure 1-4**) [44]. It should be noted that while these two intermediates give rise to the majority of the considerable diversity in triterpene scaffolds known to date, the recent discovery of the orysatinol scaffold and the confirmation of its unusual stereochemistry has led to the proposed existence of a third cationic intermediate, referred to as the orysatinyl cation [45]. The proposed existence of this additional intermediate raises the prospect of triterpene scaffold diversity being even more varied than currently thought, with many more scaffolds potentially existing but remaining currently undiscovered.

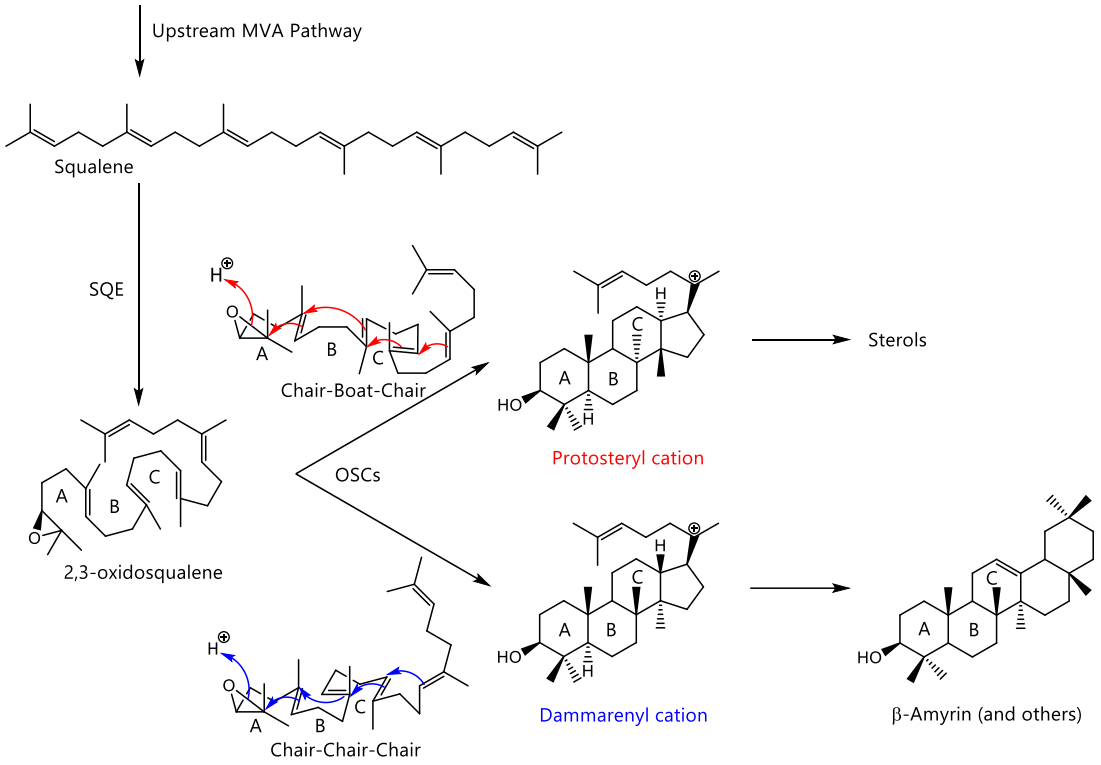


Figure 1-4: Cyclisation of 2,3-oxidosqualene, showing the two most common intermediates formed during triterpene scaffold synthesis: the protosteryl cation, formed from the chair-boat-chair configuration and leading to the formation of sterols, and the dammarenyl cation, formed from the chair-chair-chair configuration and leading to the formation of many pentacyclic non-sterol triterpenes. The A, B, and C rings are labelled for clarity.

1.3.3 Triterpene Scaffold Oxidation Leads to Further Diversification

The generation of the over 31,000 triterpenes known to nature from the over 100 known triterpene scaffolds relies upon modification of triterpene scaffolds by biosynthetic enzymes [40,46]. Oxidation of the scaffold is one such modification that is critical for generating triterpene diversity. Oxidation of a scaffold has significant effects on its polarity, can generate bioactive molecules from inactive scaffolds, and improves solubility [47]. Crucially, enzymatic oxidation functionalises usually chemically inert C-H bonds on the triterpene scaffold, therefore creating handles for further tailoring, for example through glycosylation or acylation. Enzymatic oxidation can even lead to substantial structural changes in the molecule, such as bond cleavage [48], and the formation of ether bridges or lactones [49,50].

The enzymes most commonly involved in triterpene oxidation are the cytochromes P450 (CYPs). These enzymes catalyse regio- and stereoselective oxidation of their substrates. CYPs are monooxygenases: they split molecular oxygen and incorporate

one atom into the substrate, reducing the second to yield water as a by-product [48]. Often, this results in the hydroxylation of the substrate, with the following general mechanism:

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

The catalytic centre of this process is a ferrous haem cofactor within the CYP active site, the iron atom performing a crucial role in the catalytic cycle [51]. The electrons required for this mechanism are donated by nicotinamide adenine dinucleotide phosphate (NADPH) via a partner CYP450 reductase enzyme [48]. The CYPs are a large and diverse enzyme superfamily, with similar 3D structures, including a conserved structural fold and a cysteine residue coordinated to the central haem [51]. They catalyse a myriad of different reactions, such as the aforementioned hydroxylations, bond cleavage, and bridge formation, as well as deamination, dehydration, and epoxidation [51,52]. Further oxidation of hydroxylated substrates can lead to the formation of ketones, aldehydes, and carboxylic acids. Through this diverse set of oxidative reactions, CYPs perform a variety of modifications to a basic triterpene, generating significant triterpene diversity even from the same base scaffold.

1.3.4 Tailoring Enzymes Further Increase Triterpene Diversity

Following the oxidation of a triterpene scaffold, the actions of "tailoring enzymes" create further structural diversity utilising the handles created through oxidation. Common reactions performed by tailoring enzymes include glycosylation and acylation. Glycosylation normally involves the addition of sugars to a hydroxyl or carboxylic acid group, with the C-3 hydroxyl present in most triterpenes being an especially common site for glycosylation [6]. The majority of glycosylations are carried out by family 1 glycosyltransferases (GT1), which rely on UDP-sugar donors [53]. Over 200 total GT1s have been characterised, and phylogenetic analysis shows that they cluster together based on function, with group D enzymes typically being responsible for triterpene glycosylation [54]. Despite this, other classes of enzymes have also been found to catalyse glycosylation of triterpenes, such as transglucosidases (TG) [55], and recently cellulose synthase-like enzymes (CSL) [56].

A second common decoration found in many triterpene natural products is acylation. Acyl groups are usually added to hydroxyl groups on triterpene molecules and are added by two distinct classes of enzymes. The most common type of triterpene acyltransferases are the BAHD family of enzymes, which use CoA-thioesters as acyl donors [57]. However, serine carboxypeptidase-like (SCPL) acyltransferases have also been reported to acylate triterpenes. These enzymes use O-glucose esters as their acyl donors [58]. For example, the N-methyl anthranilate group found in avenacin A-1 (**Figure 1-5**) is one example of an SCPL-catalysed acylation [59,60]. The combined actions of the triterpene biosynthetic enzymes generate substantial diversity from a single starting product (2,3-oxidosqualene). The biosynthesis of structurally complex plant triterpenes commonly involves CYPs, glycosyltransferases and acyltransferases, as can be seen in the examples in **Figure 1-5**. Although avenacin A-1 and QS-21 both originate from the β-amyrin triterpene scaffold, the different patterns of oxidation, glycosylation, and acylation found in each result in two substantially different molecules.

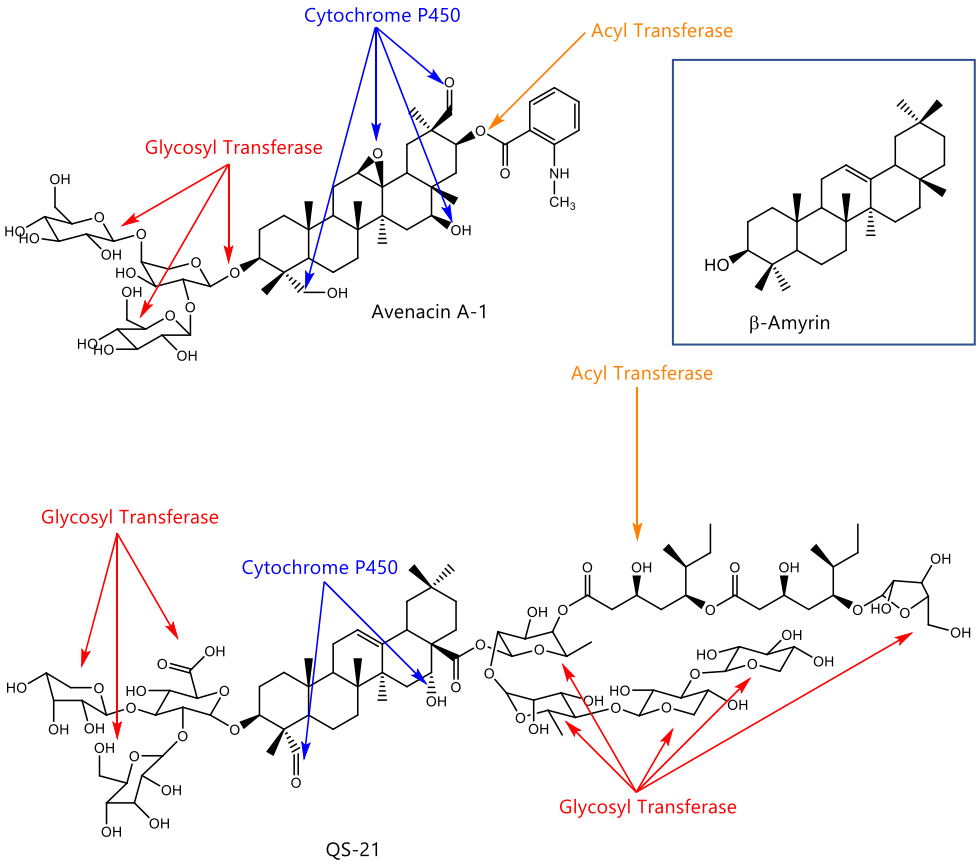


Figure 1-5: Examples of the diversity accessible through triterpene biosynthesis. Avenacin A-1 (top) and QS-21 (bottom), both based on the β-amyrin scaffold (included for reference), are shown and the modifications required to biosynthesise them are indicated with arrows (blue for CYP oxidations, red for glycosylations, orange for acylations).

1.4 Triterpene Production through Heterologous Expression

The characterisation of the genes and enzymes responsible for triterpene biosynthesis in plants has enabled the production of triterpenes in heterologous hosts. This can be achieved through the reconstitution of entire pathways, or through integrating enzymes from different pathways into a common host, a method known as combinatorial biosynthesis. Combinatorial biosynthesis can allow production of novel ('new-to-nature') compounds, so expanding on the repertoire of available triterpene diversity. Ideally, a heterologous host should be well-characterised, fast-growing, and amenable to genetic manipulation. Bacteria such as *Escherichia coli* are often used as heterologous hosts for metabolic engineering, but lack the sub-cellular localisation and post-translational processing capabilities usually required for expressing eukaryotic enzymes (for example, eukaryotic CYPs usually localise to the endoplasmic reticulum and therefore express poorly in *E. coli*

[52]). As a result, eukaryotic systems such as yeast have been used for production of terpenoids, such as the sesquiterpene anti-malarial drug, artemisinin [7]. However, yeast lacks many of the precursors and co-enzymes required for expressing plant enzymes and therefore these must be engineered into the yeast strains. These limitations have led to the development of plant-based expression systems as an alternative strategy for triterpene production and bioengineering.

1.4.1 Plant Transient Expression is Optimal for Plant Pathway Reconstitution

The use of plant transient expression for heterologous triterpene production overcomes many of the issues encountered with microbial hosts. Plants share common precursor biosynthetic pathways, codon usage, subcellular compartments, and enzyme cofactors, allowing for the expression of plant genes without significant additional engineering. They are easy to grow, requiring only water and sunlight, and production capacity can be scaled up simply by using more plants.

The wild tobacco relative *Nicotiana benthamiana* is typically used as a plant heterologous expression host due to its amenability to transient gene expression through agroinfiltration [61]. This technique involves infiltration of the bacterium *Agrobacterium tumefaciens* into the interstitial leaf spaces of *N. benthamiana*. The bacteria then transfer a section of DNA (T-DNA) containing the gene(s) of interest from a binary plasmid vector into the plant cell where they are expressed [62]. The reconstitution of biosynthetic pathways and/or the production of novel compounds can be achieved either by inserting multiple genes into a single vector or by co-infiltrating multiple *A. tumefaciens* strains, each carrying a single biosynthetic gene. Transient expression using *N. benthamiana* has been used to make a variety of triterpenes, and is the primary method used in the Osbourn lab for triterpene production [33].

Improving triterpene production through the plant transient expression system can be achieved through a variety of methods. For example, multiple biosynthetic genes can be integrated into a single binary vector, reducing the number of *A. tumefaciens* strains that are required to be co-infiltrated, which can increase yield by improving the chances of each plant cell receiving all of the required genes for the entire

biosynthetic pathway [63,64]. Integrating multiple genes into a single binary vector also allows for the incorporation of optimised promotors and terminators to improve triterpene yield, as well as for the expression of additional biosynthetic genes that remove bottlenecks in the triterpene biosynthetic pathway [65]. One such gene frequently utilised in the Osbourn lab is a N-terminal truncated, feedback-insensitive form of 3-hydroxy, 3-methylglutaryl-coenzyme A reductase (referred to as tHMGR), a rate-limiting enzyme in the upstream mevalonate pathway [66]. tHMGR is used in all of the experimental work in this thesis, as it has been found to significantly increase the yield of triterpenes produced using the transient expression system in *N. benthamiana* [50,67].

1.4.2 Large-Scale Vacuum Infiltration Provides a Scalable Platform for Triterpene Production

Scaling up of triterpene production using the *N. benthamiana* expression system can be achieved by simply increasing the number of infiltrated plants. However, with a regular hand-infiltration using a needleless syringe, the infiltration of increasing amounts of plants becomes significantly time-consuming. For infiltrations involving large (>30) numbers of plants where all are expressing the same combination of genes, vacuum infiltration provides a rapid alternative to usual hand-infiltration. Vacuum infiltration works by applying a vacuum to a chamber containing inverted *N. benthamiana* plants submerged in a suspension of *A. tumefaciens*. This draws the air out of the leaf interstitial spaces, and when atmospheric pressure is restored, the leaves are infiltrated. The number of plants that can be infiltrated at once depends on the size of the infiltrator, but the system is rapid, easily infiltrates the entire plant at once and requires minimal skill to operate. The vacuum infiltrator used in the Osbourn lab is made from a repurposed vacuum oven with an attached reservoir and can infiltrate 4 plants approximately every 1-2 minutes [68]. Combined with other efficient techniques for processing large amounts of resulting leaf material, such as pressurised solvent extraction, the Osbourn lab has developed a workflow for the rapid production of milligram to gram scales of valuable triterpenes [50,68] (see **Figure 1-6**).

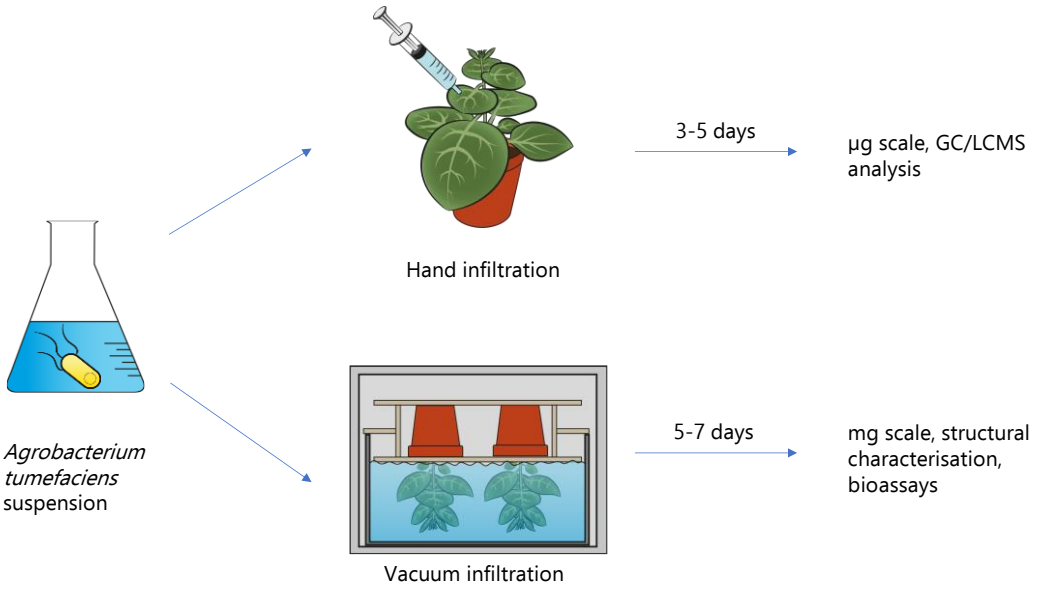


Figure 1-6: Comparison of the different methods for agro-infiltrating *N. benthamiana*. Figure produced using illustrations provided by James Reed.

1.5 Engineering Triterpene Diversity: The Triterpene Toolkit

The design and production of novel triterpenes using the *N. benthamiana* transient expression system relies upon access to a diverse range of biosynthetic enzymes capable of generating the variety of triterpene scaffolds, functional groups, and positional isomers required for a truly comprehensive suite of molecules. To enable this endeavour, the Osbourn lab has developed the triterpene toolkit, a set of genes encoding triterpene biosynthetic enzymes from a diverse range of plant sources. The toolkit exists both as a set of plasmids containing genes of interest, and as a set of stocks of *A. tumefaciens* in glycerol, each harbouring one of the plasmids. These glycerol stocks can be cultured and combined to make a suitable suspension of *A. tumefaciens* for the production of a vast diversity of triterpene molecules.

The toolkit incorporates all of the required biosynthetic enzymes for triterpene biosynthesis, including OSCs, CYPs, tailoring enzymes such as glycosyl and acyltransferases, enzymes to remove bottlenecks (such as tHMGR), and alcohol dehydrogenases. Furthermore, the triterpene toolkit is continually evolving, as newly characterised biosynthetic pathways and new sources of biosynthetic enzymes from all over the kingdoms of life provide an expanding repertoire of genetic resources to

draw from. This makes the triterpene toolkit an unparalleled resource for generating novel triterpene diversity through transient expression.

Figure 1-7: Examples of plasmids from the current triterpene toolkit.

1.6 Anti-Proliferative Effects of Triterpenes

Numerous triterpenes, either as purified compounds or as components of traditional medicines, have been tested for their anti-proliferative effects [69,70]. There is significant interest in developing triterpenes into new sources of chemotherapeutics and chemotherapy adjuvants [71-73]. As with many natural products, triterpenes are considered well tolerated compared to most existing chemotherapy agents, with lower incidences of side-effects [72,74].

1.6.1 Cancer

"Cancer" is the term for a group of diseases characterised by abnormal and uncontrolled cell division, and is one of the leading causes of death in the developed world [75]. Control of this aberrant cell division is therefore a major concern of research groups worldwide. While many think of anti-cancer agents as those causing the death of cancerous cells through programmed cell death or necrosis, cancer therapy is more varied in reality and includes any method to prevent the proliferation of cancer cells (i.e., anti-proliferative agents) as well as therapies to prevent the spread of cancer, such as anti-angiogenesis agents [76]. The current study focuses on anti-proliferation.

As previously mentioned, anti-proliferative agents are those that prevent the continued proliferation of cells, not just those that lead to cell death. There are a number of pathways that can lead to the halting of cell proliferation, the best known being activation of p53. Often called "the guardian of the genome", p53 regulates transcription of many anti-proliferative genes including p21, which inhibits cell cycle progression and therefore leads to cell cycle arrest [77]. While cell cycle arrest can be overcome under certain factors and proliferation can therefore resume, persistent and high expression of anti-proliferative factors can lead to arrest becoming stable and cells entering a state known as senescence, where they cease proliferation but remain viable [78]. Depending on the balance of pro- and anti- death signals, the cell may then enter cell death (such as through apoptosis or necrosis) or remain in senescence [77].

The activation of the pathways that lead to cell cycle arrest and ultimately halt proliferation can be brought about in a number of ways. An anti-proliferative agent may directly cause DNA damage to initiate activation of DNA repair pathways that inhibit proliferation [79], halt DNA synthesis through structurally mimicking vital metabolic substrates [80], interact with pro-apoptotic proteins to activate cell-death pathways [81] or disrupt cell division through interference with spindle proteins [82].

1.6.2 Triterpenes with Anti-Proliferative Effects

The majority of triterpenes tested for anti-proliferative effects have been evaluated for their ability to inhibit the proliferation of cell lines or shrink tumours in animal models. Some have been found to be significantly anti-proliferative in this regard, including celastrol and the boswellic acids (**Figure 1-1**) [19,71]. However, the exact mechanisms of action of these compounds are not well understood.

For some triterpenes the mechanisms of anti-proliferative activity are known. The limonoid triterpene nimbolide (**Figure 1-8**) was found to inhibit cell proliferation in two types of breast cancer cells, with a significant increase in the number of cells found to be apoptotic (measured by flow cytometry) [83]. Analysis of potential protein targets using a cysteine-reactive alkyne-functionalized iodoacetamide probe implicated RNF114, an E3 ubiquitin ligase that is known to degrade p21 [83]. Follow-up work suggested that nimbolide impairs the substrate recognition of RNF114, leading to a stabilisation of p21, increased levels in cells, and cell cycle arrest [83].

Maslinic acid (**Figure 1-8**), a pentacyclic β-amyrin derivative, is another triterpene with anti-proliferative effects for which a mechanism has been proposed. A study in HT29 colon cancer cells found that treatment with maslinic acid at the identified IC_{50} (60 μm) led to an increase in the proportion of cells found to be apoptotic [84]. Subsequent experiments with western blotting revealed that maslinic acid treatment inhibited the anti-apoptotic protein Bcl-2, activated the pro-apoptotic protein Bax, and induced release of cytochrome-c, a protein that activates caspases involved in apoptosis [69].

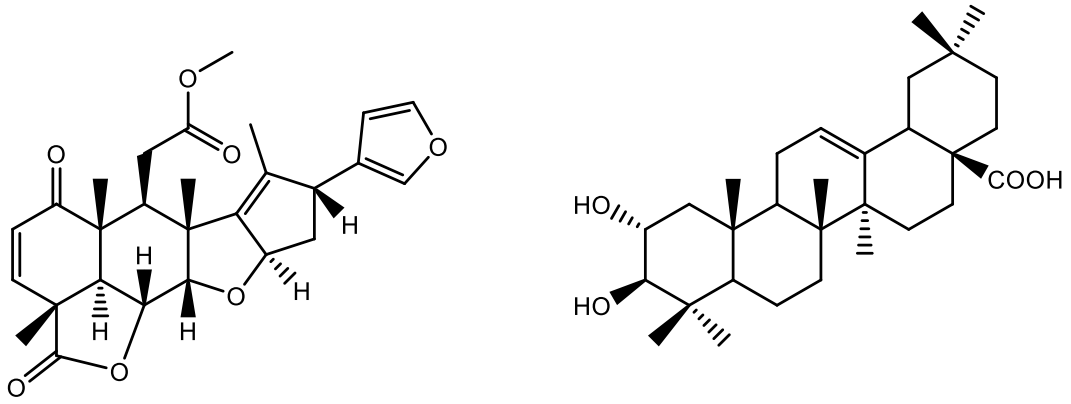


Figure 1-8: Structures of nimbolide and maslinic acid, two naturally occurring triterpenes investigated for their mechanisms of inhibiting proliferation.

1.7 Anti-Inflammatory Effects of Triterpenes

Naturally occurring triterpenes have long been used as components of traditional medicine for the treatment of a variety of diseases linked to inflammation [85], and semi-synthetic triterpene derivatives have been evaluated as anti-inflammatory agents [86]. There is therefore significant interest in the study of the anti-inflammatory effects of triterpenes.

Inflammation is a physiological response to injury, infection, or toxins in the body, and is characterised by pain, redness, heat, swelling, and loss-of-function in the affected area [87]. Physiologically, this process involves the release of pro-inflammatory cytokines such as tumour necrosis factor alpha (TNFα) and interleukin 6 (IL-6), which trigger a cascade of factors which leads to recruitment of macrophages and monocytes to the affected tissues, leading to the release of reactive oxygen or nitrogen species (ROS/NOS) and phagocytosis of invading pathogens [88].

Inflammation is controlled at the cellular level through the activation of transcription factors: pro-inflammatory factors such as nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) are activated by cytokines such as TNFα, ROS, or bacterial lipopolysaccharides (LPS) and drive inflammation [89] while factors such as nuclear factor erythroid 2-related factor 2 (Nrf2) dampen down the inflammatory response after their activation through oxidative stress [90,91]. The combination of pro- and

anti-inflammatory factors works to keep inflammation in balance, since while inflammation is beneficial in fighting disease and responding to injury, prolonged inflammation is implicated in several chronic diseases such as arthritis, asthma, and cancers [89,92]. Appropriate regulation of these transcription factors therefore provides a pathway towards controlling these debilitating conditions.

1.7.1 Key Inflammatory Pathways: Nrf2 and NFκB

Inflammation in the body is controlled by many competing pathways and transcription factors. The current study examines the effects of the triterpenes in relation to two specific and important transcription factors and the pathways involved with them: NFκB and Nrf2.

"NFκB" refers to a family of transcription factors, comprising of NFκB1 (p50), NFκB2 (p52), RelA (p65), RelB, and c-Rel [89]. In general, NFκB is used here to refer to the p65/p50 heterodimer, an abundant form of NFκB in cells which participates in the canonical NFκB pathway [89].

NFκB is constitutively present in healthy cells, where it is normally rendered inactive in the cytosol by an inhibitor known as inhibitor of κB (IκB). When activation of NFκB occurs, a kinase called IκB kinase (IKK) phosphorylates IκB such that it is targeted by the proteasome and degraded (**Figure 1-9**) [89]. NFκB is then released and moves to the nucleus [89]. Phosphorylation also typically occurs to the p65 subunit of NFκB, which improves the transactivation of p65 though interaction with co-activators such as the CREB-binding protein (CBP), ultimately leading to increased transcriptional activity [93]. The genes activated by NFκB include those involved in pro-inflammatory processes, cell survival, and immune response (**Figure 1-9**) [92]. In particular, NFκB induces the production of numerous pro-inflammatory cytokines, such as TNFα, IL-6, and COX-2 [94]. It also promotes the differentiation of $CD4^+$ T-cells into inflammatory Th1 T-cells [95].

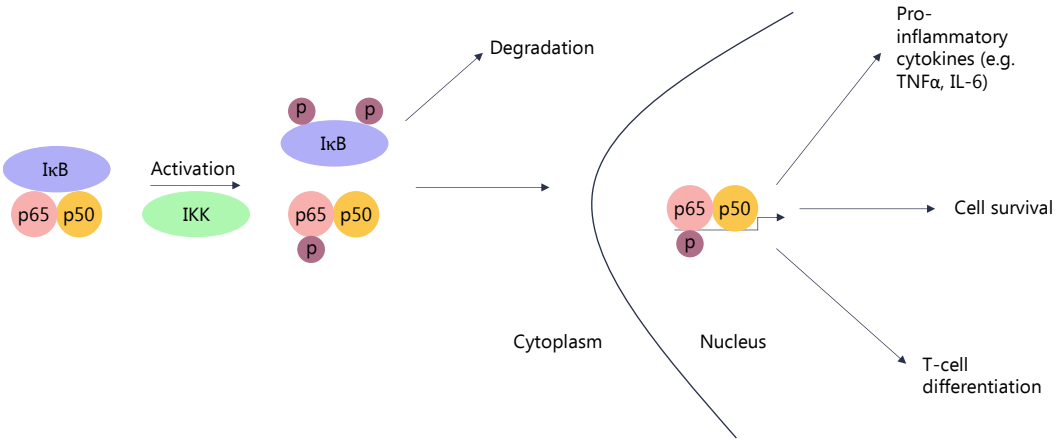


Figure 1-9: Pathway of NFκB activation and downstream effects.

Nrf2 is a master regulator of oxidative stress that controls the expression of a number of proteins involved in antioxidant production and detoxification. It is a basic leucine zipper protein from the "Cap 'n' Collar" (CNC) subfamily and contains seven Nrf2-ECH homology (Neh) domains through which it binds to DNA, conscription co-factors, and its inhibitor KEAP1 [96].

Nrf2 is usually degraded rapidly in cells through the action of Kelch-like ECH associated protein 1 (KEAP1), which recruits E3 ligase leading to the ubiquitination and subsequent proteasomal degradation of Nrf2 under normal conditions (**Figure 1-10**) [90]. Activation of Nrf2 occurs through the modification of thiol residues on KEAP1 which causes it to dissociate from Nrf2 [96]. Nrf2 then builds up in cells and eventually translocates to the nucleus, where it activates genes involved in detoxification, the production of antioxidants, and the anti-inflammatory response [96]. Notably, Nrf2 activates the transcription of HO-1, an enzyme responsible for the rate-limiting step of oxidative haem degradation, which exhibits an anti-inflammatory effect by inhibiting the release of pro-inflammatory cytokines such as TNFα or IL-6, as well as by inhibiting ROS production and activating the release of anti-inflammatory cytokines such as IL-10 (**Figure 1-10**) [96]. Nrf2 can also inhibit transcription of pro-inflammatory genes through binding to promoters [90]. Nrf2 and HO-1 can also inhibit IκB degradation and so inhibit the NFκB pro-inflammatory pathway, and Nrf2 and p65 both interact with the CBP co-activator [90,97]. Nrf2

additionally activates transcription of antioxidant genes to reduce levels of inflammatory ROS and RNS (**Figure 1-10**) [98].

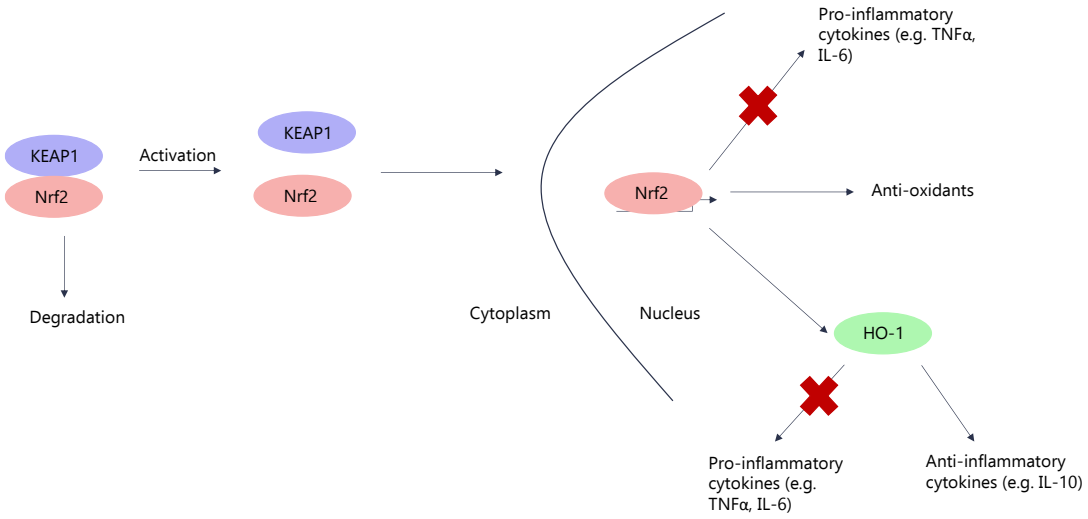


Figure 1-10: Pathway of Nrf2 activation and downstream effects.

1.7.2 Triterpenes with Anti-inflammatory Effects

As mentioned in **Section 1.2**, naturally occurring triterpenes have been found to exhibit anti-inflammatory activity. One example of this effect is celastrol, the nortriterpene derivative found in *Tripterygium wilfordii* (**Figure 1-1**). Celastrol was found to reduce the secretion of IL-6 and TNFα in LPS-stimulated RAW 264.7 murine macrophages [24]. Follow-up work through western blotting and transcription factor enzyme-linked immunosorbent assays (ELISAs) implicated inhibition of NFκB through stabilisation of IκB as a likely mechanism [24], with a separate study using immunoprecipitation suggesting inhibition of IKK as the cause [99].

The boswellic acids, a family of compounds isolated from the trees of the genus *Boswellia* and a major component of frankincense resin, have also been reported to exhibit anti-inflammatory effects [100]. The boswellic acid 3-O-acetyl-11-oxo-β-boswellic acid (**Figure 1-1**) and a β-amyrin analogue, 3-O-acetyl-α-boswellic acid, have been shown to reduce secretion of TNFα in LPS-stimulated cells (H9C2 rat heart tissue cells and peripheral monocytes) [101]. Investigation of potential mechanisms in HEK293 kidney cells again implicated inhibition of NFκB activation,

and immunoprecipitation studies again suggested inhibition of IKK as a mechanism [18].

Triterpenes containing an 11-oxo group have been found to interact with glucocorticoid signalling. Glycyrrhizin (**Figure 1-1**) and its aglycone glycyrrhetinic acid have anti-inflammatory effects through interaction with the glucocorticoid receptor, and accordingly inhibited LPS-induced secretion of TNFα in RAW264.7 macrophages [102]. Glycyrrhetinic acid is marketed as enoxolone for treatment of peptic ulcers along with a synthetic acylated derivative carbenoxolone, and both compounds have been additionally found to inhibit the 18β-hydroxy-steroid dehydrogenase 2 enzyme, therefore blocking the conversion of cortisol to cortisone and leading to higher cortisol levels, which leads to cortisol binding to the glucocorticoid receptor and a corresponding anti-inflammatory effect [103]. Carbenoxolone has also been found to reduce TNFα and IL-6 measured cytokine levels in the lung tissue of rats treated with monocrotaline [104], with western blotting on the lung tissue of asthmatic mice showing an inhibition of NFκB activation, possibly due to their interactions with glucocorticoid signalling [105].

As mentioned in **Section 1.2**, semi-synthetic triterpenes have been generated with potent biological effects, including anti-inflammatory effects. The semi-synthetic oleanane derivatives Methyl 2-cyano-3,12-dioxooleana-1,9(11)dien-28-oate (hereafter referred to as bardoxolone-methyl) and N-((4aS,6aR,6bS,8aR,12aS,14aR,14bS)-1 1-cyano-2,2,6a,6b,9,9,12a-heptamethyl-10,14-dioxo-1,2,3,4,4a,5,6,6a,6b,7,8,8a,9,10,12a,14,14a,14b- octadecahydropicen-4a-yl)-2,2-difluoropropanamide (hereafter referred to as omaveloxolone) have been found to be highly potent anti-inflammatory agents (**Figure 1-11**).

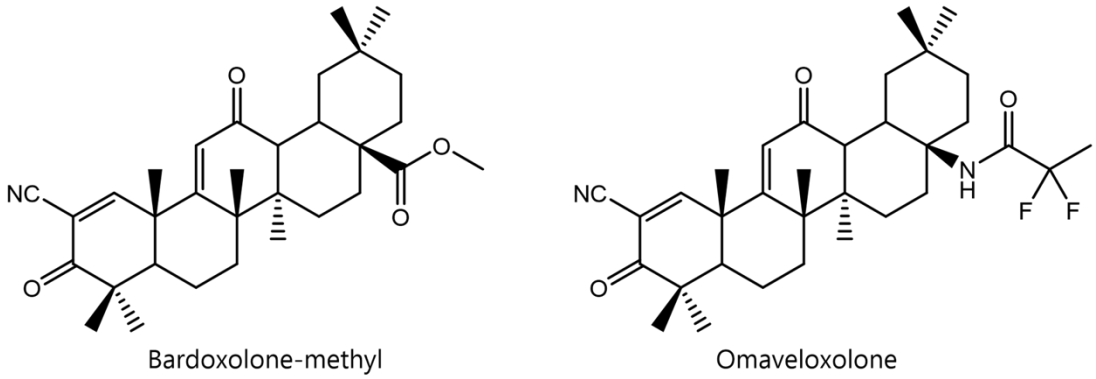


Figure 1-11: Structures of semi-synthetic derivatives of oleanolic acid (bardoxolone-methyl and omaveloxolone) produced commercially and under evaluation for potential pharmaceutical uses.

Bardoxolone has been investigated in detail for its anti-inflammatory activity and was found to inhibit activation of NFκB by immunoprecipitation in TNFα-stimulated U-937 histiocytic lymphoma cells [106]. The same study found that IκB phosphorylation was also inhibited, and implicated inhibition of IKK as a likely mechanism [106]. Mutagenesis of a specific cysteine residue of IKK, C-179, abolished this activity, further suggesting a mechanism of action through interaction with this cysteine [106]. Bardoxolone has also been found to bind to KEAP1 and therefore activate Nrf2, possibly also through interaction with cysteine residues [107]. Studies in mice and rats have shown omaveloxolone to have similar effects [108]. The combination of both Nrf2 activation and NFκB inhibition makes these compounds especially promising.

1.8 Structure-Activity Relationships: The Importance of Systematic Investigations

The promising reports of triterpene bioactivity have underlined the importance of understanding the relationships between triterpene structure and function. Such studies into structure-activity relationships are a common feature of drug development efforts. The conventional method of carrying out such a study would be to produce a library of structural analogues of a compound and test each using a known metric (such as a specific assay) to allow the bioactivities for each compound to be directly compared. Different structural features responsible for a particular bioactivity could then be identified. Multiple iterations of this process, with new

molecules designed based on the results of the last round of investigations, would eventually yield improved bioactive molecules which could be developed further into pharmaceuticals.

This process benefits from a systematic effort where the suite of molecules produced covers as many functional and positional isomers of a compound as possible. This is complicated for the triterpenes due to their structural complexity. Chemical semi-synthetic methods are limited as they can only transform chemical handles that already exist on a molecule (such as functional groups already present in naturally produced precursors). This limits chemical efforts to a few specific positions on most triterpene scaffolds. The use of heterologous expression systems, as discussed in **Section 1.4**, provides an improved source of triterpene structural diversity by allowing the use of combinations of enzymes from multiple sources to access combinations of triterpene functional groups inaccessible to chemical synthesis and unaccounted for in the triterpenes extracted from natural sources. The use of large-scale production methods allows for the synthesis of the desired molecules in suitable quantities for the bioassays required for structure-activity relationship investigations, and the creation of the triterpene toolkit (**Section 1.5**) has provided a revolutionary ease of access to the diversity of biological transformations now available.

1.9 PhD Overview

The triterpene toolkit provides an accessible route to the large-scale production of a diverse library of triterpenes for biological investigation in combination with heterologous expression in *Nicotiana benthamiana,* so enabling a detailed structure-activity relationship investigation to be carried out on pentacyclic triterpenes for the first time. The goal of this PhD is to accomplish this. In the following chapters, the use of the triterpene toolkit to generate triterpene structural diversity is evaluated, with the production of both known and previously unreported triterpenes (**Chapter 3**). Following this, the diversity accessible through the toolkit is expanded to include additional enzymes from plants as well as enzymes from beyond the Plant Kingdom, from bacteria and humans, and novel products are putatively identified (**Chapter 4**).

The large-scale production, purification, and structural determination of a suite of compounds generated by metabolic engineering is performed in **Chapter 5**. These compounds, along with a suite of triterpene standards, are then subjected to biological testing for anti-proliferative and anti-inflammatory effects (**Chapter 6**). The results of this testing are combined with previous reports of biological activity of triterpenes, from the O'Connell lab and the wider literature, to form the basis of a triterpene structure-activity evaluation. New candidate triterpenes with potentially improved bioactivities are then designed, and strategies to improve the throughput of triterpene testing are investigated (**Chapter 7**).

2 General Methods

2.1 Gateway Cloning

2.1.1 Entry Clones

pDONR207 was used for generating Gateway Entry clones. Gateway cloning was performed by incubation of the insert DNA (purified PCR product or purchased DNA fragment) with pDONR207 in equal ratios (75 ng of each) in a total volume of 4 µl EB buffer (10 mM Tris-Cl, pH 8.5, Qiagen). 1 µl of BP clonase II (Invitrogen) was added and the reaction vortexed briefly and incubated for 1 h at room temperature. Following this, the reaction was added to chemically competent *E. coli* as detailed in section **2.2.1**. Positive cells were selected for on LB agar plates plus 50 µg/ml gentamicin. Colony PCR was used to screen colonies (section **2.1.3**) and positive colonies were cultured overnight (37 °C, 200 rpm) in 10 ml liquid LB plus 50 µg/ml gentamicin. Plasmids were recovered using a miniprep kit (Qiagen) and sent for sequencing using attL-1/R-2 primers and, where necessary, custom primers as described in section **2.1.6**.

2.1.2 Expression Clones

To generate pEAQ-*HT*-DEST1 vectors carrying the gene of interest, Gateway cloning was performed by incubating the entry vector (pDONR207) with pEAQ-*HT*-DEST1 in equal ratios (usually 75 ng) in 4 µl EB buffer. 1 µl of LR clonase II (Invitrogen) was added and the reaction vortexed briefly and incubated for 6 hours at room temperature. Following this, the reaction was added to chemically competent *E. coli* as detailed in section **2.2.1**. Positive cells were selected for on LB agar plates plus 50 µg/ml kanamycin. Colony PCR was usually not necessary. A single colony was sampled from the plate and grown on overnight (37 °C, 200 rpm) in 10 ml LB media plus 50 µg/ml kanamycin. Plasmids were recovered using a miniprep kit (Qiagen) and used to transform *A. tumefaciens* LBA4404 (section **2.2.2**).

2.1.3 Colony PCR

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

Table 2-1: Standard sample set-up for colony PCR.

Reagent	Per Sample
Template	1 µl
attL ½ primers (10 µM)	1 µl
GoTaq green master mix	5 µl
H_2O	3 µl
Total	10 µl

Thermal cycling was carried out using the following programme:

Table 2-2: Standard thermal cycling programme for colony PCR.

Step	Temp (°C)	Time
Denaturation	95	30 sec
Denaturation	95	10 sec
Annealing	50	10 sec
Extension	72	1 min/kb
Extension	72	5 min

Following PCR, 5 µl DNA was analysed with 1 % agarose gel electrophoresis. Positive colonies were identified as those with a band of the expected size of the insert (plus ~200 bp due to the attL primer binding sites).

2.1.4 DNA Electrophoresis

For DNA analysis by agarose gel electrophoresis, an agarose gel was made with addition of 1 % w:v agarose (Sigma) to a suitable volume of TAE buffer using 0.01 % EtBr or GelRed (1 mg/ml). Gels were run at 120 V, 400 mA for 20 minutes with a 2-log ladder (NEB) included in the end wells. Gels were visualised under UV light.

2.1.5 Sequencing

Sequencing was performed by Eurofins. Purified plasmid templates were sent for sequencing. Sequencing primers are given in section **2.1.6**.

2.1.6 Primers

Table 2-3: Commonly used primers.

Name	Sequence
attL1-F	TCGCGTTAACGCTAGCATGGATCTC
attL2-R	ACATCAGAGATTTTGAGACACGGGC
attB1F-1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA
attB2R-1	GGGGACCACTTTGTACAAGAAAGCTGGGTA
CYP450R-M	ATAGGAAACTGAACCAGGG

attL1/2 F/R primers were used for sequencing of entry vectors while attB1/2 F/R primers were used as gateway adapter sequences added to the 5′ end of cloning primers. CYP450R-M was used as an additional primer when sequencing plasmids containing CYP450R due to the gene's size.

2.1.7 Plasmids

Table 2-4: List of plasmids used in the thesis.

Name	Vector Type	Supplier	Selection
pDONR207	Gateway Entry	Invitrogen	Gentamicin
pEAQ-*HT*-DEST-1	Gateway Destination	Lomonossoff Lab, JIC	Kanamycin

2.2 Microbiological Techniques

2.2.1 Transformation of Chemically Competent *E. coli*

Plasmid DNA was prepared as detailed in specified sections. *E. coli* was stored as 50 µl frozen aliquots of cells at -80 °C. For transformation, aliquots were removed from the freezer and placed on ice to thaw. DNA was incubated with the cells on ice for approximately 30 minutes before being heat shocked in a 42 °C water bath for 20-30 seconds. Cells were cooled on ice for 2 minutes before addition of 200 µl LB media. Cells were then incubated for 1 hour at 37 °C and 200 rpm. Cells were then removed from the incubator and plated onto LB agar plates with antibiotics for selection of positive transformants. Plates were incubated overnight at 37 °C.

2.2.2 Transformation of *A. tumefaciens* LBA4404

Chemically competent *A. tumefaciens* strain LBA4404 cells were taken from the -80 °C freezer and thawed on ice. 1-200 ng of the pEAQ-*HT*-DEST-1 plasmid containing

the gene of interest was added and the cells incubated for 30 minutes. The microcentrifuge tubes containing the cells were then transferred to liquid N_2 for 30 seconds before removal and thawing at room temperature. 200 µl of LB media was added and the cells incubated at 28 °C and 200 rpm for 4 hours. After this, tubes were vortexed briefly and bacteria plated onto LB agar plates containing rifampicin, kanamycin, and streptomycin (50/50/100 µg/ml) and incubated in a 28 °C standing incubator for 3 days until colonies could clearly be seen. A single colony was then transferred to 5 ml liquid LB media containing antibiotics/concentrations as above and incubated at 28 °C and 200 rpm overnight. Around 1 ml of the culture was then transferred to a microcentrifuge tube containing 1 ml 20 % v/v glycerol and frozen at -80 °C. These freezer stocks were used to culture the *Agrobacteria* for future infiltrations.

2.2.3 Antibiotics

Antibiotics were made as detailed in **Table 2-5**. The same concentrations were used for both liquid and solid (agar) LB media.

Table 2-5: Antibiotics used for bacterial culture.

Antibiotic	Stock	Final	Dilution Factor
Rifampicin	50 mg/ml	50 µg/ml	1/1000
Kanamycin	50 mg/ml	50 µg/ml	1/1000
Streptomycin	100 mg/ml	100 µg/ml	1/1000
Gentamicin	50 mg/ml	50 µg/ml	1/1000

All antibiotics were stored as 10 ml stocks in the specified solvent at -20 °C. All stock solutions were made with water except for rifampicin which was made with dimethylformamide.

2.3 Agroinfiltration

2.3.1 Preparation of *A. tumefaciens* for Agroinfiltration

A. tumefaciens from glycerol stocks stored at -80 °C were streaked onto LB agar plates (with rifampicin/kanamycin/streptomycin 50/50/100 µg/ml respectively) before being incubated in a 28 °C standing incubator for approximately 48 h.

A sterile pipette tip was used to transfer a small amount of bacteria to an appropriate volume of liquid LB media with antibiotics as above. The liquid cultures were incubated overnight in a 28 °C 200 rpm shaking incubator.

Cultures were centrifuged at 4000 x g to pellet the bacterial cells, the supernatant discarded and the cells resuspended in MMA solution (10 mM MES buffer, 10 mM $MgCl_2$, 150 µM 3',5'-dimethoxy-4'-acetophenone in H_2O) and left to incubate for 30 min.

The OD_{600} of the solutions was measured using a spectrophotometer set to 600 nm and blanked with MMA. Suspensions were diluted 1 part in 10 in a 1 ml cuvette to bring the concentrations within the accurate range of the spectrometer and the absorbance measured. The reading was multiplied by the dilution factor to give the true density, and the final OD_{600} adjusted to 0.2n (where n is the number of strains of *Agrobacterium* to be infiltrated). Bacterial suspensions were then combined. In the case of multiple combinations being investigated with an uneven number of bacterial strains, green fluorescent protein was used to fill in for missing strains.

2.3.2 Manual Infiltration

Manual infiltration was carried out on approximately 5-week-old *N. Benthamiana* plants. 2 leaves per plant were infiltrated. Leaves were perforated using a pipette tip and the surface of each leaf entirely infiltrated using a 1 ml needleless syringe. Plants were left under greenhouse conditions to grow for 5 days.

2.3.3 Vacuum Infiltration

Vacuum infiltration was carried out according to a published protocol [68]. Briefly, a bath was filled with 10 L infiltration culture and placed into the infiltration apparatus along with 4 inverted 5–6-week-old *N. benthamiana* plants. The vacuum was applied to draw the air out of the interstitial leaf space, and then re-established to draw the infiltration suspension into the leaves. Plants were left under greenhouse conditions to grow for 5 days.

2.4 Purification Techniques

2.4.1 General Procedure for the Evaluation of Triterpene Products by GC-MS

Leaf material was harvested between 5 and 7 days after infiltration. Care was taken to ensure only the producing parts of the plants were harvested. Leaves were frozen at -80 °C and then freeze-dried overnight. A cork borer was used to cut 12 mm discs from each leaf. Leaf discs were placed into 1.5 ml microcentrifuge tubes along with a tungsten carbide bead. The tubes were placed into a ball mill and the leaves ground at 1000 rpm for 1 min.

500 µl ethyl acetate was then added to each tube and the tubes shaken in a thermomixer at 1400 rpm for 30 min.

150 µl of the resulting solution was transferred into a vial and evaporated to dryness. 30 µl of Tri-Sil-Z derivatisation solution was added to each vial and GC-MS spectra recorded for each sample.

2.4.2 GC-MS Analysis

The GC system used was an Agilent 7890B fitted with a Zebron ZB-5HT Inferno column (35 m x 250 µm x 0.1 µm, Phenomenex). Injections were 1 µL performed in pulsed split mode (30 psi pulse pressure, 20:1 split ratio) or splitless mode depending on the concentration of the sample with the inlet temperature set to 250 °C. The GC oven was coupled to an Agilent 5977 A Mass Selective Detector set to scan mode from 60 to 800 mass units (solvent delay 8 min unless stated otherwise) using electron impact (EI) ionisation running at 3.9 scans/sec.

15-Minute Method:

The GC oven temperature gradient began at 170 °C (held 2 min) then ramped to 300 °C at a rate of 20 °C/min and held for 6.5 mins (total run time 15 mins).

20-Minute Method:

The GC oven temperature gradient began at 170 °C (held 2 min) then ramped to 300 °C at a rate of 20 °C/min and held for 11.5 mins (total run time 20 mins).

2.4.3 General Procedure for the Evaluation of Triterpene Products by LC-MS

Leaf material was harvested between 5 and 7 days after infiltration. Care was taken to ensure only the producing parts of the plants were harvested. Leaves were frozen at -80 °C and then freeze-dried overnight. A cork borer was used to cut 12 mm discs from each leaf. Leaf discs were placed into 1.5 ml microcentrifuge tubes along with a tungsten carbide bead. The tubes were placed into a ball mill and the leaves ground at 1000 rpm for 1 min.

550 µl 80% methanol: water (containing 4 µl/ml digoxin standard solution) was added to each tube and the tubes shaken in a thermomixer at 1400 rpm for 30 min. 400 µl of each sample was then transferred to a 1.5 ml microcentrifuge tube and 400 µl hexane added. Tubes were capped, vortexed briefly, then centrifuged at 13,000 x g for 30 s to separate phases. 300 µl hexane was removed from each sample and discarded. Samples were then evaporated using a Genevac.

Samples were re-suspended in 75 µl methanol, transferred to a filter column, and spun at 13,000 x g for 1 min. 50 µl of flow-through was added to 50 µl 50% methanol in water in glass inserts in vials. Vials were capped and LC-MS spectra recorded for each sample.

2.4.4 LC-MS Analysis

The LC system used was a Thermo Vanquish LC fitted with a Kinetex 2.6 µm XB-C18 (100 Å, 50 x 2.1 mm, Phenomenex) column unless otherwise specified. Injections were 5 µl with autosampler temperature set to 10 °C and column oven set to 40 °C. The LC flow passed through a photodiode array (PDA) detector and then was split between a charged aerosol detector (CAD) and a Thermo Qexactive orbitrap mass spectrometer using an atmospheric pressure chemical ionisation (APCI) source.

General LC-MS Method

Pump A was run using MilliQ water + 0.1% formic acid, and pump B was run using MeCN + 0.1 % formic acid. Flow rate was 0.6 ml/min. The gradient for this method started at 35 % B and held for 1 min before increasing to 100 % B over 9 min, holding for a further 5 min at 100 % B, then decreasing to 35 % B over 0.5 min and holding for a final 2 min for a final run time of 17.5 min.

2.4.5 Specifics for Preparative HPLC

Preparative HPLC was carried out using an Agilent 1290 Infinity II preparative system fitted with a Luna 5 µm C18 (100 Å 21.2 x 250 mm, Phenomenex) column. The system used a flow rate of 25 ml/min, and the flow was split 1:1000 between the detectors and the fraction collector. The flow to the detectors passed through a diode array detector (DAD) and was then split between an evaporative light scattering detector (ELSD) and an Agilent InfinityLab LC/MDS XT mass detector using a dual atmospheric pressure chemical ionisation/electrospray ionisation (APCI/ESI) source with a 1.2 ml/min make-up flow for the mass spectrometer.

2.4.6 General Procedure for the Extraction of Triterpene Products from Large Scale Infiltrations

Leaf material was harvested between 5 and 7 days after infiltration. Care was taken to ensure only the producing parts of the plants were harvested. Leaves were frozen at -80 °C and then freeze-dried for 3 days. The leaves were then mechanically crushed, mixed 2:1 with quartz sand, and extracted using a speed extractor (Buchi) at 130 bar pressure and 100 °C with the following programme:

Table 2-6: Standard programme for pressurised solvent extraction using the Buchi speed extractor.

Cycle	Heat-Up Time (min)	Hold Time (min)	Discharge Time (min)	Solvent
1	5	0	3	Ethyl Acetate
2	1	5	3	Ethyl Acetate
3	1	5	3	Ethyl Acetate

The discharge from all the cells of the speed extractor was pooled together through being collected in a bottle attached to the waste outlet. The extract from the three cycles was dried before being further treated with Ambersep (see section **2.4.7**) and subjected to flash chromatography to isolate the compound(s) of interest. Following flash chromatography, decolourisation and recrystallisation were frequently performed to remove minor impurities.

2.4.7 General Procedure for Ambersep Resin Treatment

The ion-exchange resin treatment was carried out according to a published protocol. [68] Dried plant extract was re-dissolved in ethanol and 50 ml Ambersep 900 hydroxide form ion exchange resin (Sigma) was added. The mixture was stirred for 30 minutes during which time the resin changed colour from pink to green. This step was repeated twice more until the colour change of resin was no longer observed, during which time the extract colour changed from green to yellow-brown. The mixture was then filtered through a column of diatomaceous earth and the column washed with 200 ml ethanol, 200 ml 1:1 ethanol hexane mixture, and 200 ml hexane. The washes were combined and dried.

2.4.8 Flash Chromatography

Flash chromatography steps were carried out using an Biotage Isolera One. Samples were dry loaded on silica gel. Flash chromatography was performed in normal phase. Details of the various gradients and mobile phases is given in **Table 2-7**.

2.4.9 Decolourisation

Compounds were dissolved in several ml of a suitable solvent (typically ethyl acetate) and decolourisation was performed by adding a small amount of activated charcoal powder (Sigma). This was incubated for several minutes at room temperature with occasional agitation. The suspension was then transferred to a flash column cartridge emptied of stationary phase and packed with diatomaceous earth (Celite, Sigma). The suspension was passed through the column and the eluent collected in a round bottom flask. The column was washed multiple times with a suitable solvent to ensure maximum recovery of product and the washes combined with the eluent and dried using rotary evaporation.

2.4.10 Flash Chromatography Programmes

Table 2-7: Flash chromatography programmes.

Programme	Mobile Phase (Weak: Strong)	Strong Solvent	Column Volumes
P-1	Hexane: Ethyl Acetate	5-100 % 100 %	10 1
P-2	Hexane: Ethyl Acetate	5-100 %	40
P-3	Hexane: Ethyl Acetate	30 %	45
P-4	Hexane: Ethyl Acetate	20-40%	45
P-5	Dichloromethane: Ethyl Acetate	0 % 100 %	20 3
P-6	Hexane: Ethyl Acetate	40-80%	45
P-7	Dichloromethane: Ethyl Acetate	10 %	20
P-8	Dichloromethane	100 %	25
P-9	Dichloromethane	100 %	30
P-10	Dichloromethane: Ethyl Acetate	0 % 20 % 100 %	5 15 10
P-11	Dichloromethane: Ethyl Acetate	5 %	30
P-12	Dichloromethane: Ethyl Acetate	10 %	15
P-13	Dichloromethane Ethyl Acetate	20 %	25
P-14	Dichloromethane Ethyl Acetate	0 % 10 % 20 % 100 %	15 15 15 10
P-15	Dichloromethane Ethyl Acetate	20 % 50 % 100 %	20 20 20
P-16	Dichloromethane Ethyl Acetate	10	25
R-1	Water Acetonitrile	50 % 50-100 % 100 %	3 15 10

The solvent system used in each programme is listed. The volume of solvent is given in column volumes due to differences in column size. Details of columns used are given in the

relevant sections. Flow rate was always the maximum recommended by the manufacturer (Biotage).

2.4.11 Recrystallisation

A flask containing the solvent to be used in recrystallisation was placed into a water bath set to the boiling point of the solvent in question (usually methanol). Where an anti-solvent was required, this was also placed into the water bath. Compounds to be recrystallised were dissolved in the minimal amount of hot solvent, such that removal from the water bath would cause the solution to become turbid. Where an anti-solvent was required, hot anti-solvent was added dropwise, allowing any precipitate to re-dissolve between drops. This was repeated until the solution became turbid. The flasks were then removed from the water bath and the solution left at room temperature until crystals were visibly observed. The flask was further transferred to ice for ~1 hour to ensure maximum precipitation of compounds. Crystals were filtered from the suspension using vacuum filtration with a sintered funnel covered with filter paper (Wattman). Crystals were washed with cold recrystallisation solvent and dried under vacuum for 1 hour before being transferred from the filter paper to a vial.

2.4.12 NMR

Nuclear magnetic resonance (NMR) spectra were recorded using a Bruker AVIII 400 MHz with broadband probe (400 MHz) or a Bruker NEO 600 MHz with cryoprobe (600 MHz) spectrometer. NMR spectra were recorded in Fourier transform mode at a nominal frequency of 400 MHz ^{1}H/100 MHz ^{13}C or 600 MHz ^{1}H/125 MHz ^{13}C. All NMR spectra were recorded using $CDCl_3$ unless stated otherwise. Chemical shifts were recorded in ppm and referenced to an internal TMS standard or the residual solvent peak. Coupling constants are reported as observed and not corrected for second order effects. Multiplicities are described as: s = singlet; d = doublet; t -= triplet; dd = doublet of doublets; dt = doublet of triplets; td = triplet of doublets; qd = quartet of doublets; br = broad. Assignments were made via a combination of ^{1}H, ^{13}C, distortionless enhancement by polarisation transfer (DEPT), DEPT-edited heteronuclear single quantum coherence (HSQC), heteronuclear multiple bond

coherence (HMBC), correlation spectroscopy (COSY), and 2D rotational frame overhauser enhancement spectroscopy (ROESY) experiments. Where signals overlap, ranges of values are given based on the width of HSQC cross-peaks. Tables of spectra can be found in **Section A.1** along with selected spectra.

2.5 Common Reagents and Solutions

Coprostanol- 1 mg/ml (Sigma, in ethanol or ethyl acetate).

Digoxin- 5 mg/ml in methanol.

RMPI Media- 1640 (Sigma) + Fetal Bovine Serum (Gibco) + 1 % 200 mM L-glutamine (Gibco) + 1 % 10,000 µg/ml Streptomycin/Penicillin (Gibco).

MMA- 10 mM 2-(N-morpholino)-ethanesulphonic acid buffer (MES, Sigma), pH 5.5 (NaOH) + 10 mM $MgCl_2$ + 150 µl 3'5'-dimethoxy-4'-acetophenone (Acetosyringone, Sigma).

3 Combinatorial Biosynthesis for the Generation of Novel Triterpene Diversity

3.1 Introduction

The triterpene toolkit (TTK) in the Osbourn lab is an expansive and innovative platform for the production of valuable triterpenes through biosynthesis in a heterologous host. At the time of writing (December 2022) the toolkit consists of 125 enzymes encompassing the major enzyme families involved in triterpene biosynthesis- the scaffold synthesising oxidosqualene cyclase (OSC) enzymes, scaffold oxidising cytochrome P450s (CYPs), and other types of decorating enzymes such as glycosyl and acyl transferases- the TTK enables a large number of enzymatic transformations. The toolkit has evolved through the introduction of enzymes reported in the literature in a continuous process of evolution as well as through the integration of entire biosynthetic pathways for known natural products with sought-after bioactivities [55,109,110]. This allows for the reconstruction of biosynthetic pathways in their entirety to make specific known triterpene products. Additionally, the significant size of the TTK, which includes enzymes from numerous plant species, also allows for the creation of new to nature compounds with significant diversity. The diversity of triterpenes stems from the modification of a limited number of scaffolds in a vast variety of patterns. The use of enzymes from different plant species that act on similar base scaffolds allows for the combination of modifications in new patterns, expanding the available triterpene diversity to a potentially limitless degree (**Figure 3-1**).

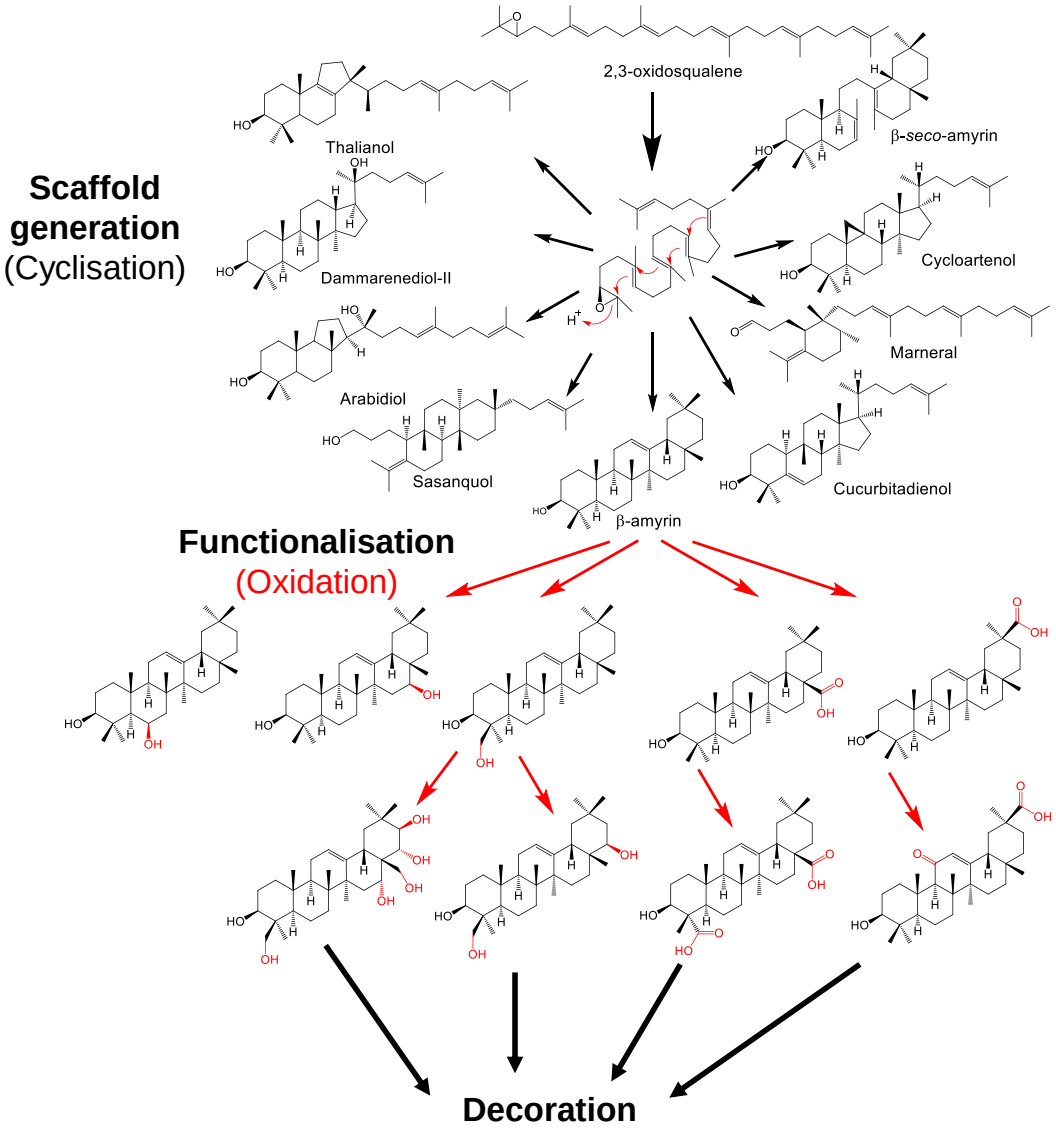


Figure 3-1: Examples of triterpene diversity generated through biosynthetic enzymes. A relatively limited number of scaffolds formed through cyclisation can be functionalised in a potentially unlimited way through the action of enzymes such as CYPs, from simple single oxidations up to more complicated patterns. CYP oxidations are shown in red.

Combinatorial biosynthesis enables expansion of the chemical space of molecules available for evaluation of biological activity and ultimately allows for the direction of biosynthesis for the specific introduction of desired modifications to a scaffold. Therefore, it provides a pathway towards the production of entire suites of molecules and thus greater understanding of triterpene activity. Transient expression systems are especially valuable for this type of work, as combinations of interest can be generated rapidly. Small-scale investigation of potential combinatorial biosynthesis products is important for determining which of the

theoretical products of multiple co-expressed enzymes are producible in practice and which are not, due to the inability of particular combinations of enzymes to work well together. Considerations of which enzymes will and will not work effectively together can then be taken into account when designing molecules containing particular modifications, by enabling optimisation of production. James Reed in the Osbourn group previously carried out a preliminary experiment through combinatorial biosynthesis in which five CYPs (which were from different plant species) from the TTK (previously shown to modify a common triterpene scaffold at different positions) were each co-expressed with a β-amyrin synthase enzyme by combinatorial expression in *N. benthamiana* [50]. This initial proof of concept experiment yielded 9 oxygenated triterpenes with the expected modifications [50]. Subsequent expression of different pairings of the five CYPs with β-amyrin synthase yielded over 40 oxygenated triterpene products, some of which were previously reported and some that were new to nature [50], illustrating the power of combinatorial biosynthesis with just this small subset of enzymes. Importantly, for those compounds that had not been previously reported, it was possible to carry out large-scale agro-infiltration and generate sufficient amounts of compound for NMR structural determination and for use in bioassays. Some examples of the triterpenes generated are shown in **Figure 3-2.**

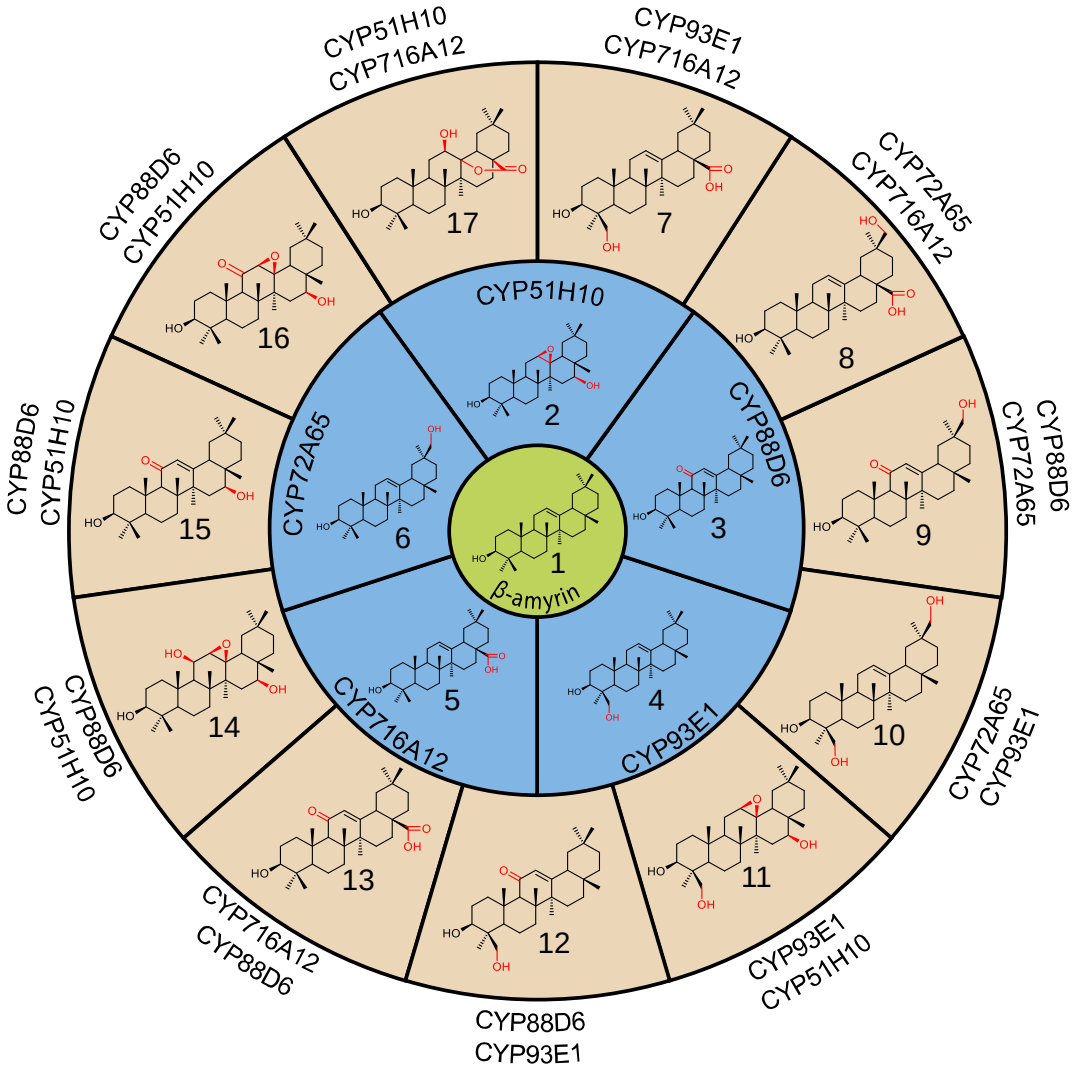


Figure 3-2: Previous combinatorial biosynthesis products. Structures of the produced triterpenes are shown along with the combinations of CYPs used to produce them. A total of 41 compounds were generated, 11 of which were structurally characterised (numbers 7-17 in the figure). Of these, compounds 7-10 were putative structures based on GC-MS data, and structures 11-17 were purified and characterised by NMR. Two of the seven characterised compounds were found to have previously been reported, and five were new-to-nature. Reproduced from [50] with numbering from original source.

The success of this initial study has raised the possibility of expanding this combinatorial biosynthesis approach to include other enzymes. This will enable a fuller understanding of the properties of the CYPs within the TTK and the variety of triterpene products that can be produced by co-expression of OSCs and CYPs. This information will then feed into the design of novel triterpenes with specific structural modifications for biological evaluation.

For simple oxygenated triterpenes, such as those produced through the action of either one or two CYPs on a triterpene scaffold, the resulting products are usually detectable by gas chromatography-mass spectrometry (GC-MS) [111]. This analytical technique, like most forms of chromatography, separates molecules based on their affinity for the stationary phase of a column (in this case a capillary coiled inside the instrument). [112] Molecules are then analysed using a mass spectrometer. This allows for the identification of molecules based on both their mass and retention times. The combination of the high temperatures and hard ionisation technique (election impact or EI) used in the GC-MS instruments at JIC lead to the fracturing of most molecules into fragments [111]. Therefore, the mass ion of a given molecule is likely to be detected in low abundance while fragment ions make up the majority of the MS spectrum produced. This is advantageous in identifying relatively simple hydrophobic triterpene structures, as the fragmentation patterns observed in combination with knowledge of the chemistry underlying the fragmentation reactions can often enable the structures to be inferred. For example, triterpenes based on the β-amyrin scaffold typically undergo a retro Diels-Alder reaction to produce two fragments, one of which is positively charged and therefore shows up with high intensity in the EI mass spectrum of molecules of this type [111].

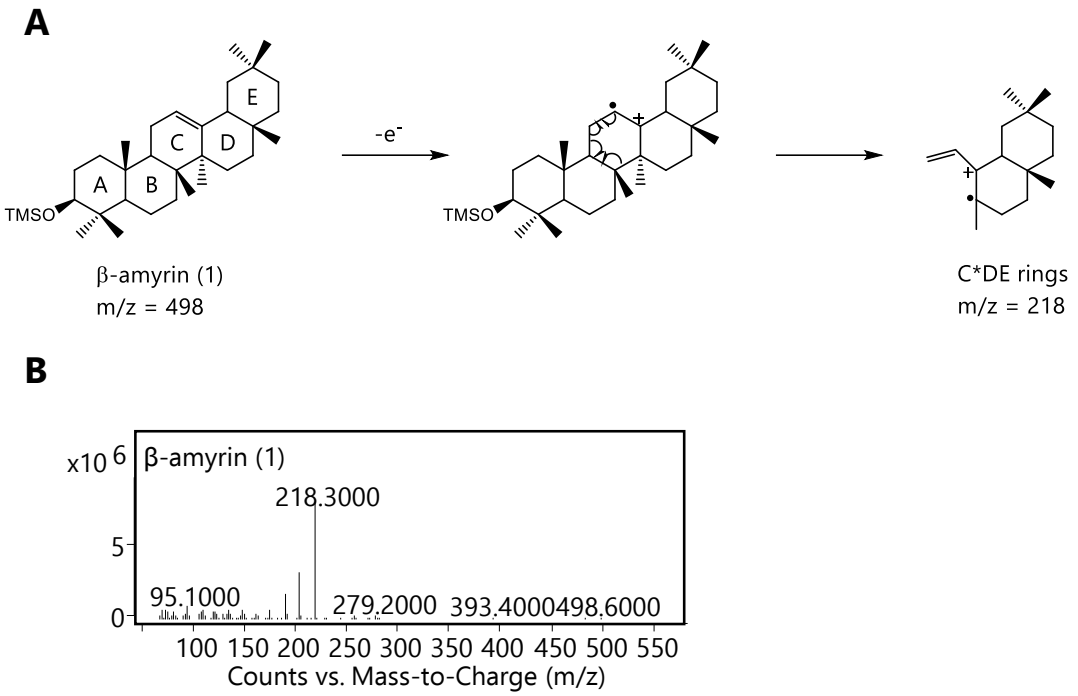


Figure 3-3: The retro Diels-Alder reaction forms characteristic triterpene fragments. A) Schematic of the retro Diels-Alder reaction for β-amyrin (**1**), and the positively charged fragment formed, with m/z values for the major ions shown. B) EI mass spectrum for β-amyrin (**1**) with major ions labelled.

Where the triterpene in question additionally contains an 11-oxo group, the molecule also undergoes a McLafferty rearrangement, which leads to the formation of a different positively charged fragment [113].

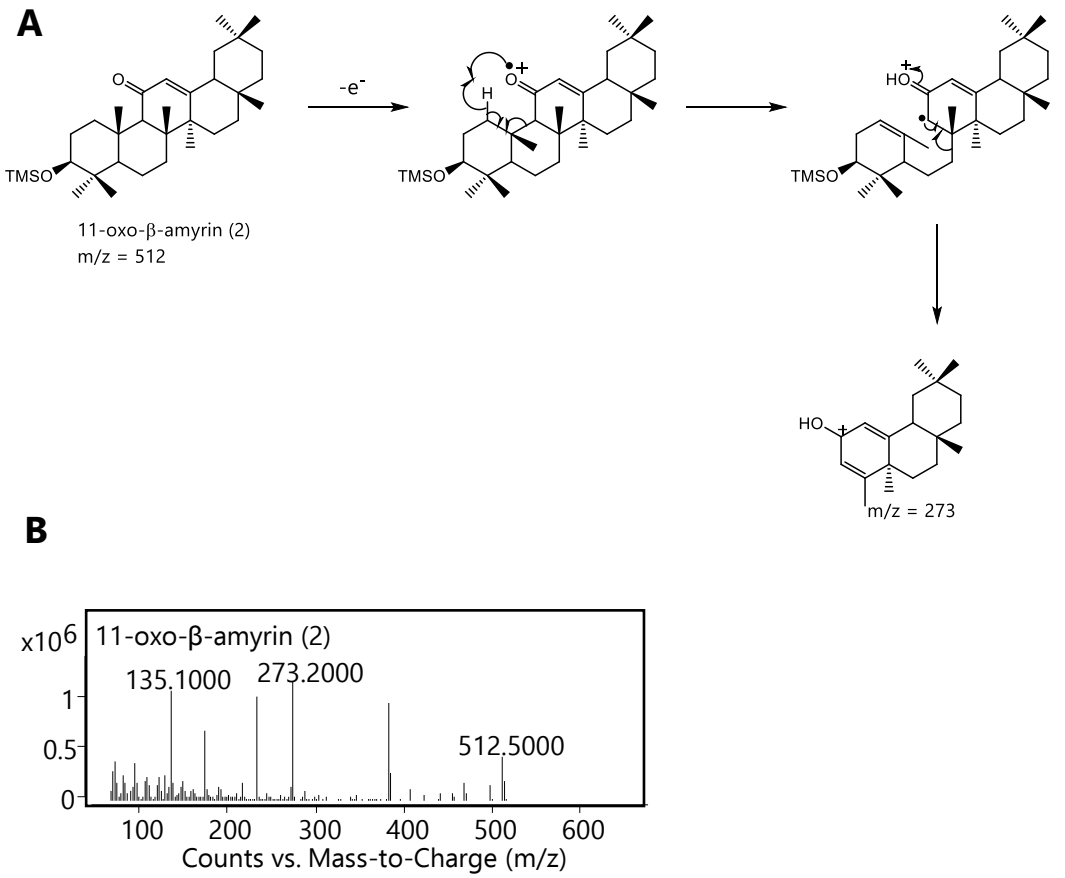


Figure 3-4: The McLafferty rearrangement is commonly observed with ketone-containing triterpenes. A) Schematic of the McLafferty rearrangement for 11-oxo-β-amyrin (**2**), and the positively charged product formed, with m/z values for the major ions shown. B) EI mass spectrum for 11-oxo-β-amyrin (**2**) with major ions labelled.

The introduction of additional functional groups to different parts of a triterpene compound will result in a shift of the masses of these parent fragments in the EI mass spectrum. For example, the addition of a hydroxyl group to the C-, D-, or E-rings of β-amyrin will cause the absence of the original retro Diels-Alder fragment ion at *m/z* = 218 and the observation of a new ion with a mass corresponding to 218 plus a hydroxyl group [111]. Knowledge of the likely identities of the fragment ions within an EI mass spectrum will therefore give clues to the type and location of any additional modifications.

3.2 Aims

In this chapter, a more comprehensive experiment to combine a diverse variety of CYPs from the current TTK in order to determine their ability to oxygenate triterpene scaffolds is carried out. The CYPs are combined pairwise with β-amyrin synthase, the

most common triterpene scaffold, through transient expression and the resulting plant extracts analysed to determine the number and types of new products generated. This work will both evaluate the flexibility of the toolkit enzymes to work in combination and highlight the kinds of products likely to be produced as a result of combinatorial biosynthesis in a pairwise fashion.

3.3 Results and Discussion

3.3.1 Evaluation of Triterpene Combinatorial Biosynthesis via Pairwise Co-expression

In order to gain a better understanding of how well the various CYPs in the TTK are able to act in combinatorial biosynthesis, a pairwise screen of CYPs within the toolkit was carried out. A total of 11 toolkit CYPs, including the five from the previous study [50] (CYP716A12, CYP716A14, CYP716A141, CYP716E26, CYP716S5, CYP72A65, CYP72A69, CYP51H10, CYP88D6, CYP93E1, CYP94D65) were chosen due to their coverage over the range of different positions on the β-amyrin scaffold that can be modified by CYPs within the TTK. Details of the CYPs that were chosen is provided in **Table 3-1** and **Figure 3-5**.

Table 3-1: CYPs used for combinatorial biosynthesis in the present study. The name of the CYP, the species it is derived from, its known substrates and the major products as observed in *N. benthamiana* leaf extracts are given.

Name	Species	Substrate	Major Product	Ref
CYP716A12	*Medicago truncatula*	α/β-amyrin and lupeol	Ursolic/oleanolic/betulinic acid	[114]
CYP716A14v2	*Artemesia annua*	α/β/δ-amyrin	α/β/δ-amyrone	[115]
CYP716A141	*Platycodon grandiflorus*	β-amyrin/ oleanolic acid	16β-hydroxy-β-amyrin	[109]
CYP716E26	*Solanum lycopersicum*	α/β-amyrin	6β-hydroxy-β-amyrin	[116]
CYP716S5	*Platycodon grandiflorus*	β-amyrin	12,13-epoxy-β-amyrin	[117]
CYP72A65	*Medicago truncatula*	β-amyrin	30-hydroxy-β-amyrin	[118]
CYP72A69	*Glycine max*	β-amyrin	21β-hydroxy-β-amyrin	[119]
CYP51H10	*Avena strigosa*	β-amyrin	12,13β-epoxy-16β-hydroxy-β-amyrin	[120]
CYP88D6	*Glycyrrhiza uralensis*	β-amyrin	11-oxo-β-amyrin	[121]
CYP93E1	*Glycine max*	β-amyrin	24-hydroxy-β-amyrin	[122]
CYP94D65	*Avena strigosa*	β-amyrin	23-hydroxy-β-amyrin	[123]

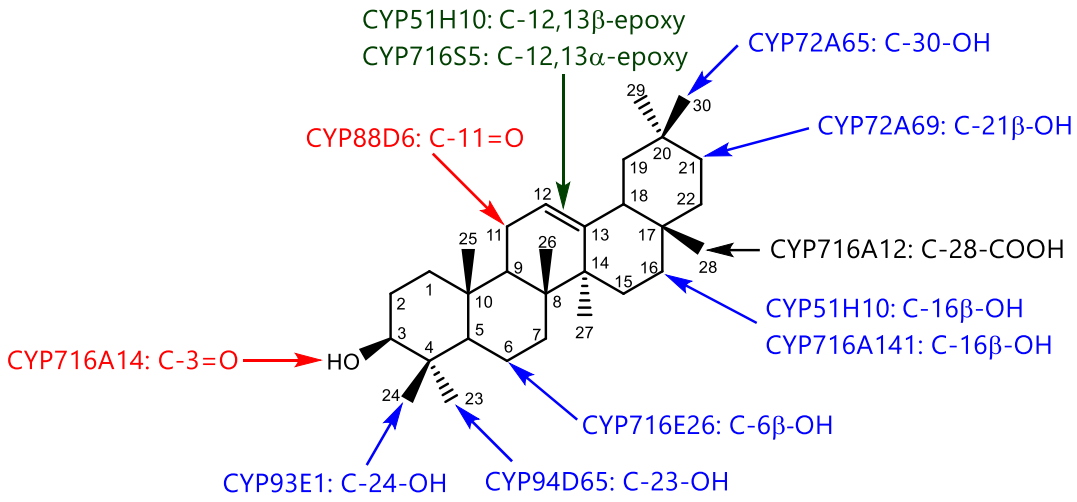


Figure 3-5: Positions on the β-amyrin scaffold modified by the CYPs chosen for combinatorial biosynthesis in the present work.

The CYPs were expressed singly and in pairwise combinations along with the yield boosting enzyme tHMGR (see **Chapter 1**), and the *Avena strigosa* β-amyrin synthase OSC SAD1 (a total of four strains per infiltration) [124]. *A. tumefaciens* strains harbouring the relevant constructs were infiltrated into *N. benthamiana* leaves. Five days later the leaves were harvested, and leaf extracts were analysed by GC-MS (see **3.5.1**). Other treatments included tHMGR and SAD1 only. Where necessary (where fewer than four gene combinations were used), an *A. tumefaciens* strain harbouring a green fluorescent protein (GFP) expression construct was included in order to keep the number of *A. tumefaciens* cells per infiltration constant.

A summary of the combinations, indicating which were successful, and the numbers of new products generated, is provided in **Table 3-2.**

Table 3-2: Pairwise combinations showing numbers of observed products detected by GC-MS.

	CYP716A12	CYP716A14	CYP716A141	CYP716S5	CYP716E26	CYP72A65	CYP72A69	CYP51H10	CYP88D6	CYP93E1	CYP94D65
CYP716A12		1	1	0	1	1	2	0	1	2	1
CYP716A14	1		1	2	0	1	0	2	1	1	0
CYP716A141	1	1		0	1	2	1	0	1	2	3
CYP716S5	0	2	0		1	0	0	0	0	2	2
CYP716E26	1	0	1	1		1	0	1	2	2	0
CYP72A65	1	1	2	0	1		1	0	4	2	0
CYP72A69	2	0	1	0	0	1		0	2	2	0
CYP51H10	0	2	0	0	1	0	0		2	5	5
CYP88D6	1	1	1	0	2	4	2	1		2	1
CYP93E1	2	1	2	2	2	3	2	5	2		0
CYP94D65	1	0	3	2	0	0	0	5	1	0	

Some conclusions can be drawn about the flexibility of the different enzymes when used in combination. A total of 37 out of 66 combinations were successful in producing at least one new product. Certain enzymes appear to be more likely to work in combination with others; CYP716A12 (C-28 oxidase), CYP88D6 (C-11 oxidase), and CYP93E1 (C-24 oxidase) were found to be especially productive, resulting in the detection of new products in almost all of the combinations tested.

Example GC-MS spectra from the combinatorial biosynthesis are shown below. Further spectra can be found in **Section A.2.1**.

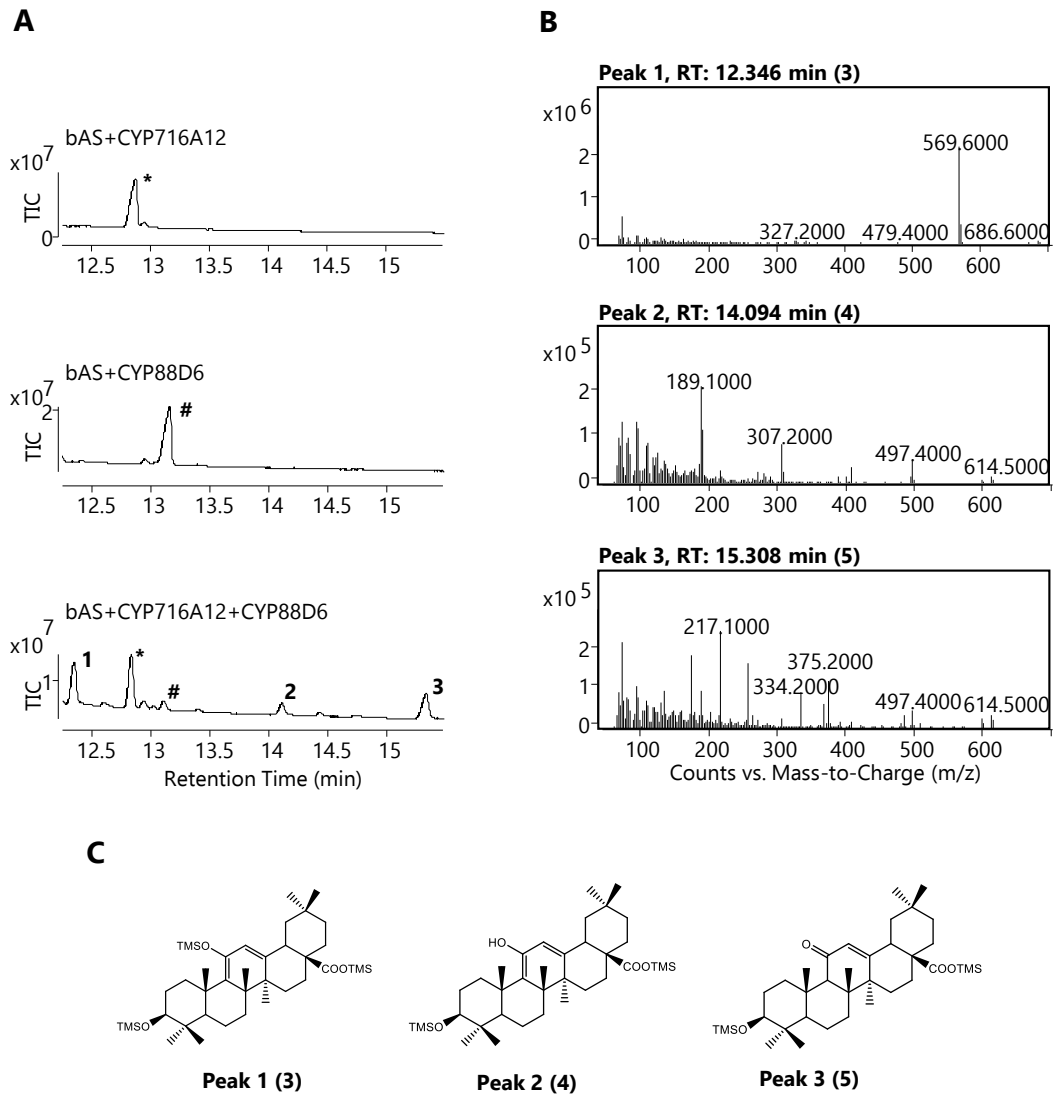


Figure 3-6: CYP716A12+CYP88D6 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP88D6 only (middle); and CYP716A12+CYP88D6 (bottom). B) Mass spectra for the peaks at 12.3 (peak 1), 14.1 (peak 2), and 15.3 (peak 3) minutes. C) Predicted structures for each of the peaks. TIC = total ion chromatogram. * = oleanolic acid (**18**). # = 11-oxo-β-amyrin (**2**).

Figure 3-6 shows the GC-MS trace from the combination of CYP716A12 and CYP88D6. A number of peaks can be seen in the combined trace when compared to the control traces. Based on the mass spectrum, the peak at 15.3 minutes is consistent with 11-oxo-oleanolic acid (**5**), with a molecular ion (M^+) at *m/z* = 614, and major fragments at *m/z* = 497 $[M\text{-}COOTMS]^+$, m/z = 375 (McLafferty rearrangement fragment), and *m/z* = 334 (retro Diels-Alder rearrangement fragment). The peak at 14.1 minutes represents an enol tautomer of the same product (**4**), with the same M^+ ion at *m/z* = 614, and a major fragment peak at

m/z = 497 representing the loss of C-28 and its associated carboxylic acid group. This tautomerism is common in ketone containing compounds but is usually visible in only small amounts; it is especially noticeable here due to the use of pyridine as both a solvent and catalyst in the derivatisation reaction carried out on the samples prior to analysis [125] which induces base-catalysed tautomerism [126]. The peak at 12.3 minutes, meanwhile, is the enol tautomer that has been stabilised through derivatisation in the GC-MS inlet (**3**). This peak has a molecular ion at *m/z* = 686, and a major fragment peak at *m/z* = 569. As the enol and keto tautomers readily interconvert under standard conditions (see **Figure 3-7**), all three of these peaks can be considered as the same product [126].

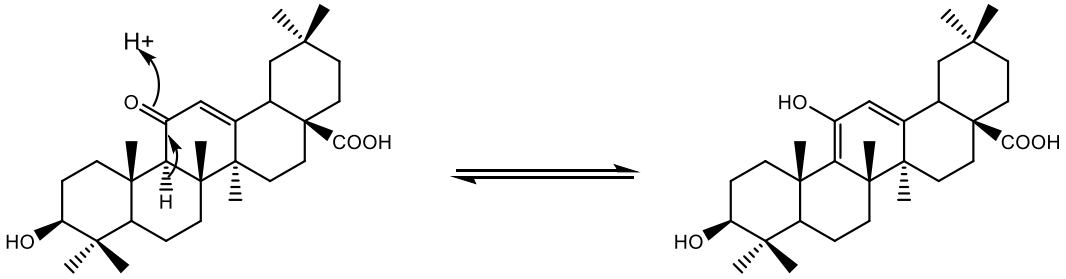


Figure 3-7: Keto-enol tautomerisation.

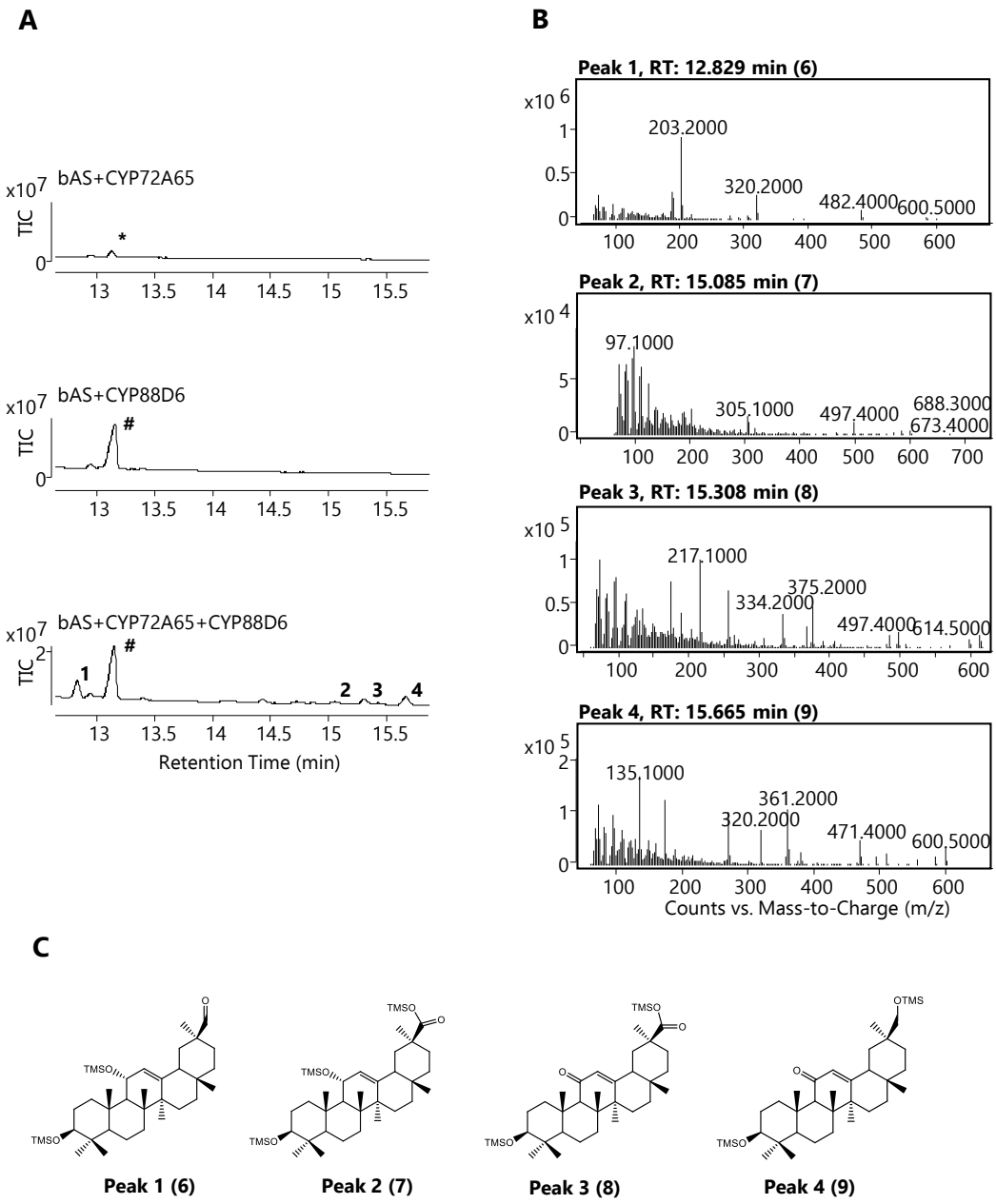


Figure 3-8: CYP72A65+CYP88D6 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP72A65 only (top); CYP88D6 only (middle); and CYP72A65+CYP88D6 (bottom). B) Mass spectra for the peaks at 12.8 (peak 1), 15.1 (peak 2), 15.3 (peak 3), and 15.7 (peak 4) minutes. C) Predicted structures for each of the peaks. TIC = total ion chromatogram. * = 30-hydroxy-β-amyrin (**19**). # = 11-oxo-β-amyrin (**2**).

Figure 3-8 shows the GC-MS trace from the combination of CYP72A65 and CYP88D6. Among the peaks visible in the combined trace compared to the control traces, there are four distinct new peaks. The peak at 15.7 minutes has a retention time, M^+ ion, and fragmentation pattern consistent with 30-hydroxy-11-oxo-β-amyrin (**9**). The peak has a molecular ion at *m/z* = 600, and major fragment ions at

m/z = 361 (McLafferty rearrangement fragment) and *m/z* = 320 (retro Diels-Alder rearrangement fragment). The peak at 12.8 minutes has the same molecular ion, but lacks the McLafferty rearrangement fragment, instead only containing the retro Diels-Alder rearrangement fragment at *m/z* = 320. This suggests that the compound represented by this peak lacks the 11-oxo group but otherwise has the same mass, which would be consistent with an 11α-hydroxy variant of glycyrrhetinic aldehyde (**6**). This is to be expected, since CYP88D6 is known to produce 11α-hydroxy-β-amyrin (**10**) as an intermediate along with 11-oxo-β-amyrin (**2**) [121].

The peak at 15.3 minutes has a molecular ion at *m/z* = 614, consistent with glycyrrhetinic acid (**8**). This peak also has major fragments at *m/z* = 375 (McLafferty rearrangement fragment), and *m/z* = 334 (retro Diels-Alder rearrangement fragment). The peak at 15.1 minutes has a molecular ion at *m/z* = 688, and again does not contain a fragment consistent with McLafferty rearrangement. This suggests that the peak is the 11α-hydroxy variant of glycyrrhetinic acid (**7**).

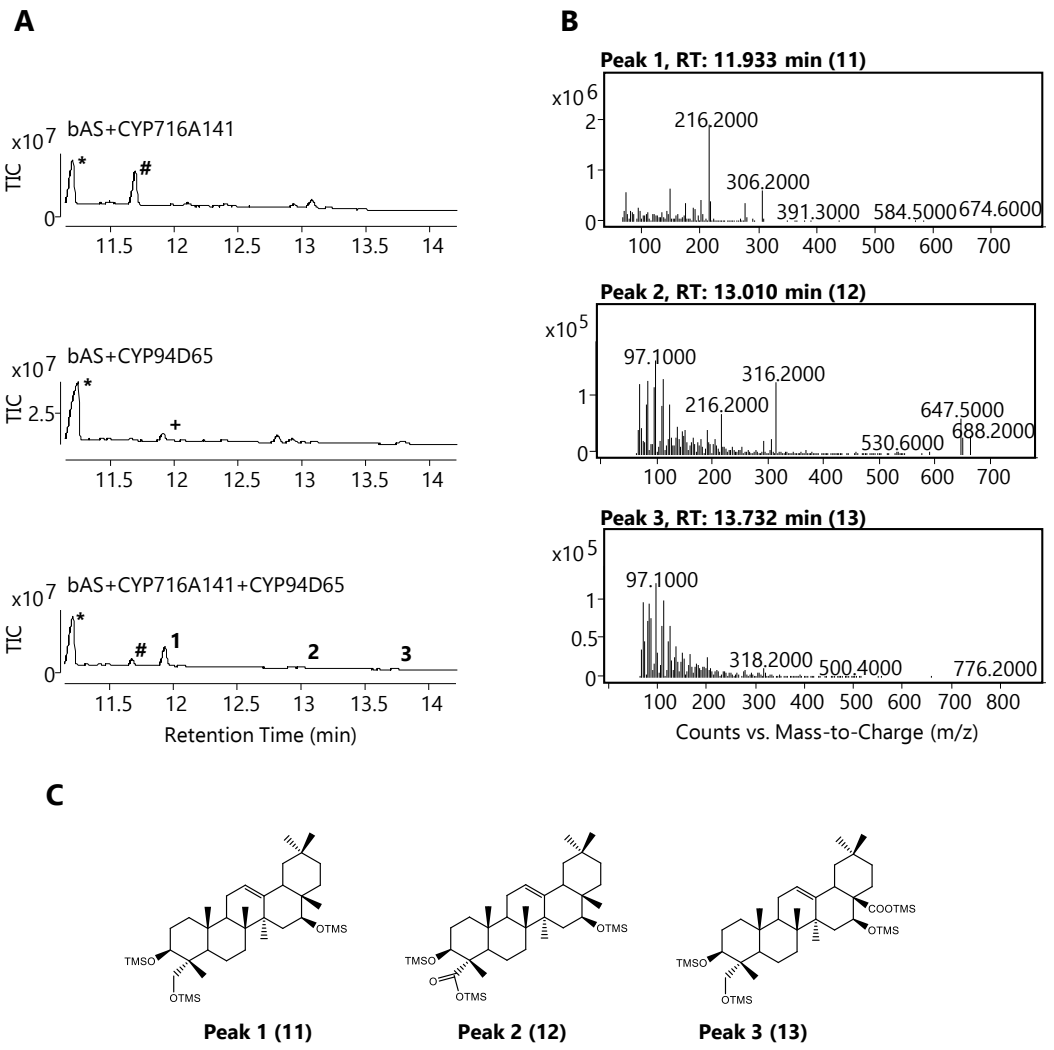


Figure 3-9: CYP716A141+CYP94D65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A141 only (top); CYP94D65 only (middle); and CYP716A141+CYP94D65 (bottom). B) Mass spectra for the peaks at 11.9 (peak 1), 13.0 (peak 2), and 13.7 (peak 3) minutes. C) Predicted structures for each of the peaks. TIC = total ion chromatogram. * = β-amyrin (**1**). # = 16β-hydroxy-β-amyrin (**20**). + = 23-hydroxy-β-amyrin (**21**).

Figure 3-9 shows the GC-MS trace for the combination of CYP716A141 and CYP94D65. The peak visible at 11.9 minutes is consistent with the expected product, 16β,23-dihydroxy-β-amyrin (**11**). The peak has a molecular ion at *m/z* = 674, and a fragment at *m/z* = 306 corresponding to the expected retro Diels-Alder rearrangement fragment. Two smaller peaks can be observed at 13.0 and 13.7 minutes. The peak at 13.0 minutes has a molecular ion at *m/z* = 688, consistent with a carboxylic acid group at C-23 and an alcohol at C-16 (**12**). The peak at 13.7 minutes has a molecular ion at *m/z* = 776 and is consistent with the mass of 16,23-

dihydroxy-oleanolic acid (**13**). This product could also be predicted based on the enzymes used in combination in this experiment, since CYP716A141 is known to produce small amounts of 16-hydroxy oleanolic acid (**14**) [109], which would then be hydroxylated by the action of CYP94D65 to produce the product observed here.

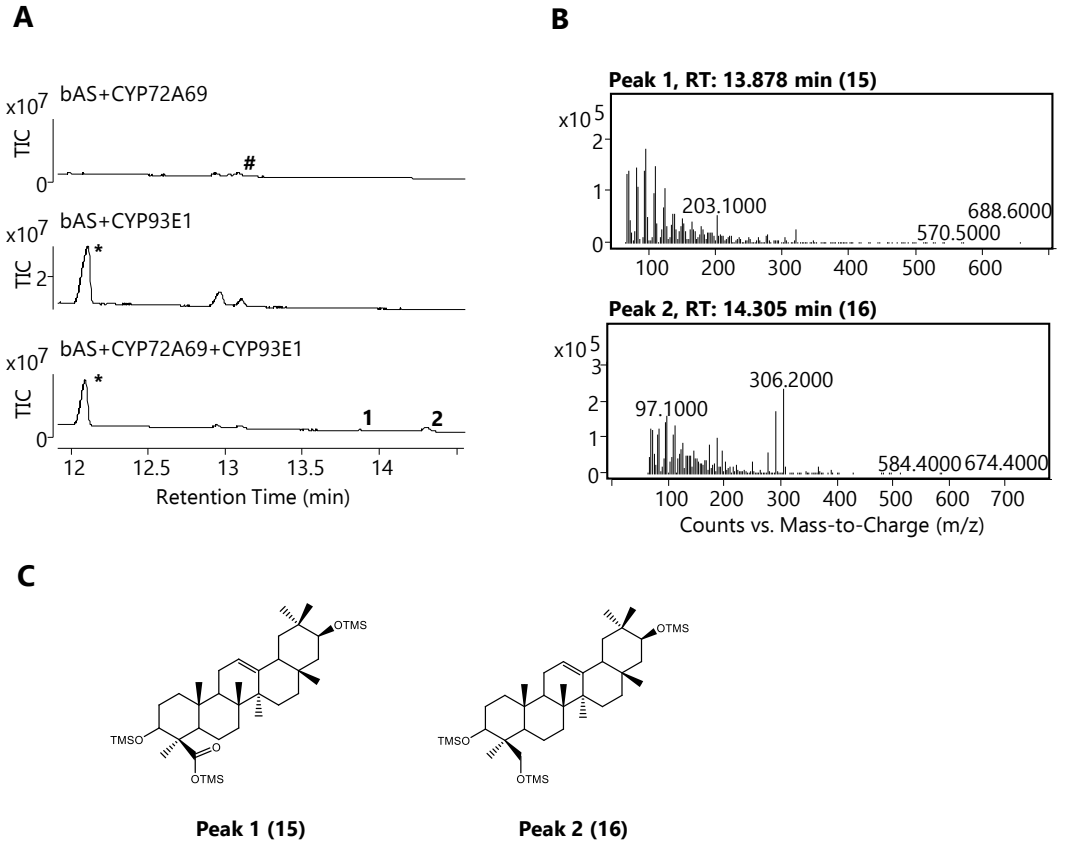


Figure 3-10: CYP72A69+CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP72A69 only (top); CYP93E1 only (middle); and CYP72A69+CYP93E1 (bottom). B) Mass spectra for the peaks at 13.9 (peak 1) and 14.3 (peak 2) minutes. C) Predicted structures for each of the peaks. TIC = total ion chromatogram. * = 24-hydroxy-β-amyrin (**22**). # = 21β-hydroxy-β-amyrin (**23**).

Figure 3-10 shows the GC-MS trace for the combination of CYP72A69 and CYP93E1. Compared to the control traces, a peak can be seen at 14.3 minutes that corresponds to the expected product of 21β,24-dihydroxy-β-amyrin (**16**). This peak has a molecular ion at *m/z* = 674, and a retro Diels-Alder rearrangement fragment at *m/z* = 306. An additional, if small, peak can be seen at 13.9 minutes. This peak has a molecular ion at *m/z* = 688, which would be expected of a minor product of 21β-hydroxy-epi-boswellic acid (**15**). CYP93E1 has been reported previously to produce

small amounts of epi-boswellic acid (**17**), so this minor product is not unexpected [127].

The above examples represent only a fraction of the successful combinations tested, chosen due to being a representative sample of the types of modifications produced in the current study with a good level of accumulation of products in the samples. Similar analyses of all of the combinatorial data were carried out to determine the number and types of products formed. A summary of the full dataset is provided in **Table 3-3**, structures for the compounds are in **Figure 3-11**, and the full GC-MS data can be found in **Section A.2.1**.

Table 3-3: Summary of CYP combinations. Numbers are given for each successful combination corresponding to the putative structures of the compounds produced.

	CYP716A12	CYP716A14	CYP716A141	CYP716S5	CYP716E26	CYP72A65	CYP72A69	CYP51H10	CYP88D6	CYP93E1	CYP94D65
CYP716A12		(**60**)	(**37**)	N/A	(**61**)	(**62**)	(**63, 64**)	N/A	(**5**)	(**65, 66**)	(**52**)
CYP716A14	(**60**)		(**67**)	(**69, 70**)	N/A	(**71**)	N/A	(**73, 74**)	(**35**)	(**75**)	N/A
CYP716A141	(**37**)	(**67**)		N/A	(**76**)	(**77, 78**)	(**79**)	N/A	(**80**)	(**81, 82**)	(**11-13**)
CYP716S5	N/A	(**69, 70**)	N/A		(**83**)	N/A	N/A	N/A	N/A	(**84, 85**)	(**86, 87**)
CYP716E26	(**61**)	N/A	(**76**)	(**83**)		(**88**)	N/A	(**89**)	(**90, 91**)	(**92, 93**)	N/A
CYP72A65	(**62**)	(**71**)	(**77, 78**)	N/A	(**88**)		(**94**)	N/A	(**6-9**)	(**95, 96**)	N/A
CYP72A69	(**63, 64**)	N/A	(**79**)	N/A	N/A	(**94**)		N/A	(**38,97**)	(**15,16**)	N/A
CYP51H10	N/A	(**73, 74**)	N/A	N/A	(**89**)	N/A	N/A		(**80, 98**)	(**99-103**)	(**104-108**)
CYP88D6	(**5**)	(**35**)	(**80**)	N/A	(**90, 91**)	(**6-9**)	(**38, 97**)	**80, 98**)		(**30, 109**)	(**110**)
CYP93E1	(**65, 66**)	(**75**)	(**81 ,82**)	(**84, 85**)	(**92, 93**)	**95, 96**)	(**15, 16**)	(**99-103**	(**30, 109**)		N/A
CYP94D65	(**52**)	N/A	(**11-13**)	(**86,87**)	N/A	N/A	N/A	(**104-108**)	(**110**)	N/A	

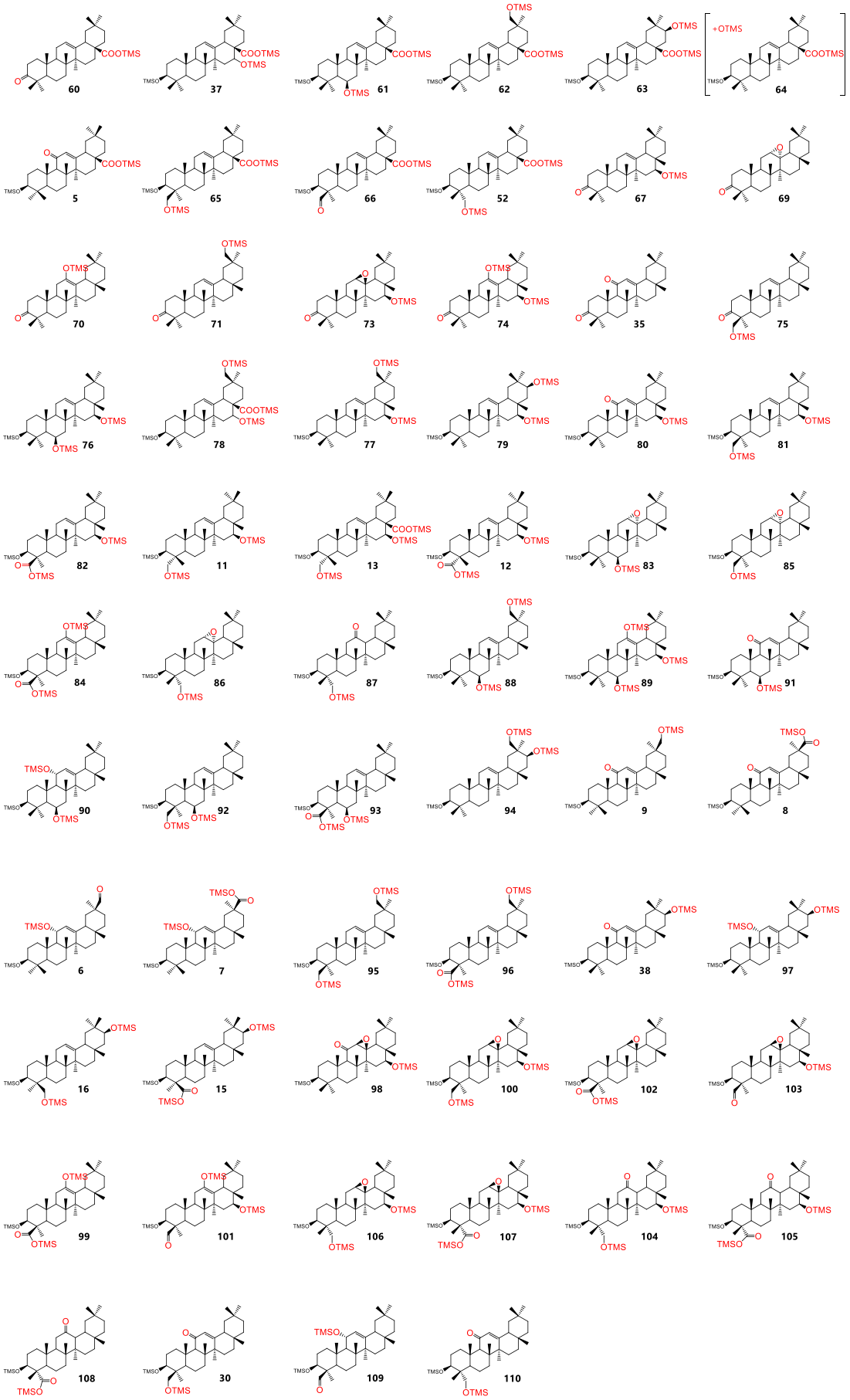

60
37
61
62
63
64
5
65
66
52
67
69
70
71
73
74
35
75
76
78
77
79
80
81
82
11
13
12
83
85
84
86
87
88
89
91
90
92
93
94
9
8
6
7
95
96
38
97
16
15
98
100
102
103
99
101
106
107
104
105
108
30
109
110

***Figure 3-11*:** Structures of the compounds produced using combinatorial biosynthesis. Major functional groups are shown in red.

Overall, more than half of the tested combinations were successful in generating new products. The compounds generated represent considerable chemical diversity, highlighting the versatility of the toolkit for accessing both known and new-to-nature triterpenes, and demonstrating the robust nature of the combinatorial biosynthesis strategy. A total of 64 products have been putatively identified, including 19 products from the previous study [50] along with an additional 45 new compounds. Of these 45, 14 putative structures are for compounds that have been previously reported and 31 are potentially new-to-nature triterpenes not accessible before.

Evaluation of combinations at small scale to identify those that work well together is an essential step prior to scaling up for preparative-scale production of desired triterpenes for full structural analysis and evaluation of bioactivities. These results give important insight into the combinations of enzymes that are likely to work well together, and those that do not. Combinatorial biosynthesis relies on utilised enzymes having a relaxed substrate specificity, but this is not always achievable [128]. Enzymes may fail to work in combination, even when expressed in a system in which they are known to express correctly, where the substrate has been structurally altered [128]. This is particularly the case for enzymes which are structurally specific about their substrates. However, the large number of successful combinations seen here is a cause for optimism in future co-expression studies, showing that combinations of multiple enzymes are more likely to be active together than not.

These results also provide some insight into the flexibility of CYPs involved in triterpene biosynthesis towards different substrates. CYPs are generally specific in the products they produce. The majority of successful combinations produced only a single product representing the oxidations carried out by both CYPs, suggesting that the enzymes have maintained their regiospecificity. Where multiple products were formed, these represented expected side activities, such as CYPs known to oxidise in 2 positions instead of only one (such as CYP716A141 [109], and CYP51H10

[120]), or CYPs that can catalyse multiple rounds of oxidation to a single position forming small amounts of aldehyde or carboxylic acids (such as CYP93E1 [127]). Despite this specificity in the products produced, CYPs can be more promiscuous in the substrates they accept. Of particular note, the CYP716 family (including the enzymes CYP716A12, CYP716A14, CYP716A141, CYP716S5, and CYP716E26 used in the present study) are known to accept as substrates a variety of pentacyclic triterpenes, including α- and β-amyrin, lupeol, germanicol [114], as well as derivatives such as oleanolic acid [117,129]. This raises the potential for expanding the combinatorial approach seen here to additional scaffolds beyond β-amyrin (**1**). Additionally, the order of reactions that occur during the generation of these products is not known, but in order for so many of the combinations investigated here to be successful, the majority of the CYPs tested must be able to accept various oxidised derivatives of β-amyrin (**1**) as a substrate. Another avenue for future development of this approach would therefore be to test combinations of CYPs with more oxidised β-amyrin derivatives.

3.4 Conclusions

Here the CYPs within the triterpene toolkit have been systematically evaluated through pairwise combination to reveal the numbers and types of products produced by multiple CYPs in co-expression experiments. Over half of the potential pairwise combinations were successful in producing at least one new product detectable by GC-MS, with 18 combinations producing multiple new products due to one or both enzymes in the combination producing multiple products in isolation.

Of the 64 total products formed, 31 had putative structures that are currently unrepresented in the literature, showing the potential of the chemical space accessible through combinatorial biosynthesis. A selection of the combinations has been examined here in more detail to discuss the methodology of the GC-MS analysis and showcase the variety of the products produced using a simple combination of CYPs with β-amyrin synthesis genes. This variety of compounds would have been difficult to produce synthetically and represents a major advance

in the ability to harness this major class of compounds, as well as showcasing the production of molecules with a variety of potential handles for further modification through either further combinatorial biosynthesis efforts or semi-synthetic chemistry techniques.

The pattern of products generated through this approach is representative of the specific nature of the products formed by triterpene-oxidising CYPs, as generated products are consistent with the expected major products and known side-products of the CYPs involved. The products also show the flexibility of the triterpene-oxidising CYPs towards the substrates they accept, as the number and variety of products generated implies that the investigated CYPs are capable of accepting multiple oxidised derivatives of β-amyrin as substrates as well as β-amyrin (**1**) itself.

Building on this foundation, future work will utilise this knowledge and flexibility by using the toolkit in the design and production of a variety of molecules, both known and novel. It will also be important to perform scale-up and purification on the successful combinations reported here in order to verify their putative structures by NMR and provide certainty as to the identities of the produced compounds.

3.5 Materials and Methods for Chapter 3

3.5.1 Procedure for Combinatorial Biosynthesis

11 CYPs from the triterpene toolkit (CYP716A12, CYP716A14, CYP716A141, CYP716S5, CYP716E26, CYP72A65, CYP72A69, CYP51H10, CYP88D6, CYP93E1, CYP94D65) along with tHMGR, SAD1, and GFP, were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined such that each combination contained tHMGR, SAD1, then either one CYP + GFP or two CYPs combined pairwise (for a full list of combinations, see **Table 3-4**). The combined strains were used to infiltrate 1 leaf of 5-week-old *N. benthamania* plants using the manual infiltration method (see **Section 2.3.2**). The plants were grown for 5 days under greenhouse conditions before leaves were harvested and processed using the standard method for GC-MS analysis (see **Section 2.4.1**) using the 20-minute method.

Table 3-4: Combinations used in the pairwise evaluation of the triterpene toolkit CYPs. All combinations contain tHMGR and SAD1.

tHMGR+GFP only	CYP716A14+ CYP88D6	CYP88D6+ CYP716E26	CYP716S5+ CYP716A12	CYP716A12+ CYP72A69
tHMGR+SAD1 +GFP	CYP88D6+ CYP51H10	CYP88D6+ CYP716A141	CYP716S5+ CYP716E26	CYP716A12+ CYP93E1
CYP716A14+ GFP	CYP716A14+ CYP51H10	CYP88D6+ CYP72A69	CYP716S5+ CYP716A141	CYP716A12+ CYP94D65
CYP88D6+GFP	CYP716A14+ CYP716S5	CYP88D6+ CYP93E1	CYP716S5+ CYP72A69	CYP716E26+ CYP716A141
CYP51H10+ GFP	CYP716A14+ CYP72A65	CYP88D6+ CYP94D65	CYP716S5+ CYP93E1	CYP716E26+ CYP72A69
CYP716S5+ GFP	CYP716A14+ CYP716A12	CYP51H10+ CYP716S5	CYP716S5+ CYP94D65	CYP716E26+ CYP93E1
CYP72A65+ GFP	CYP716A14+ CYP716E26	CYP51H10+ CYP72A65	CYP72A65+ CYP716A12	CYP716E26+ CYP94D65
CYP716A12+ GFP	CYP716A14+ CYP716A141	CYP51H10+ CYP716A12	CYP72A65+ CYP716E26	CYP716A141+ CYP72A69
CYP716E26+ GFP	CYP716A14+ CYP72A69	CYP51H10+ CYP716E26	CYP72A65+ CYP716A141	CYP716A141+ CYP93E1
CYP716A141+ GFP	CYP716A14+ CYP93E1	CYP51H10+ CYP716A141	CYP72A65+ CYP72A69	CYP716A141+ CYP94D65
CYP72A69+ GFP	CYP716A14+ CYP94D65	CYP51H10+ CYP72A69	CYP72A65+ CYP93E1	CYP72A69+ CYP93E1
CYP93E1+ GFP	CYP88D6+ CYP716S5	CYP51H10+ CYP93E1	CYP72A65+ CYP94D65	CYP72A69+ CYP94D65
CYP94D65+ GFP	CYP88D6+ CYP72A65	CYP51H10+ CYP94D65	CYP716A12+ CYP716E26	CYP93E1+ CYP94D65
CYP88D6+ CYP716A12		CYP716S5+ CYP72A65		CYP716A12+ CYP716A141

4 New Functionalisations through Incorporation of Novel Enzymes from Plant, Bacterial, and Mammalian Sources

4.1 Introduction

As demonstrated in **Chapter 3**, the triterpene toolkit is a versatile platform for the production of triterpenes, especially of the β-amyrin scaffold, oxidised at a variety of positions. However, the toolkit as it stands remains incomplete, as it is not yet possible to access all of the positions on the β-amyrin scaffold with existing enzymes, and scaffolds other than β-amyrin (**1**) remain largely unserved. To expand the accessible chemical space and therefore the diversity of the triterpenes that can be bioengineered, it is important to augment the existing toolkit with new enzymes with different functions.

The emergence of organised databases showing triterpene biosynthetic pathways and the enzymes involved has greatly improved the accessibility of the vast amounts of data on triterpene biosynthesis now becoming available due to genome mining efforts. For example, the TriForC (triterpenes for commercialisation) database [130], which was set up in 2017 by the EU TriForC consortium, in which the Osbourn lab was a partner. The database comprises a collection of triterpene biosynthetic pathways, enzymes, and compounds, fully searchable by keywords, DNA sequence, compound (sub)structure, or plant species, and interlinked for ease of use [131]. Upon its inception, the database comprised 271 enzymes from 14 different enzyme classes and 70 plant species [131], with the aim of updating the database twice a year. At time of writing, the database consists of 307 enzymes.

The β-amyrin scaffold is one of the most common triterpene scaffolds found in plants and is also the scaffold for a number of molecules with desirable functions, such as Avenacin A-1, QS-21, and glycyrrhizin (see **Chapter 1**) [27,60,118,132]. For this reason, the characterisation of biosynthetic enzymes capable of modifying β-amyrin (**1**) or its derivatives has been a particular focus within the Osbourn lab. It was previously shown by Michael Stephenson in the Osbourn lab that the majority of accessible carbon atoms within β-amyrin (that is, those that are not quaternary) are oxidised to at least some degree by plants in nature. A list of cytochromes P450 (CYPs) that have been characterised in the literature has also been assembled as part of a literature review by Malhotra and Franke covering all triterpene scaffolds

[129] and cross referenced against a subset of toolkit CYPs acting on β-amyrin (**1**) identified as part of the current work (see **Section A.4**). Therefore, it is possible to build up an understanding of the "orphan positions" on the β-amyrin scaffold- those that it should be possible to oxidise but for which no oxidative enzyme has currently been characterised- as well as to identify positions for which a characterised enzyme exists but is not currently represented within the toolkit.

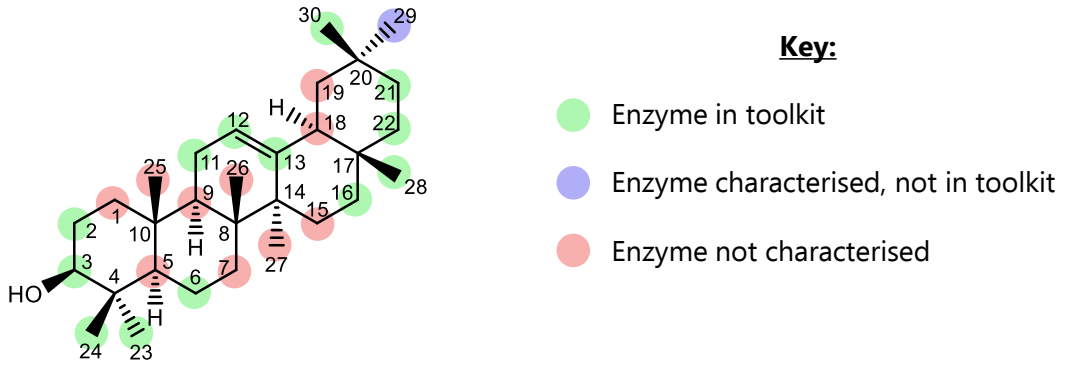


Figure 4-1: The β-amyrin scaffold, showing the oxidisable positions and their status in terms of characterised enzymes, inclusion within the triterpene toolkit, and orphan positions.

The "orphan" positions (those for which no characterised enzyme has yet been reported) shown in **Figure 4-1** are of particular interest as they represent novel chemistry relative to the toolkit's current capabilities. In addition to these positions, another focus for toolkit expansion is enzymes that modify positions already covered by the existing toolkit but in new ways. For example, the secondary carbons within the β-amyrin ring can be hydroxylated in either an α or β configuration, each of which is different due to the rigid ring-structure of pentacyclic triterpenes (see **Figure 4-2**). Secondary carbons can also be oxidised either by hydroxylation or formation of a ketone, while the primary carbons on the methyl groups of the ring can be oxidised with hydroxylation, or formation of an aldehyde or carboxylic acid group. Not every possible oxidation on every possible position of the scaffold is accessible with the current set of known characterised enzymes, but where multiple enzymes exist that oxidise the same position in different ways, this opens up opportunities for additional versatility and triterpene diversity.

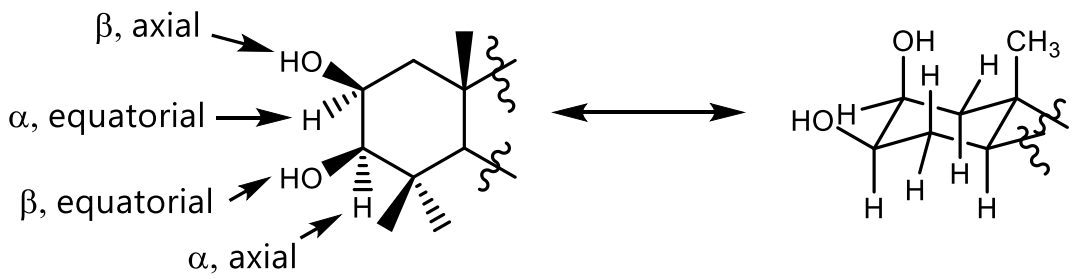


Figure 4-2: Diagram showing the different arrangements in space of substituents on the A-ring of a triterpene molecule. These spatial arrangements mean that a C-2α and C-2β oxidised triterpene, for instance, will represent two unique and different molecules.

Previous toolkit expansion has focused on plant biosynthetic enzymes. While the increased availability of plant genomic data and the corresponding increase in the discovery of plant enzymes will likely bring forth novel plant enzymes for some time to come, plants represent only one branch of the tree of life. There is therefore value in looking beyond the plant kingdom to capture the range of triterpene chemistry carried out by enzymes from other organisms.

Non-eukaryotic CYPs are generally soluble and expressed in the cytoplasm of the cell rather than being targeted to a membrane, as is often the case for eukaryotic CYPs [133]. Bacterial CYPs have also evolved that metabolise steroids and sterol-like molecules, which are themselves a subset of triterpenes [134]. Therefore, bacterial CYPs that metabolise steroids are a clear target for evaluation as potential toolkit additions. The only downside to be considered is the potential requirement for bacterial redox partners for bacterial CYPs due to their cellular localisation being different to that of the largely membrane bound plant CYPs and relevant co-factors.

The animal kingdom contains different CYP families compared to the plant kingdom, some of which show significant promiscuity in terms of the substrates they will accept. The most widely known of these are the CYPs found within the human liver, where they have important roles in detoxification of substances. They are therefore a source of promiscuous CYPs that are able to act on a wide range of molecules in a variety of ways. Particular attention should be drawn to CYP3A4, an especially versatile CYP that is known to show activity towards steroids and related molecules [135].

The above introduction has largely focused on CYPs due to their diversity as an enzyme superfamily. In addition to CYPs, other classes of biosynthetic enzymes that are able to modify triterpene scaffolds include alcohol dehydrogenases, glucosyl transferases, acyl transferases, and isomerases. All of these are currently under-represented in the toolkit and therefore offer further dimensions to enzymatic diversification of biosynthesised triterpenes.

4.2 Aims

To expand and develop the triterpene toolkit, in this chapter new enzymes are evaluated for their ability to enzymatically modify and diversify triterpenes. These include a newly characterised plant CYP450 and CYP450s and relevant co-factors from bacterial and mammalian sources. The products generated by these enzymes in combinatorial biosynthesis experiments are characterised through preliminary structure inference through GC-MS. Finally, two alcohol dehydrogenase enzymes from *Arabidopsis thaliana* are evaluated with the aim of generating derivatives of β-amyrin (**1**) with alternate stereochemistry at C-3, a modification observed in some naturally occurring bioactive triterpenes such as boswellic acids, produced by the frankincense tree. Collectively these advances unlock new chemical space, so enabling us to broaden our investigations of the relationships between structure and bioactivity.

4.3 Results and Discussion

4.3.1 Investigation of the Activity of New Plant Enzyme CYP716C11 Through Transient Expression

A significant goal of the triterpene toolkit is to be able to functionalise as many positions as possible around the triterpene scaffold, and to generate different stereochemistries. New sources of chemical diversity for combinatorial biosynthesis are therefore continually incorporated. A C-2β hydroxylase (CYP72A67) from *Medicago truncatula* already exists within the toolkit [136]. This enzyme produces C-2β hydroxylated products in combination with oleanolic acid (**18**) or derivatives and requires oleanolic acid (**18**) or a derivative as a substrate. A newly reported enzyme, CYP716C11 from *Centella asiatica*, which catalyses a C-2α hydroxylation, therefore represents a new stereochemistry [117]. The gene encoding this enzyme was synthesised from a commercial source and a binary expression construct made as detailed in section **4.5.1**.

Once the binary expression construct for CYP716C11 had been produced, the next step was to validate its reported activity through transient expression. As CYP716C11 had been reported to produce maslinic acid (2α-oleanolic acid) (**24**) using oleanolic acid (**18**) as a substrate, co-expression was carried out with the required enzymes to produce oleanolic acid (**18**). *Agrobacterium* strains harbouring the combined constructs were agro-infiltrated into *N. benthamiana* (see **Section 2.3.2**). Five days later, samples obtained from the freeze-dried infiltrated leaves were extracted with ethyl acetate and extracts analysed by GC-MS (see **Section 2.4.1**). **Figure 4-3** shows the GC-MS spectra from the infiltration.

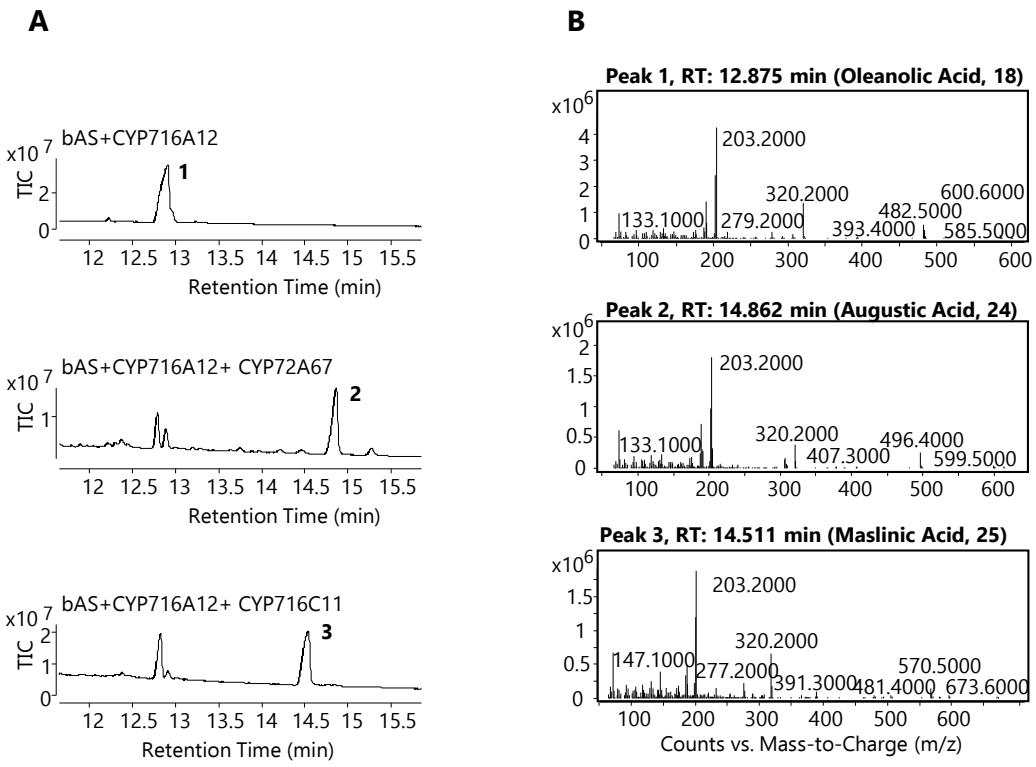


Figure 4-3: CYP72A67 and CYP716C11 produce two different products. A) GC-MS traces for tested C-2 hydroxylases co-expressed with tHMGR, β-amyrin synthase (SAD1), and C-28 oxidase CYP716A12: CYP72A67 (middle), known to hydroxylate oleanolic acid (**18**) at the C-2β position [136], and CYP716C11 (bottom), expected to hydroxylate at C-2α. The GC-MS trace for the combination of tHMGR, SAD1, and CYP716A12, producing oleanolic acid (**18**), is also shown (top). B) The mass spectra of the peaks seen at 12.9 (peak 1), 14.9 (peak 2), and 14.5 (peak 3) minutes. TIC = Total ion chromatogram.

Analysis of these spectra show that the combination expressing CYP72A67 produces a peak with a retention time of 14.9 min and a *m/z* of 599, corresponding to the known product of augustic acid (**25**) (599.5 = $[M\text{-}OH]^+$). Meanwhile, the expected C-2α hydroxylation activity of the combination expressing CYP716C11 is seen by the peak at the different retention time of 14.5 min having a *m/z* of 673.6 corresponding to maslinic acid (**24**) (673.6 = $[M\text{-}CH_3]^+$). This represents an oxidation of oleanolic acid (**18**) with a new stereochemistry.

4.3.2 Evaluation of the Ability of Bacterial Cytochromes P450 to Modify Triterpene Scaffolds in a Plant Expression System

In addition to newly reported plant enzymes with relevant activity, another avenue for expanding the triterpene toolkit is the incorporation of non-plant enzymes. Cytochromes P450 exist in most kingdoms of life [137]. Bacterial enzymes have been

well-studied due to their solubility and ease of expression in hosts such as *E. coli*, and mutagenesis studies have generated variants with promiscuous and wide-ranging activity, with mutants of the *Bacillus megaterium* enzyme BM3 being one prominent example [138,139].

Two other relevant enzymes have been recently reported from *Bacillus megaterium*: CYP106A1 and CYP106A2 [140,141]. These enzymes have previously been evaluated for their ability to use triterpenes as substrates (e.g. oleanolic acid (**18**) and boswellic acid (**26**)), and shown to oxidise positions at C-7 and C-15 [142,143]. As these are positions not currently served by existing toolkit enzymes, they have particular potential for improving the available diversity of triterpene structures. The genes coding for these enzymes were codon-optimised for *N. benthamiana*, synthesised, and inserted into binary expression constructs according to section **4.5.2.**

Other bacterial CYPs have previously been expressed in transgenic plants [144,145]. However, CYP106A1 and CYP106A2 have previously only been investigated in vitro or in microbial-based expression systems [141]. It was therefore not known whether plant reductases would be able to efficiently act as co-enzymes and drive expression of these CYPs. Therefore, a *Bacillus megaterium* ferredoxin, BmFdx2, was also inserted into a binary expression construct, since this has been shown to most effectively support the activity of CYP106A1 and CYP106A2, compared to other ferredoxins (see **4.5.2**) [141].

This ferredoxin in itself requires a co-enzyme to work effectively. Its redox partner in *Bacillus megaterium* has not been identified, so alternative co-reductases capable of working with ferredoxins were also inserted into binary expression constructs (FDR, BphA4, Arh1) as described in **4.5.3**.

CYP106A1 and CYP106A2 were then tested in the transient plant expression system. As these enzymes have reported activity towards a range of triterpenes, including oleanolic acid, they were initially co-expressed with either the genes required to synthesise β-amyrin (**1**) or oleanolic acid (**18**), without co-enzymes and with different combinations of co-enzymes (BmFdx2 + FDR, BmFdx2 + BphA4, BmFdx2 + Arh1).

The combinations tested are shown in **Table 4-4**. Agro-infiltration of these combinations was carried out (see **4.5.3**). Leaf extracts were analysed by GC-MS.

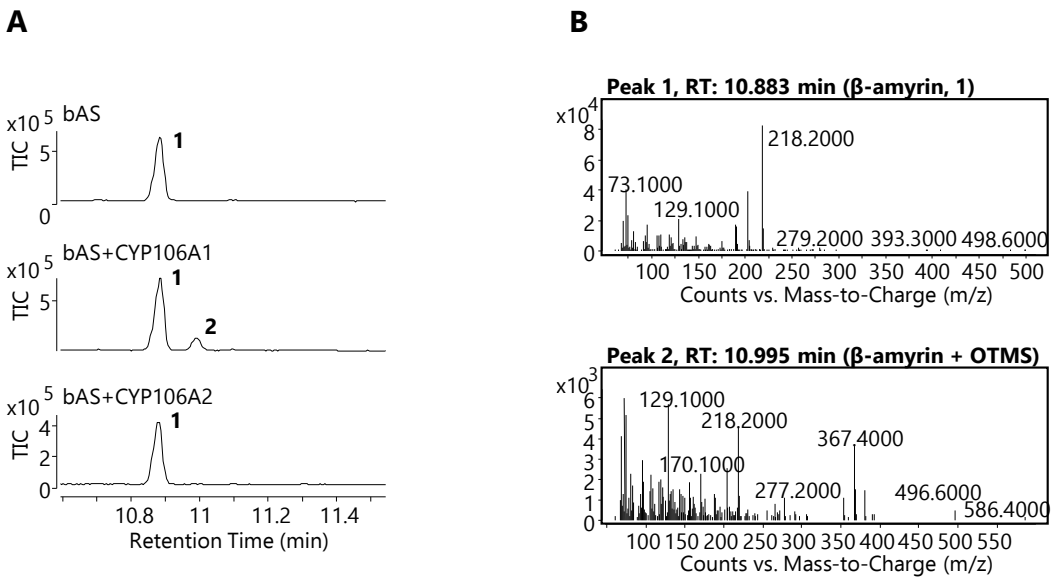


Figure 4-4: Co-expression of CYP106A1 and bAS produces a new product. A)GC-MS traces for CYP106A1 (middle) and CYP106A2 (bottom), co-expressed with tHMGR and β-amyrin synthase (SAD1), and tHMGR and β-amyrin synthase only (top). B) Mass spectra of the peaks at 10.9 (peak 1) and 11.0 (peak 2) minutes. TIC = Total ion chromatogram.

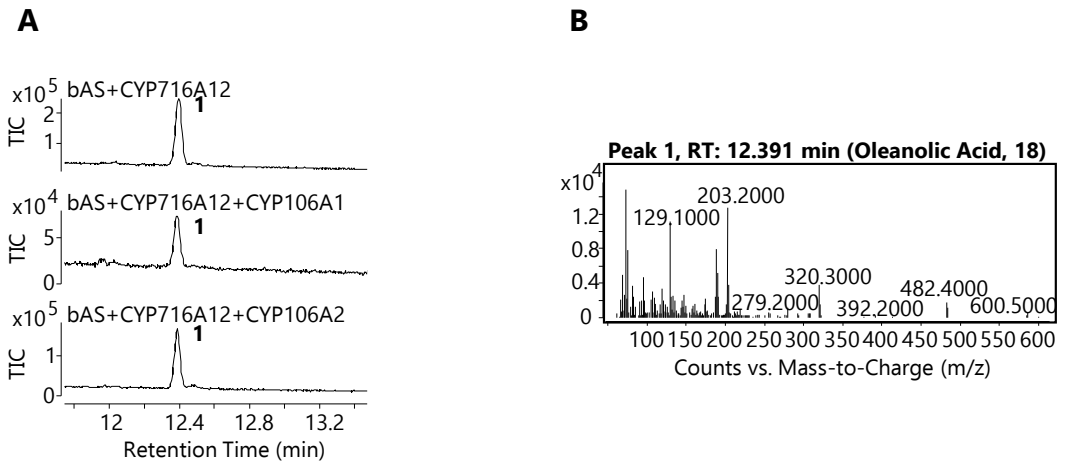


Figure 4-5: Co-expression of CYP106A1 or CYP106A2 and CYP716A12 produces no new products visible by GC-MS. A) GC-MS traces for CYP106A1 (middle) and CYP106A2 (bottom), co-expressed with tHMGR, β-amyrin synthase (SAD1), and C-28 oxidase CYP716A12. The GCMS trace for tHMGR, SAD1, and CYP716A12, producing oleanolic acid (**18**), is also shown (top). B) Mass spectrum of the peak at 12.4 minutes (peak 1). TIC = Total ion chromatogram.

The spectra show a new peak that appears at 11 min when CYP106A1 is added to the β-amyrin-producing combination (see **Figure 4-4**) but not in combination with oleanolic acid (**Figure 4-5**). Additionally, no new peaks are seen in combination with CYP106A2. The mass spectrum of the peak (m/z = 586 for the M^+ ion) suggests a hydroxylated β-amyrin product. The co-reductases do not seem to have a significant effect on the amount of product produced by CYP106A1, suggesting that plant co-enzymes are capable of working with the bacterial cytochrome P450. This is interesting when it is considered that plant CYP450s tend to localise to the endoplasmic reticulum while bacterial CYP450s are cytosolic in nature.

4.3.3 Evaluation of CYP106A1 and CYP106A2 Through Co-expression with Existing Toolkit Enzymes

In order to gain a clearer understanding of the activities of CYP106A1 and CYP106A2, these enzymes were co-expressed with other cytochromes P450 in the TTK. To that end, a set of *A. tumefaciens* strains containing the expression constructs for the enzymes CYP716A12, CYP716A14v2, CYP716A141, CYP716C11, CYP716E26, CYP72A65, CYP72A67, CYP72A69, CYP88D6, CYP93E1, and CYP94D65 were cultured and agro-infiltrated as detailed in **4.5.4** and **Table 4-1** (see **Section A.4** for details of these CYPs). Extracts of leaves from these co-infiltrations were analysed by GC- and LC-MS (see **Section 2.4.1** and **2.4.3**). Representative spectra are shown in **Figure 4-6, Figure 4-7,** and **Figure 4-8**.

Table 4-1: Combinations of vectors used to combinatorially evaluate CYP106A1 and CYP106A2

tHMGR, SAD1	tHMGR, SAD1, CYP106A1	tHMGR, SAD1, CYP106A2	tHMGR, SAD1, CYP716A14
tHMGR, SAD1, CYP716A141	tHMGR, SAD1, CYP716E26	tHMGR, SAD1, CYP72A65	tHMGR, SAD1, CYP72A69
tHMGR, SAD1, CYP88D6	tHMGR, SAD1, CYP93E1	tHMGR, SAD1, CYP94D65	tHMGR, SAD1, CYP716A12
tHMGR, SAD1, CYP716A12, CYP716C11	tHMGR, SAD1, CYP716A12, CYP72A67	tHMGR, SAD1, CYP106A1, CYP716A14	tHMGR, SAD1, CYP106A1, CYP716A141
tHMGR, SAD1, CYP106A1, CYP716E26	tHMGR, SAD1, CYP106A1, CYP72A65	tHMGR, SAD1, CYP106A1, CYP72A69	tHMGR, SAD1, CYP106A1, CYP88D6
tHMGR, SAD1, CYP106A1, CYP93E1	tHMGR, SAD1, CYP106A1, CYP94D65	tHMGR, SAD1, CYP106A1, CYP716A12	tHMGR, SAD1, CYP106A1, CYP716A12, CYP716C11
tHMGR, SAD1, CYP106A1, CYP716A12, CYP72A67	tHMGR, SAD1, CYP106A2, CYP716A14	tHMGR, SAD1, CYP106A2, CYP716A141	tHMGR, SAD1, CYP106A2, CYP716E26
tHMGR, SAD1, CYP106A2, CYP72A65	tHMGR, SAD1, CYP106A2, CYP72A69	tHMGR, SAD1, CYP106A2, CYP88D6	tHMGR, SAD1, CYP106A2, CYP93E1
tHMGR, SAD1, CYP106A2, CYP94D65	tHMGR, SAD1, CYP106A2, CYP716A12	tHMGR, SAD1, CYP106A2, CYP716A12, CYP716C11	tHMGR, SAD1, CYP106A2, CYP716A12, CYP72A67

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

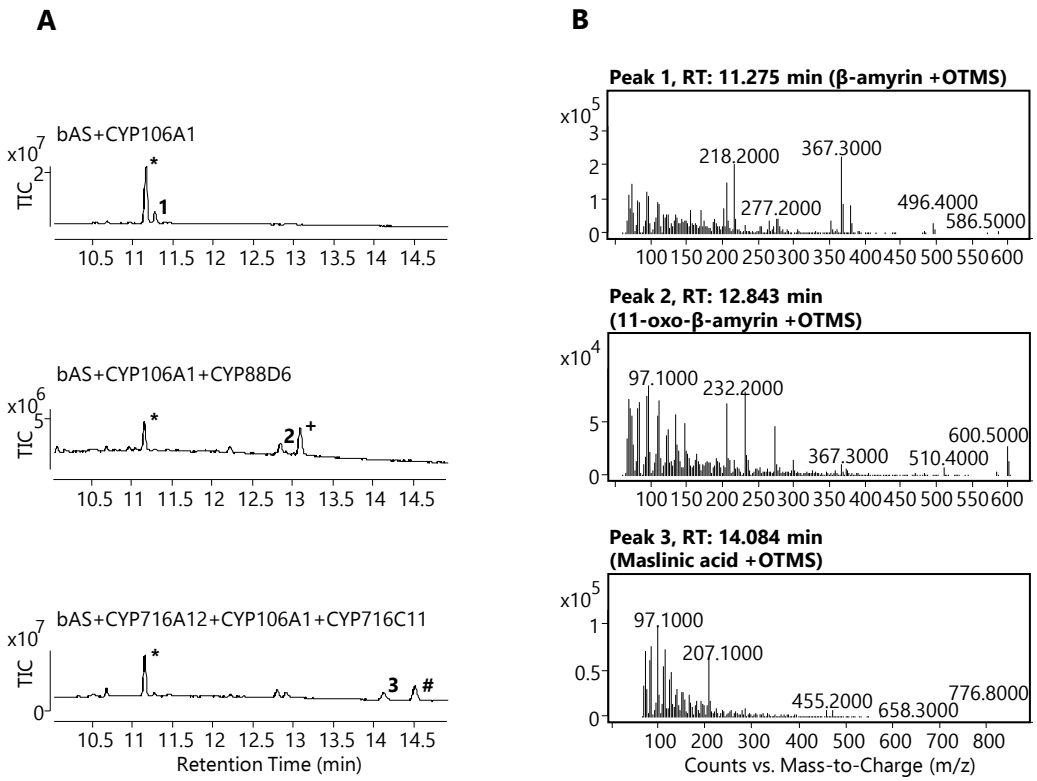


Figure 4-6: CYP106A1 produces new products in co-expression with toolkit enzymes. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP106A1 (top); tHMGR, SAD1, CYP106A1, and CYP88D6 (middle); and tHMGR, SAD1, CYP106A1, CYP716A12, and CYP716C11 (bottom). B) Mass spectra for the observed products at 11.3 (peak 1), 12.8 (peak 2), and 14.1 (peak 3) minutes. TIC = total ion chromatogram. * = β-amyrin (**1**), + = 11-oxo-β-amyrin (**2**), # = maslinic acid (**24**).

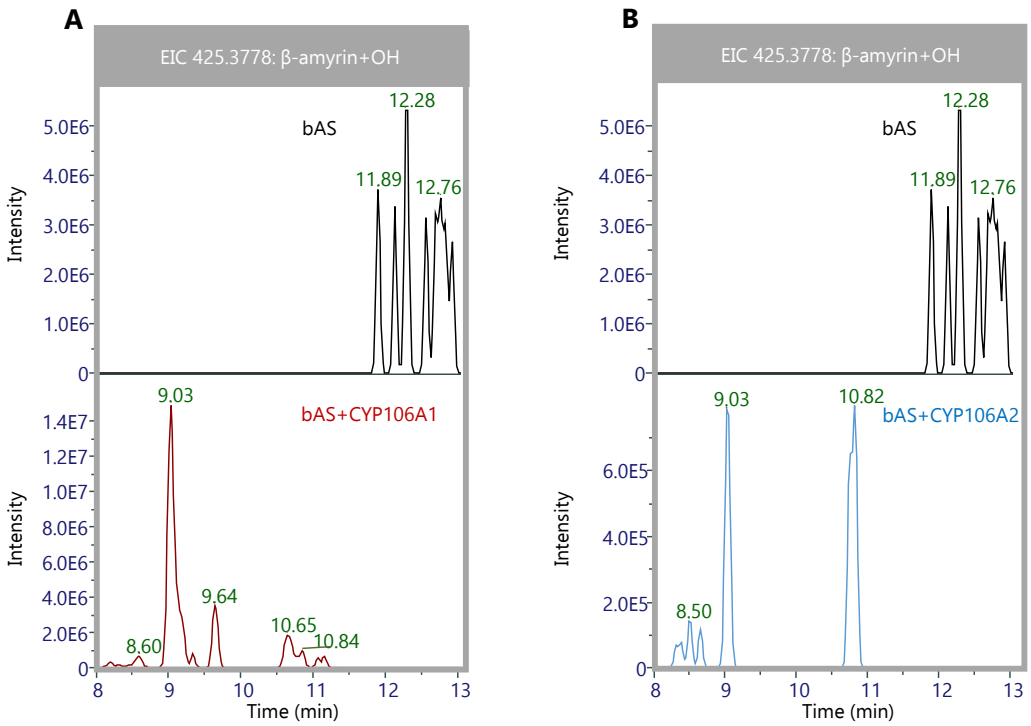


Figure 4-7: Co-expression of CYP106A1 or CYP106A2 and bAS shows two hydroxylated products by LC-MS. Extracted ion chromatograms (EIC) for m/z = 425.3378, corresponding to hydroxylated β-amyrin $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR and β-amyrin synthase (SAD1) as a control (black), and with addition of either A) CYP106A1 (red), or B) CYP106A2 (blue).

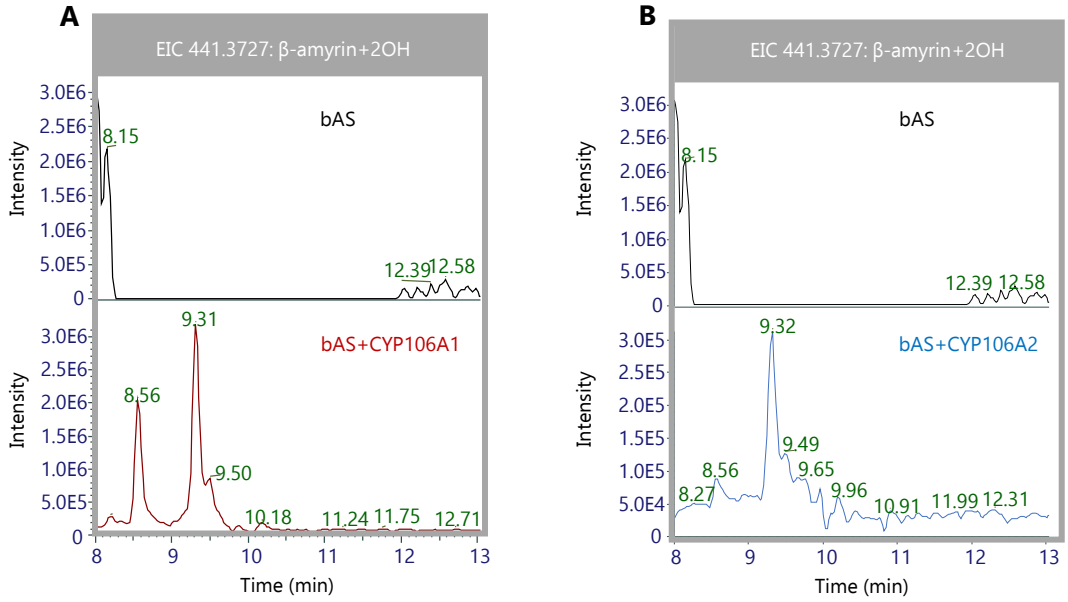


Figure 4-8: Co-expression of CYP106A1 or CYP106A2 and bAS shows a double hydroxylation product by LC-MS. Extracted ion chromatograms (EIC) for *m/z* = 441.3727, corresponding to twice-hydroxylated β-amyrin $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR and β-amyrin synthase (SAD1) as a control (black), and with addition of either A) CYP106A1 (red), or B) CYP106A2 (blue).

The spectra reveal that CYP106A1 oxidises the β-amyrin scaffold alone and in conjunction with nine other toolkit enzymes, with six of the combinations yielding visible products by GC-MS and the remainder yielding products visible by LC-MS (**Section A.3**). Analysis of the GC-MS fragmentation patterns suggests that the oxidation consists of the addition of a single hydroxyl-group to the A or B ring of the scaffold, as the unmodified *CDE ring retro Diels-Alder product can be seen with m/z = 218. The LC-MS traces specifically show evidence for a second oxidation, again the addition of a hydroxyl group, and the presence of an additional minor product consisting of a double oxidation. These other products are likely not visible in the GC-MS spectrum due to their low abundance relative to the first, major product. Based on the reported properties of this enzyme [141], it is likely that the major product is a C-7 hydroxylation, the minor single-oxidation product a C-15 hydroxylation, and the second minor product a double C-7 and C-15 hydroxylation.

No products were visible for CYP106A2 based on GC-MS. However, the LC-MS spectra reveal that CYP106A2 likely produces the same products as CYP106A1 but in relatively equal abundance, i.e. the two single hydroxylations and the double hydroxylation occur in roughly the same amounts. The overall levels of any one product are presumably too low to be readily detected by GC-MS, but are sufficient for detection by the more sensitive LC-MS method.

Hydroxylation of the C-7 position represents a new regiochemistry for the triterpene toolkit, C-7 having previously been considered one of the "orphan" positions. In addition to expanding our ability to selectively functionalise the triterpene scaffold, CYP106A1 is also the first example of the use of plant-based transient expression with bacterial CYPs for metabolic engineering, as well as their first use for the biosynthesis of triterpenes for pharmaceutical purposes (stable transgenic plants expressing bacterial CYPs have previously been reported in relation to bioremediation studies [52,145]). The scale-up of the major product of CYP106A1 for structural confirmation through NMR is carried out in **Chapter 5**.

4.3.4 Investigation of the Activity of CYP1A2, CYP2D6, and CYP3A4 Through Transient Expression

Another well-known source of promiscuous cytochromes P450 with varied activity is the human liver. A handful of human CYP450s are responsible for the oxidative metabolism of 70-80 % of drug molecules in clinical use [146], among them CYP1A2, CYP2D6, and CYP3A4. The genes encoding these enzymes were codon-optimised for *N. benthamiana* and then inserted into binary expression constructs as described in **4.5.5**. A mammalian co-enzyme, CYP450R from *Rattus norvegicus* was also codon-optimised, synthesised, and inserted into a binary expression construct, to investigate whether its co-expression with the human CYP450s would improve their activity. This was chosen due to its previous use as a fusion protein with CYP3A4 in *in vitro* studies of its activity [147].

The expression constructs were introduced into *A. tumefaciens*, and agro-infiltration was then carried out to evaluate the activities of CYP1A2, CYP2D6, and CYP3A4 towards pentacyclic triterpenes. Each of these CYPs was co-expressed with either β-amyrin synthase or the gene set needed for biosynthesis of oleanolic acid (**18**) (see **Section 2.3**). The gene combinations that were co-expressed are provided in **4.5.5.** GC-MS spectra of the resulting leaf extracts are provided in **Figure 4-9** and **Figure 4-10**.

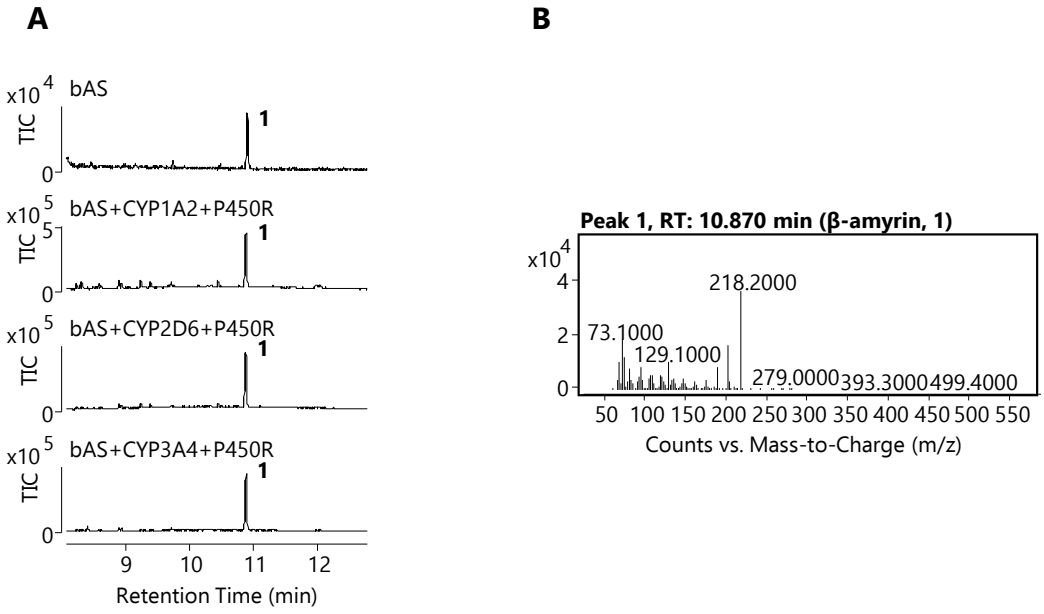


Figure 4-9: Co-expression of human CYPs and bAS shows no new products by GC-MS. A) GC-MS traces showing tHMGR, and β-amyrin synthase (SAD1) only control (first trace), and in combination with; CYP1A2 and CYP450R (second trace), CYP2D6 and CYP450R (third trace), and CYP3A4 and CYP450R (fourth trace) B) Mass spectrum of the peak at 10.9 minutes (peak 1). TIC = total ion chromatogram.

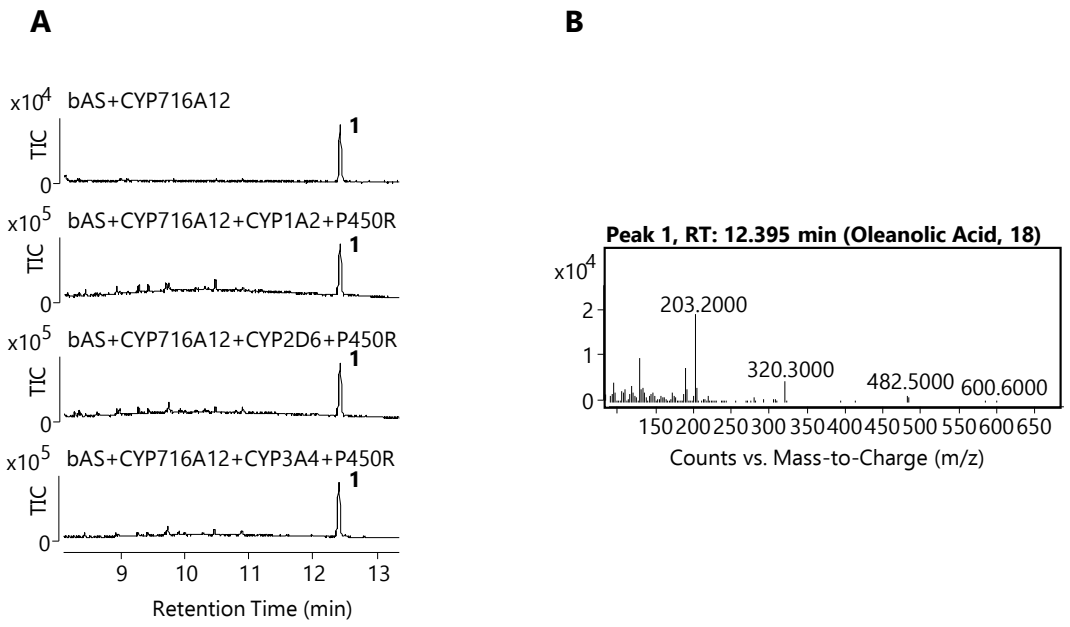


Figure 4-10: Co-expression of human CYPs and CYP716A12 shows no new products by GC-MS. A) GC-MS traces showing tHMGR, β-amyrin synthase (SAD1), and CYP716A12 only control (first trace), and in combination with; CYP1A2 and CYP450R (second trace), CYP2D6 and CYP450R (third trace), and CYP3A4 and CYP450R (fourth trace) B) Mass spectrum of the peak at 12.4 minutes (peak 1). TIC = total ion chromatogram.

As shown above, no additional products were detected when the mammalian enzymes were co-expressed with β-amyrin synthase with or without the addition of the C-28 oxidase CYP716A12. This could be because the mammalian enzymes are not active towards beta-amyrin (**1**) or oleanolic acid (**18**). Alternatively, it could be due to problems with expression, folding, localisation, and/or lack of appropriate cofactors (reductases). However, mammalian CYPs CYP1A2, CYP2D6, and CYP3A4 have previously been successfully expressed in functional form in plant-based expression systems including *N. benthamiana* [148-150].

4.3.5 Evaluation of CYP1A2, CYP2D6, and CYP3A4 Through Co-expression with Existing Toolkit Enzymes

In order to gain a more comprehensive understanding of the activity of CYP1A2, CYP2D6, and CYP3A4 towards triterpenes, these enzymes were co-expressed with

other CYP450s in the TTK, namely CYP716A12, CYP716A14v2, CYP716A141, CYP716C11, CYP716E26, CYP72A65, CYP72A67, CYP72A69, CYP88D6, CYP93E1, and CYP94D65, as detailed in **4.5.6** and **Table 4-2**. Examples of the resulting GC-MS and LC-MS spectra are shown in **Figure 4-11**, **Figure 4-12**, **Figure 4-13**, **Figure 4-14**, and **Figure 4-15.**

Table 4-2: Combinations of vectors used to combinatorially evaluate CYP1A2, CYP2D6, and CYP3A4

tHMGR, SAD1	tHMGR, SAD1, CYP1A2	tHMGR, SAD1, CYP2D6	tHMGR, SAD1, CYP3A4
tHMGR, SAD1, CYP716A14	tHMGR, SAD1, CYP716A141	tHMGR, SAD1, CYP716E26	tHMGR, SAD1, CYP72A65
tHMGR, SAD1, CYP72A69	tHMGR, SAD1, CYP88D6	tHMGR, SAD1, CYP93E1	tHMGR, SAD1, CYP94D65
tHMGR, SAD1, CYP716A12	tHMGR, SAD1, CYP716A12, CYP716C11	tHMGR, SAD1, CYP716A12, CYP72A69	tHMGR, SAD1, CYP1A2, CYP716A14
tHMGR, SAD1, CYP1A2, CYP716A141	tHMGR, SAD1, CYP1A2, CYP716E26	tHMGR, SAD1, CYP1A2, CYP72A65	tHMGR, SAD1, CYP1A2, CYP72A69
tHMGR, SAD1, CYP1A2, CYP88D6	tHMGR, SAD1, CYP1A2, CYP93E1	tHMGR, SAD1, CYP1A2, CYP94D65	tHMGR, SAD1, CYP1A2, CYP716A12
tHMGR, SAD1, CYP1A2, CYP716A12, CYP716C11	tHMGR, SAD1, CYP1A2, CYP716A12, CYP72A67	tHMGR, SAD1, CYP2D6, CYP716A14	tHMGR, SAD1, CYP2D6, CYP716A141
tHMGR, SAD1, CYP2D6, CYP716E26	tHMGR, SAD1, CYP2D6, CYP72A65	tHMGR, SAD1, CYP2D6, CYP72A69	tHMGR, SAD1, CYP2D6, CYP88D6
tHMGR, SAD1, CYP2D6, CYP93E1	tHMGR, SAD1, CYP2D6, CYP94D65	tHMGR, SAD1, CYP2D6, CYP716A12	tHMGR, SAD1, CYP2D6, CYP716A12, CYP716C11
tHMGR, SAD1, CYP2D6, CYP716A12, CYP72A67	tHMGR, SAD1, CYP3A4, CYP716A14	tHMGR, SAD1, CYP3A4, CYP716A141	tHMGR, SAD1, CYP3A4, CYP716E26
tHMGR, SAD1, CYP3A4, CYP72A65	tHMGR, SAD1, CYP3A4, CYP72A69	tHMGR, SAD1, CYP3A4, CYP88D6	tHMGR, SAD1, CYP3A4, CYP93E1
tHMGR, SAD1, CYP3A4, CYP94D65	tHMGR, SAD1, CYP3A4, CYP716A12	tHMGR, SAD1, CYP3A4, CYP716A12, CYP716C11	tHMGR, SAD1, CYP3A4, CYP72A67

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

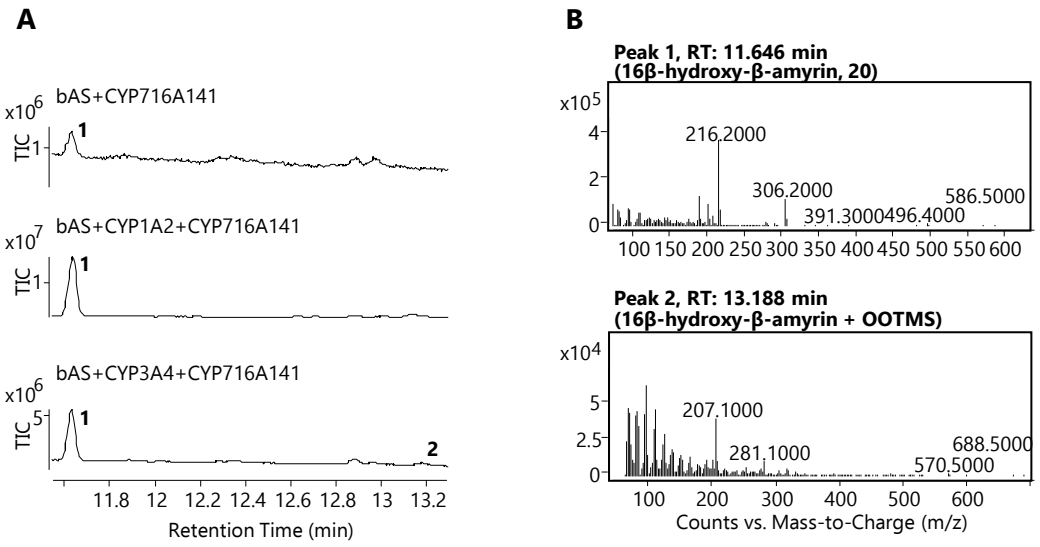


Figure 4-11: CYP3A4 and CYP716A141 produce new products in co-expression. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP716A141 (top); tHMGR, SAD1, CYP716A141, and CYP1A2 (middle); and tHMGR, SAD1, CYP716A141, and CYP3A4 (bottom). B) Mass spectra for the observed products at 11.6 (peak 1) and 13.2 (peak 2) minutes. TIC = total ion chromatogram.

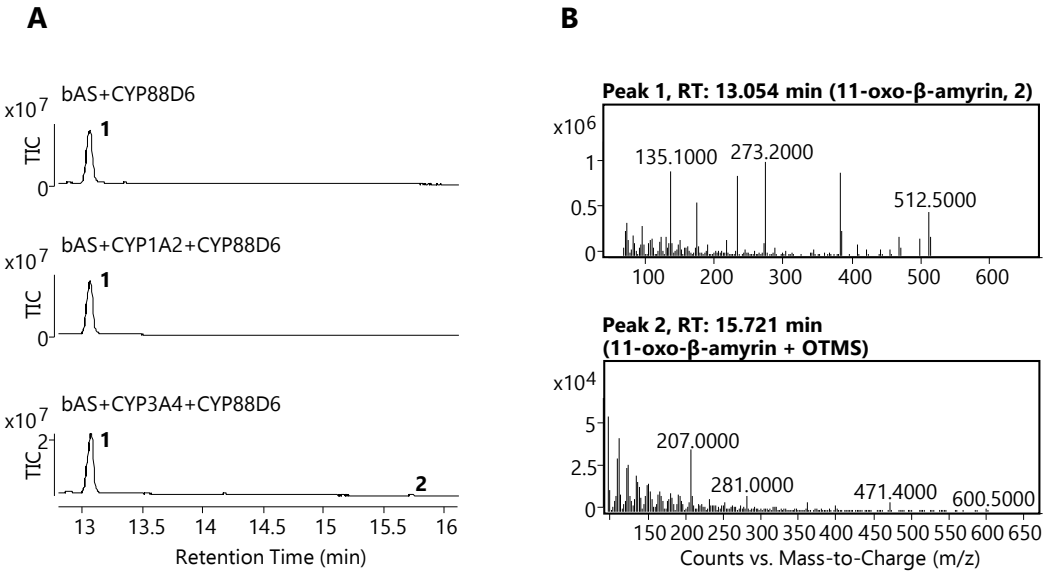


Figure 4-12: CYP3A4 and CYP88D6 produce new products in co-expression. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP88D61 (top); tHMGR, SAD1, CYP88D6, and CYP1A2 (middle); and tHMGR, SAD1, CYP88D6, and CYP3A4 (bottom). B) Mass spectra for the observed products at 13.1 (peak 1) and 15.7 (peak 2) minutes. TIC = total ion chromatogram.

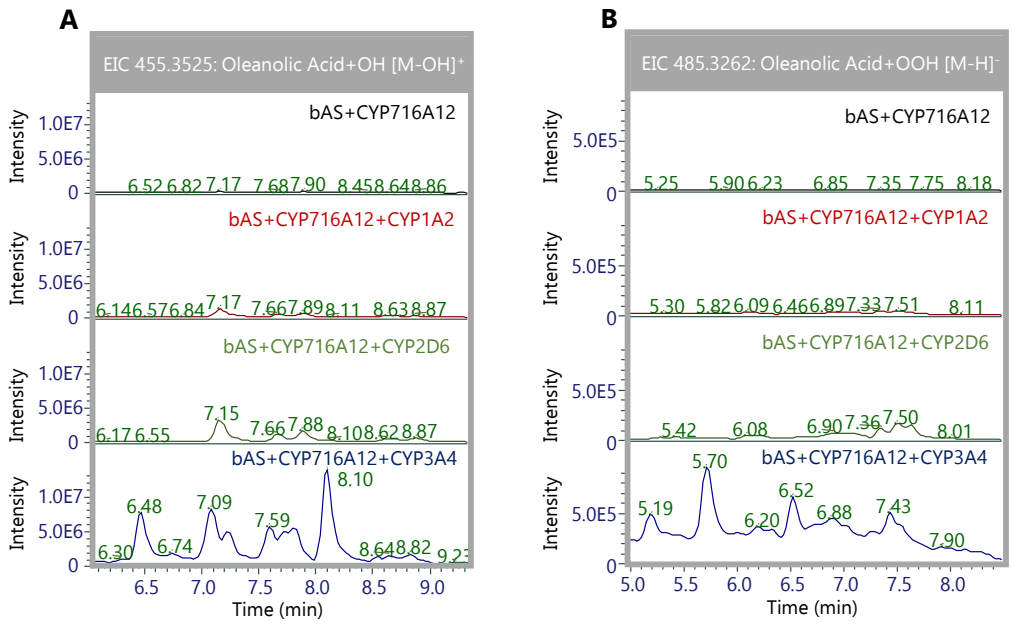


Figure 4-13: CYP1A2 and CYP716A12, CYP2D6 and CYP716A12, or CYP3A4 and CYP716A12 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 455.3525, corresponding to hydroxylated oleanolic acid (**18**) $[M\text{-}OH]^+$, or B) *m/z* = 485.3262, corresponding to oleanolic acid (**18**) plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP716A12 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

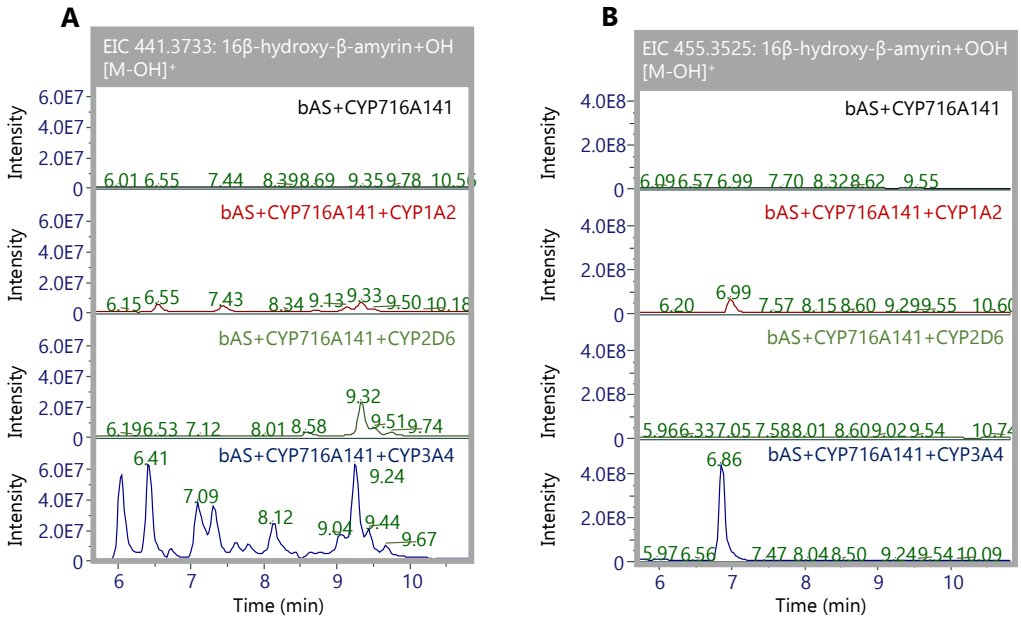


Figure 4-14: CYP1A2 and CYP716A141, CYP2D6 and CYP716A141, or CYP3A4 and CYP716A141 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3733, corresponding to hydroxylated 16β-hydroxy-β-amyrin (**20**) $[M\text{-}OH]^+$, or B) *m/z* = 455.3525, corresponding to 16β-hydroxy-β-amyrin (**20**) plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase

(SAD1), and CYP716A141 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

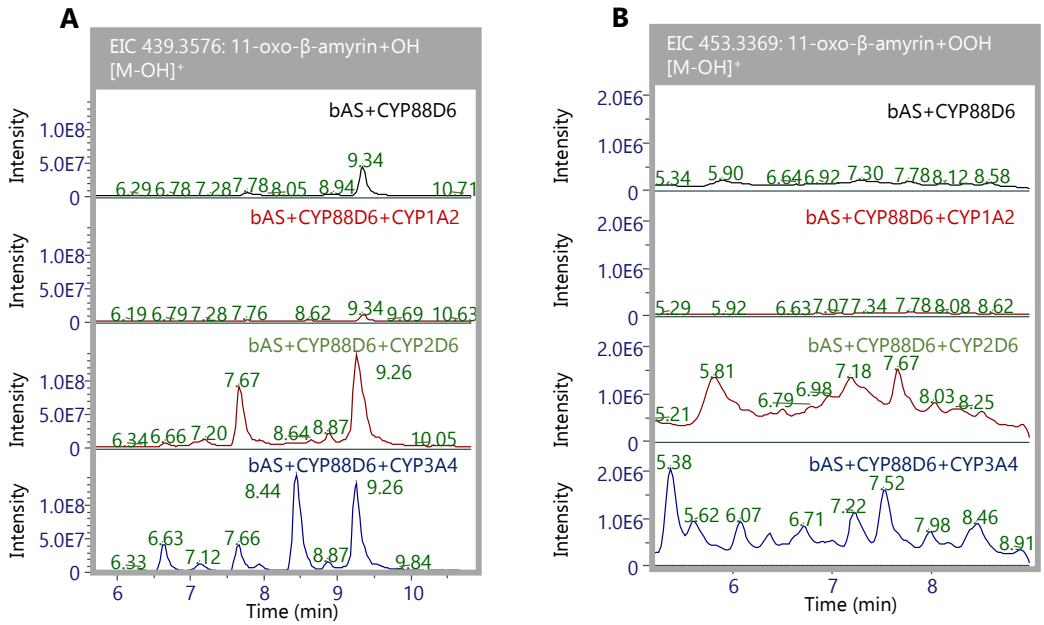


Figure 4-15: CYP2D6 and CYP88D6 or CYP3A4 and CYP88D6 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 439.3576, corresponding to hydroxylated 11-oxo-β-amyrin (**2**) $[M\text{-}OH]^+$, or B) *m/z* = 453.3369, corresponding to 11-oxo-β-amyrin (**2**) plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP88D6 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

The co-expression of β-amyrin synthase, CYP716A141, and CYP3A4 shows a new peak at 13.2 minutes (see **Figure 4-11**) with *m/z* = 688, consistent with the mass of 16β-hydroxy-β-amyrin plus a carboxylic acid group. Meanwhile, the co-expression of β-amyrin synthase, CYP88D6, and CYP3A4 shows a new peak at 15.7 minutes with *m/z* = 600, consistent with 11-oxo-β-amyrin plus a trimethylsilyated hydroxyl group (OTMS) (see **Figure 4-12**).

In the LC-MS results, co-expression of β-amyrin synthase, CYP716A12, and either CYP1A2 or CYP2D6 leads to the formation of an additional peak at 7.2 min with a mass consistent with the addition of a single hydroxyl group, while co-expression of β-amyrin synthase, CYP716A12, and CYP3A4 leads to the formation of additional peaks at 6.5 and 8.1 min with a mass consistent with the addition of a single hydroxyl group (see **Figure 4-13A**). The co-expression of β-amyrin synthase, CYP716A12, and CYP3A4 additionally leads to the formation of peaks at 5.2, 5.7, and

6.5 min with a mass consistent with the addition of a carboxylic acid group (**Figure 4-13B**).

Co-expression of β-amyrin synthase, CYP716A141, and CYP1A2, meanwhile, led to the formation of new peaks at 6.6, 7.4, and 9.3 min with a mass consistent with 16β-hydroxy-β-amyrin plus a single hydroxyl group. The peak at 9.3 min is also formed with co-expression of CYP2D6, while co-expression of β-amyrin synthase, CYP716A141, and CYP3A4 leads to the formation of peaks at 6.4, 7.1, 8.1, and 9.2 min with mass consistent with the addition of a single hydroxyl group (**Figure 4-14A**). Co-expression of β-amyrin synthase with CYP716A141 and CYP1A2 also leads to the formation of a peak at 7.0 min consistent with 16β-hydroxy-β-amyrin plus a carboxylic acid group, while co-expression of β-amyrin synthase, CYP716A141, and CYP3A4 leads to the formation of a peak at 6.9 min with the same mass (**Figure 4-14B**).

Co-expression of β-amyrin synthase, CYP88D6, and CYP2D6 leads to the formation of new peaks at 7.7 and 9.3 min consistent with the mass of 11-oxo-β-amyrin plus a single hydroxyl group. The same two peaks are seen with co-expression of β-amyrin synthase, CYP88D6, and CYP3A4, as well as two additional peaks with the same mass at 6.6 and 8.4 min (**Figure 4-15A**). Co-expression of β-amyrin synthase, CYP88D6, and CYP2D6 additionally leads to the formation of peaks at 5.8 and 7.7 min consistent with 11-oxo-β-amyrin plus a carboxylic acid group, while co-expression of β-amyrin synthase, CYP88D6, and CYP3A4 leads to the formation of peaks with the same mass at 5.4, 6.1, 7.2, 7.5, and 8.5 min (**Figure 4-15**).

The spectra show that the human CYP450s are active towards oxidised β-amyrin based triterpenes rather than β-amyrin itself, with the different enzymes showing activity towards different substrates and producing different products. CYP3A4 is overall the most active of the tested enzymes, showing activity towards the greatest number of substrates and producing products in the highest quantity. Overall, 8 of the potential substrates showed some activity in combination with at least one of the tested enzymes. Notably, one combination which did not yield new products is that with CYP72A69, which produces 21β-hydroxy-β-amyrin (**23**).

The products generated by the new enzymes are varied but mostly consist of single hydroxylations at multiple positions, shown by the appearance of more than one new peak with a mass consistent with the addition of a single hydroxyl group in the LC-MS spectra. This is especially noticeable with CYP3A4 (see **Figure 4-13** and **Figure 4-14**) but can also be seen with the other two enzymes at lower intensity (in **Figure 4-15** this can be seen with CYP2D6, and in **Figure 4-14** for CYP1A2 there are multiple peaks visible at an intensity an order of magnitude lower than for CYP3A4). In addition, there are some new peaks suggesting the addition of a carboxylic acid group, especially in combination with CYP3A4. The pattern of oxidations seen is consistent with the known function of these CYPs as promiscuous detoxification enzymes.

In particular, the combination of CYP3A4 with CYP88D6, which oxidises β-amyrin (**1**) at the C-11 position, is a particularly well accumulated major product. This makes this combination the most suitable for scaling-up, especially as 11-oxo-β-amyrin (**2**), CYP88D6's primary product, is known to have bioactivity in cell-based assays [50].

4.3.6 THAR1 and THAR2 Can Act Together to Produce an Unusual C-3α Hydroxyl Moiety

In addition to CYPs, other classes of enzymes with unusual or novel activities were also evaluated for their abilities to modify β-amyrin and oleanolic acid. These include two recently reported triterpene alcohol dehydrogenases, THAR1 and THAR2, from *Arabidopsis thaliana,* originally reported for their activity towards the tricyclic triterpene scaffold thalianol [151]. THAR1 oxidises the C-3 alcohol in the triterpenes to a ketone, while THAR2 reduces this ketone back to an alcohol, but with the alternate α-stereochemistry [151]. It had been reported that these enzymes showed promiscuous activity towards triterpene scaffolds, with activity also reported towards arabidiol and the tetracyclic triterpene tirucalladienol [151], likely due to the similarity of most triterpenes in the A-ring, where these transformations occur. The activity of these enzymes towards pentacyclic scaffolds such as β-amyrin has not previously been reported.

CYP716A14v2 from *Artemisia annua* is also known to catalyse formation of β-amyrone (**27**), the C-3 ketone of β-amyrin [115]. It was therefore decided to investigate whether THAR2 can work together with CYP716A14v2 to catalyse the same transformation seen in conjunction with THAR1. This was tested with substrates with increasing numbers of oxidations. CYP88D6 and CYP93E1 were chosen for these co-expression studies, as these catalyse oxidations at positions found in 11-oxo-boswellic acid, a known bioactive molecule. Although boswellic acid has a carboxylic acid group at C-24, CYP93E1 produces mostly the alcohol. However, it was used as a corresponding C-24 acid oxidase is not present in the toolkit.

Therefore, co-expression was carried out with β-amyrin synthase and a series of additional enzymes as shown in **Table 4-8**. Extracts of the infiltrated leaves were processed for standard GC-MS analysis (see **Section 2.4.1**and **4.5.7**).

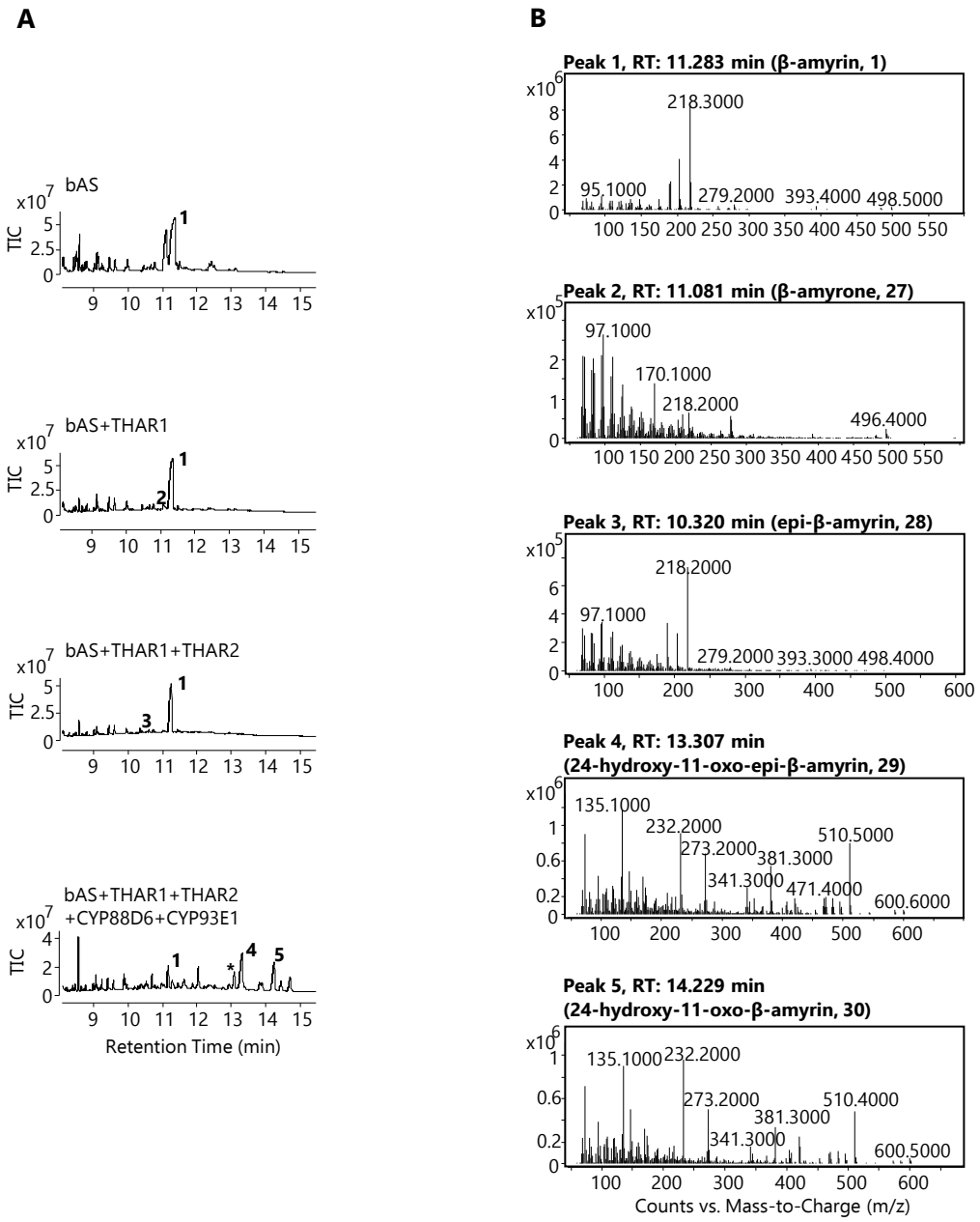


Figure 4-16: THAR1 and THAR2 work together to produce epi-β-amyrin (**28**) and derivatives. A) GC-MS traces showing tHMGR and β-amyrin synthase (SAD1) only control (first trace),and in combination with; THAR1 (second trace), THAR1 and THAR2 (third trace), and THAR1, THAR2, CYP88D6, and CYP93E1 (fourth trace) B) Mass spectra of the peaks at 11.3 (peak 1), 11.1 (peak 2), 10.3 (peak 3), 13.3 (peak 4), and 14.3 (peak 5) minutes. TIC = total ion chromatogram. * = 11-oxo-β-amyrin (**2**)

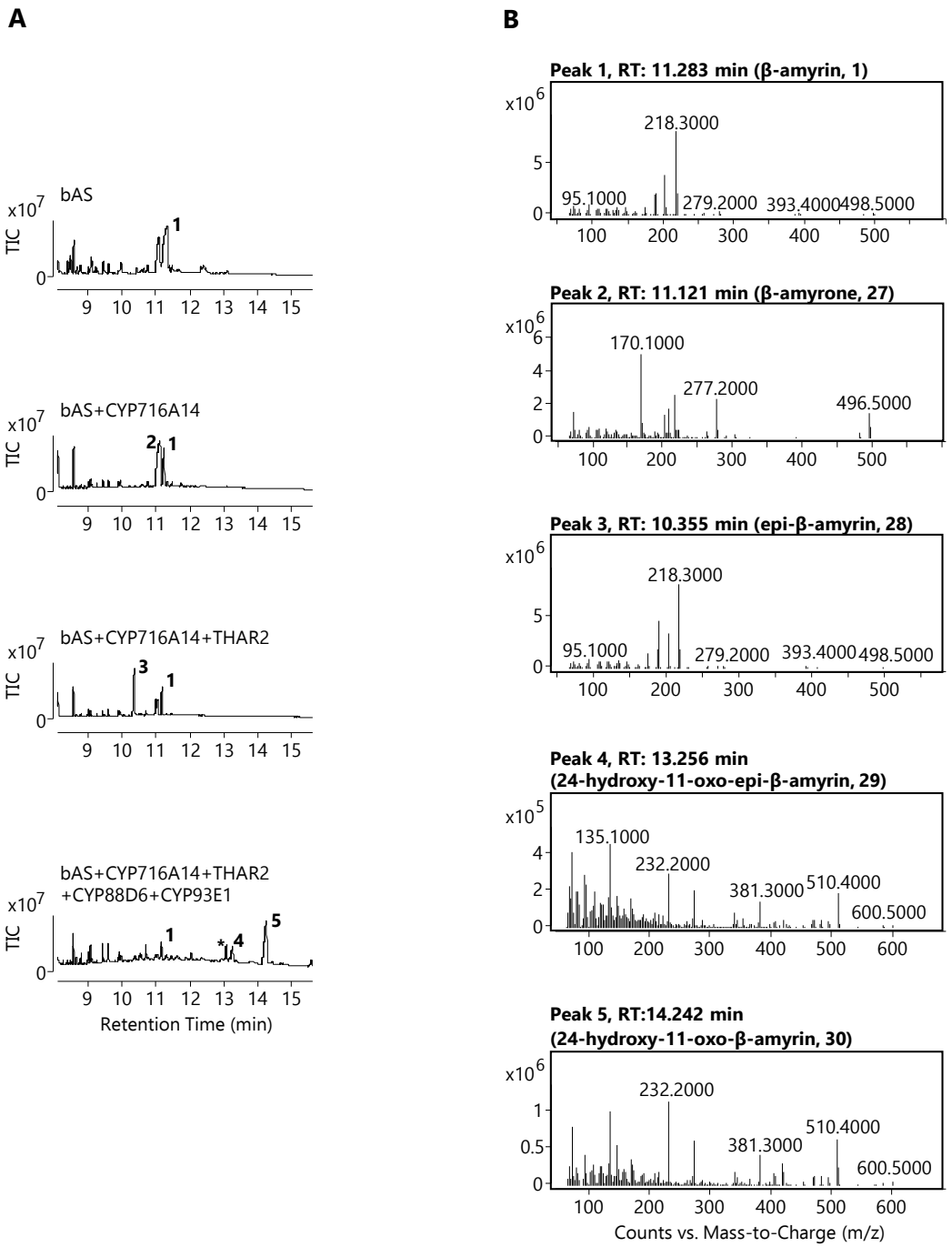


Figure 4-17: CYP716A14v2 and THAR2 work together to produce epi-β-amyrin (**28**) and derivatives. A) GC-MS traces showing tHMGR and β-amyrin synthase (SAD1) only control (first trace),and in combination with; CYP716A14v2 (second trace), CYP716A14v2 and THAR2 (third trace), and CYP716A14v2, THAR2, CYP88D6, and CYP93E1 (fourth trace) B) Mass spectra of the peaks at 11.3 (peak 1), 11.1 (peak 2), 10.3 (peak 3), 13.3 (peak 4), and 14.3 (peak 5) minutes. TIC = total ion chromatogram. * = 11-oxo-β-amyrin (**2**)

The spectra show that co-expression of either THAR1 or CYP716A14v2 with β-amyrin synthase causes the appearance of a new peak at 11.1 minutes with a *m/z* of 496, corresponding to β-amyrone (**27**). The additional co-expression of THAR2 leads

to the disappearance of the β-amyrone peak and the appearance of a new peak at 10.3 minutes, with an *m/z* of 498 and identical mass spectrum to β-amyrin but with a different retention time, suggesting that the new peak is epi-β-amyrin (**28**), the C-3α-OH epimer of β-amyrin. This is potentially exciting due to the C-3α moiety being present in bioactive molecules such as the boswellic acids. When combined with additional enzymes, more oxidised products are observed, such as 24-hydroxy-11-oxo-epi-β-amyrin (**29**) (13.3 min), showing that this process is compatible with combinatorial biosynthesis.

4.3.7 Investigation of THAR1 vs CYP716A14v2 Together with THAR2: A Biosynthetic Pathway to Boswellic Acid?

The results of the small-scale experiments showed differing amounts of products between CYP716A14v2 and THAR2. As future experiments would likely involve scaling up the products of these combinations, the relative efficiencies of each enzyme were investigated. Either THAR1 or CYP716A14v2 was co-expressed with THAR2 and the biosynthetic genes for β-amyrin (**1**), 11-oxo-β-amyrin (**2**), and 24-hydroxy-11-oxo-β-amyrin (**30**) (see **Table 4-9**) to evaluate the relative amounts of C-3β and C-3α products produced by each enzyme co-expression.

Combinations of *A. tumefaciens* strains harbouring the relevant constructs were made as shown in section **4.5.7** and used to agro-infiltrate *N. benthamiana*. Infiltrated leaves were processed for GC-MS analysis in a quantitative manner (see **4.5.7**), with the relative size of product peaks compared to an internal coprostanol standard. Representative GC-MS spectra are shown in **Figure 4-18**, **Figure 4-19**, and **Figure 4-20**, and a table of analysis is given in **Table 4-3**.

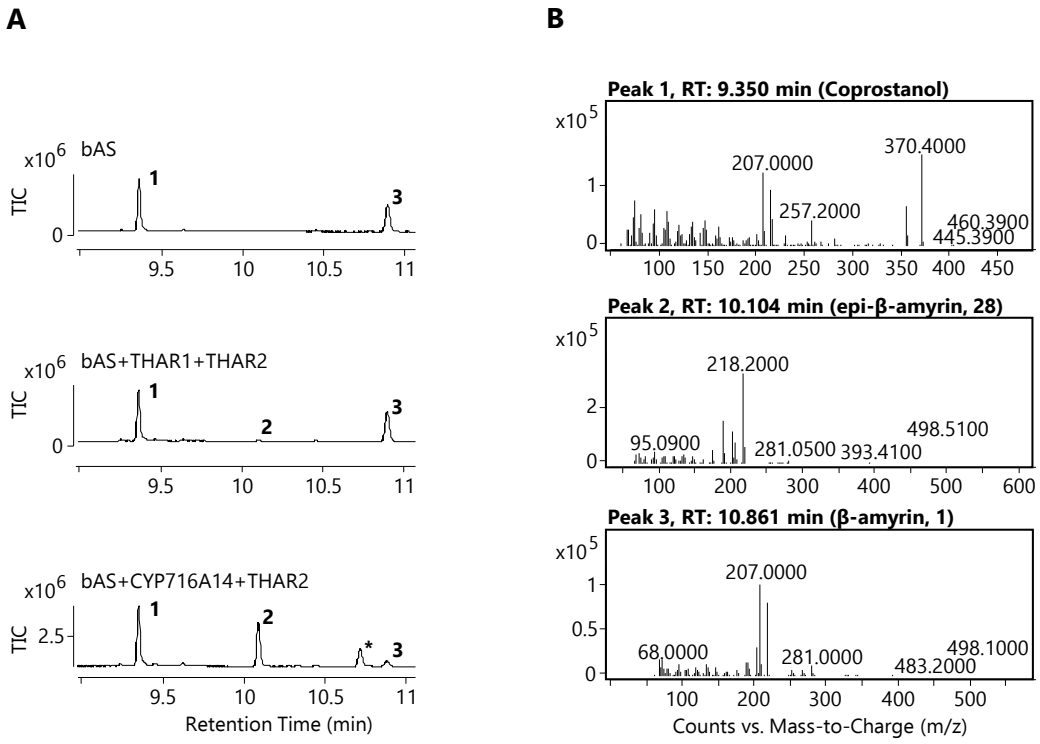


Figure 4-18: Example from one replicate of the quantitative analysis for β-amyrin. A) Example GC-MS traces of one replicate of the quantitative analysis showing tHMGR and β-amyrin synthase (SAD1) only (top) and in combination with THAR1 and THAR2 (middle) and CYP716A14v2 combined with THAR2 (bottom). B) Mass spectra of the peaks at 9.4 (peak 1), 10.1 (peak 2), 10.9 (peak 3) minutes. TIC = total ion chromatogram. * = β-amyrone (**27**).

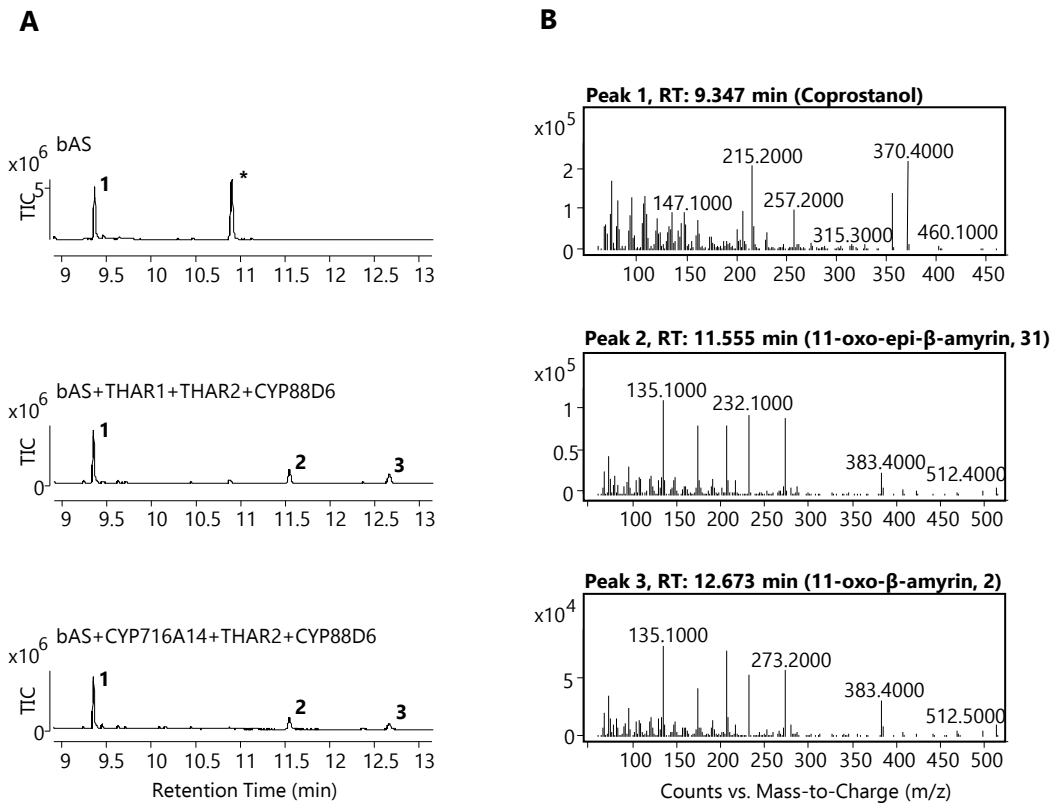


Figure 4-19: Example from one replicate of the quantitative analysis for 11-oxo-β-amyrin. A) Example GC-MS traces of one replicate of the quantitative analysis showing tHMGR and β-amyrin synthase (SAD1) only (top) and in combination with THAR1, THAR2, and CYP88D6 (middle) and CYP716A14v2, THAR2, and CYP88D6 (bottom). B) Mass spectra of the peaks at 9.3 (peak 1), 11.6 (peak 2), 12.7 (peak 3) minutes. TIC = total ion chromatogram. * = β-amyrin (**1**).

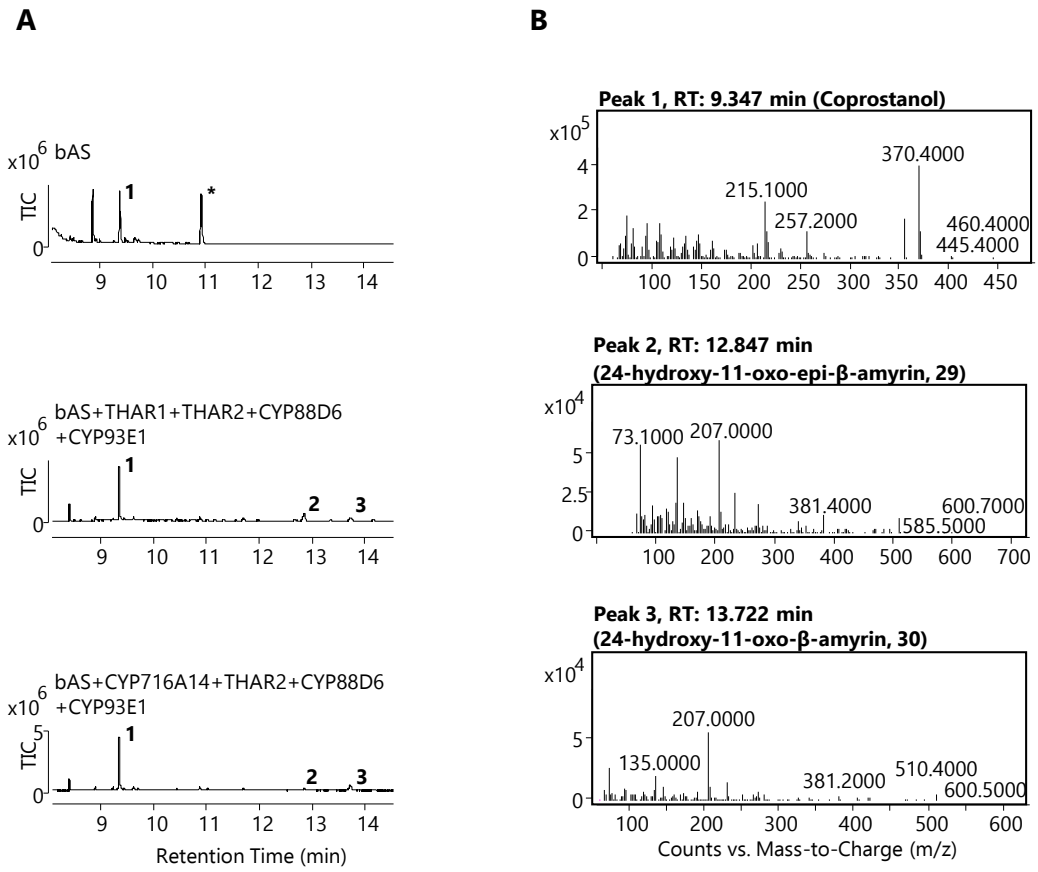


Figure 4-20: Example from one replicate of the quantitative analysis for 24-hydroxy-11-oxo-β-amyrin. A) Example GC-MS traces of one replicate of the quantitative analysis showing tHMGR and β-amyrin synthase (SAD1) only (top) and in combination with THAR1, THAR2, CYP88D6, and CYP93E1 (middle) and CYP716A14v2, THAR2, CYP88D6, and CYP93E1 (bottom). B) Mass spectra of the peaks at 9.3 (peak 1), 12.8 (peak 2), 13.7 (peak 3) minutes. TIC = total ion chromatogram. * = β-amyrin (**1**).

***Table 4-3*:** Peak area of triterpene compounds relative to coprostanol internal standard

Compound	**Relative Peak Area**	
	THAR1	**CYP716A14v2**
β-amyrin (**1**)	0.87 ± 0.07	0.10 ± 0.03
epi-β-amyrin (**28**)	0.03 ± 0.01	1.05 ± 0.31
11-oxo-β-amyrin (**2**)	0.23 ± 0.14	0.18 ± 0.04
11-oxo-epi-β-amyrin (**31**)	0.28 ± 0.17	0.32 ± 0.09
24-hydroxy-11-oxo-β-amyrin (**30**)	0.10 ± 0.02	0.18 ± 0.04
24-hydroxy-11-oxo-epi-β-amyrin (**29**)	0.27 ± 0.03	0.17 ± 0.01

Data represent mean ± standard deviation of 3 biological replicates.

The spectra and table show an increase in the efficiency of CYP716A14v2 over THAR1 at producing larger amounts of the desired product when the scaffold contains fewer oxidations, and an increased efficiency for THAR1 when scaffolds are

more heavily oxidised. This will be important information for future co-expressions of enzymes from the triterpene toolkit where a C-3α configuration is desired.

4.4 Conclusions

Here the triterpene toolkit has been expanded through the incorporation of multiple CYP450s, one from a plant source, two from a bacterial source, and three from a human source. These CYP450s have been evaluated for their activity towards the pentacyclic triterpenes β-amyrin and oleanolic acid along with potential redox partners. These experiments demonstrate that the *N. benthamiana* chassis is capable of supporting the expression of CYP450s from bacterial and animal sources in functional form, and that these enzymes can be integrated into biosynthetic pathways with other plant enzymes from the toolkit. This will be of benefit for future toolkit expansion and use as it suggests that specific co-enzymes and redox partners do not need to be co-expressed with CYPs from the native sources in order for them to be effective.

Putative structures for several of the products generated in co-expression experiments that include these enzymes have been determined by GC- and LC-MS analysis. These products included maslinic acid (**24**), which contains a 2α-hydroxylation previously unattainable using the original enzyme toolkit; a structure putatively identified as 7β-hydroxy-β-amyrin (**32**), representing oxidation of a previously "orphan" position of the β-amyrin scaffold; and a hydroxylation of the bioactive triterpene 11-oxo-β-amyrin (**2**). These experiments have thus significantly expanded the breadth of the triterpene toolkit and provide proof-of-concept for its future expansion by drawing on enzymes from both plant and non-plant sources. The putative structures will be confirmed through scale-up, purification, and NMR analysis in **Chapter 5**.

In addition to new CYPs, two alcohol dehydrogenase enzymes, THAR1 and THAR2, were also evaluated and shown to alter the stereochemistry of β-amyrin (**1**), producing C-3α hydroxyl derivatives that bear structural resemblance to the naturally occurring and biologically relevant boswellic acids found in *Boswellia* species. A semi-quantitative evaluation of the THAR1 enzyme in competition with

the equivalently acting CYP716A14v2 is also detailed, showing that the specific conditions of combinatorial biosynthesis can favour differing combinations of enzymes. This will be important for future combinatorial efforts, especially when maximum yield is required. These interesting derivatives will be purified at preparative scale in **Chapter 5**, and their biological properties will be investigated in **Chapter 6**.

The number and variety of potential CYPs now present in the toolkit unlocks the further chemical potential of the β-amyrin scaffold by presenting suitable handles for further modification by other biosynthetic enzymes. The potential exists for further expanding the diversity of the toolkit, accessing the full range of potential modifications including "orphan positions" as well as moving beyond CYPs and into a larger variety of biosynthetic enzymes in future. These include previously unmentioned classes of tailoring enzymes such as glucosyl- and acyl- transferases which will be instrumental for bioengineering structurally diverse triterpenes for evaluation of bioactivities, a strategy that could eventually enable the generation of designer triterpenes optimised for desirable therapeutic properties.

4.5 Materials and Methods for Chapter 4

4.5.1 Procedure for Production of the CYP716C11 Agrobacterium Strain

Gene fragments for the known P450 CYP716C11 were purchased from IDT and prepared according to the manufacturer's instructions. Entry cloning into pDONR207 using the gene fragments was performed according to the standard protocol (see **Section 2.1**), and plasmids were used to transform chemically competent *E coli* according to the standard protocol (see **Section 2.2.1**), with positive colonies selected for.

8 clearly separated colonies of *E. coli* were sampled and used as a template for PCR using the standard colony PCR procedure (see **Section 2.1.3**).

Following PCR, DNA was analysed using the standard DNA electrophoresis procedure (see **Section 2.1.4**). Positive colonies were identified as those with a band between the 1.5 and 2.0 kb markers in the ladder.

Positive colonies were cultured, and the plasmids extracted and sent for sequencing according to the standard procedure (see **Section 2.1.5**). Two of the 8 colonies showed the correct gene sequence.

Plasmids from one of the colonies with the correct sequence was used to generate destination vectors according to the standard procedure (see **Section 2.1.2**) then the vectors used to transform chemically competent *E. coli* (see **Section 2.2.1**). Positive colonies were selected for and cultured, and the plasmids recovered. Plasmids were then used to transform chemically competent *A. tumefaciens* (LBA4404) according to the standard procedure (see **Section 2.2.2**).

Positive colonies were selected for and cultured according to the standard procedure (see **Section 2.3**). Manual infiltration of *N. benthamiana* was then carried out to validate the strain.

4.5.2 Procedure for Production of the CYP106A1, CYP106A2, and BmFdx2 Agrobacterium Strains

Gene fragments for the known P450s CYP106A1, and CYP106A2 as well as the ferredoxin BmFdx2 were purchased from IDT and prepared according to the manufacturer's instructions. Entry cloning into pDONR207 using the gene fragments was performed according to the standard protocol (see **Section 2.1**), and plasmids were used to transform chemically competent *E coli* according to the standard protocol (see **Section 2.2.1**), with positive colonies selected for.

8 clearly separated colonies of *E. coli* per gene were sampled and used as a template for PCR using the standard colony PCR procedure (see **Section 2.1.3**).

Following PCR, DNA was analysed using the standard DNA electrophoresis procedure (see **Section 2.1.4**). Positive colonies were identified as those with a band between the 1.5 and 2.0 kb markers in the ladder.

Positive colonies were cultured, and the plasmids extracted and sent for sequencing according to the standard procedure (see **Section 2.1.5**). Two of the four colonies of CYP106A1, one of the four colonies of CYP106A2, and all of the four colonies of BmFdx2 showed the correct gene sequence.

Plasmids isolated from one of the colonies of each candidate gene with the correct sequence were used to generate destination vectors according to the standard procedure (see **Section 2.1.2**), then the vectors used to transform chemically competent *E. coli* (see **Section 2.2.1**). Positive colonies were selected for and cultured, and the plasmids recovered. Plasmids were then used to transform chemically competent *A. tumefaciens* (LBA4404) according to the standard procedure (see **Section 2.2.2**).

Positive colonies were selected for and cultured according to the standard procedure (see **Section 2.3**). Manual infiltration of *N. benthamiana* was then carried out to validate the strains using the combinations in **Table 4-4**. GC-MS analysis was performed using the 15-minute method.

4.5.3 Procedure for Production of Binary Expression Constructs for Arh1, BphA4, and FDR

Gene fragments for the ferredoxin reductases Arh1, BphA4, and FDR were purchased from IDT and prepared according to the manufacturer's instructions. Entry cloning into pDONR207 using the gene fragments was performed according to the standard protocol (see **Section 2.1**), and plasmids were used to transform chemically competent *E coli* according to the standard protocol (see **Section 2.2.1**), with positive colonies selected for.

8 clearly separated colonies of *E. coli* per gene were sampled and used as a template for PCR using the standard colony PCR procedure (see **Section 2.1.3**).

Following PCR, DNA was analysed using the standard DNA electrophoresis procedure (see **Section 2.1.4**). Positive colonies were identified as those with a band between the 1.5 and 2.0 kb markers in the ladder (Arh1, BphA4), or between the 0.7-0.8 kb markers (FDR).

Positive colonies were cultured, and the plasmids extracted and sent for sequencing according to the standard procedure (see **Section 2.1.5**). One of the four colonies of Arh1, one of the three colonies of BphA4, and one colony of FDR showed the correct gene sequence.

Plasmids isolated from one of the colonies of each candidate gene with the correct sequence were used to generate destination vectors according to the standard procedure (see **Section 2.1.2**), then the vectors used to transform chemically competent *E. coli* (see **Section 2.2.1**). Positive colonies were selected for and cultured, and the plasmids recovered. Plasmids were then used to transform chemically competent *A. tumefaciens* (LBA4404) according to the standard procedure (see **Section 2.2.2**).

Positive colonies were selected for and cultured according to the standard procedure (see **Section 2.3**). Manual infiltration of *N. benthamiana* was then carried out to validate the strains using the combinations in **Table 4-4.**

Table 4-4: Combinations of vectors used to evaluate CYP106A1 and CYP106A2

tHMGR, SAD1	tHMGR, SAD1, CYP716A12	tHMGR, SAD1, CYP106A1	tHMGR, SAD1, CYP106A1, BmFdx2, Arh1
tHMGR, SAD1, BmFdx2, BphA4	tHMGR, SAD1, BmFdx2, FDR	tHMGR, SAD1, CYP716A12, CYP106A1	tHMGR, SAD1, CYP716A12, CYP106A1, BmFdx2, Arh1
tHMGR, SAD1, CYP716A12, CYP106A1, BmFdx2, BphA4	tHMGR, SAD1, CYP716A12, CYP106A1, BmFdx2, FDR	tHMGR, SAD1, CYP106A2	tHMGR, SAD1, CYP106A2, BmFdx2, Arh1
tHMGR, SAD1, CYP106A2, BmFdx2, BphA4	tHMGR, SAD1, CYP106A2, BmFdx2, FDR	tHMGR, SAD1, CYP716A12, CYP106A2	tHMGR, SAD1, CYP716A12, CYP106A2, BmFdx2, Arh1
tHMGR, SAD1, CYP716A12, CYP106A2, BmFdx2, BphA4		tHMGR, SAD1, CYP716A12, CYP106A2, BmFdx2, FDR	

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

4.5.4 Procedure for Combinatorial Testing of CYP106A1, CYP106A2, and Toolkit Enzymes

A. tumefaciens strains containing tHMGR, SAD1, CYP716A12, CYP716A14v2, CYP716A141, CYP716C11, CYP716E26, CYP72A65, CYP72A67 CYP72A69, CYP88D6, CYP93E1, CYP94D65, CYP106A1, and CYP106A2 from the triterpene toolkit were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined as shown in **Table 4-5** and used to infiltrate 1 leaf each of 5-week-old *N. benthamania* plants using the manual infiltration method (see **Section 2.3.2**). Plants were grown for 5 days under greenhouse conditions before leaves were harvested and processed according to the standard procedure for GC-MS and LC-MS analysis (see **Section 2.4.1**). GC-MS was performed using the 20-minute method.

Table 4-5: Combinations of vectors used to combinatorially evaluate CYP106A1 and CYP106A2 (table is identical to **Table 4-1**)

tHMGR, SAD1	tHMGR, SAD1, CYP106A1	tHMGR, SAD1, CYP106A2	tHMGR, SAD1, CYP716A14
tHMGR, SAD1, CYP716A141	tHMGR, SAD1, CYP716E26	tHMGR, SAD1, CYP72A65	tHMGR, SAD1, CYP72A69
tHMGR, SAD1, CYP88D6	tHMGR, SAD1, CYP93E1	tHMGR, SAD1, CYP94D65	tHMGR, SAD1, CYP716A12
tHMGR, SAD1, CYP716A12, CYP716C11	tHMGR, SAD1, CYP716A12, CYP72A67	tHMGR, SAD1, CYP106A1, CYP716A14	tHMGR, SAD1, CYP106A1, CYP716A141
tHMGR, SAD1, CYP106A1, CYP716E26	tHMGR, SAD1, CYP106A1, CYP72A65	tHMGR, SAD1, CYP106A1, CYP72A69	tHMGR, SAD1, CYP106A1, CYP88D6
tHMGR, SAD1, CYP106A1, CYP93E1	tHMGR, SAD1, CYP106A1, CYP94D65	tHMGR, SAD1, CYP106A1, CYP716A12	tHMGR, SAD1, CYP106A1, CYP716A12, CYP716C11
tHMGR, SAD1, CYP106A1, CYP716A12, CYP72A67	tHMGR, SAD1, CYP106A2, CYP716A14	tHMGR, SAD1, CYP106A2, CYP716A141	tHMGR, SAD1, CYP106A2, CYP716E26
tHMGR, SAD1, CYP106A2, CYP72A65	tHMGR, SAD1, CYP106A2, CYP72A69	tHMGR, SAD1, CYP106A2, CYP88D6	tHMGR, SAD1, CYP106A2, CYP93E1
tHMGR, SAD1, CYP106A2, CYP94D65	tHMGR, SAD1, CYP106A2, CYP716A12	tHMGR, SAD1, CYP106A2, CYP716A12, CYP716C11	tHMGR, SAD1, CYP106A2, CYP716A12, CYP72A67

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

4.5.5 Procedure for Production of Binary Expression Constructs for CYP1A2, CYP2D6, CYP3A4 and CYP450R

Gene fragments for the known P450s CYP1A2, CYP2D6, and CYP3A4 as well as the cytochrome P450 reductase CYP450R were purchased from IDT and prepared according to the manufacturer's instructions. Entry cloning into pDONR207 using the gene fragments was performed according to the standard protocol (see **Section 2.1**), and plasmids were used to transform chemically competent *E coli* according to the standard protocol (see **Section 2.2.1**), with positive colonies selected for.

8 clearly separated colonies of *E. coli* per gene were sampled and used as a template for PCR using the standard colony PCR procedure (see **Section 2.1.3**).

Following PCR, DNA was analysed using the standard DNA electrophoresis procedure (see **Section 2.1.4**). Positive colonies were identified as those with a band between the 1.5 and 2.0 kb markers in the ladder (CYP1A2, CYP2D6, CYP3A4), or at approximately the same level as the 2.0 kb marker (CYP450R).

Positive colonies were cultured, and the plasmids extracted and sent for sequencing according to the standard procedure (see **Section 2.1.5**). CYP450R was additionally sequenced using a primer designed for the middle of the gene due to its size (CYP450R-M, see **Section 2.1.6**). One of the two colonies of CYP1A2, one of the two colonies of CYP2D6, both of the two colonies of CYP3A4, and one of the two colonies of CYP450R showed the correct gene sequence.

Plasmids were isolated from one of the colonies of each candidate gene with the correct sequence and used to generate destination vectors according to the standard procedure (see **Section 2.1.2**), then the vectors used to transform chemically competent *E. coli* (see **Section 2.2.1**). Positive colonies were selected for and cultured, and the plasmids recovered. Plasmids were then used to transform chemically competent *A. tumefaciens* (LBA4404) according to the standard procedure (see **Section 2.2.2**).

Positive colonies were selected for and cultured according to the standard procedure (see **Section 2.3**). Manual infiltration of *N. benthamiana* was then carried out to validate the strains using the combinations shown in **Table 4-6**. GC-MS analysis was performed using the 15-minute method.

Table 4-6: Combinations of vectors used to evaluate CYP1A2, CYP2D6, and CYP3A4

tHMGR, SAD1	tHMGR, SAD1, CYP716A12	tHMGR, SAD1, CYP1A2	tHMGR, SAD1, CYP1A2, CYP450R
tHMGR, SAD1, CYP2D6	tHMGR, SAD1, CYP2D6, CYP450R	tHMGR, SAD1, CYP3A4	tHMGR, SAD1, CYP3A4, CYP450R
tHMGR, SAD1, CYP716A12, CYP1A2	tHMGR, SAD1, CYP716A12, CYP1A2, CYP450R	tHMGR, SAD1, CYP716A12, CYP2D6	tHMGR, SAD1, CYP716A12, CYP2D6, CYP450R
tHMGR, SAD1, CYP716A12, CYP3A4		tHMGR, SAD1, CYP716A12, CYP3A4, CYP450R	

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

4.5.6 Procedure for Combinatorial Testing of CYP3A4, CYP2D6, and CYP3A4 and Toolkit Enzymes

A. tumefaciens strains containing tHMGR, SAD1, CYP716A12, CYP716A14v2, CYP716A141, CYP716C11, CYP716E26, CYP72A65, CYP72A67 CYP72A69, CYP88D6, CYP93E1, CYP94D65, CYP1A2, CYP2D6, and CYP3A4 from the triterpene toolkit were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined as shown in **Table 4-7** and used to infiltrate 1 leaf each of 5-week-old *N. benthamania* plants using the manual infiltration method (see **Section 2.3.2**). Plants were grown for 5 days under greenhouse conditions before leaves were harvested and processed according to the standard procedure for GC-MS and LC-MS analysis (see **Section 2.4.1**and **2.4.3**). GC-MS was performed using the 20-minute method.

Table 4-7: Combinations of vectors used to combinatorially evaluate CYP1A2, CYP2D6, and CYP3A4 (table is identical to **Table 4-2**).

tHMGR, SAD1	tHMGR, SAD1, CYP1A2	tHMGR, SAD1, CYP2D6	tHMGR, SAD1, CYP3A4
tHMGR, SAD1, CYP716A14	tHMGR, SAD1, CYP716A141	tHMGR, SAD1, CYP716E26	tHMGR, SAD1, CYP72A65
tHMGR, SAD1, CYP72A69	tHMGR, SAD1, CYP88D6	tHMGR, SAD1, CYP93E1	tHMGR, SAD1, CYP94D65
tHMGR, SAD1, CYP716A12	tHMGR, SAD1, CYP716A12, CYP716C11	tHMGR, SAD1, CYP716A12, CYP72A69	tHMGR, SAD1, CYP1A2, CYP716A14
tHMGR, SAD1, CYP1A2, CYP716A141	tHMGR, SAD1, CYP1A2, CYP716E26	tHMGR, SAD1, CYP1A2, CYP72A65	tHMGR, SAD1, CYP1A2, CYP72A69
tHMGR, SAD1, CYP1A2, CYP88D6	tHMGR, SAD1, CYP1A2, CYP93E1	tHMGR, SAD1, CYP1A2, CYP94D65	tHMGR, SAD1, CYP1A2, CYP716A12
tHMGR, SAD1, CYP1A2, CYP716A12, CYP716C11	tHMGR, SAD1, CYP1A2, CYP716A12, CYP72A67	tHMGR, SAD1, CYP2D6, CYP716A14	tHMGR, SAD1, CYP2D6, CYP716A141
tHMGR, SAD1, CYP2D6, CYP716E26	tHMGR, SAD1, CYP2D6, CYP72A65	tHMGR, SAD1, CYP2D6, CYP72A69	tHMGR, SAD1, CYP2D6, CYP88D6
tHMGR, SAD1, CYP2D6, CYP93E1	tHMGR, SAD1, CYP2D6, CYP94D65	tHMGR, SAD1, CYP2D6, CYP716A12	tHMGR, SAD1, CYP2D6, CYP716A12, CYP716C11
tHMGR, SAD1, CYP2D6, CYP716A12, CYP72A67	tHMGR, SAD1, CYP3A4, CYP716A14	tHMGR, SAD1, CYP3A4, CYP716A141	tHMGR, SAD1, CYP3A4, CYP716E26
tHMGR, SAD1, CYP3A4, CYP72A65	tHMGR, SAD1, CYP3A4, CYP72A69	tHMGR, SAD1, CYP3A4, CYP88D6	tHMGR, SAD1, CYP3A4, CYP93E1
tHMGR, SAD1, CYP3A4, CYP94D65	tHMGR, SAD1, CYP3A4, CYP716A12	tHMGR, SAD1, CYP3A4, CYP716A12, CYP716C11	tHMGR, SAD1, CYP3A4, CYP72A67

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

4.5.7 Procedure for Qualitative Evaluation of THAR Enzymes

A. tumefaciens strains containing tHMGR, SAD1, THAR1, THAR2, CYP716A14v2, CYP88D6, and CYP93E1 were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined as shown in **Table 4-8** and used to infiltrate 1

leaf each of 5-week-old N. benthamiana plants using the manual infiltration method (see **Section 2.3.2**). Plants were grown for 5 days under greenhouse conditions before leaves were harvested and processed according to the standard procedure for GC-MS analysis using the 20-minute method (see **Section 2.4.1**).

Table 4-8: Combinations of vectors used to qualitatively investigate THAR constructs

tHMGR + SAD1 only	tHMGR, SAD1, THAR1	tHMGR, SAD1, CYP716A14
tHMGR, SAD1, THAR1, THAR2	tHMGR, SAD1, CYP716A14, THAR2	
tHMGR, SAD1, THAR1, THAR2, CYP88D6, CYP93E1	tHMGR, SAD1, CYP716A14, THAR2, CYP88D6, CYP93E1	

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

4.5.8 Procedure for Semi-Quantitative Evaluation of THAR Enzymes

A. tumefaciens strains containing tHMGR, SAD1, THAR1, THAR2, CYP716A14v2, CYP88D6, and CYP93E1 were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined as shown in **Table 4-9** and used to infiltrate 4-week-old N. benthamiana plants using the manual infiltration method (**Section 2.3.2**). Each combination was used to infiltrate 3 plants.

Table 4-9: Combinations of vectors used to quantitatively investigate THAR constructs

tHMGR + SAD1 only	tHMGR, SAD1, THAR1, THAR2	tHMGR, SAD1, CYP716A14, THAR2
tHMGR, SAD1, THAR1, THAR2, CYP88D6	tHMGR, SAD1, CYP716A14, THAR2, CYP88D6	
tHMGR, SAD1, THAR1, THAR2, CYP88D6, CYP93E1	tHMGR, SAD1, CYP716A14, THAR2, CYP88D6, CYP93E1	

All constructs were cultured using the standard procedure (see **Section 2.3.1**).

Plants were grown for 5 days under greenhouse conditions before infiltrated leaves were harvested.

Harvested leaves were frozen at –80 °C before being dried in a freeze drier for 24 h. 5 mg of dry leaf powder from each leaf was measured out and processed for GC-MS analysis using a Gerstel MPS liquid handler as follows:

500 µl of ethyl acetate containing a coprostanol internal standard (10 µg/ml) was transferred to each sample. Samples were shaken before being centrifuged for 30 s.

50 µl of each sample was then transferred into a sample vial. The solvent was evaporated, and derivatisation of each sample with 30 µl of Tri-Sil-Z solution was carried out. GC-MS spectra were recorded for each sample using the 15-minute method.

5 Large-Scale Synthesis of Triterpene Derivatives for Structural Characterisation

Acknowledgements:

23-hydroxy-β-amyrin was purified by Dr. James Reed and the NMR spectrum was recorded and interpreted by Dr. Michael Stephenson. 11-oxo-β-amyrin, 24-hydroxy-β-amyrin, and 12,13β-epoxy-β-amyrin were provided by members of the Osbourn lab as noted. All other work in this chapter is the author's own.

5.1 Introduction

The design and production of novel and bioactive triterpenes requires an in-depth understanding of the structural features underlying triterpene activity (or structure-activity relationship of triterpenes). This has proved to be difficult, partly as a result of the types of biological investigations typically carried out on triterpene natural products. Many of the studies into triterpene biological activity consist of the testing of crude plant extracts instead of individual molecules. When a crude extract provides a bioactive "hit", the study will then seek to identify potentially bioactive constituents of the plant extracts involved using high pressure liquid chromatography (HPLC) or other methods. While these studies do frequently uncover potent bioactivities, the delineation of these activities (identifying exactly which components are responsible for any particular bioactive effect) given the complexity of the extracts is not trivial. Additionally, while the components of a mixture can often be identified, the precise quantification of the amount of a bioactive component present in an extract (and therefore the concentration of said component that is eliciting the biological effects) often cannot be determined accurately. Additionally, the other components of a mixture can interfere with the assays typically carried out to investigate potential bioactivity.

When studies are carried out into purified triterpene compounds, they often feature a relatively small number of molecules purified from a given plant source. Different studies will therefore test small sets of different molecules with potentially different methodology, hindering comparison between studies. As the compounds studied are extracted from nature, the set of molecules under investigation will therefore feature an assortment of structural features and complexity instead of being rationally designed. These factors lead to a fragmented picture of triterpene structure-activity relationships.

Further investigation of triterpene structure-activity relationships would benefit from a systematic approach. This has previously proved difficult due to the difficulty of accessing most positions on a triterpene scaffold such as β-amyrin through chemical means. The use of the triterpene toolkit to functionalise the β-amyrin scaffold

enzymatically has previously been demonstrated by James Reed in the Osbourn Group [50]. With the improved diversity of the triterpene toolkit over the intervening years, it is now possible to produce a wider variety of functionalised triterpenes with a range of different scaffold positions covered. This presents a new opportunity to revisit the previous work and expand upon it to produce a larger suite of molecules, in order to deepen the understanding of triterpene structure-activity relationships.

One consideration which warrants further investigation is the stereochemistry of triterpenes. As mentioned in **Chapter 4**, pentacyclic triterpenes are rigid structures made of fused cyclohexane rings. The existence of multiple trans-fused ring junctions in a triterpene scaffold such as β-amyrin means ring-flip is impossible [126]. The positions of functional groups on the rings of a triterpene are therefore clearly defined. A pair of triterpene epimers, differentiated by a single substituent existing in either the α or β configuration, are therefore entirely separate molecules with different 3D structures. As the 3D structure of a molecule potentially has implications for its binding to protein targets [126], an understanding of the effects of triterpene stereochemistry on bioactivity is therefore a valuable insight to be gained.

5.2 Aims

In order to investigate the structure-activity relationships of triterpenes based on the β-amyrin scaffold, in this chapter a suite of molecules was created through large-scale infiltrations and subsequent purification. Particular attention was paid to singly oxidised derivatives of β-amyrin, and to triterpenes with differences in stereochemistry. Selected compounds that had been putatively identified in **Chapter 4** are also produced at mg scale for full structural characterisation. Specific and unusual methods utilised in the purification process are discussed. The structures and stereochemistry of the generated molecules were confirmed through the use of NMR and GC-MS techniques. The completed suite of molecules will be subjected to biological testing in **Chapter 6,** supplemented with triterpenes from commercial sources.

5.3 Results and Discussion

5.3.1 Production of a Suite of Molecules for Testing through Large Scale Infiltrations

In order to take a systematic approach to investigating the structure-activity relationship of triterpenes, it was necessary to biosynthesise a suite of molecules for testing. This was achieved largely through the transient plant expression system, using β-amyrin synthase and single CYP450s from the triterpene toolkit. The methods used for the biosynthesis of these molecules are detailed in **Section 5.5** and **Section 2.4**, and the positions oxidised across the range of molecules are detailed in **Figure 5-1**. The structures of all of the biosynthesised molecules generated in this work were confirmed by NMR analysis; the NMR spectra can be found in **Section A.1**.

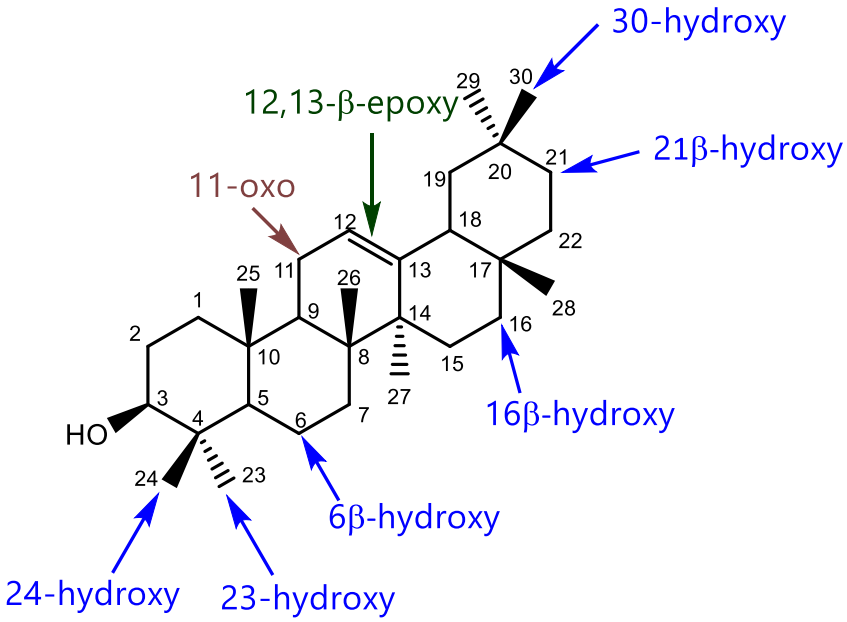


Figure 5-1: Positions on the β-amyrin scaffold modified in the present study across the set of simple triterpenes produced through one modification of β-amyrin.

Table 5-1 details the molecules that were biosynthesised and the combinations of constructs that produced them, as well as detailing where molecules were purified by other members of the Osbourn lab.

Table 5-1: Simple triterpenes with a maximum of one modification to the β-amyrin scaffold.

Compound	Enzymes/Source	# Plants (DW)	Yield (Total, mg)	Yield (Per g DW)
11-oxo-β-amyrin (**2**)	James Reed (tHMGR, SAD1, CYP88D6)	23 (10.4 g)	10.4	1
24-hydroxy-β-amyrin (**22**)	Michael Stephenson	274 (84 g)	170	2.02
12,13-β-epoxy-β-amyrin (**33**)	James Reed (tHMGR, SAD1, CYP51H10-I471M)	95 (43.7 g)	11*	0.25
23-hydroxy-β-amyrin (**21**)	This study (tHMGR, SAD1, CYP94D65)	100 (91.2 g)	79.36	0.87
30-hydroxy-β-amyrin (**19**)	This study (tHMGR, SAD1, CYP72A65)	100 (74 g)	8.49	0.11
16β-hydroxy-β-amyrin (**20**)	This study (tHMGR, SAD1, CYP716A141)	100 (117 g)	203.69	1.74
6β-hydroxy-β-amyrin (**34**)	This study (tHMGR, SAD1, CYP716E26)	100 (53 g)	16.26	0.31
21β-hydroxy-β-amyrin (**23**)	This study (tHMGR, SAD1, CYP72A69)	100 (152 g)	5.6	0.04

Each molecule is listed with the enzymes co-expressed to produce it, the number of plants used, dry leaf weight, and details on yield.

* Only part of a 140 mg semi-pure product was recrystallised for use due to the instability of this compound.

5.3.2 Production of Epi-β-Amyrin and Related Derivatives through Large Scale Infiltrations

Of particular interest in the determination of structure-activity relationships is the contribution of stereochemistry. Triterpenes are generally rigid molecules, with the pentacyclic triterpenes in particular being made of 5 fused rings that make the molecules highly resistant to ring-flip [126]. Therefore, the configuration of each constituent (often described as α or β) is of particular importance.

As shown in **Chapter 4**, alcohol dehydrogenase enzymes from *Arabidopsis thaliana* have been found to catalyse the formation of triterpenes with a hydroxyl group in the C-3α configuration, an unusual orientation found in nature in molecules like bile acids and the bioactive boswellic acids [152]. Therefore, epi-β-amyrin (**28**) and a number of derivatives inspired by the substituents in the boswellic acids were synthesised (see **Figure 5-2**).

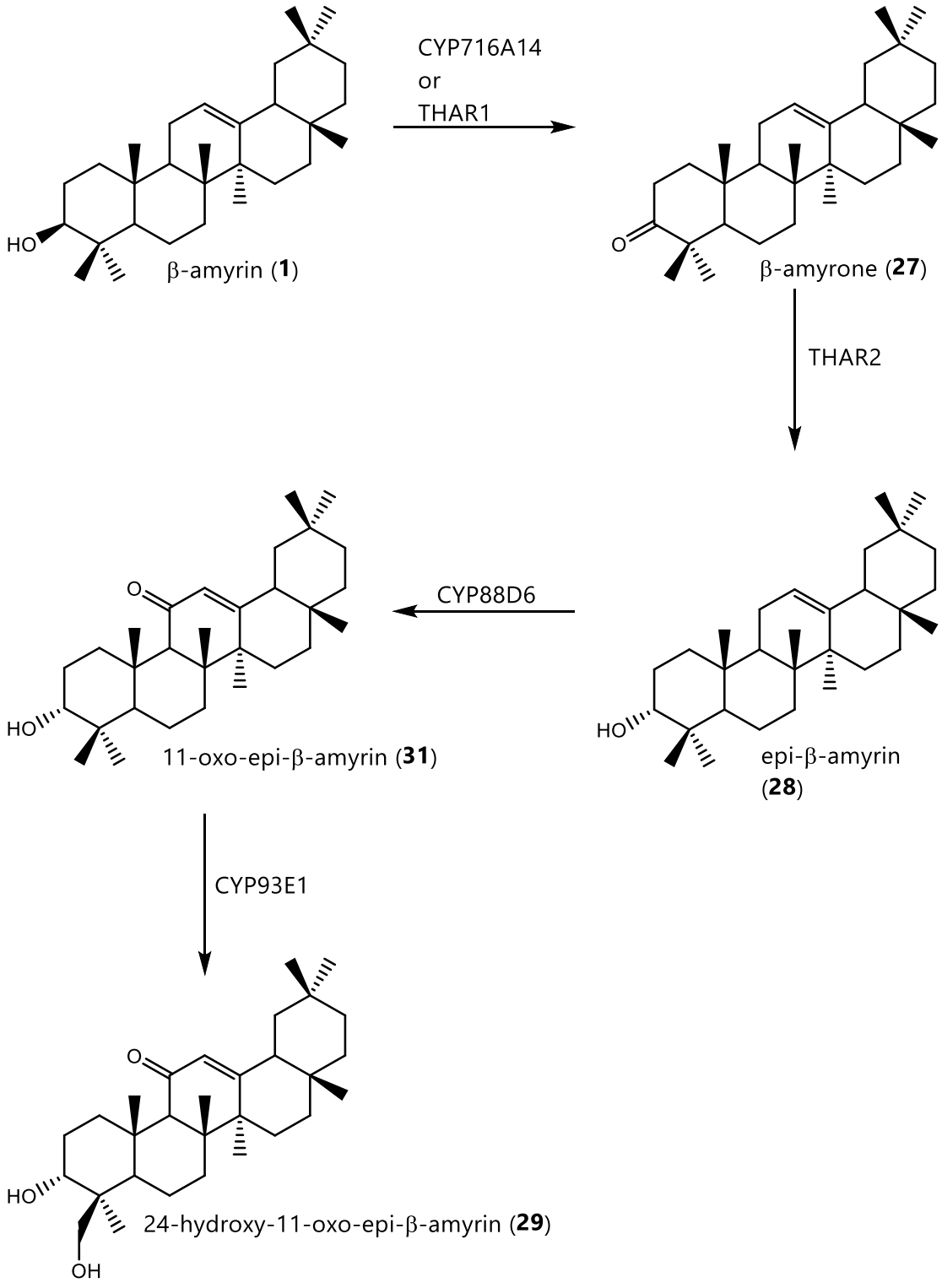


Figure 5-2: Molecules synthesised using alcohol dehydrogenase enzymes from *Arabidopsis thaliana*, inspired by the substituents present in the boswellic acids. Note that this sequence is illustrative of the molecules purified shown in a linear order of reactions for simplicity.

The synthesised molecules had their configurations confirmed by NMR analysis. When spectra were compared to those of the C-3β analogues, there was a

noticeable shift in the chemical shift, especially of the C-3 proton of the molecule (an example of this can be seen in **Figure 5-3**). The full spectra of each molecule are in **Section A.1**.

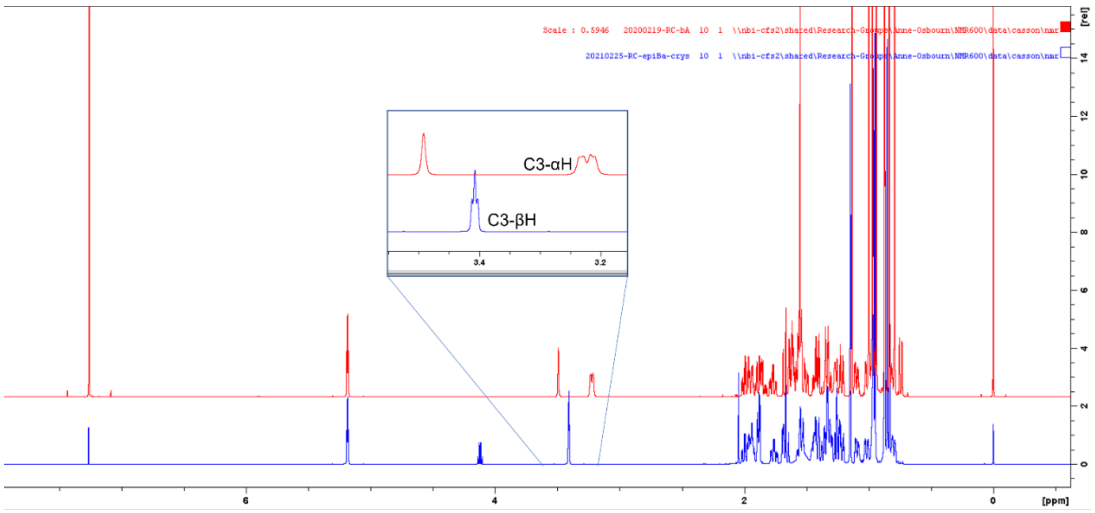


Figure 5-3: ^{1}H NMR spectra of β-amyrin (**1**) (red) and epi-β-amyrin (**28**) (blue), with the region containing the C-3 protons enlarged. The difference in chemical shift between the protons can be seen.

Overall, 3 C-3α molecules were produced along with their C-3β analogues: epi-β-amyrin (**28**), its 11-oxo derivative (**31**), and a 24-hydroxy, 11-oxo derivative (**29**). 11-oxo-β-amyrone (**35**) was also produced. A summary of the molecules produced, and the constructs used to synthesise them is given in **Table 5-2**.

Table 5-2: Compounds synthesised through production of C-3α-hydroxyl triterpenes.

Compound	Enzymes	# Plants (DW)	Yield (Total, mg)	Yield (Per g DW)
β-amyrin (**1**)	tHMGR, SAD1, CYP716A14, THAR2	100 (50 g)	10.29	0.21
epi-β-amyrin (**28**)	tHMGR, SAD1, CYP716A14, THAR2	100 (50 g)	167.20	3.34
11-oxo-β-amyrin (**2**)	tHMGR, SAD1, THAR1, THAR2, CYP88D6	100 (54 g)	5.65	0.10
11-oxo-β-amyrone (**35**)	tHMGR, SAD1, THAR1, THAR2, CYP88D6	100 (54 g)	32.11	0.59
11-oxo-epi-β-amyrin (**31**)	tHMGR, SAD1, THAR1, THAR2, CYP88D6	100 (54 g)	13.62	0.25
24-hydroxy-11-oxo-β-amyrin (**30**)	tHMGR, SAD1, THAR1, THAR2, CYP88D6, CYP93E1	100 (74 g)	20.72	0.28
24-hydroxy-11-oxo-epi-β-amyrin (**29**)	tHMGR, SAD1, THAR1, THAR2, CYP88D6, CYP93E1	100 (74 g)	59.66	0.81

The constructs combined to produce each molecule are given along with yield values, number of plants used, and dry weight of leaves.

5.3.3 Generating Augustic Acid for Comparison with Maslinic Acid

As mentioned in **Section 5.3.2**, the rigid structure of triterpenes places a relevance on investigating derivatives with the same regiochemistry, but different stereochemistry, as the different 3D structures of these molecules could have impacts on biological activity. As well as C-3, the C-2 position features well-known examples of molecules with alternate stereochemistry that are easily accessible using the triterpene toolkit. Triterpene C-2 hydroxylase enzymes are usually active on oleanolic acid (**18**). 2α-hydroxy oleanolic acid, also known as maslinic acid (**24**), can be produced by expression of the enzyme CYP716C11, as previously reported (see **Chapter 4**) [117]. This triterpene is also produced commercially, and a commercial sample of maslinic acid was purchased for testing.

2β-hydroxy oleanolic acid, known as augustic acid (**25**), can also be produced using the toolkit enzymes, namely CYP72A67 [136]. A scale-up synthesis of augustic acid was therefore carried out so that its activity could be compared to that of maslinic acid. The full details of the synthesis are given in **5.5.9**. Notably, this purification made use of an innovative "capture-release" mechanism using the Ambersep resin usually employed to clean-up purification of triterpenes in the Osbourn lab. Ambersep has previously been considered incompatible with acid-containing triterpenes due to its highly basic nature; however, it was discovered that when triterpenes containing carboxylic acid groups were mixed with Ambersep resin, the molecules adhered to the resin, allowing for much of the crude extract to be washed away with solvent. The adhered triterpenes could crucially then be released from the resin by treatment with acid, namely a 50 % solution of acetic acid in ethanol, as shown in **Figure 5-4**. This method sped up the purification of augustic acid without a loss of yield. The total amount of augustic acid (**25**) purified was 35.93 mg, corresponding to a yield per g of dry leaf weight of 0.51 mg/g.

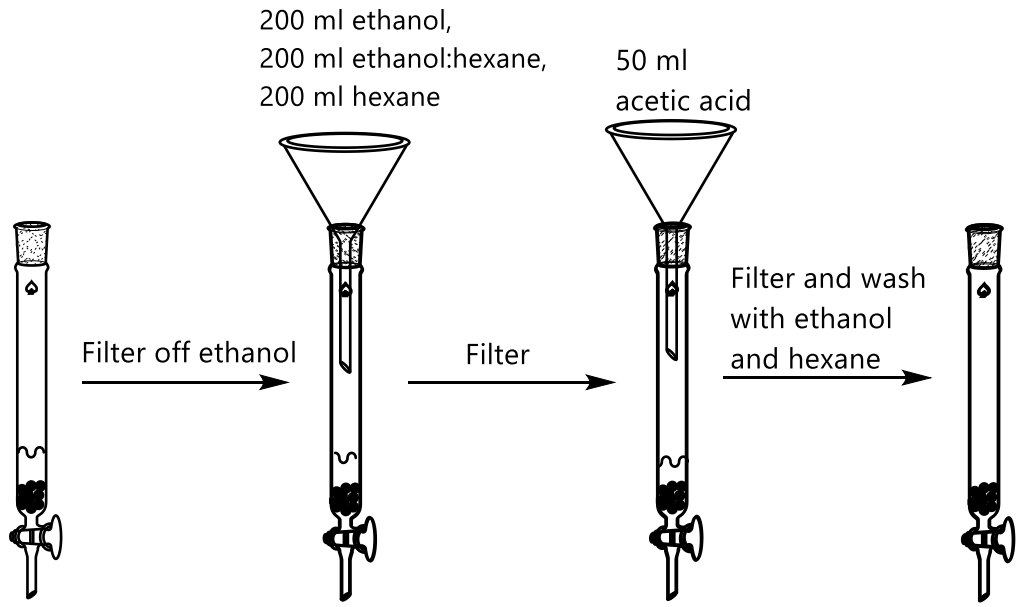


Figure 5-4: The process for capture-release of carboxylic acid-containing triterpenes through Ambersep treatment.

After purification of augustic acid (**25**), its identity and configuration were confirmed by NMR analysis. The full NMR spectra are shown in **Section A.1**. Crucially, the configuration of this molecule as C-2β-OH could be confirmed through the

observation of a cross peak in the ROSEY spectrum of the sample as seen in **Figure 5-5**, indicating the interaction of the C-2H and C-3H of the compound and showing those relevant protons to be in the syn configuration.

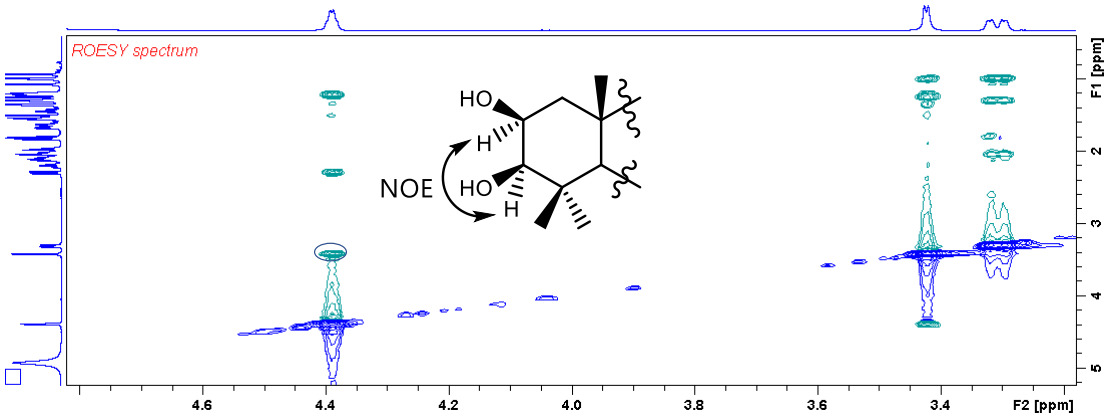


Figure 5-5: Magnification of the ROSEY ^{1}H-^{1}H spectrum of augustic acid (**25**). The cross-peak between C-2H and C3-H is circled. Full spectrum can be found in (**Section A.1**).

5.3.4 Scale-up of the Combination of CYP716A12 and CYP716A141 Produces Echinocystic and not Cochalic Acid

Another possible position on the β-amyrin scaffold that can be hydroxylated with alternate stereochemistry is C-16. Multiple enzymes exist to oxidise C-16, with a number existing within the triterpene toolkit. Curiously, all known C-16α-hydroxylases are reported to only be active on oleanolic acid (**18**), not the basic β-amyrin scaffold, while C-16β-hydroxylases are active on β-amyrin (**1**). In order to investigate the effects of C-16 stereochemistry on biological activity, the combination of CYP716A12, which produces oleanolic acid (**18**), and CYP716A141, a C-16β-hydroxylase, was scaled up using the vacuum infiltration method in an attempt to produce cochalic acid (16β-hydroxy-oleanolic acid) (**36**). Echinocystic acid (16α-hydroxy-oleanolic acid) (**37**) is already well-reported, and a commercial standard of this compound had already been purchased. The yield of the obtained product was 12.7 mg corresponding to a yield per g of dry leaf weight of 0.31 mg/g.

After purification of this compound (see **5.5.10**), NMR spectroscopy was carried out to confirm its structure, and, most importantly, its stereochemistry. The first initial sign that the produced compound may not be cochalic acid (**36**) was the exact and complete agreement of its ^{1}H NMR spectrum with a literature source for

echinocystic acid (**37**), something that is not usually the case for stereoisomers. Indeed, when ROSEY spectroscopy was used to attempt to confirm the C-16β configuration of the hydroxyl group, the expected cross-peaks between the C-16*H* proton and the protons at C-27 were absent. A ^{1}H spectrum was therefore collected using the purchased commercial echinocystic acid (**37**) standard (see **Figure 5-6**), which compared exactly to that of the produced compound, proving that it was in fact echinocystic acid (**37**) and not cochalic acid (**36**) which had been produced by the combination of CYP716A12 and CYP716A141.

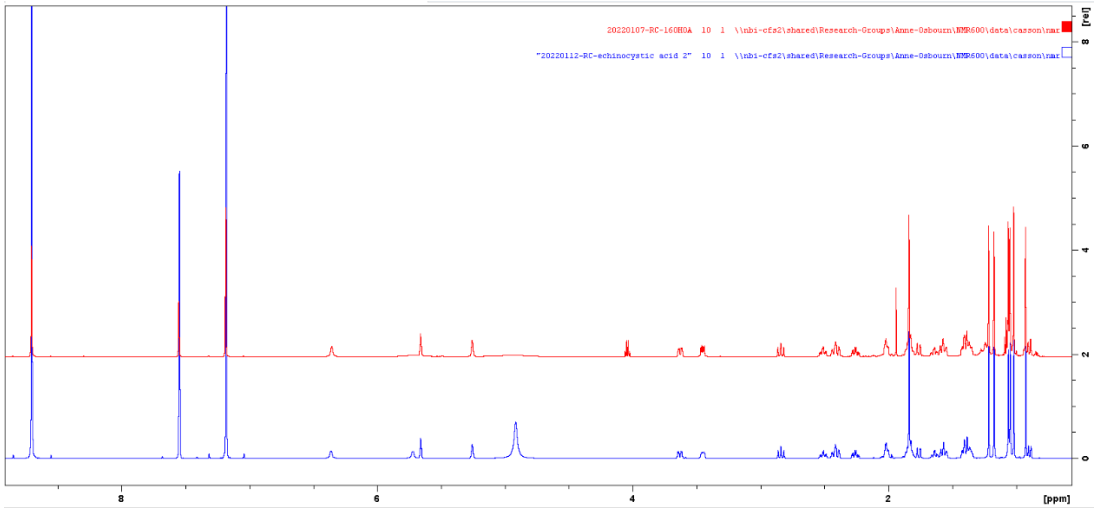


Figure 5-6: ^{1}H NMR spectra of an echinocystic acid (**37**) standard (blue) and the biosynthesised product suspected to be echinocystic acid (red). The peaks corresponding to the two compounds overlap exactly, showing that they are chemically identical.

As CYP716A141 had previously been reported as a C-16β-hydroxylase and had been treated as such in all previous work [109], investigations were carried out to ensure that the usual product of CYP716A141 when acting on the β-amyrin scaffold was indeed 16β-hydroxy-β-amyrin (**20**) as had been previously assumed. This was confirmed with ROSEY spectroscopy on the sample of 16β-hydroxy-β-amyrin (**20**) previously produced for biological testing (see **Section A.1**).

These results suggest that CYP716A141 produces a different stereochemistry for its hydroxylations depending on whether β-amyrin (**1**) or oleanolic acid (**18**) are the substrate. This is an unusual activity not previously seen in other enzymes. To investigate if this occurs with other C-16 hydroxylases, a number of C-16 hydroxylase enzymes from the triterpene toolkit and within the Osbourn lab more

generally were tested on the β-amyrin scaffold. As can be seen in **Figure 5-7**, on the β-amyrin scaffold, all of the tested hydroxylases produce products with identical GC-MS retention times and mass spectra, suggesting they are the same product. The mass ion at m/z = 586.5 is consistent with a single hydroxylation on β-amyrin, suggesting that the products are 16-hydroxy-β-amyrin (**20**) based on the known activity of these enzymes. Although it cannot be conclusively proven that these are all C-16β without purification and NMR spectra, the fact that stereoisomers of β-amyrin and its derivatives have previously been shown to have different retention times using the same GC-MS method (see **Chapter 4**) lends credence to the theory that they are.

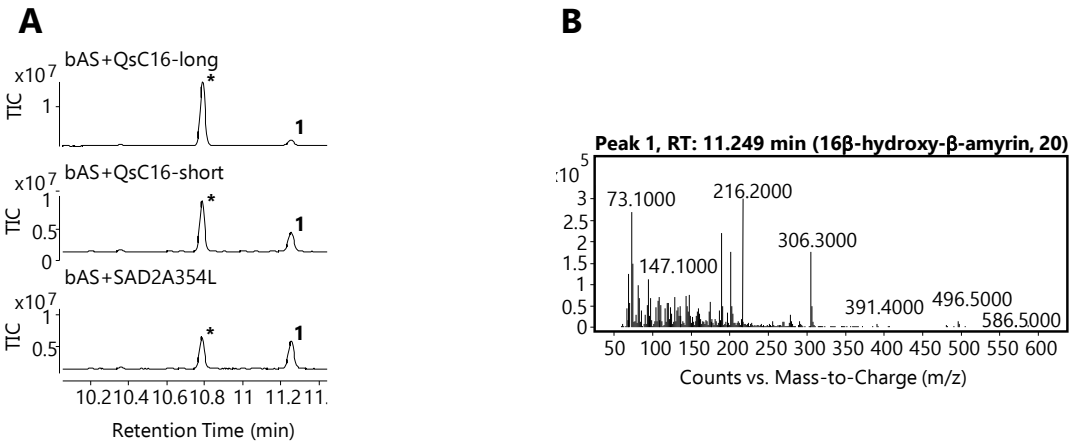


Figure 5-7: Comparison of C-16 oxidases. A) tested C-16 hydroxylases co-expressed with tHMGR and β-amyrin synthase: long (top) and short (middle) versions of a Quillaja Saponaria C-16 oxidase and a mutated form of the Avena strigosa CYP51H10 CYP450 that is known to hydroxylate only at C-16β (Sad2-A354L, bottom). B)The mass spectrum of the peak seen at 11.2 minutes (peak 1). TIC = Total ion chromatogram.

The *Q. saponaria* C-16 oxidases are known to produce echinocystic acid (**37**) in combination with oleanolic acid-producing constructs (J. Reed, submitted) but appear to be producing 16β-hydroxy-β-amyrin (**20**) here, as the product of these enzymes has the same retention time and mass spectrum as that for SAD2-A345L, a SAD2 mutant enzyme known to produce 16β-hydroxy-β-amyrin (**20**) [153]. Overall, this suggests that many C-16 hydroxylases will oxidise in the α configuration when oleanolic acid is the scaffold but in the β configuration when the scaffold is β-amyrin, raising the question of whether this in fact occurs more often with cytochromes P450 acting towards the rigid fused-ring triterpenes of a tetracyclic or

pentacyclic nature and has previously been overlooked due to assumptions about the predictable nature of enzyme-based synthetic biology methods. This therefore serves as a cautionary tale for future combinatorial biosynthesis work, for one to make sure to evaluate carefully the potential products of enzyme combinations and be prepared for unexpected synergies to occur.

Table 5-3: Compounds synthesised through production of triterpenes with alternate stereochemistry at C-2 and C-16.

Compound	Enzymes	# Plants (DW)	Yield (Total, mg)	Yield (Per g DW)
Augustic Acid (**25**)	tHMGR, SAD1, CYP716A12, CYP72A67	100 (69.5 g)	35.93	0.51
Echinocystic Acid (**37**)	tHMGR, SAD1, CYP716A12, CYP716A141	100 (40.8 g)	12.7	0.31

The constructs combined to produce each molecule are given along with yield values, number of plants used, and dry weight of leaves.

5.3.5 Scale-up of Production of the CYP106A1 Product

In **Chapter 4**, two bacterial enzymes, CYP106A1 and CYP106A2, were evaluated for their ability to modify β-amyrin. Based on literature reports of these enzymes' activity [141,142] the putative structure of the major product of CYP106A1 was assigned as 7β-hydroxy-β-amyrin (**32**). Given the relatively high abundance of the major product of CYP106A1 expression relative to the other products produced by CYP106A1 and CYP106A2, this product was selected for scale-up and purification to enable structural confirmation. Co-expression of the enzymes necessary to produce β-amyrin and the CYP106A1 construct was therefore carried out using the vacuum infiltration method. Details of the methods used for purification of the product are provided in section **5.5.12**. The yield of the product was 29.63 mg, corresponding to a yield per g of dry leaf weight of 0.59 mg/g.

Full analysis of this product by NMR was then carried out. The spectra and full assignment for this molecule are given in **Section A.1** and confirm the structure as 7β-hydroxy-β-amyrin (**32**). Key to this assignment are the observation of a peak in

the COSY spectrum between C-7H and C6-H of the compound, as well as a peak in the HMBC between C-7 and C-26.

Additionally, the β-configuration of the hydroxyl group can be shown definitively with ROSEY spectroscopy through the cross peak visible between the C-5H and C-7H of this compound (see **Figure 5-8**).

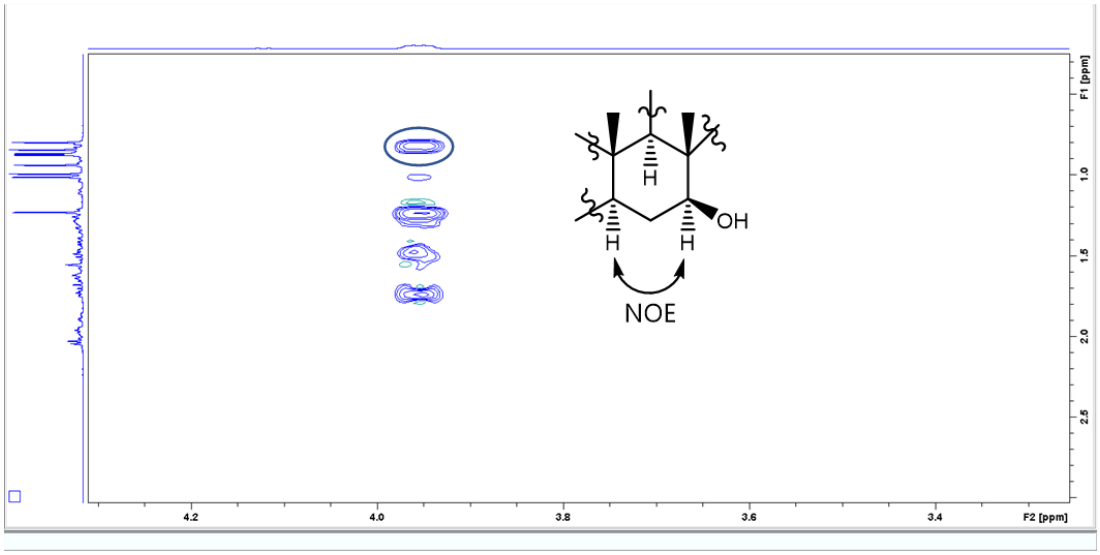


Figure 5-8: Magnification of the ROSEY ^{1}H-^{1}H spectrum of 7β-hydroxy-β-amyrin (**32**). The cross-peak between C-5H and C7-H is circled. Full spectrum can be found in (**Section A.1**).

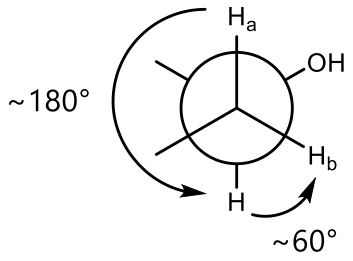
5.3.6 CYP3A4 and CYP88D6 Together Generate New Products Including 11-oxo-erythrodiol

Chapter 4 discussed the evaluation of the human enzymes CYP1A2, CYP2D6, and CYP3A4 in combinatorial biosynthesis with toolkit enzymes. One combination in particular showed promise. The combination of CYP3A4 and CYP88D6 was particularly well accumulated in leaf extracts and was interesting from a bioactivity perspective, due to the known bioactive properties of the related compound 11-oxo-β-amyrin (**2**). Therefore, this was selected for scale-up to the mg scale. Co-infiltration of β-amyrin scaffold production enzymes, CYP88D6, and CYP3A4 was therefore carried out using the vacuum infiltration method. Two major products were produced following co-infiltration. The procedure used for their purification is described in **5.5.13**. The yield of the two products was 4.49 mg of 21β-hydroxy-11-oxo-β-amyrin (**38**) and 4.26 mg of 11-oxo-erythrodiol (**39**),

corresponding to a yield per g of dry leaf weight of 0.06 mg/g and 0.05 mg/g respectively.

NMR analysis of the two products (spectra and assignment given in **Section A.1**) showed one of them to have a good agreement to a literature spectrum for 11-oxo-erythrodiol (**39**) (see **Section A.1**), with major signals such as those for the methyl groups and hydroxyl groups matching the literature source. This compound has previously shown significant anti-proliferative activity in literature reports [154].

The other product was assigned as 21β-hydroxy-11-oxo-β-amyrin (**38**). Key signals in this assignment include the COSY coupling visible between C-21H and C-22H and HMBC coupling visible between C-21, C-29, and C-30. The assignment of the C-21 hydroxyl in the β-configuration is made due to the size of the ^{3}J coupling constant between C-21H and both hydrogens at C-22. The signal for C-21 in the ^{1}H NMR spectrum for this compound is a doublet of doublets with J values of 12 and 4.8 Hz. The presence of a 12 Hz coupling constant suggests a H-H diaxial interaction between C-21H and one of the hydrogen atoms on C-22, an interaction that would only occur where the C-21 hydroxyl is in the β-configuration (which is equatorial) [126]. Similarly, the 4.8 Hz coupling constant between C-21H and the other hydrogen atom at C-22 is consistent with an axial-equatorial interaction (**Figure 5-9**) [126].



^{3}J H, H_a = 12 Hz
^{3}J H, H_b = 4.8 Hz

Figure 5-9: Newman projection along the C-22-C-21 carbon-carbon bond of the E-ring of 21β-hydroxy-11-oxo-β-amyrin. C-22 is in the foreground and C-21 in the background. The approximate dihedral angles between the hydrogen at C-21 and those at C-22, along with the ^{3}J coupling constants, are shown.

The production of these two products in reasonable yield demonstrates the successful addition of human enzymes to the triterpene toolkit. Further investigation of the bioactivity of these compounds is detailed in **Chapter 6**.

The identity of these products provides a possible explanation for the apparent inactivity of CYP3A4 towards 21β-hydroxy-β-amyrin (**23**) seen in **Chapter 4**. Given that C-21 is already oxidised, the addition of a C-21 oxidising enzyme would not be expected to produce additional products.

One unidentified product of combinations of CYP3A4 with enzymes from the TKK seen in **Chapter 4** was addition of a carboxylic acid group. While the identity of this product has not been positively confirmed, the discovery of the C-28 oxidation ability of CYP3A4 suggests that the product of these combinations could be an oleanolic acid derivative.

Table 5-4: Compounds synthesised by scale-up of the products of the new enzymes characterised in Chapter 4.

Compound	Enzymes	# Plants (DW)	Yield (Total, mg)	Yield (Per g DW)
7β-hydroxy-β-amyrin (**32**)	tHMGR, SAD1, CYP106A1	100 (50.4 g)	29.63	0.59
11-oxo-erythrodiol (**39**)	tHMGR, SAD1, CYP88D6, CYP3A4	100 (81.0 g)	4.26	0.05
21β-hydroxy-11-oxo-β-amyrin (**38**)	tHMGR, SAD1, CYP88D6, CYP3A4	100 (81.0 g)	4.49	0.06

The constructs combined to produce each molecule are given along with yield values, number of plants used, and dry weight of leaves.

5.4 Conclusions

Here a suite of triterpenes based on the β-amyrin scaffold were synthesised for biological testing, including singly oxidised triterpenes on a number of scaffold positions, triterpenes derived from the C-3 epimer of β-amyrin (epi-β-amyrin (**28**)), and oleanolic acid derivatives with differing stereochemistry at C-2. Scale-up, purification, and structural determination through NMR of selected compounds putatively identified in **Chapter 4** was carried out. New techniques developed in the production of these compounds are discussed alongside analytical methods for

proving the stereochemistry of produced triterpenes. The compounds produced will be subjected to biological testing alongside a range of triterpenes from commercial sources in **Chapter 6**.

5.5 Materials and Methods for Chapter 5

5.5.1 Procedure for Large Scale Synthesis of 21β-hydroxy-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, and CYP72A69 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 6-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 5 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 48 h. The dried leaves (151.53 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried and Ambersep treatment carried out according to the standard procedure (see **Section 2.4.7**).

The crude extract (2.28 g) was dry-loaded onto a 50 g Biotage SNAP ultra column for column chromatography (Programme P-1). Fractions were analysed using GC-MS (20-minute method) and those containing the compound of interest were pooled and dry loaded onto a 25 g Biotage SNAP Kp-Sil column for column chromatography (Programme P-2). Fractions were analysed using GC-MS (20-minute method) and those containing the compound of interest were pooled and dried to give 30 mg of semi pure product which was treated with decolourising charcoal and filtered, then recrystalised from methanol to give 5.6 mg of 21-hydroxy-β-amyrin.

5.5.2 Procedure for Large Scale Synthesis of 6β-hydroxy-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1 and CYP716E26 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 6-week-old *N. benthamiana* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 5 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried using a freeze-drier for 48 h. Dried leaves (53.42 g) were crushed and extracted using a speed extractor (see

Section 2.4.6). Ambersep ion exchange resin was used to treat the dried extract according to the standard procedure (see **Section 2.4.7**).

The crude extract (2.86 g) was dry-loaded onto a 50 g Biotage SNAP ultra column for column chromatography (Programme P-1). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dry loaded onto a 25 g Biotage SNAP Kp-Sil column for column chromatography (Programme P-2). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dry-loaded onto a 10 g Biotage Kp-Sil column for column chromatography (Programme P-3). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dried to give 28 mg of semi pure product, which was recrystallised from ethanol with water as an anti-solvent to give 16.26 mg of 6β-hydroxy-β-amyrin.

5.5.3 Procedure for Large Scale Synthesis of 16β-hydroxy-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, and CYP716A141 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 6-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 6 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 72 h. The dried leaves (117.20 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 12.33 g of crude extract which was treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract (8.65 g) was dry-loaded onto a 50 g Biotage SNAP ultra column for column chromatography (Programme P-1). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dry loaded onto a 25 g Biotage SNAP Kp-Sil column for column chromatography (Programme P-2). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and

treated again with Ambersep ion exchange resin until no additional colour change was observed. The semi-pure extract was then dry-loaded onto a 25 g Biotage Kp-Sil column for column chromatography (Programme P-5). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dried to give 550 mg of semi pure product, which was recrystallised from methanol to give 203.69 mg of 16β-hydroxy-β-amyrin.

5.5.4 Procedure for Large Scale Synthesis of 30-hydroxy-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, and CYP72A65 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 5-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 6 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 96 h. The dried leaves (74.05 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 17.62 g of crude extract which was treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract (6.72 g) was dry-loaded onto a 50 g Biotage SNAP ultra column for column chromatography (Programme P-1). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dry loaded onto a 25 g Biotage SNAP Kp-Sil column for column chromatography (Programme P-6). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dry-loaded onto a 25 g Biotage Kp-Sil column for column chromatography (Programme P-7). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dried to give 22 mg of semi pure product, which was recrystallised from methanol to give 8.49 mg of 30-hydroxy-β-amyrin.

5.5.5 Procedure for Large Scale Synthesis of 23-hydroxy-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, and CYP94D65 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 6-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 6 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 48 h. The dried leaves (91.21 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 11.72 g of crude extract which was treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract (3.82 g) was dry-loaded onto a 50 g Biotage SNAP ultra column for column chromatography (Programme P-1). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled to give a semi pure product (0.37 g) which was purified further by James Reed.

5.5.6 Procedure for Large Scale Synthesis of epi-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, CYP716A14, and THAR2 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 6-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 5 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 96 h. The dried leaves (49.9 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 6.6 g of crude extract which was treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract was dry-loaded onto a 50 g Biotage SNAP Ultra column for column chromatography (Programme P-1). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dry loaded onto a 25 g Biotage Sfär D Duo column for column chromatography

(Programme P-8). Fractions were analysed using GC-MS (15-minute method) and pooled into three different semi-pure extracts: an epi-β-amyrin extract (260 mg), a β-amyrin extract (30 mg), and an extract containing a mixture of both (250 mg).

The mixed extract was dry-loaded onto a 25 g Biotage Sfär Silica D Duo column for chromatography as before and fractions monitored using GC-MS (15-minute method). Fractions containing solely epi-β-amyrin or β-amyrin were added to the semi-pure extracts while the mixed fractions were pooled and dry-loaded onto a 10 g Biotage SNAP Ultra column for chromatography (Programme P-9). Fractions were monitored by GC-MS (15-minute method) and fractions containing solely epi-β-amyrin or β-amyrin were added to the semi-pure extracts to give final amounts of 280 mg epi-β-amyrin and 140 mg β-amyrin.

The two semi-pure extracts were treated with activated charcoal and filtered through a celite plug, washed with 50 ml ethyl acetate, then dried. The epi-β-amyrin semi-pure extract was recrystallised from hot methanol to give 167.2 mg of epi-β-amyrin. The β-amyrin fraction was recrystallised from hot methanol to give 10.29 mg of β-amyrin.

5.5.7 Procedure for Large Scale Synthesis of 11-oxo-epi-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, THAR1, THAR2, and CYP88D6 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 6-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 7 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 72 h. The dried leaves (54.24 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried and treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract (3.04 g) was dry-loaded onto a 50 g Biotage SNAP Ultra colulmn for column chromatography (Programme P-1). Fractions were analysed by TLC, GC-MS, and UV absorbance at 254 nm. Fractions containing the compound of interest were

pooled and dry-loaded onto a 25 g Biotage Sfär D Duo column for chromatography (Programme P-10). Fractions were monitored by GC-MS (15-minute method) and UV absorbance at 254 nm and fractions of interest were pooled and dry-loaded onto a 25 g Biotage Sfär D Duo column for chromatography (Programme P-11). Fractions were monitored by GC-MS (15-minute method) and UV absorbance at 254 nm and those containing pure product were collected and retained while those containing a mixture of products were dry-loaded onto a 10 g Biotage SNAP Ultra column for chromatography (Programme P-12). Fractions were monitored by GC-MS (15-minute method) and UV absorbance at 254 nm and pure fractions were retained while those containing a mixture of compounds were pooled and dry-loaded onto a 10 g Biotage SNAP Ultra column for chromatography (Programme P-12). This process was continued for a further 7 columns, resulting in three semi pure products: product 1 (30 mg), product 2 (80 mg), and product 3 (40 mg), which were recrystallised from methanol to give 13.62 mg product 1 (identified as 11-oxo-epi-β-amyrin), 32.11 mg product 2 (identified as 11-oxo-β-amyrone), 5.65 mg product 3 (identified as 11-oxo-β-amyrin).

5.5.8 Procedure for Large Scale Synthesis of 24-hydroxy-11-oxo-epi-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, THAR1, THAR2, CYP88D6, and CYP93E1 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 5-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 7 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 72 h. The dried leaves (74.1 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 5.5 g of crude extract which was treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract was dry-loaded onto a 50 g Biotage SNAP Ultra column for column chromatography (Programme P-1). Fractions were analysed using TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing the

compound of interest were pooled to give two different semi-pure extracts: an 11-oxo-24-hydroxy-β-amyrin extract (200 mg), and an 11-oxo-24-hydroxy-epi-β-amyrin extract (100 mg). The 11-oxo-24-hydroxy-β-amyrin extract was dry loaded onto a 25 g Biotage Sfär Silica D Duo column for chromatography (Programme P-13). Fractions were analysed by TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing pure product were collected and retained while those containing a mixture of products were dry-loaded onto a 10 g Biotage SNAP Ultra column for chromatography (Programme P-14). Fractions were analysed by TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing pure product were collected and retained to give a total of 115 mg semi-pure 11-oxo-24-hydroxy-β-amyrin.

The 11-oxo-24-hydroxy-epi-β-amyrin extract was dry-loaded onto a 10 g Biotage SNAP Ultra column for chromatography (Programme P-14). Fractions were analysed by TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing pure product were collected and retained while those containing a mixture of products were dry-loaded onto a 10 g Biotage SNAP Ultra column for chromatography (Programme P-15). Fractions were analysed by TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing pure product were collected and retained. The product fractions were then treated with activated charcoal, filtered, and washed with 20 ml ethyl acetate before drying, to give 129 mg semi-pure 11-oxo-24-hydroxy-epi-β-amyrin.

Both the semi-pure 11-oxo-24-hydroxy-β-amyrin and 11-oxo-24-hydroxy-epi-β-amyrin were recrystallised from methanol with water as an anti-solvent to give 20.72 mg 11-oxo-24-hydroxy-β-amyrin and 59.66 mg 11-oxo-24-hydroxy-epi-β-amyrin.

5.5.9 Procedure for Large Scale Synthesis of Augustic Acid

A. tumefaciens strains containing tHMGR, SAD1, CYP716A12, and CYP72A67 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 5-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 6 days under

greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 72 h. The dried leaves (69.5 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 4.9 g of crude extract.

The extract was divided into two and half dry-loaded onto a 50 g Biotage SNAP Ultra column for column chromatography (Programme P-1). Fractions were analysed using GC-MS (15-minute method) and those containing the compound of interest were pooled and dry-loaded onto a 10 g Biotage SNAP Ultra column for column chromatography (Programme P-14). Fractions were analysed by TLC and GC-MS (15-minute method) and those containing the product of interest were pooled, treated with activated charcoal, washed with 50 ml ethyl acetate, then dried to give 83 mg of a semi-pure product which was recrystallised from methanol to give 11.46 mg augustic acid.

The other half of the extract was treated with Ambsersep according to the standard procedure (see **Section 2.4.7**), except that the filtrate and washes were discarded instead of retained. The resin bed was then treated with 50 ml glacial acetic acid and washed with 100 ml ethanol and 200 ml hexane. The filtrate and washes were then combined and dried before being dry-loaded onto a 25 g Biotage Sfär Silica D Duo column for chromatography (Programme P-14). Fractions were analysed by TLC and GC-MS (15-minute method) and those containing the product of interest were pooled and recrystallised from methanol to give 24.47 mg augustic acid

The total amount of augustic acid obtained over both purification methods was 35.93 mg.

5.5.10 Procedure for Large Scale Synthesis of Echinocystic Acid

A. tumefaciens strains containing tHMGR, SAD1, CYP716A12, and CYP716A141 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 50 6-week-old and 50 7-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown

for 6 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze-drier for 72 h. The dried leaves (40.8 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 7.0 g of crude extract. This was treated with Ambersep according to the standard procedure (see **Section 2.4.7**), except that the filtrate and washes were discarded instead of retained. The resin bed was then washed with 100 ml 1:1 glacial acetic acid: ethanol, then washed with 200 ml ethanol, 200 ml 1:1 ethanol: hexane, 200 ml hexane. The washes were combined and dried. The extract (2 g) was then loaded onto a 50 g Biotage SNAP Ultra column for column chromatography (Programme P-1). Fractions were analysed by TLC and GC-MS (15-minute method) and those containing the product of interest were pooled and dry-loaded onto a 10 g Biotage SNAP KP-Sil cartridge for column chromatography (Programme P-13). Fractions were analysed by TLC and GC-MS (15-minute method) and those containing the compound of interest were pooled, dried, and recrystallised from methanol with water as anti-solvent to give 12.7 mg of a white solid that was revealed by NMR (^{1}H, ^{13}C, 2D, ROSEY) to be echinocystic acid rather than cochalic acid.

5.5.11 Procedure for Testing C-16 oxidases

A. tumefaciens strains containing tHMGR, SAD1, CYP(Qs-C16 short), CYP(Qs-C16 long), and SAD2-A354-L were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined as shown in **Table 5-5** and used to infiltrate 1 leaf each of 5-week-old *N. benthamiana* plants using the manual infiltration method (**Section 2.3.2**). Plants were grown for 6 days under greenhouse conditions before leaves were harvested and processed according to the standard procedure for GC-MS analysis using the 15-minute method (see **Section 2.4.1**).

Table 5-5: Combinations of vectors used to investigate C-16 oxidases

tHGMR+SAD1	tHMGR+SAD1+CYPQS-s
tHMGR+SAD1+CYPQs-L	tHMGR+SAD1+SAD2-A354-L

5.5.12 Procedure for Large Scale Synthesis of 7β-hydroxy-β-amyrin

A. tumefaciens strains containing tHMGR, SAD1, and CYP106A1 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 6-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 6 days under greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 72 h. The dried leaves (50.4 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 6.1 g of crude extract which was treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract (2.2 g) was dry-loaded onto a 50 g Biotage SNAP ultra column for column chromatography (Programme P-1). Fractions were analysed using TLC and GC-MS (15-minute method) and those containing the compound of interest were pooled and dry loaded onto a 25 g Biotage Sfär Silica D Duo column for column chromatography (Programme P-16). Fractions were analysed using TLC and GC-MS (15-minute method) and those containing the compound of interest were pooled and dry-loaded onto a 10 g Biotage Sfär Silica HC D Duo column for column chromatography (Programme P-14). Fractions were analysed using TLC and GC-MS (15-minute method) and those containing the compound of interest were pooled, treated with activated charcoal according to the standard procedure (see **Section 2.4.9**), and dried to give 81 mg of semi pure product, which was recrystallised from methanol with water as an anti-solvent to give 29.63 mg of 7β-hydroxy-β-amyrin.

5.5.13 Procedure for Large Scale Synthesis of 21β-hydroxy-11-oxo-β-amyrin and 11-oxo-erythrodiol

A. tumefaciens strains containing tHMGR, SAD1, CYP88D6, and CYP3A4 were cultured according to the standard procedure (see **Section 2.3.1**). All strains were combined and used to infiltrate 100 7-week-old *N. benthamania* plants using the vacuum infiltration method (see **Section 2.3.3**). Plants were grown for 7 days under

greenhouse conditions before leaves showing signs of 'necrotic phenotype' were harvested.

Harvested leaves were frozen at -80 °C before being dried in a freeze drier for 72 h. The dried leaves (81.0 g) were crushed and extracted using a speed extractor (see **Section 2.4.6**). The extract was dried to give 9.5 g of crude extract which was treated with Ambersep according to the standard procedure (see **Section 2.4.7**).

The extract (6.8 g) was dry-loaded onto a 50 g Biotage SNAP ultra column for column chromatography (Programme P-1). Fractions were analysed using TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing the compound of interest were pooled and dry loaded onto a 10 g Biotage KP-Sil column for column chromatography (Programme P-15). Fractions were analysed using TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing the compound of interest were pooled and dry-loaded onto a 10 g Biotage KP-Sil column for column chromatography (Programme P-14). Fractions were analysed using TLC, GC-MS (15-minute method), and UV absorbance at 254 nm and those containing the compound of interest were pooled and dry loaded onto 12 g Biotage Sfär C18 D Duo column for column chromatography (Programme R-1). Fractions were analysed by UV absorbance at 254 nm and those containing the compound of interest were combined and subjected to preparative HPLC using the following gradient:

Table 5-6: Preparative HPLC gradient used in the purification of the products of CYP3A4 and CYP88D6. Solvent A was 100 % water and solvent B was 100 % MeCN. A 3-minute post-time with 42 % B was used between runs to equilibrate the column.

Time	%B
0	42
1	42
1.5	70
11.5	95
16.5	95
17.0	42

Fractionation was carried out based on ELSD detection, molecular ion at m/z = 547 ($[M+H]^+$), and UV absorbance at 254 nm and yielded two products: 4.49 mg of 21β-hydroxy-11-oxo-β-amyrin and 4.26 mg of 11-oxo-erythrodiol.

6 Biological Evaluation of Oxidised Triterpene Derivatives

Acknowledgements:

Anti-proliferation assays for 23 hydroxy-β-amyrin and 11-oxo-epi-β-amyrin were performed by Maya Valmiki. The representative Western blot in **Figure 6-10** was carried out by Emily Hobson, and the representative Western blot in **Figure 6-9** was carried out by Prof. Maria O'Connell. All other work in this chapter is the author's own.

6.1 Introduction

Plant triterpenes, from sources such as frankincense resin and ginseng [85,155], have been found to have a range of biological activities but have shown particular promise as anti-proliferative and anti-inflammatory agents [19].

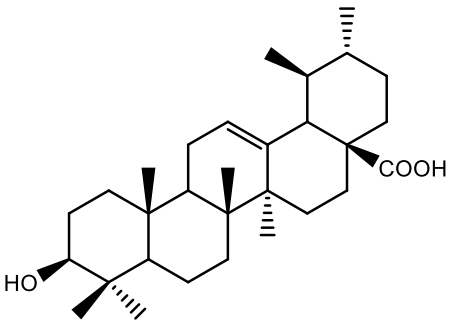
As mentioned in **Chapter 1**, inflammation is a physiological response to injury, infection, or toxins in the body and is implicated in multiple diseases when prolonged and/or uncontrolled [87]. The regulation of inflammation occurs through a balance of pro- and anti-inflammatory factors. The process of inflammation involves the release of pro-inflammatory cytokines such as TNFα and IL-6, and the measurement of secretion of these factors therefore provides a measure for determining inflammation within cells *in vitro*. The anti-inflammatory potential of a given compound could therefore be determined through assessing the secretion of cytokines by appropriately stimulated cells.

The regulation of inflammation occurs through transcription factors such as NFκB. As mentioned in **Chapter 1**, the current study focuses on the p50/p65 heterodimer form of NFκB, which regulates the transcription of pro-inflammatory cytokines such as TNFα and IL-6 [88,94]. Inhibition of the activation of this factor is therefore one potential mechanism for triterpene anti-inflammatory activity. The p65 subunit of NFκB is typically phosphorylated when activated. Measurement of the abundance of phosphorylated p65 therefore allows for the determination of NFκB activation in cell culture. Another relevant transcription factor, Nrf2, accumulates in cells when its inhibitor is degraded, activating an anti-inflammatory response and supressing expression of pro-inflammatory cytokines such as TNFα and IL-6 (see **Chapter 1** for the mechanism of activation) [96] As Nrf2 levels in cells under normal conditions are low, abundance of Nrf2 can be measured to determine its activation within cells.

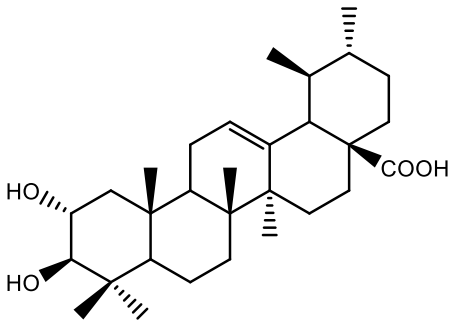
Naturally occurring triterpenes in general have weak anti-inflammatory activity, although there are studies to suggest they exert this activity through inhibition of NFκB, potentially by inhibiting IKK [89]. Triterpenes shown to act in this manner include ursolic and corosolic acid (**Figure 6-1**) [156,157]. More examples of naturally

occurring and semi-synthetic triterpenes that have shown biological activity, and their potential mechanisms of action, are discussed in **Chapter 1**.

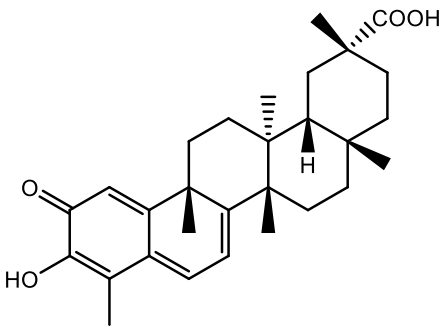
Naturally occurring triterpenes can show promising anti-inflammatory activity when heavily functionalised. For example, the friedelin derivative celastrol (**Figure 6-1**), produced by the medicinal plant *Tripterygium wilfordii*, inhibits TNFα and IL-6 cytokine secretion at nM concentrations [24]. This activity is significantly more potent than that seen for more simple triterpenes. However, the bioactivity of celastrol is broad and unfocused, with a broad spectrum of reported bioactivities rather than a focus on any specific application [158]. Understanding the features behind triterpene bioactivity, and defining which modifications are responsible for the activity of the most potent natural triterpenes, would enable the production of precisely engineered triterpene derivatives.



Ursolic Acid



Corosolic Acid



Celastrol

Figure 6-1: Triterpenes found to have anti-inflammatory effects.

6.2 Aims

In this chapter, the suite of molecules generated in **Chapter 5**, along with triterpenes from commercial sources, are subjected to biological testing through anti-proliferative and anti-inflammatory assays to determine properties such as IC_{50} values and cytokine release inhibition. Further investigations were carried out by Western blot analysis to investigate the potential mechanisms underlying observed anti-inflammatory effects, by measuring protein levels of relevant transcription factors to track their activation or repression during treatment with triterpenes. The insights gained through this study will inform the design of more bioactive triterpenes moving forward.

6.3 Results and Discussion

6.3.1 Investigating the Anti-Proliferative Activity of Triterpenes

Initial tests of biological activity of pharmaceutically relevant compounds often involve evaluation of their anti-proliferative effects: the ability to inhibit cell growth when incubated with cells. The anti-proliferative activities of the triterpenes described in **Chapter 5** were evaluated using the human cancer cell lines HL-60 (promyelocytic leukaemia) and THP-1 (monocytic leukaemia). These cell lines were chosen because of their ease of culture (they are non-adherent cells) and the availability of a substantial body of existing data for the HL-60 cell line relating to triterpenes, so allowing comparisons to be made with previous work.

The assay chosen to evaluate the anti-proliferative effects of the triterpenes in this study was the MTS assay (**Figure 6-2**). MTS (5-[3-(carboxymethoxy)phenyl]-3-(4,5-dimethyl-2-thiazolyl)-5-[(phenylamino)-carbonyl]-2H-tetrazolium inner salt), is a tetrazolium reagent that is reduced to a formazan product with a corresponding colour change in the presence of NADH (or NADPH) from actively respiring cells [159,160]. As MTS has low cell permeability, the reaction is assisted through the use of PES (phenazine ethyl sulphate) as an intermediate electron acceptor. The formation of the formazan product can be monitored by absorption at 490 nm, providing a simple means for investigating cell proliferation [160].

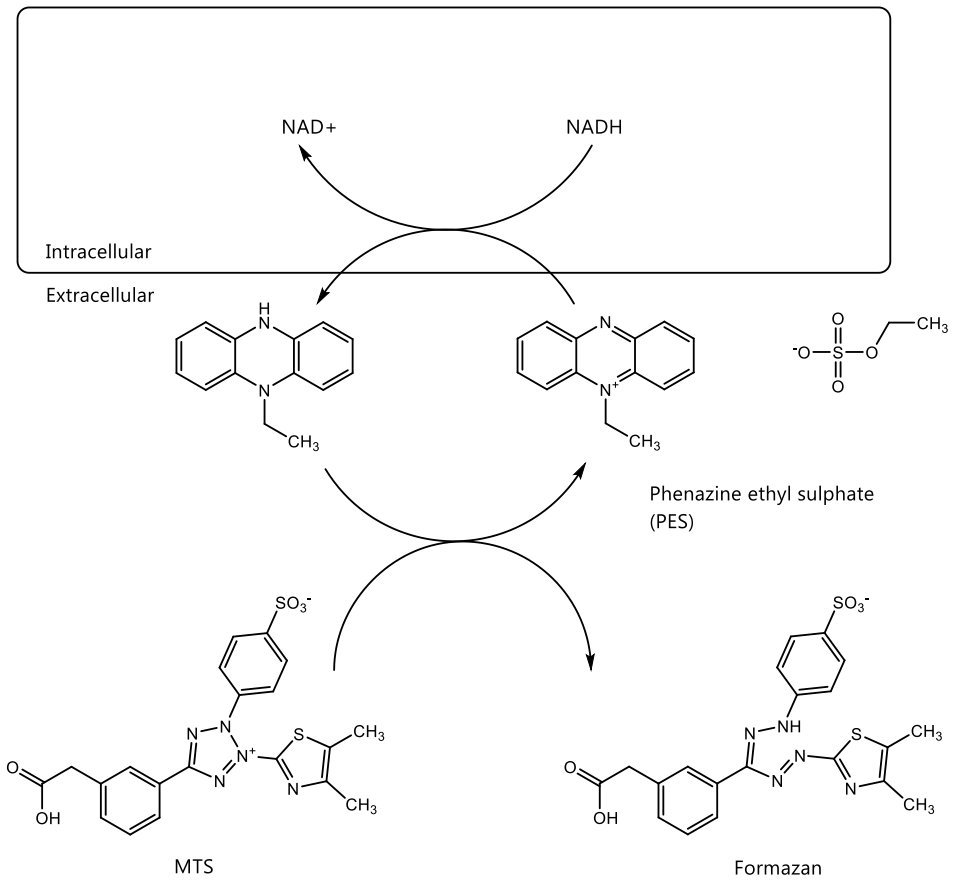


Figure 6-2: MTS assay used to detect actively respiring cells. The reduction of MTS using NADH (via PES) creates a formazan product that absorbs strongly at 490 nm. The relative abundance of living cells (in treated vs untreated cells) can be calculated by measuring absorbance of a cell culture plate at this wavelength.

Due to the lipophilicity of the triterpenes, dimethyl sulfoxide (DMSO) was selected as the solvent used to make stock solutions and serial dilutions (see **6.5.1**). An initial qualitative test involved stock solutions of 1 mM and 10 mM for each compound, allowing MTS assays to be run at two final concentrations: 10 μM and 100 μM. The solvent alone (DMSO) was used as a control. Cells were incubated for 72 hours in the presence of the compounds, and cell viability was then measured using the MTS assay (see **6.5.4**). The results are shown in **Table 6-1**.

Table 6-1: Inhibition of the tested triterpenes at two concentrations across two cell lines.

Compound	**Average Inhibition vs DMSO (%)**			
	HL-60		THP-1	
	100 µM	10 µM	100 µM	10 µM
β-amyrin (**1**)	-43%	7%	-69%	-66%
β-amyrone (**27**)	-17%	2%	14%	17%
Oleanolic Acid (**18**)	**89%**	17%	**94%**	9%
Erythrodiol (**40**)	**91%**	33%	**100%**	1%
11-oxo-β-amyrin (**2**)	**99%**	1%	**99%**	7%
24-hydroxy-β-amyrin (**22**)	-33%	-27%	-91%	-49%
12,13-β-epoxy-β-amyrin (**33**)	-5%	2%	-17%	-5%
30-hydroxy-β-amyrin (**19**)	-61%	-2%	-50%	-2%
16β-hydroxy-β-amyrin (**20**)	17%	-2%	**100%**	-7%
6β-hydroxy-β-amyrin (**34**)	**90%**	-5%	**100%**	0%
21β-hydroxy-β-amyrin (**23**)	-24%	-8%	-34%	-8%
Ellarinacin (**41**)	-29%	7%	**60%**	23%
Epi-β-amyrin (**28**)	-33%	-21%	NT	NT
11-oxo-epi-β-amyrin (**31**)	**77%**	-52%	NT	NT
24-hydroxy-11-oxo-epi-β-amyrin (**29**)	**98%**	19%	NT	NT
Maslinic Acid (**24**)	**91%**	-1%	NT	NT
Augustic Acid (**25**)	**53%**	34%	NT	NT
Echinocystic Acid (**37**)	**86%**	10%	NT	NT
7β-hydroxy-β-amyrin (**32**)	**100%**	8%	NT	NT
23-hydroxy-β-amyrin (**21**)	**66%**	-28%	NT	NT
3-O-Acetyl-α-boswellic Acid (**42**)	**98%**	-47%	NT	NT
α-boswellic Acid (**43**)	20%	-17%	NT	NT
β-boswellic Acid (**44**)	**93%**	-41%	NT	NT
3-O-Acetyl-β-boswellic Acid (**45**)	**95%**	-28%	NT	NT
11-oxo-β-boswellic Acid (**46**)	**78%**	-45%	NT	NT
3-O-Acetyl-11-oxo-β-boswellic Acid (**47**)	**95%**	-16%	NT	NT
Caulophyllogenin (**48**)	48%	-5%	NT	NT
Medicagenic Acid (**49**)	-5%	-19%	NT	NT
11-oxo-erythrodiol (**39**)	-27%	-2%	NT	NT
21β-hydroxy-11-oxo-β-amyrin (**38**)	**89%**	-11%	NT	NT

The percentages shown are the mean of 3 technical replicates only (each normalised to mean DMSO vehicle control). Compounds with >50 % inhibition at 100 μM, were selected for further investigation (shown in bold). All compounds were incubated for 72 h.

Ten of the tested triterpenes have little to no activity (<20 % inhibition) even at the highest concentration tested, and may even increase cell proliferation (although the apparently increased cell proliferation values may also be due to crystallisation of the tested triterpenes interfering with absorbance). These included β-amyrin (**1**), which has been shown to be inactive in previous tests in the O'Connell lab despite reports of its cytotoxic effects [161]. Additionally, some compounds showed possibly increased proliferation at 10 μM despite showing inhibition of proliferation at 100 μM. These include 3-O-Acetyl-α-boswellic acid (**42**) and 11-oxo-epi-β-amyrin (**31**). There is variability in the raw data for these compounds, potentially due to differences in pipetting, cell density, or the vehicle control between different wells. Repeating these experiments would determine whether there was a true pro-proliferative effect: this was not carried out due to the focus of the present study being on anti-proliferative activity. A further two compounds, 16β-hydroxy-β-amyrin (**20**) and ellarinacin (**41**), showed activity only towards the THP-1 cell line. The remaining sixteen compounds showed significant activity towards the HL-60 cell line at the highest concentration tested.

IC_{50} values were calculated for those compounds that showed at least 50 % inhibition in at least one cell line- the required concentration for a test compound to inhibit cell growth by 50 % relative to controls, which acts as a useful point of comparison between molecules. A dilution series was made for each compound, such that the final assay concentrations were between 1 μM and either 100 μM or 200 μM depending on solubility in DMSO. DMSO alone was used as the control. Cells were incubated for 72 hours in the presence of the compounds and cell viability then measured using the MTS assay (see **Section 6.5.4**). The IC_{50} values are shown in **Table 6-2**.

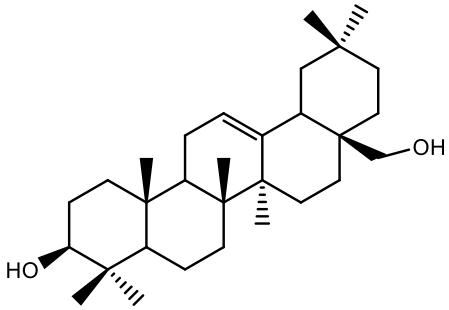
Table 6-2: IC_{50} values for the tested triterpenes.

Compound	**HL-60**	**THP-1**
Doxorubicin	0.01	NT
Erythrodiol (**40**)	14.32 ± 2.46	62.76 ± 18.04
3-O-Acetyl-11-oxo-β-boswellic Acid (**47**)	15.62 ± 5.48	NT
Echinocystic Acid (**37**)	16.14 ± 4.70	NT
Augustic Acid (**25**)	16.52 ± 4.64	NT
6β-hydroxy-β-amyrin (**34**)	17.56 ± 3.68	49.89 ± 2.40
24-hydroxy-11-oxo-epi-β-amyrin (**29**)	18.82 ± 3.83	NT
23-hydroxy-β-amyrin (**21**)	19.14 ± 2.04	NT
11-oxo-β-amyrin (**2**)	19.66 ± 4.68	60.68 ± 2.86
Maslinic Acid (**24**)	24.12 ± 3.30	NT
3-O-Acetyl-α-boswellic Acid (**42**)	24.22 ± 1.02	NT
Oleanolic Acid (**18**)	25.57 ± 1.62	49.52 ± 0.61
7β-hydroxy-β-amyrin (**32**)	28.13 ± 6.81	NT
α-boswellic Acid (**43**)	30.16 ± 5.89	NT
11-oxo-epi-β-amyrin (**31**)	30.44 ± 2.03	NT
3-O-Acetyl-β-boswellic Acid (**45**)	36.32 ± 6.40	NT
β-boswellic Acid (**44**)	53.99 ± 2.83	NT
11-oxo-β-boswellic Acid (**46**)	55.22 ± 3.64	NT
21β-hydroxy-11-oxo-β-amyrin (**38**)	66.04 ± 6.66	NT
Caulophyllogenin (**48**)	78.10 ± 13.07	NT
16β-hydroxy-β-amyrin (**20**)	NT	25.64 ± 0.21
Ellarinacin (**41**)	NT	27.51 ± 0.99

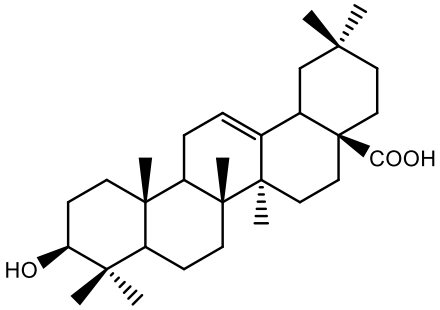
IC_{50} values (μM) are shown for the two cell lines used. Values represent the mean of 3 biological replicates (each was calculated using 3 technical replicates by a log (inhibitor) vs response-variable slope curve), ± SD, except for doxorubicin, which is for 3 technical

replicates only. 23 hydroxy-β-amyrin and 11-oxo-epi-β-amyrin were assayed by Maya Valmiki.

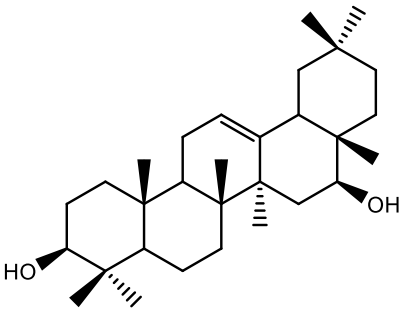
Many of the tested triterpenes show IC_{50} values within a range of values (~20-50 μM) which is common among many triterpenes previously tested despite varying levels of oxidation, indicating that more oxidised or 'complicated' structures do not always have improved activity. Erythrodiol (**40**) (**Figure 6-3**) had the strongest anti-proliferative activity in the HL-60 cell line, consistent with previous results [50], but had the weakest activity in the THP-1 cell line out of all of the triterpenes tested. 16β-hydroxy-β-amyrin (**20**) (**Figure 6-3**) meanwhile shows the strongest activity in the THP-1 cell line but is inactive in the HL-60 cell line. Two new positions on the β-amyrin scaffold with anti-proliferative activity have therefore been discovered: C-6β and C-7β (**Figure 6-3**).



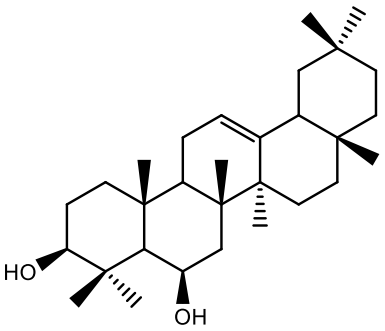
Erythrodiol **40** IC_{50} = 14.32 (HL-60)



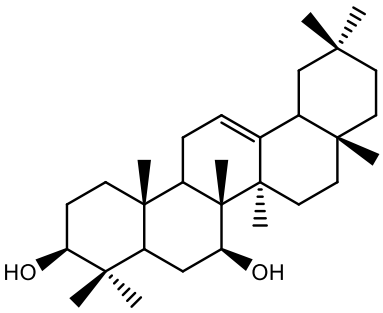
Oleanolic Acid **18** IC_{50} = 25.57 (HL-60)



16β-hydroxy-β-amyrin **20** IC_{50} = 25.64 (THP-1)



6β-hydroxy-β-amyrin **34** IC_{50} = 17.56 (HL-60)

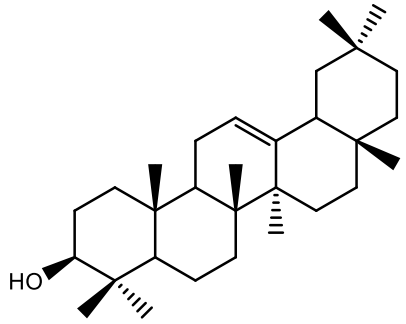


7β-hydroxy-β-amyrin **32** IC_{50} = 28.13 (HL-60)

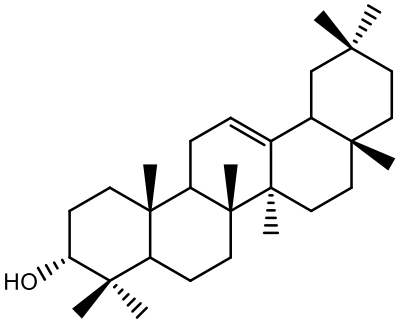
Figure 6-3: Selected compounds found to have an anti-proliferative effect.

These results also give broader insights into the structure-activity relationships of the triterpenes. C-3 epimerisation generally has little effect on anti-proliferative activity: neither β-amyrin (**1**) nor epi-β-amyrin (**28**) show anti-proliferative effects (**Figure 6-4**), and 11-oxo-epi-β-amyrin (**31**) has weaker activity than 11-oxo-β-amyrin (**2**) **(Figure 6-4**). However, 24-hydroxy-11-oxo-epi-β-amyrin (**29**) has anti-proliferative activity effects towards the HL-60 cell line, while 24-hydroxy-11-oxo-β-amyrin (**30**) does not (**Figure 6-4**) [50]. This compound also compares favourably with the boswellic acids, closely related molecules differing mainly in the further

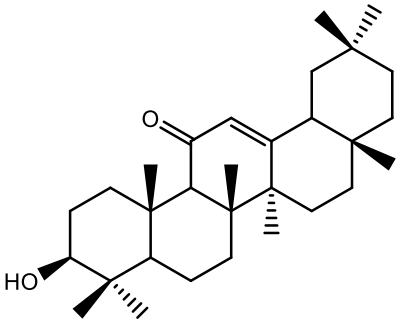
oxidation of C-24 to a carboxylic acid. 24-hydroxy-11-oxo-epi-β-amyrin (**29**) shows stronger anti-proliferative activity than most of the boswellic acids, suggesting that oxidation of C-24 to a carboxylic acid is not required for potent activity, a suggestion backed up by erythrodiol's (**40**) greater potency compared to oleanolic acid (**18**) (**Figure 6-3**).



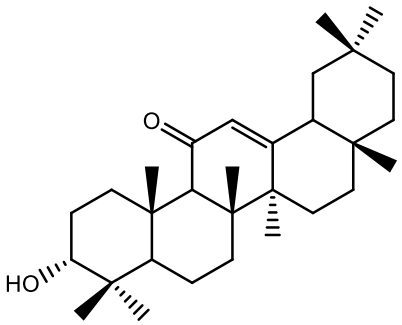
β-amyrin **1** (inactive)



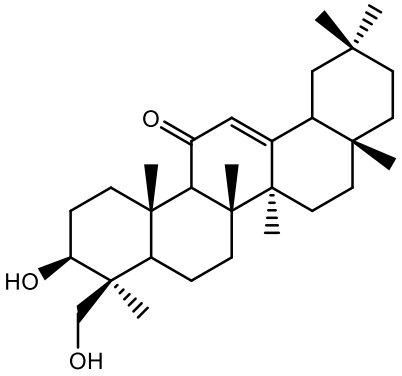
epi-β-amyrin **28** (inactive)



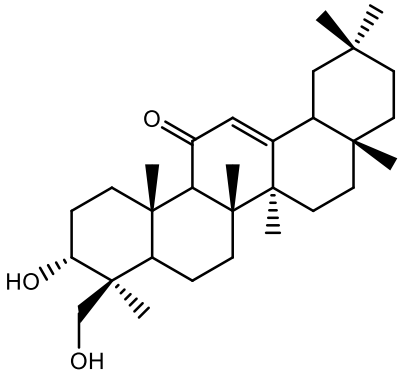
11-oxo-β-amyrin **2** IC_{50} = 19.66 (HL-60)



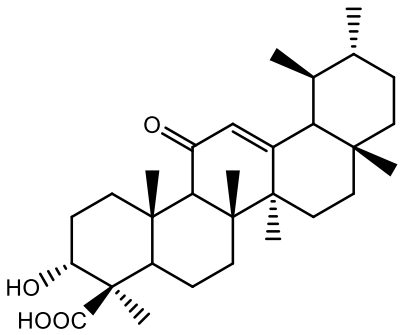
11-oxo-epi-β-amyrin **31** IC_{50} = 30.44 (HL-60)



24-hydroxy-11-oxo-β-amyrin **30** (inactive)



24-hydroxy-11-oxo-epi-β-amyrin **29**
IC_{50} = 18.82 (HL-60)



11-oxo-β-boswellic acid **46** IC_{50} = 55.22 (HL-60)

Figure 6-4: C-3 epimers investigated for anti-proliferative effects. 11-oxo-β-boswellic acid (**46**) is also included for reference.

Previously, it has been suggested that triterpenes based on the β-amyrin scaffold show stronger activity than those based on the α-amyrin scaffold [162]. This assertion is backed up by the current results, which show that α-boswellic acid (**43**), based on β-amyrin, has a lower IC_{50} value than β-boswellic acid (**44**), based on α-amyrin. 3-O-acetyl-α-boswellic acid (**42**) also shows greater potency than 3-O-acetyl-β-boswellic acid (**45**). 3-O-acetyl-11-oxo-β-boswellic acid (**47**), however, has a lower IC_{50} value than any of the other boswellic acids, suggesting the importance of both the 11-oxo group and C-3 acetylation in boswellic acid bioactivity. Future experiments to evaluate C-3 acetylated β-amyrin-derived triterpenes, especially 24-hydroxy-11-oxo-epi-β-amyrin, would provide confirmation of the importance of this modification to the bioactivity of boswellic acid-like compounds.

Some effects of C-2 stereochemistry on anti-proliferative effects were also observed. Augustic acid (**25**), with a C-2β stereochemistry, shows greater anti-proliferative effects than maslinic acid (**24**), with C-2α stereochemistry. This insight could also be taken forwards in the future design of more potent molecules.

6.3.2 Investigating Anti-Inflammatory Activity of Triterpenes

As inflammation is another important target of drug-related research, the anti-inflammatory effects of the triterpenes in this study were evaluated using an enzyme-linked immunosorbent assay (ELISA) to measure the secretion of TNFα and IL-6 in THP-1 cells. LPS stimulation of THP-1 cells has been previously used as an *in vitro* model for assessing inflammation in the O'Connell lab. Monocytes are recruited during the inflammatory response and play a key role in inflammation. As a monocytic cell line, THP-1 provides a comparable response to stimulation to primary monocytes [163]. LPS is used to induce a response as it known to be a potent monocyte activator [163,164]. TNFα and IL-6 were investigated because of their key roles as pro-inflammatory cytokines released during the inflammatory response.

To begin with, 12 triterpenes were incubated with THP-1 cells at one of two concentrations (20 μM for triterpenes showing no anti-proliferative effects towards

the THP-1 cell line; and 10 µM for triterpenes showing some anti-proliferative effects). Additionally included was 10 µm Bay 11-7082 (a known anti-inflammatory compound that inhibits IKK and therefore blocks NFκB activation and subsequent cytokine release [165]) as a positive control. All of the added compounds were incubated for 30 minutes prior to LPS stimulation. After 3 hours, supernatants were analysed for TNFα release by ELISA (see **6.5.7**, **Figure 6-5**).

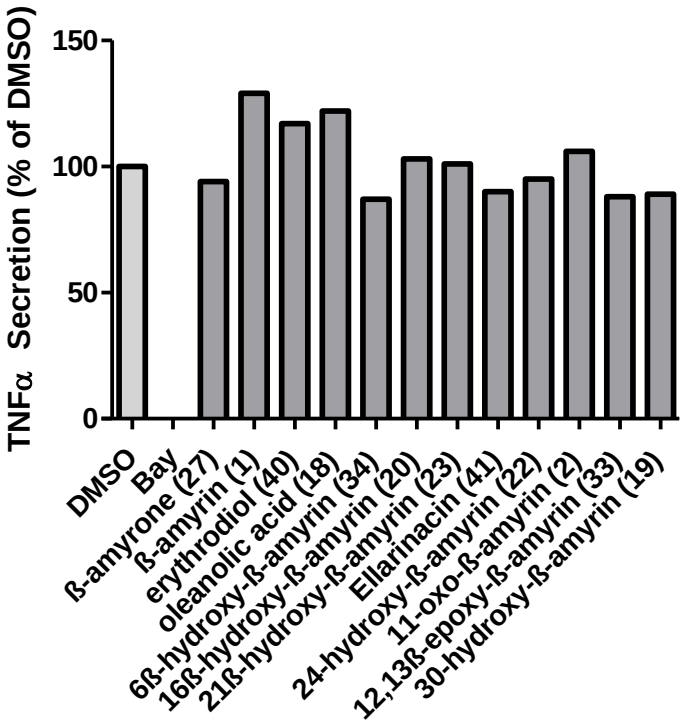


Figure 6-5: TNFα secretion in LPS-stimulated THP-1 cells incubated with the tested triterpenes. Cells were pre-incubated with triterpenes (20 µM except for 6β-hydroxy-β-amyrin (**34**), ellarinacin, 16β-hydroxy-β-amyrin (**20**), 11-oxo-β-amyrin (**2**), erythrodiol (**40**), and oleanolic acid (**18**), which were used at 10 µM) prior to LPS stimulation for 3 hours. TNFα secretion was quantified by ELISA and is shown relative to DMSO (vehicle) treated cells. No error bars are shown due to the experiment being performed with only one biological replicate. Experiment was not repeated due to lack of observed activity, which was corroborated by previous results from the O'Connell lab.

The results show that the triterpenes tested had little effect on LPS-induced TNFα secretion in THP-1 cells. This is in line with previous results obtained by the O'Connell lab under comparable conditions which included the additional triterpenes not tested here (personal communication). Several of the triterpenes tested were then selected for use as negative controls in tests of a wider range of triterpenes (**Figure 6-6**).

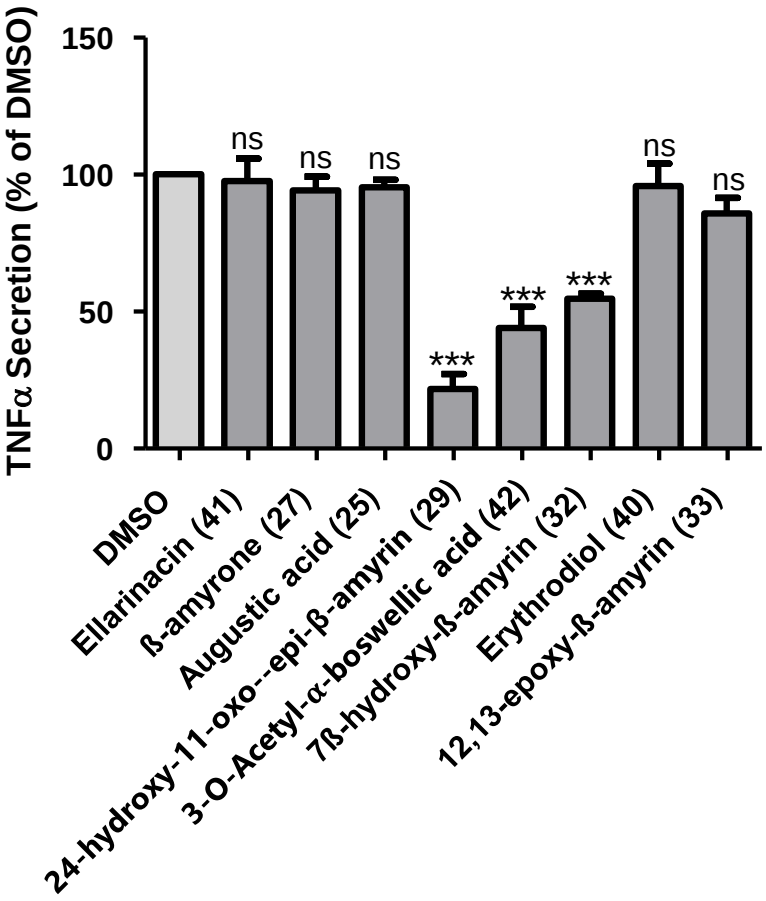


Figure 6-6: TNFα secretion in LPS-stimulated THP-1 cells incubated with the tested triterpenes. Cells were pre-incubated with triterpenes (20 μM) prior to LPS stimulation for 3 hours. TNFα production was quantified by ELISA and is shown relative to DMSO (vehicle) treated cells. Values represent the mean of 3 biological replicates ± SD. Significance values (calculated by ANOVA) are denoted as follows: P<0.001(***) or not significant (ns).

Three of the tested triterpenes significantly inhibited LPS-induced TNFα secretion in THP-1 cells: 24-hydroxy-11-oxo-epi-β-amyrin (**29**), 3-O-acetyl-α-boswellic acid (**42**), and 7β-hydroxy-β-amyrin (**32**), suggesting that these compounds may have anti-inflammatory effects. The inhibition of TNFα secretion by 3-O-acetyl-α-boswellic acid (**42**) is especially noteworthy since previous studies in the O'Connell lab had found that its β-boswellic acid counterpart, 3-O-acetyl-β-boswellic acid (**45**), did not inhibit TNFα secretion in THP-1 cells at 20 μM (personal communication). 3-O-acetyl-β-boswellic acid (**45**) does not appear to have been evaluated for TNFα secretion, however, its derivative 3-O-acetyl-11-oxo-β-boswellic acid (**47**) has been found to inhibit TNFα secretion in human peripheral blood mononuclear cells

(PMBCs) at 10 μM [18]. 3-O-Acetyl-α-boswellic acid (**42**) has also been previously reported to inhibit TNFα secretion in human peripheral monocytes at 10 μM [18].

The same set of triterpenes was then analysed by ELISA for IL-6 release. Compounds were incubated at 20 μM with THP-1 cells for 30 minutes before LPS stimulation and were then incubated for a further 24 hours before the supernatants were removed for assay by ELISA (see **6.5.8**). The results are shown in **Figure 6-7**.

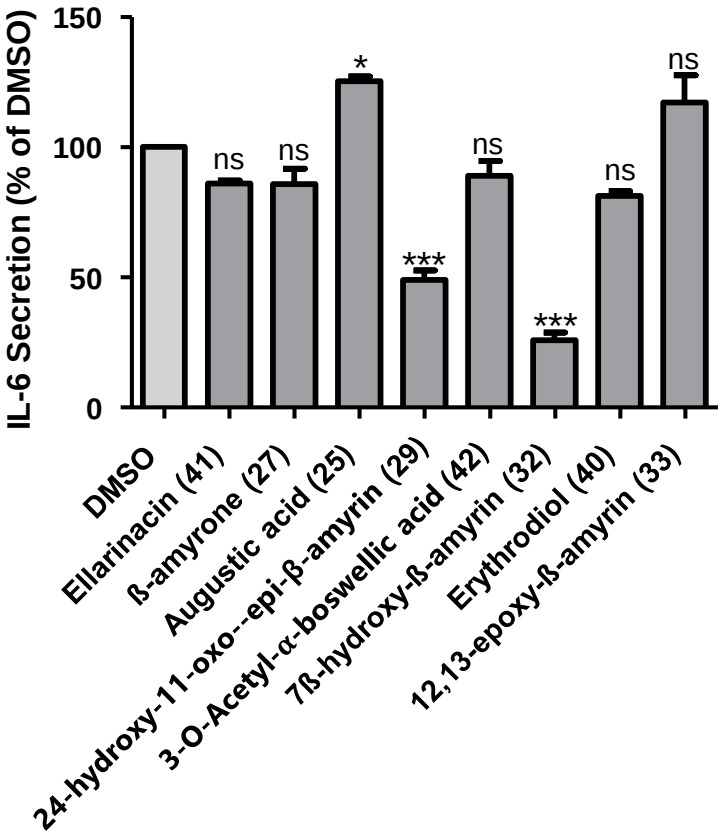


Figure 6-7: IL-6 secretion in LPS-stimulated THP-1 cells. Cells were pre-incubated with triterpenes (20 μM) prior to LPS stimulation for 24 hours. IL-6 secretion was quantified by ELISA and is shown relative to DMSO (vehicle) treated cells. Values represent the mean of 3 biological replicates ± SD. Significance values (calculated by ANOVA) are denoted as follows: $P<0.05$ (*), $P<0.001$(***), or not significant (ns).

Two of the triterpenes that were shown to inhibit release of TNFα also inhibited release of IL-6: 24-hydroxy-11-oxo-epi-β-amyrin (**29**) and 7β-hydroxy-β-amyrin (**32**), providing further evidence of potential anti-inflammatory effects. 24-hydroxy-β-amyrin (**30**), which is structurally related to 24-hydroxy-11-oxo-epi-β-amyrin (**29**), has also been found to inhibit IL-6 secretion,

but not TNFα secretion, at 20 μM in previous studies in the O'Connell lab (personal communication), although the effect was small.

However, as these triterpenes were also shown to be anti-proliferative towards the HL-60 cell line and had not been tested for anti-proliferative effects on the THP-1 cell line, they were further tested to ensure that observed effects were not due to cytotoxicity. The compounds were therefore incubated at 20 μM with THP-1 cells for 24 h, the longest timeframe used in the anti-inflammatory assays, before cell viability was measured by MTS assay (**Figure 6-8**).

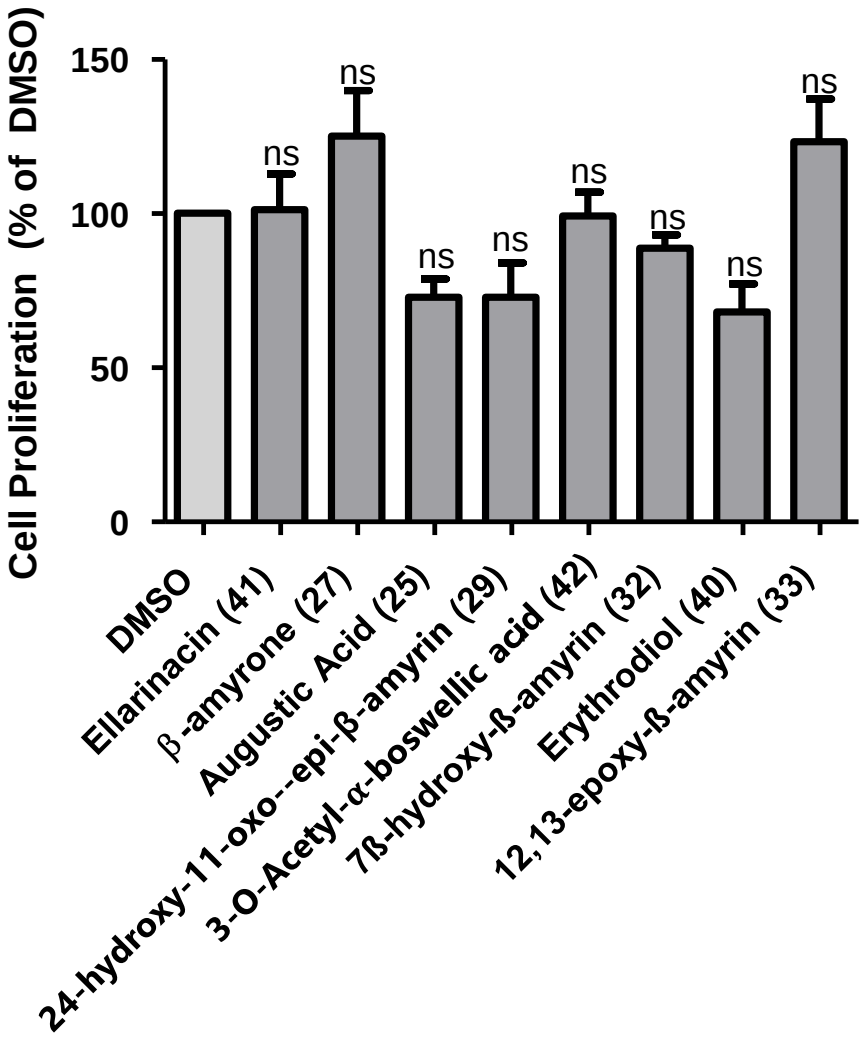


Figure 6-8: Impact of the tested triterpenes on cell proliferation at the concentration and timeframe relative to the anti-inflammatory testing. Compounds were tested at 20 μM in THP-1 cells for 24 hours prior to evaluation via MTS assay. Viability is shown relative to DMSO (vehicle control) Values are mean averages of three technical replicates ± SD. Significance values (calculated by ANOVA) are denoted as follows: not significant (ns).

The results show that there is no significant difference in cell proliferation when the compounds were incubated with THP-1 cells over the timeframe of the anti-inflammatory assays. Therefore, any changes in cytokine production are not a result of effects on cell viability.

6.3.3 Investigation of Potential Anti-Inflammatory Mechanisms by Analysis of Levels of Inflammation-Related Transcription Factors

To further investigate the mechanisms behind the effects of the tested triterpenes on cytokine release, the levels of inflammation-relevant transcription factors within triterpene-treated THP-1 cells were determined by Western blotting. One of these was Nrf2, which regulates the expression of cytoprotective proteins and when activated exhibits an anti-inflammatory effect [96]. The triterpenes to be tested were therefore incubated with THP-1 cells at 20 μM for 4 hours along with DMF (dimethyl fumarate, a known Nrf2 activator). Cells were harvested and lysed (see **Section 6.5.10**) and the protein levels of Nrf2 visualised by western blot (see **Section 6.5.12**). A representative blot and results of analysis of the intensity of the protein bands are shown in **Figure 6-9**.

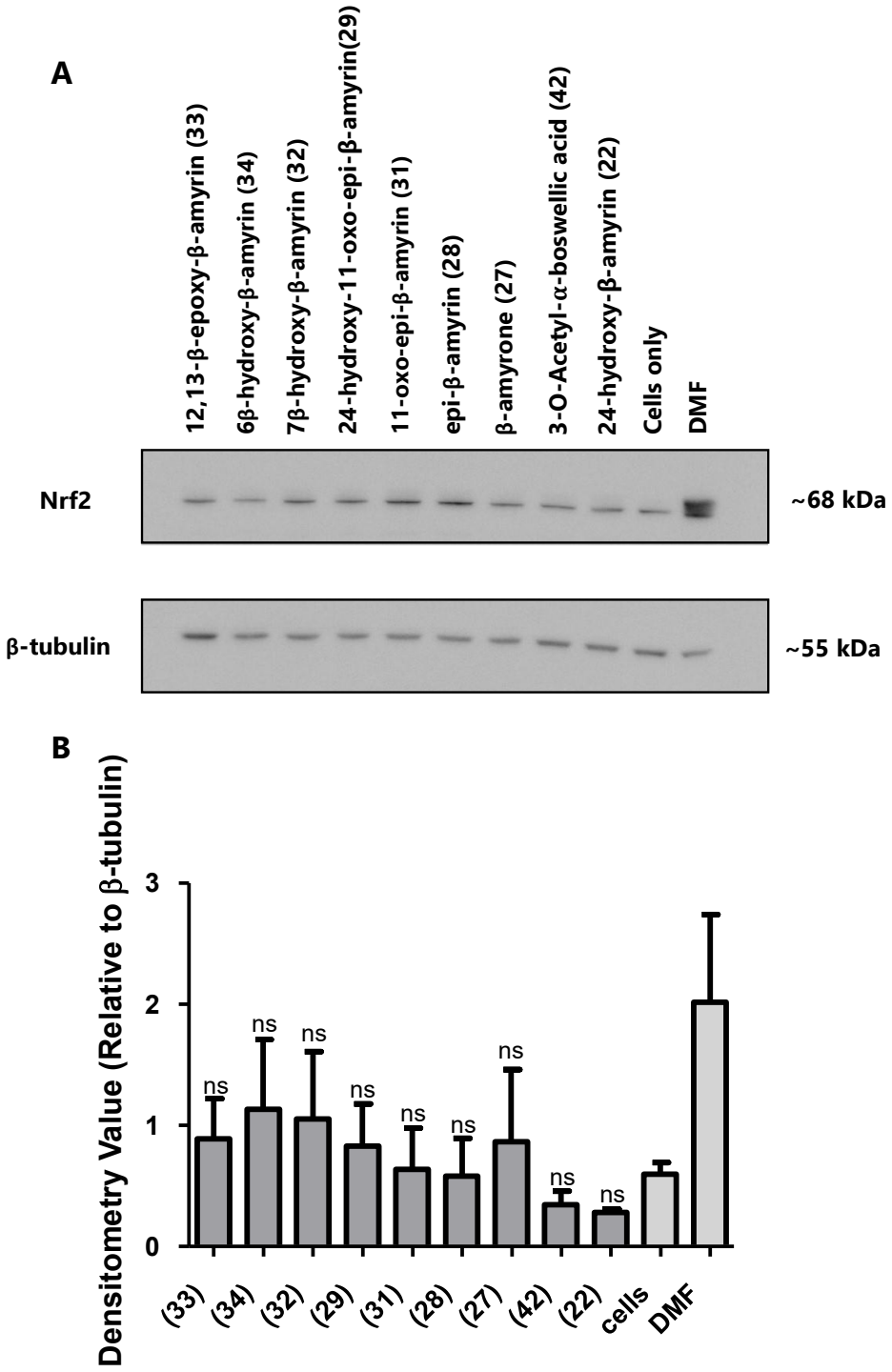


Figure 6-9: Western blot analysis for Nrf2. A) Representative blot for Nrf2, B) densitometric analysis. Cells were incubated with triterpenes (20 μM) for 4 hours before they were harvested and lysed. Extracts were then separated by SDS-PAGE and transferred onto a PVDF membrane, and proteins (Nrf2 and β-tubulin control) visualised by western blotting using chemiluminescence. Data are for three biological replicates ± SD. Significance values (calculated by ANOVA) are denoted as follows: not significant (ns).

As can be seen in the above figure, the triterpenes tested here showed no effect on levels of Nrf2 present in the THP-1 cells, suggesting that the potential anti-inflammatory effects are mediated via another mechanism.

Another transcription factor investigated for its role in inflammatory pathways is NFκB. This transcription factor is activated through phosphorylation in response to stimulation and then triggers a pro-inflammatory response including the release of cytokines such as TNFα and IL-6. [89,93] The triterpenes to be tested were therefore incubated with THP-1 cells for 30 minutes before LPS stimulation to induce NFκB activation. After 2 hours cells were harvested and lysed (see **6.5.10**) and levels of NFκB visualised by western blot analysis(see **6.5.12** and **6.5.13**). A representative blot and results of analysis of the intensity of the protein bands are shown in **Figure 6-10**.

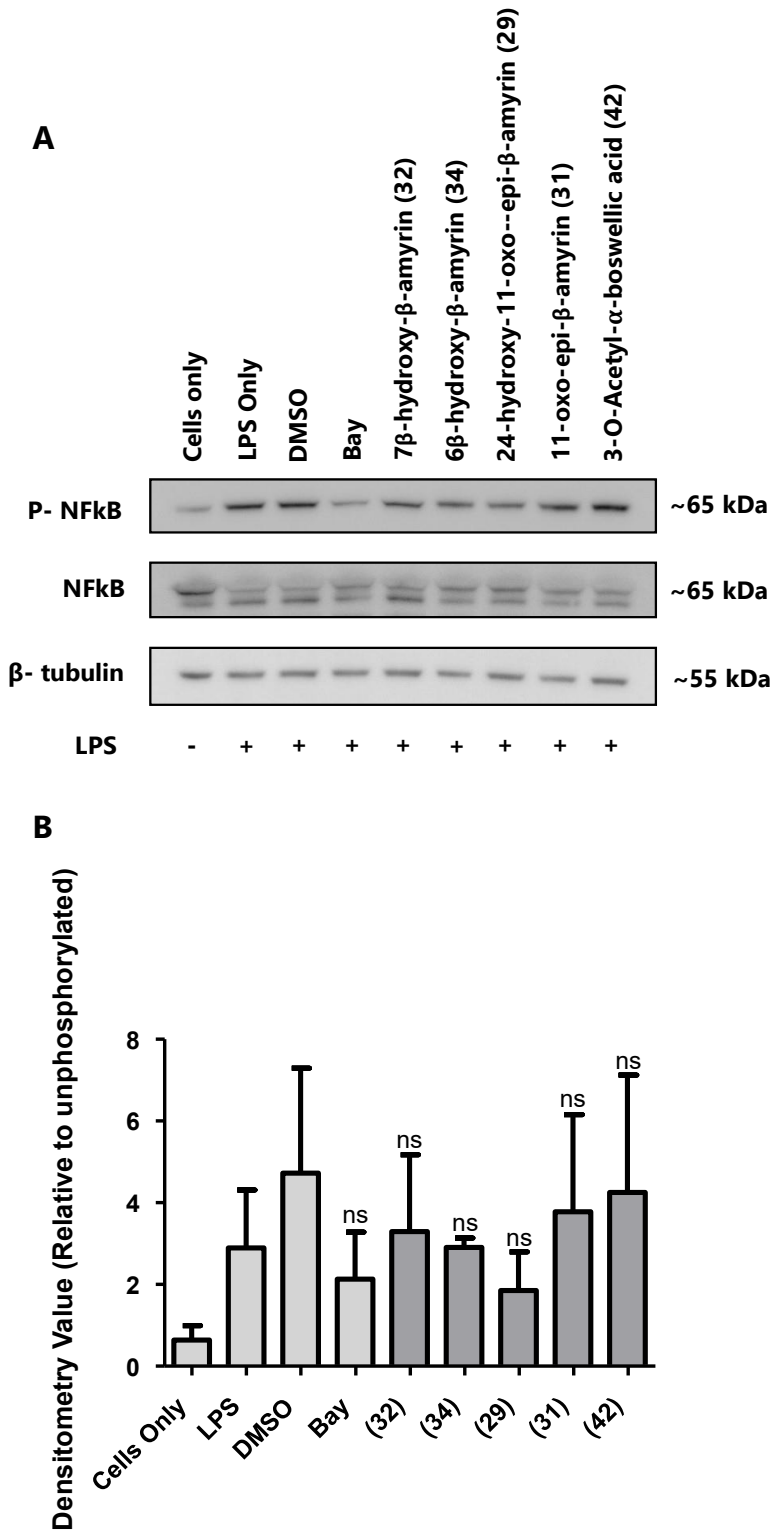


Figure 6-10: Western blot analysis of NfκB. A) Representative blot for NFκB, B) densitometric analysis. Cells were pre-incubated with triterpenes (20 μM) prior to LPS stimulation for 2 hours before cells were harvested and lysed. Extracts were then separated and transferred onto a PVDF membrane via western blotting, and proteins (p-NFκB, NFκB, β-tubulin) visualised using chemiluminescence. Data are for three biological replicates ± SD. Significance values (calculated by ANOVA) are denoted as follows: not significant (ns) vs vehicle control (DMSO).

In **Figure 6-10A** a visible difference can be seen for two of the triterpenes inhibiting the activation of NFκB compared to DMSO (vehicle control): 24-hydroxy-11-oxo-epi-β-amyrin (**29**) and 7β-hydroxy-β-amyrin (**32**), but this difference does not reach statistical significance. This could be due to the high variability of western blotting resulting in the assay being too insensitive to pick up small differences with statistical significance. These are the same triterpenes that inhibited production of both TNFα and IL-6, cytokines that are regulated by NFκB. 24-hydroxy-β-amyrin (**22**), which had been found to inhibit IL-6 secretion in a previous study in the O'Connell lab, was not found to inhibit NFκB in this way at 10 μM (personal communication). 3-O-acetyl-α-boswellic acid (**42**), which only inhibited TNFα secretion, does not appear to inhibit NFκB activation in this experiment, suggesting that it works via a different mechanism. Interestingly, the study on human peripheral monocytes which found that 3-O-acetyl-α-boswellic acid (**42**) inhibited TNFα secretion reported through a luciferase gene reporter assay in HEK293 cells that 3-O-acetyl-α-boswellic acid inhibited NFκB activation at 10 μM [18].

The Western blot results are not particularly clear and it was not possible to determine the statistical significance of any effects observed due to the variable nature of Western blotting. Future investigations of the anti-inflammatory nature of these compounds will utilise other methods to probe their mechanisms further, such as quantitative PCR to measure transcription levels of cytokines, or flow cytometry.

6.4 Conclusions

Here a suite of triterpenes based on the β-amyrin scaffold as detailed in **Chapter 5** were subjected to biological testing. Anti-proliferative screening of these molecules confirmed previously detailed activities and revealed two new modifications that confer anti-proliferative effects: the presence of a hydroxyl group at either C-6β or C-7β. Additional insights were also gained into the structure-activity relationships of the triterpenes, summarised in **Figure 6-11**.

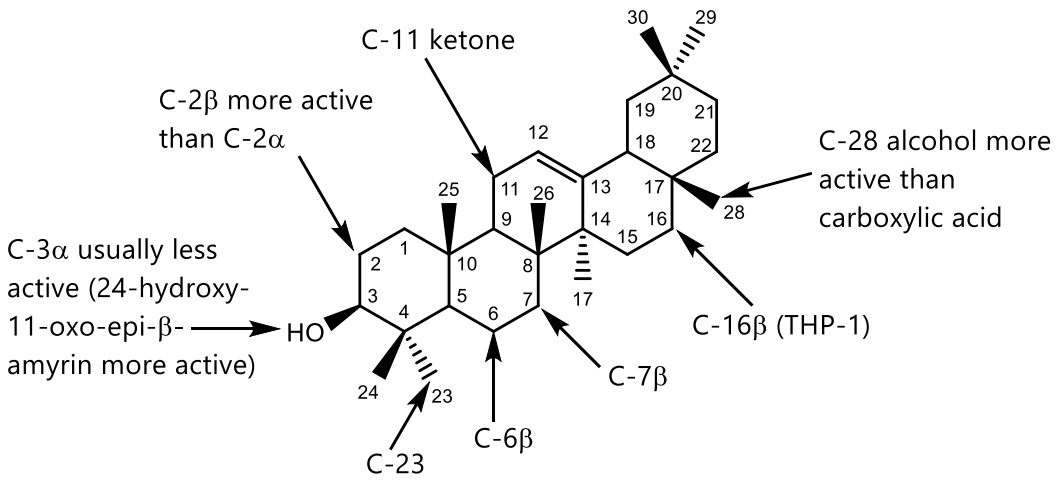


Figure 6-11: Summary of known and new modifications to the β-amyrin scaffold associated with triterpene activities. The positions on the scaffold where modifications confer anti-proliferative effects are indicated with arrows.

Anti-inflammatory assays revealed three molecules with potentially important activity: 7β-hydroxy-β-amyrin (**32**), 24-hydroxy-11-oxo-epi-β-amyrin (**29**), and 3-O-acetyl-α-boswellic acid (**42**). Further evaluation revealed that these activities were not associated with anti-proliferative effects, and western blotting suggested a potential mechanism of action for 7β-hydroxy-β-amyrin (**32**) and 24-hydroxy-11-oxo-epi-β-amyrin (**29**) through inhibition of NFκB. Previous reports of the mild anti-inflammatory activity of 24-hydroxy-β-amyrin (**22**), and the more potent anti-inflammatory activities of multiple α- and β-boswellic acid derivatives [19,85], suggest the importance of a combination of modifications at C-24 along with the C-3 epimerisation of triterpenes in anti-inflammatory activity. More research is needed to determine the exact contribution of each of these modifications to the observed activities. Additionally, the conflicting results between the current study and literature reports relating to the ability of 3-O-acetyl-α-boswellic acid (**42**) to inhibit NFκB highlight the need for further research into the mechanisms of action of these molecules.

Taken together, these results will feed into the design and production of more potent triterpenes in the future.

6.5 Materials and Methods for Chapter 6

6.5.1 Preparation of Compounds for Assays

Stock solutions of each of the compounds were made in sterile DMSO. Where possible, these were made to 20 mM, but some were used at 10 mM due to poor solubility. Stock solutions were vortexed vigorously to ensure solubility prior to dilutions and use. A series of six or seven stocks were made for each compound by dilutions in DMSO to give stock concentrations ranging from 0.1 mM to either 20 mM or 10 mM depending on the solubility of the compounds in DMSO. These stock compounds were used as detailed below in the relevant assays.

Table 6-3: Compounds used in the biological assays.

Compound	Formula	Mass	Stock	Supplier
β-amyrin	$C_{30}H_{50}O$	426.7	10 mM	Extrasynthese
β-amyrone	$C_{30}H_{48}O$	424.7	10 mM	Chemfaces
Erythrodiol	$C_{30}H_{50}O_2$	442.7	10 mM	Extrasynthese
Oleanolic Acid	$C_{30}H_{48}O_3$	456.7	20 mM	Extrasynthese
11-oxo-β-amyrin	$C_{30}H_{48}O_2$	440.7	10 mM	J. Reed
24-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	10 mM	M. Stephenson
12,13-β-epoxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	10 mM	-
30-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	20 mM	-
16β-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	10 mM	-
6β-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	10 mM	-
21β-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	20 mM	-
Ellarinacin	$C_{30}H_{46}O_3$	454.7	20 mM	G. Polturak
Epi-β-amyrin	$C_{30}H_{50}O$	426.7	20 mM	-
11-oxo-epi-β-amyrin	$C_{30}H_{48}O_2$	440.7	20 mM	-
24-hydroxy-11-oxo-epi-β-amyrin	$C_{30}H_{48}O_3$	456.4	20 mM	-
Maslinic Acid	$C_{30}H_{48}O_4$	472.7	20 mM	Extrasynthese
Augustic Acid	$C_{30}H_{48}O_4$	472.7	20 mM	-
Echinocystic Acid	$C_{30}H_{48}O_4$	472.7	20 mM	Extrasynthese
7β-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	20 mM	-
23-hydroxy-β-amyrin	$C_{30}H_{50}O_2$	442.7	20 mM	-
3-O-Acetyl-α-boswellic Acid	$C_{32}H_{50}O_4$	498.4	20 mM	Extrasynthese
α-boswellic Acid	$C_{30}H_{48}O_3$	456.4	20 mM	Extrasynthese
β-boswellic Acid	$C_{30}H_{48}O_3$	456.4	20 mM	Extrasynthese
3-O-Acetyl-β-boswellic Acid	$C_{32}H_{50}O_4$	498.4	20 mM	Extrasynthese
11-oxo-β-boswellic Acid	$C_{30}H_{46}O_4$	470.3	20 mM	Extrasynthese
3-O-Acetyl-11-oxo-β-boswellic Acid	$C_{32}H_{48}O_5$	512.4	20 mM	Extrasynthese
Caulophyllogenin	$C_{30}H_{48}O_5$	488.7	20 mM	Extrasynthese
Medicagenic Acid	$C_{30}H_{46}O_6$	502.7	20 mM	Extrasynthese

Suppliers of compounds are given for commercially sourced compounds and for those provided by other researchers. Where no supplier is given, compounds were purified from *N. benthamiana*. Molecular formulae, masses, and the highest stock concentrations for the compounds are also given.

6.5.2 Culture of Human Cell Lines

All cell lines were obtained from the European Collection of Authenticated Cell Cultures (ECACC). HL-60 and THP-1 cells were thawed and centrifuged at 2000 rpm for 5 minutes. The supernatant was discarded and 1 ml complete RPMI 1640 media (see **Section 2.5** for additives) added. Cells were added to 7 ml RPMI media in flasks. Cell cultures were maintained at 37 °C and 5 % CO_2. Cells were passaged twice weekly and used between passages 5 and 20.

6.5.3 Cell Density Measurements

Cell density was measured using a Neubauer haemocytometer. A 1:1 mixture of PBS buffer and cells was made in an Eppendorf tube. 10 µl of this solution was added to a Neubauer haemocytometer for counting under a cell microscope according to the manufacturer's instructions. After measuring, cells were adjusted to a density of 3×10^5 cells/ml (for MTS assays), 5×10^5 cells/ml (for preparation of samples for Nrf2 Western blot) or 1×10^6 cells/ml (for anti-inflammatory assay and preparation of samples for NfkB Western blot) on the day of the assay by dilution in RPMI 1640 media.

6.5.4 MTS Antiproliferation Assay

MTS assays were carried out in 96 well plates under sterile conditions. Both HL-60 and THP-1 cells were tested. Cells were aliquoted into 96 well plates at 100 µl per well to give a total of 3×10^4 cells/well. For each plate of a "qualitative" assay, up to 8 compounds were assayed in triplicate at 2 final concentrations: 100 µM and 10 µM. For determination of IC_{50} values, up to 2 compounds were assayed in triplicate across 7-8 final concentrations depending on solubility (1 µM, 5 µM, 10 µM, 25 µM, 50 µM, 100 µM, and 200 µM where solubility allowed). For determination of cell viability under the same conditions as the anti-inflammatory assays, up to 8 compounds per plate were assayed twice in triplicate at 20 µM final concentration. Untreated and vehicle (DMSO) treated controls were also included for reference. Doxorubicin positive control was assayed in triplicate using a 10-fold serial dilution from 100 µM to 0.01 nM. Three wells containing media only were included in order to subtract media absorbance during measurement.

Plates were returned to the incubator after the addition of compounds and maintained at 37 °C/5 % CO_2 for 24 hours (for the 20 µM assay only) or 72 hours (all other assays). 10 µl MTS reagent (Cell Titer 96 Aqueous non-radioactive assay, Promega) was then added to each well in the dark (excluding the outermost wells). Plates were returned to the incubator for 4 hours, after which absorbance at 492 nm was measured using a plate reader (POLARstar OPTIMA, BMG Labtech) and the averaged media only reading subtracted from all wells. Data analysis was performed using the GraphPad Prism software (Version 5). For the "qualitative" assays, corrected data were normalised against DMSO-treated cells to obtain % proliferation. IC_{50} values were generated by using non-linear regression (log of treatment concentration versus response, variable slope, least squares method). For the 24 h assay, statistical analysis was carried out via one-way analysis of variance (ANOVA) with post-hoc analysis performed by Tukey's multiple comparison test.

6.5.5 Procedure for Anti-Inflammatory Assay

THP-1 cells were grown as detailed previously. Cell density was measured and the cells were pelleted by centrifugation (1200 rpm, 5 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 500 µL per well to give $5x10^5$ cells per well. Test compounds in DMSO were added to the wells (10 µM or 20 µM final concentration) and incubated for 30 mins at 37 °C (5 % CO2). After this time, to stimulate cytokine production, LPS was added to the appropriate wells (10 ng/mL LPS final concentration in RPMI media for TNFα, 1 µg/mL LPS final concentration in RPMI for IL-6). Cells were incubated as above for a further 3 hours (TNFα) or 24 hours (IL-6). Supernatants were collected on ice and stored at -80 °C until needed.

6.5.6 ELISA Solutions

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

Phosphate Buffered Saline (PBS) solution

Per litre:

8 g NaCl, 1.16 g Na2HPO4, 0.2 g KH2PO4, 0.2 g KCl

pH was set to 7.0 for the TNF ELISA and 7.2-7.4 for the IL-6 ELISA

Coating buffer (TNFα)

Per litre (pH 9.5):

7.13 g $NaHCO_3$, 1.59 g Na_2CO_3

pH to 9.5 with 5 M NaOH

Coating buffer (IL-6)

PBS (see above)

Wash buffer

1 L PBS (as above), 500 µL Tween-20

Assay diluent (TNFα)

90 % Phosphate buffered saline solution (PBS), 10 % Foetal bovine serum (FBS)

pH to 7.0

Assay Diluent (IL-6)

1 % bovine serum albumin (BSA) in PBS (pH 7.2-7.4), filtered

Stop solution

2 N H_2SO_4

6.5.7 TNFα ELISA

Determination of TNFα secretion was carried out in 96 well plates using a BD OptEIA™ Human TNF ELISA set (BD Biosciences). Plates were coated in 100 µL of human TNF capture antibody diluted in coating buffer, sealed and stored overnight at 4 °C. Wells were aspirated three times with wash buffer, blocked with >200µL assay diluent and incubated at room temperature for 1 hour. Wells were washed again three times in wash buffer before addition of samples. 100 µL of supernatants from THP-1 cells were added to the plate. Supernatants derived from LPS-treated cells were first diluted three-fold in assay diluent 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 500 pg/mL to 7.8 pg/mL (diluted in PBS). This was used to generate the standard curve for each assay for

determination of TNFα secretion. Plates were incubated at room temperature for 2 hours and then washed five times in wash buffer. 100 µL of the detection antibody in assay diluent 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. 100 µL 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 50 µL stop solution. Plates were measured by absorption at both 450 nm and 570 nm and the value for 570 nm was subtracted from the value for 450 nm.

6.5.8 IL-6 ELISA

Determination of IL-6 secretion was carried out in 96 well plates using a Human IL-6 DuoSet ELISA kit (R+D Systems). Plates were coated in 100 µL of human IL-6 capture antibody diluted in coating buffer, sealed and stored overnight at room temperature. Wells were aspirated three times with wash buffer, blocked with >200µL assay diluent and incubated at room temperature for 1 hour. Wells were washed again three times in wash buffer before addition of samples. 100 µL of supernatants from THP-1 cells were added to the plate. An IL-6 standard (R+D Systems) was included across a two-fold serial dilution range from 600 pg/mL to 9.4 pg/mL (diluted in PBS). This was used to generate the standard curve for each assay for determination of IL-6 production. Plates were incubated at room temperature for 2 hours and then washed three times in wash buffer. 100 µL of the detection antibody in assay diluent was added before sealing and incubating the plate at RT for 2 hours. The plate was aspirated and washed three times in wash buffer. 100 µl of streptavidin-HRP (diluted in assay diluent) was added to each well and the plate was sealed and incubated in the dark for 20 minutes. After this period, the plate was washed again three times with wash buffer, leaving the wash buffer in the wells for at least 30 seconds each time. 100 µL of substrate solution (substrate reagent pack, R+D systems) was added to each well and left in the dark for 10 minutes before addition of 50 µL stop solution. Plates were measured by absorption at both 450 nm and 570 nm and the value for 570 nm was subtracted from the value for 450 nm.

6.5.9 Post-ELISA Data Analysis

Using the serial dilutions of TNFα and IL-6 standards provided with the ELISA kits, calibration curves for each assay were determined and these were used to convert the corrected absorbance values of test samples to pg/mL. The values were divided by the value for DMSO + LPS-treated control to show relative cytokine secretion. 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 using the GraphPad Prism software (Version 5).

6.5.10 Preparation of Samples for Western Blotting

THP-1 cells were grown as detailed previously. Cell density was measured and the cells were pelleted by centrifugation (1200 rpm, 5 min). Cells were resuspended at $1x10^6$ cells/mL (NFκB) or 5 $x10^5$ cells/mL (Nrf2) and aliquoted into 24 well plates at 1 mL (NFκB) or 2 mL (Nrf2) per well to give $1x10^6$ cells/well. For Nrf2 only the plate was incubated overnight at 37 °C (5 % CO_2) before 1 mL of media was removed from the top of each well. The test compounds in DMSO were then added to the wells (20 μM final concentration) and incubated for 30 mins at 37 °C (5 % CO_2). After this time, to stimulate the inflammatory response, LPS in RPMI media was added to the appropriate wells (1 μg/mL LPS final concentration) for NFκB only. Cells were incubated as above for a further 2 hours (NFκB) or 4 hours (Nrf2). Finally, cells were harvested by transferring cells to microcentrifuge tubes and centrifuging at 2000 rpm for 5 mins. Pellets were washed in PBS and centrifuged at 2000 rpm for 5 minutes. Pellets were re-suspended in 100 μL ice-cold lysis buffer (1 x Tris-glycine SDS (Novex, Invitrogen) in PBS) and boiled for 5 minutes before being stored at -80 °C until needed.

6.5.11 Western Blot Solutions

Running Buffer

40 mL 20 x MOPS (NuPage, Invitrogen), 760 mL DiH_2O

Transfer Buffer (10 x)

24.26 g Tris Base, 112.6 g Glycine, 1 g SDS

Make up to 1 L with DiH_2O

Transfer Buffer (1 x)

200 mL methanol, 100 mL 10 x transfer buffer, 700 mL DiH_2O

TBST (20 x)

Per litre:

48.4 g Tris base, 160 g NaCl, 62 ml 5 M HCl, 20 ml Tween-20

Adjust pH to 7.6 then make up to 1 L with DiH_2O

TBST (1 x)

Per litre:

50 ml 20x TBST, 950 ml DiH_2O

5 % Milk

Per litre:

50 g skimmed Milk powder (Marvel), 1 L 1x TBST

5 % BSA

Per litre:

50 g Bovine Serum Albumin, 1 L 1x TBST

Stripping Solution:

Per litre:

15 g glycine, 1 g SDS, 10 ml Tween-20, 990 ml DiH_2O

pH to 2.2 with 6 M HCl

6.5.12 Western Blotting

Extracts were taken from the -80 °C freezer, defrosted, and disrupted using a sonic dismembrator for 10 s. Protein concentration was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) by absorbance at 280 nm. Fresh tubes were labelled and 18 µL of each sample made up, with extra lysis buffer being used to adjust the concentration of higher-concentrated samples to be equal to the concentration of the least-concentrated sample. 2 µL 10X sample reducing agent

(NuPage, Invitrogen) was added to each tube and the samples centrifuged briefly before heating at 70-95 °C for 5 minutes prior to running on the gel.

A 4-12% Bis-Tris gel (Invitrogen) was removed from its packing, washed in DiH_2O, and locked into a gel chamber (Novex XCellSureLock Mini-Cell system, Invitrogen). The gel chamber was filled with running buffer to the top of the middle chamber and 2/3rds of the outer chamber. 500 µL antioxidant (NuPage, Invitrogen) was added to the middle chamber. The comb was removed from the gel and the wells washed by pipetting the buffer in the middle chamber up and down.

5 µL PageRuler pre-stained protein ladder (Thermo Scientific) was added to one well of the gel. Samples were centrifuged briefly and 18 µL of each sample added to the middle wells of the gel. Where there were fewer samples than wells, 20 µL lysis buffer was added to the remaining wells at the edges of the gel. The gel was run at 150 V for 5 minutes before the voltage was increased to 200 V until the blue band within each sample reached the bottom of the gel.

A PVDF membrane was washed in methanol for 30 s, rinsed in DiH_2O, and shaken in transfer buffer for 20 minutes. Spacing pads for the blotting module were soaked in transfer buffer to remove any air.

The gel was removed from the gel tank and the top and bottom cut off. The plastic of the gel casing was separated and the PVDF membrane applied to the side of the gel containing the protein. The gel was placed into the blotting module (Novex XCell II Blot Module, Invitrogen) with spacing pads (2 or 3 depending on the pads used) and a piece of filter paper on either side of the membrane and gel. The gel tank was drained, washed with DiH_2O, and the blotting module locked into the tank. The middle chamber of the tank was filled with transfer buffer to the bottom of the module's screws. The outer chamber was filled ¾ with DiH_2O. The tank was run at 33 V for 1 h.

The membrane was removed from the blotting module and the side containing protein marked. The membrane was rinsed with 1 x TBST and stirred with 5 % milk for 1 hour to block the membrane. The membrane was then stored in the fridge for future use.

6.5.13 Probing of Western Blot Membranes

As the molecular weight of the two proteins to be probed was far enough apart for both to be probed simultaneously, Nrf2 membranes to be probed were cut into two and the two pieces marked on the side of the membranes containing the proteins. NFκB membranes were incubated with p-NFκB rabbit anti-human monoclonal antibody (Cell Signalling Technology) at 1:500 dilution in 5 % BSA while the upper piece of Nrf2 membranes were incubated with Nrf2 rabbit anti-human monoclonal antibody (Affinity Biosciences) at 1:750 dilution in 5 % milk and the lower piece of Nrf2 membranes in β-tubulin rabbit anti-human monoclonal antibody (Cell Signalling Technology) at 1:1000 dilution in 5 % milk at 60 rpm for 1 hour 30 minutes at room temperature. All membranes were then washed three times with 5 % milk for 5 minutes per wash, and three times with 1x TBST for 5 minutes per wash. The membranes were then incubated with goat anti-rabbit HRP conjugated monoclonal antibody (Cell Signalling Technology) at 1:1000 dilution in 5 % milk at 60 rpm for 45 minutes at room temperature. Membranes were then washed three times in 5 % milk for 5 minutes per wash, and three times in 1x TBST for 5 minutes per wash. The membranes were then visualised with an electrochemiluminescence (ECL) HRP Chemiluminescent substrate reagent kit (Novex, Invitrogen) and an ImageQuant LAS4000 (GE healthcare) using chemiluminescence and epi-illumination. ECL reagent was incubated with the membranes for 1 minute prior to development of the membranes.

To probe for unphosphorylated NFκB and β-tubulin, the membrane was first incubated in stripping solution twice for 10 minutes, then washed with PBS twice for 10 minutes per wash, and 1x TBST twice for 5 minutes per wash. The membrane was blocked with 5 % milk for 30 minutes, then probed as per the method above using either NFκB rabbit anti-human monoclonal antibody (Cell Signalling Technology) at 1:500 dilution in 5 % BSA or with β-tubulin rabbit anti-human monoclonal antibody (Cell Signalling Technology) at 1:1000 dilution in 5 % milk.

Following visualisation of the membranes, optical density was calculated using ImageJ software and statistical analysis on the average of three biological replicates

carried out by a one-way ANOVA with post-hoc analysis performed by Tukey's multiple comparison test using the GraphPad Prism software (Version 5).

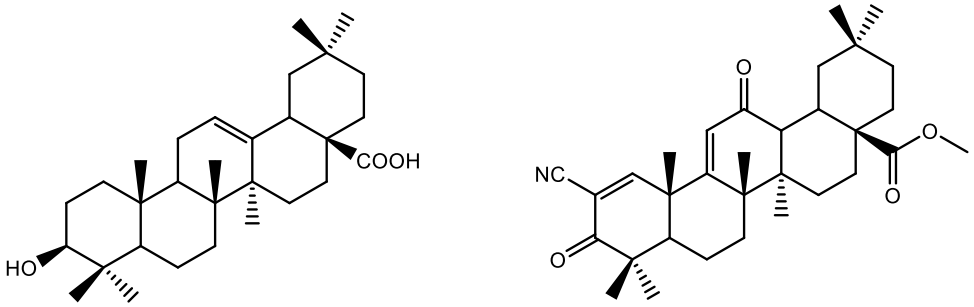
7 Structure-Activity Relationship Insights and Methods for Improving Throughput of Triterpene Bioactivity Investigation

7.1 Introduction

As discussed in **Chapter 6**, understanding the relationship between triterpene structure and function will be critical for the production of more potent molecules for use as pharmaceuticals. This would be best achieved through a systematic approach to evaluate triterpene structure-activity relationships. However, until now this has not been possible because the majority of triterpenes evaluated are sourced from their producing plants, rather than being designed and generated by metabolic engineering.

Some progress has been made using chemical modification of naturally occurring triterpenes. One such study of structure-activity relationships as they relate to anti-proliferative and anti-inflammatory effects (the two bioactive properties most relevant to the current study) was carried out in 2011 by the Sporn and Gribble laboratories, and led to the development of the semi-synthetic triterpene bardoxolone methyl [162] (see **Chapter 6**). This study involved generation of semi-synthetic derivatives of oleanolic and ursolic acid, and the evaluation of the bioactivities of these molecules, leading to the identification of bardoxolone methyl as a promising drug candidate [162]. Bioactivity was assessed using an assay for anti-inflammatory activity based on the inhibition of induction of inducible nitric oxide synthase (iNOS), although the authors also claimed in the same study a correlation between the results of this assay and observed anti-proliferative effects as well (the exact values of the anti-proliferative effects were not published). A summary of the structure-activity relationships of the compounds is shown in **Figure 7-1**.

A



Oleanolic Acid

Bardoxolone-methyl

B

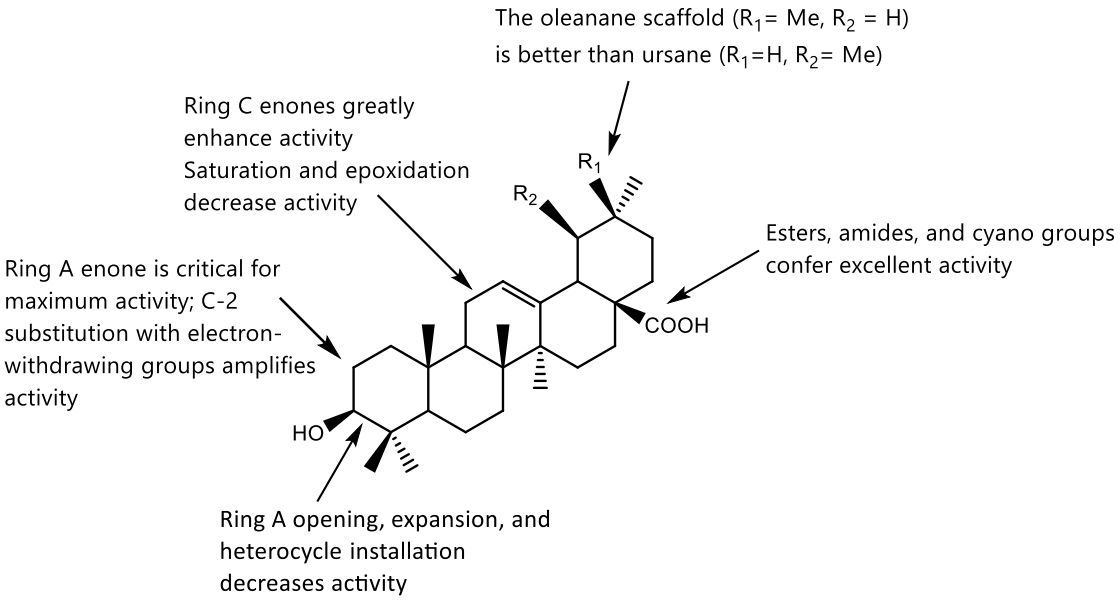


Figure 7-1: Bardoxolone-methyl was generated following a structure-activity relationship investigation of semi-synthetic oleanolic acid derivatives. A) Structures of oleanolic acid and bardoxolone-methyl. B) Summary of the findings from the structure-activity relationship investigation of oleanolic and ursolic acid derivatives that gave rise to the drug candidate bardoxolone methyl. Adapted from [162].

The insights gained in this study have proven largely representative, as conclusions made in other subsequent studies using molecules obtained from plant sources

have been generally in agreement with this [154,166]. This is not just the case for anti-proliferative and anti-inflammatory effects; one study echoed the finding that ester and amide groups at C-28 improved activity when applied to anti-viral effects as well [167].

One aspect of modification where significant insights can start to be gleaned is the glycosylation patterns of triterpenes. For instance, one study comparing four closely related saponins found that while glycosylation at the C-3 of oleanolic acid derivatives conferred potent anti-proliferative activity, additional glycosylation at C-28 greatly reduced anti-proliferative effects [168]. This is in agreement with previous studies in the O'Connell lab which have found that C-28 glycosylation of oleanolic acid abolishes activity (personal communication). Interestingly, the sugar chain at C-3 in this study consisted of an L-arabinose as the first sugar, with the four saponins (as seen in **Figure 7-2**) consisting of the monosaccharide, a disaccharide with D-glucose as the second sugar, a branched trisaccharide with both D-glucose and D-xylose attached to the arabinose, and a saponin with the branched trisaccharide at C-3 and an additional D-glucose at C-28. This branched trisaccharide chain bears some resemblance to those seen in naturally occurring molecules such as the avenacins [60] and could be an important clue to the bioactivity of saponins, as sugar chains containing only D-glucose at C-3 were also found not to be active (O'Connell lab, personal communication).

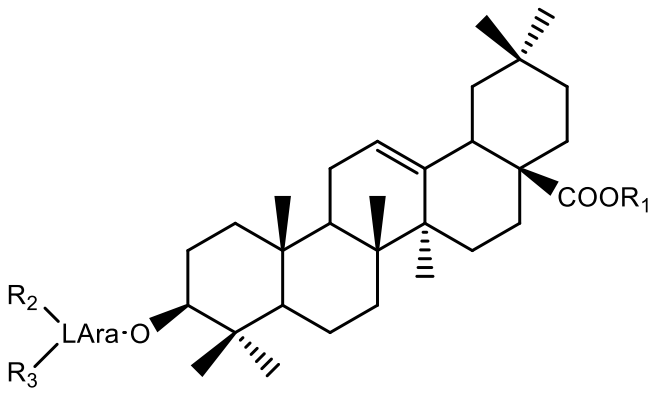


Figure 7-2: Structure of the saponins investigated in [168]. R_1 = D-Glc or H, R_2 = D-Glc or H, R_3 = D-Xyl or H.

As a general overview, the majority of research into structure-activity relationships, especially with regards to pentacyclic triterpenes, has consisted of functionalising either the C-3 or C-28 positions of oleanolic or ursolic acid. A large variety of functional groups have been added, from simple amides, methyl groups, or acetyl groups, to PEGylation, dimer formation, or the conjugation of heterocycles. The majority of these modifications have significantly increased the potency of oleanolic acid as an anti-proliferative triterpene [169]. However, these experiments consist of synthetic derivatisation of a natural molecule in a fashion that is highly unlikely to occur in nature, and therefore the utility of these studies from the perspective of triterpene metabolic engineering is limited. Such chemical modification strategies are also restricted because of the lack of availability of accessible 'handles' for chemical modification, since chemical derivatisation generally requires modifiable functional groups to already exist on the molecule. This limits the available diversity of structures compared to the rich source of potential bioactivities provided by natural products. Additionally, oleanolic acid, while abundant in nature, is only mildly anti-proliferative, and more potent examples of naturally occurring triterpenes exist. The breadth of molecules that have been biosynthesised using the triterpene toolkit in the current study, coupled with previous investigations of bioactivity in collaboration with the O'Connell laboratory, now opens up unprecedented opportunities to expand the investigations of the structure-activity relationships of triterpenes with the ultimate aim of generating new triterpenes with optimal bioactive properties using metabolic engineering approaches.

An improved understanding of triterpene structure-activity relationships will allow for the design of potent molecules with greater bioactivity, the testing of which will create even greater insights and feed into the design of yet more potent bioactive compounds. As this process is iterative, there is significant interest in increasing the throughput of triterpene testing through innovative new technology. This chapter therefore also explores the potential for using fractionated extracts from agro-infiltrated *N. benthamiana* leaves in bioassays.

Solid-phase extraction (SPE) is an extractive technique for purifying compounds, using similar principles to standard chromatography techniques, and has previously been used to increase the throughput and speed of investigations of biological activity [170]. The National Cancer Institute (NCI) programme for natural products discovery (NPNPD) has generated a pre-fractionated library of extracts from classical sources of natural products (bacteria, plants, and marine organisms) through use of an automated SPE system that splits organic and aqueous extracts into 7 fractions of decreasing polarity [171]. These fractions are publicly available in the form of 384-well plates for high-throughput screening. Preliminary testing of fractions from the library using the NCI-60 human tumour cell line panel identified 34 fractions with significant activity. Second stage HPLC fractionation of these fractions and dereplication of their active components led to the identification of 28 active compounds from the 34 active fractions [172]. Currently, 326,000 fractions are available for screening through this process; the eventual goal of the project is to generate up to 1,000,000. [173] A major limitation of this approach is the use of multiple diverse species as sources for the initial extracts. This makes extraction of bioactive molecules after an initial hit complicated due to the presence of different contaminants and matrix components that may vary depending on the source species, making a single extraction method inefficient. It also creates difficulties in distinguishing between single bioactive components, synergistic mixtures, and assay-interfering contaminants. It would therefore be beneficial to use a single species as a source of extracts in order to provide a stable background signal that can be easily controlled for. Small-scale biosynthesis of triterpene candidates by transient expression in *N. benthamiana* provides a route to overcoming these downsides, with the added advantage that the platform also enables preparative-scale production of discovered compound "hits".

Although the NPNPD relies upon an automated system for its high-throughput library creation, the principles of the project could be applied to the evaluation of products from the triterpene toolkit without the need for the same high-tech equipment. The average manifold for processing SPE cartridges via vacuum allows for between 8 and 20 samples to be processed in parallel, depending on size, but

this could still enable a substantial improvement in throughput when compared to purification and evaluation of individual compounds of interest one at a time. The development of SPE techniques to enable the testing of semi-purified extracts from agro-infiltrated *N. benthamiana* leaves could thus markedly accelerate our ability to rapidly evaluate triterpenes generated by combinatorial biosynthesis.

7.2 Aims

In this chapter, a comprehensive review of the available information about triterpene structure-activity relationships relating to anti-proliferative data is carried out, with a focus on pentacyclic triterpene scaffolds and the HL-60 cell line, as this is where the majority of data exists and thus where insights will be most impactful. Additionally, the capacity for triterpene biological testing is improved through the development of a semi-pure evaluation method for triterpene-containing *N. benthamiana* extracts utilising solid-phase extraction methods to allow for rapid screening of novel compounds for potential anti-proliferative effects.

7.3 Results and Discussion

7.3.1 Integration of Existing Triterpene Cell Viability Data for Analysis

More data has currently been obtained for cell viability than for anti-inflammation effects within the studies carried out in the O'Connell lab. This data had been generated in 11 individual small projects over the course of eight years, and a meta-analysis of the results had yet to be carried out. Therefore, cell viability data were focused on in the first instance to search for trends and structure-activity relationships. The data were first tabulated based on the current and previous results from the O'Connell lab. The tabulated data represented 135 individual tests carried out on 75 compounds over the course of the 11 projects. IC_{50} values for each compound were available for nine human cell lines of both adherent and non-adherent cell types, encompassing breast, lung, skin, bone, blood, and brain derived cells. All values over 100 μM were recorded as 100 μM for ease of data analysis, and a value of 100 μM considered "inactive" for the purpose of the current analysis.

A literature search was then carried out to find additional cell viability data for other triterpenes that were not included in the analyses carried out in the O'Connell Lab, as well as to add in missing data for triterpenes that had been included but were missing results in the relevant cell lines. The search was carried out by querying Web of Science in January 2021 for the topics "triterpenes" and "IC_{50}" and manually filtering the results to find the cell lines of interest. These new data were also incorporated into the table. As well as the IC_{50} values, each compound was categorised based on the nature of the scaffold, the number of modifications anticipated to be required to convert the base scaffold to the compound, the types of these modifications, and an overall classification of the compound (triterpene, saponin, acetylated triterpene, steroid, steroid saponin, acetylated steroid). Overall, 141 compounds were analysed. The full table is shown in **Section A.5**.

These data are summarised in **Figure 7-3**, **Figure 7-4, Figure 7-5,** and **Figure 7-6.**

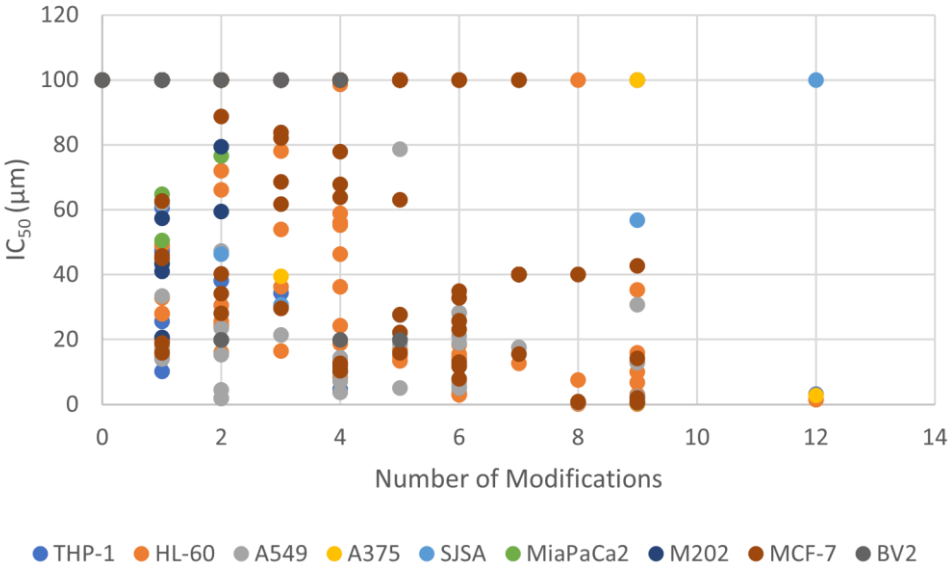
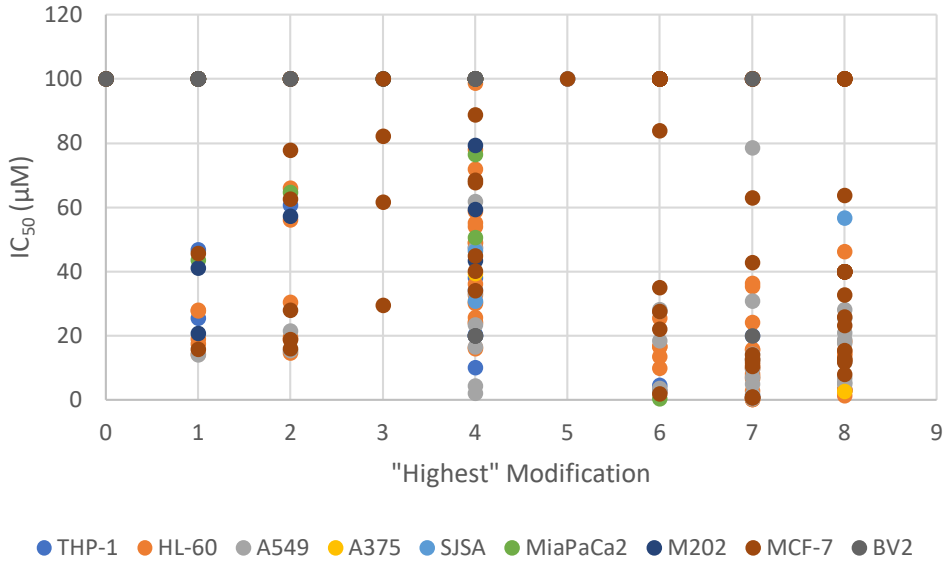


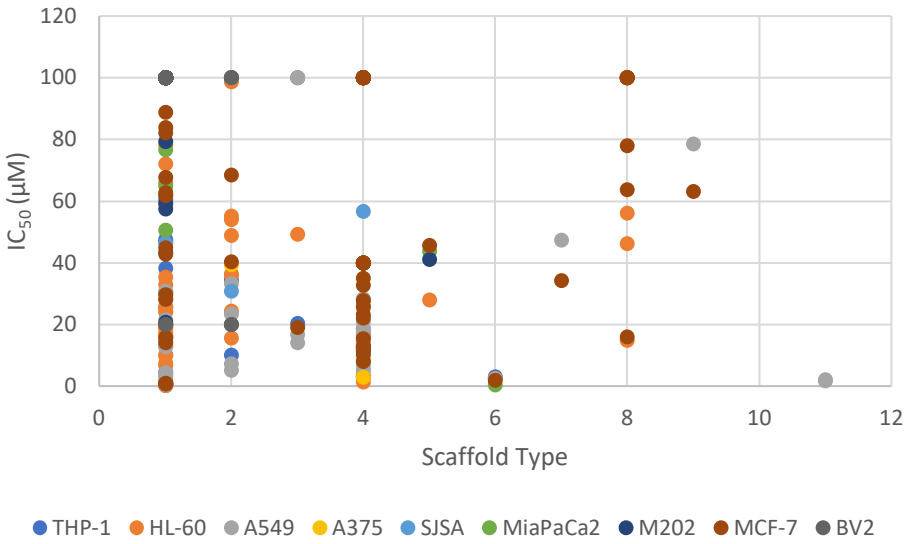
Figure 7-3: Total number of modifications to a triterpene scaffold(where an OSC product = 1 modification) vs IC_{50} value. Each dot represents one compound.



Key:

1 = hydroxylation
2 = aldehyde/ketone
3 = epoxidation
4 = carboxylic acid
5 = lactone
6 = glycosylation
7 = acylation
8 = other

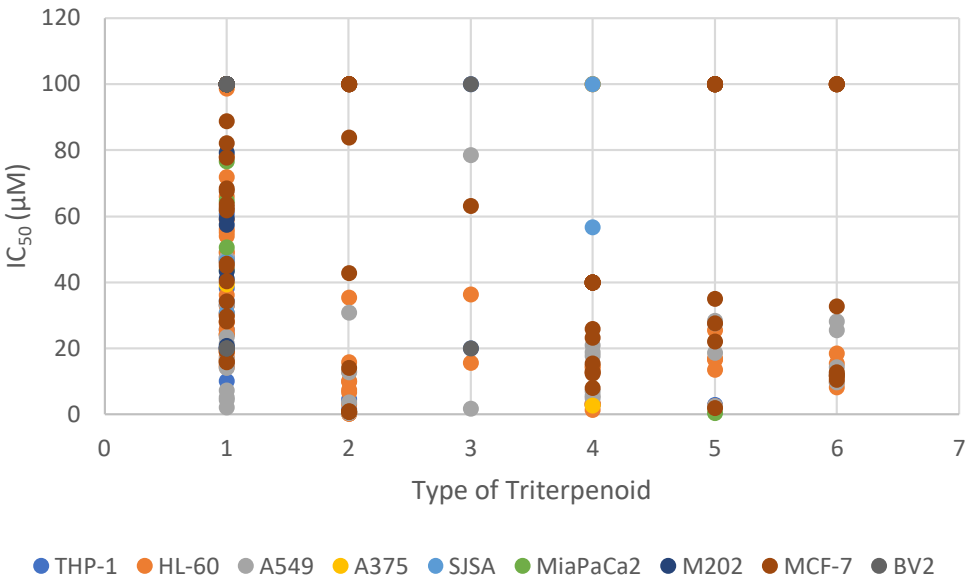
Figure 7-4: Compounds grouped by type of modification vs IC_{50} value. Each potential modification was given a rank and compounds were plotted based on the highest ranked modification they contain. Each dot represents one compound.



Key:

1 = β-amyrin
2 = α-amyrin
3 = lupeol
4 = lanosterol
5 = dammaranediol-II
6 = iso-lanosterol
7 = nor-lupeol
8 = iso-malabaricane
9 = dammardienol
10 = curcubitadienol
11 = germanicol

Figure 7-5: Compounds grouped by triterpene scaffold vs IC_{50} value. Each dot represents one compound.



Key:

1 = aglycone
2 = saponin
3 = acylated triterpene
4 = steroid
5 = steroid saponin
6 = acylated steroid

Figure 7-6: Compounds grouped by type vs IC_{50} value. Each compound was categorised based on whether it was a steroidal or non-steroidal triterpene, and whether it contained sugars (and therefore was a saponin) or was acylated. Each dot represents one compound.

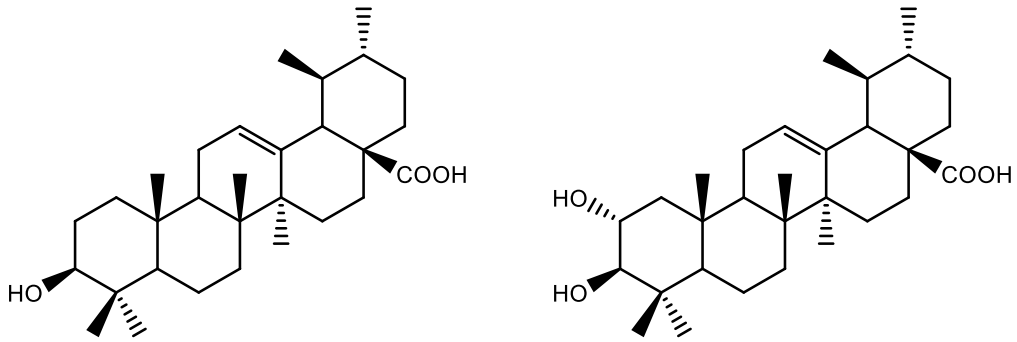
Based on these graphs, some preliminary insights can be gained into triterpene activity. For example, a general trend can be seen where an increased number of scaffold modifications is associated with a lower IC_{50} value (**Figure 7-3**). Triterpenes that have been glycosylated or acylated also appear to have overall lower IC_{50} values compared to those that have not (**Figure 7-4**), although the number of these types of compounds represented in the dataset is relatively small. It is difficult to draw conclusions about the effects of the scaffold itself on the activity of triterpenes; while some scaffolds such as iso-lanosterol appear to be associated with lower IC_{50} values (**Figure 7-5**), the sample size for these scaffolds is small. As a general point, it must be remembered that the dataset may be biased, since molecules with high bioactivity may be more likely to be reported.

7.3.2 Pairwise Comparisons of Pentacyclic Triterpenes in the HL-60 Cell Line

In order to design and build triterpenes with improved bioactive properties, a more focused consideration of the structure-activity relationship is needed. A large number of reported bioactive triterpenes in the literature are derived from the β-amyrin scaffold, such as oleanolic acid (**18**) [174], glycyrrhetinic acid (**8**) [102], and maslinic acid (**24**) [69]. Because of this, the development of the triterpene toolkit has had a particular focus on biosynthetic enzymes that are active towards β-amyrin and β-amyrin derivatives. Therefore, this scaffold was chosen as a focus for triterpene design in the present study. Given that the largest amount of triterpene bioactivity data has been generated for the HL-60 cell line within the O'Connell lab, the data resource for this cell line was focused on to enable direct comparison of the bioactivities of sets of molecules.

The O'Connell lab has tested 75 triterpene compounds to date comprising of 49 biosynthesised compounds and 26 purchased standards. The available data from these compounds were therefore filtered for compounds based on the β-amyrin scaffold (with compounds based on the related scaffolds α-amyrin and lupeol also considered) that had been tested on the HL-60 cell line, a dataset of 27 biosynthesised molecules and 23 commercial standards. Firstly, single modifications of the scaffold were considered, and active positions on the scaffold identified. Next, double modifications were considered, with each compound compared to related single-modified compounds to investigate whether additional modifications have raised, lowered, or maintained the activity. Regio- and stereo-isomers were also compared with each other.

For example, the single hydroxylation required to change ursolic acid (**50**) into corosolic acid (**51**) is associated with a significant increase in anti-proliferative effect, as seen in **Figure 7-7**.

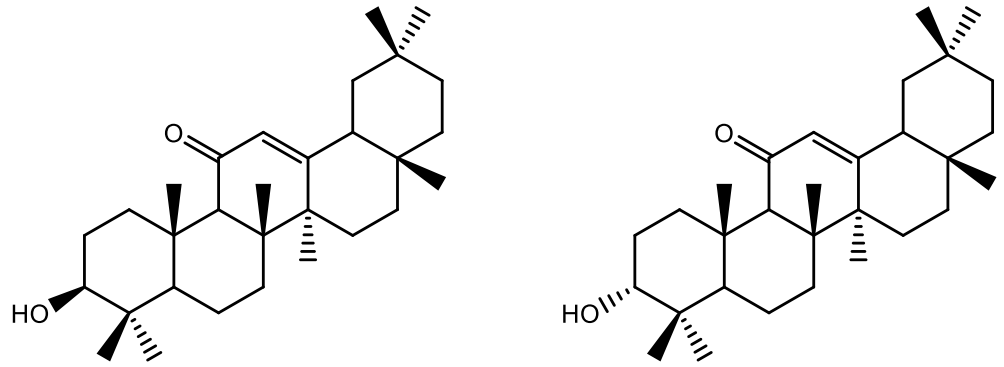


Ursolic Acid (**50**)
IC_{50} = 48.96

Corosolic Acid (**51**)
IC_{50} = 24.4

Figure 7-7: Structures of ursolic acid (**50**) and corosolic acid (**51**) and their corresponding IC_{50} values for the HL-60 cell line.

The transformation of the C-3 hydroxyl in 11-oxo-β-amyrin (**2**) to its C-3α epimer does not increase anti-proliferative effect, and in fact reduces it, as seen in **Figure 7-8**.

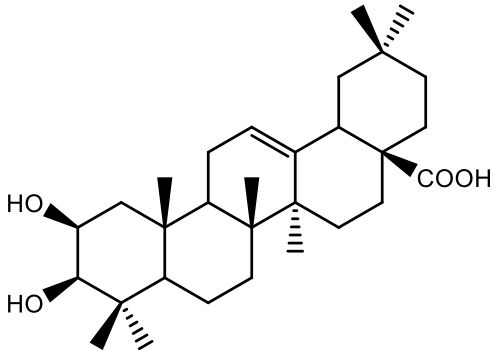


11-oxo-β-amyrin (**2**)
IC_{50} = 19.66

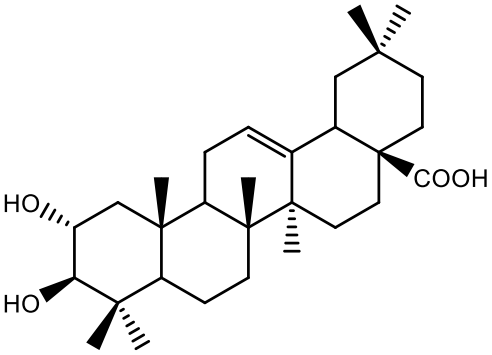
11-oxo-epi-β-amyrin (**31**)
IC_{50} = 30.44

Figure 7-8: Structures of 11-oxo-β-amyrin (**2**) and 11-oxo-epi-β-amyrin (**31**) and their corresponding IC_{50} values for the HL-60 cell line.

The C-2β hydroxyl in augustic acid (**25**) is associated with a slight improvement in the anti-proliferative effect in the HL-60 cell line over the C-2α hydroxyl in its isomer maslinic acid (**24**), as seen in **Figure 7-9.**



Augustic Acid (**25**)
IC_{50} = 16.52



Maslinic Acid (**24**)
IC_{50} = 24.12

Figure 7-9: Structures of augustic acid (**25**) and maslinic acid (**24**) and their corresponding IC_{50} values for the HL-60 cell line.

The glycosylation that occurs when α-hederagenin (**52**) is transformed into the saponin α-hederin (**53**) causes a substantial increase in the anti-proliferative effect, as seen in **Figure 7-10** . Notably, this is not seen when the inactive aglycone quillaic acid (**54**) is glycosylated.

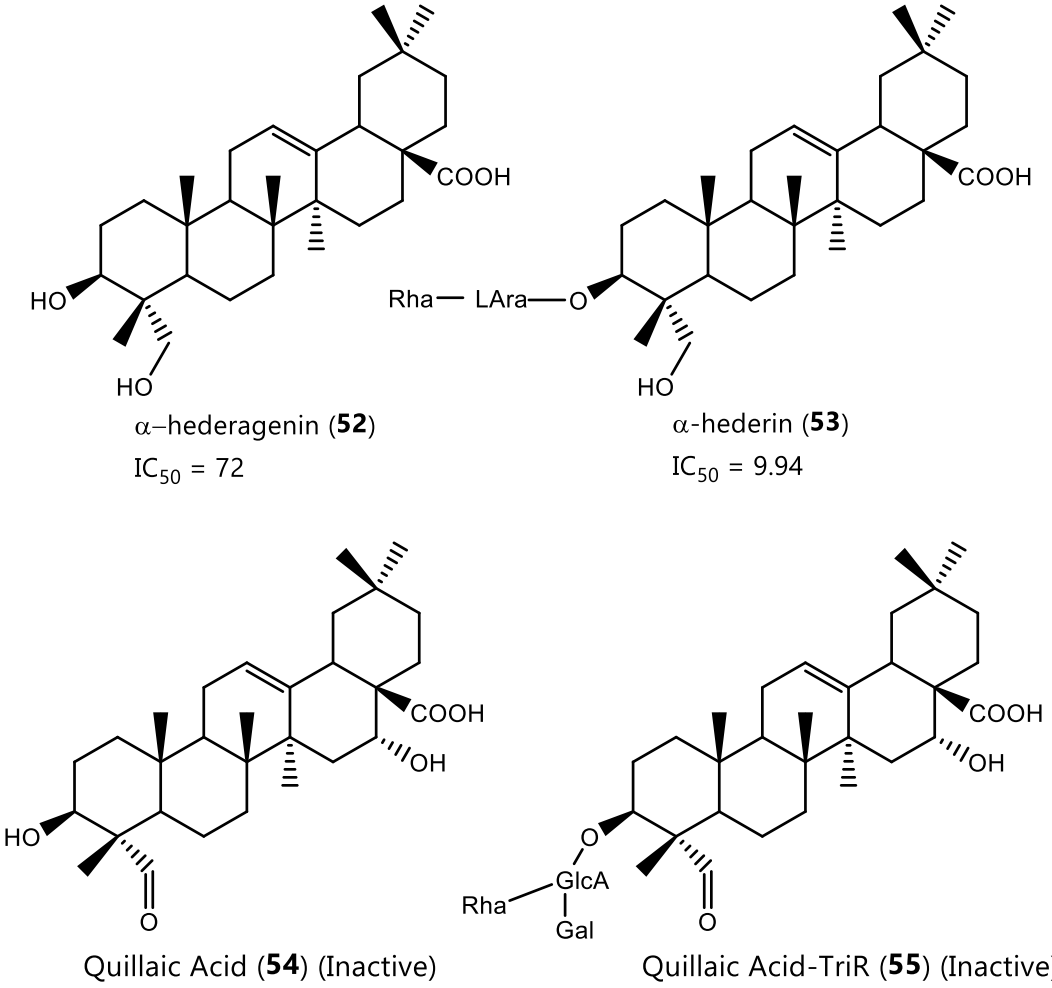


Figure 7-10: Structures of α-hederagenin (**52**), α-hederin (**53**), quillaic acid (**54**), and Quillaic acid-TriR (**55**) and their corresponding IC_{50} values for the HL-60 cell line. Rha =L-rhamnose, LAra = L-arabinose, GlcA = D-glucuronic acid, Gal = D-galactose.

This process was repeated with increasing numbers of modifications to build a series of conclusions about structure-activity relationships, which are summarised in **Figure 7-11**.

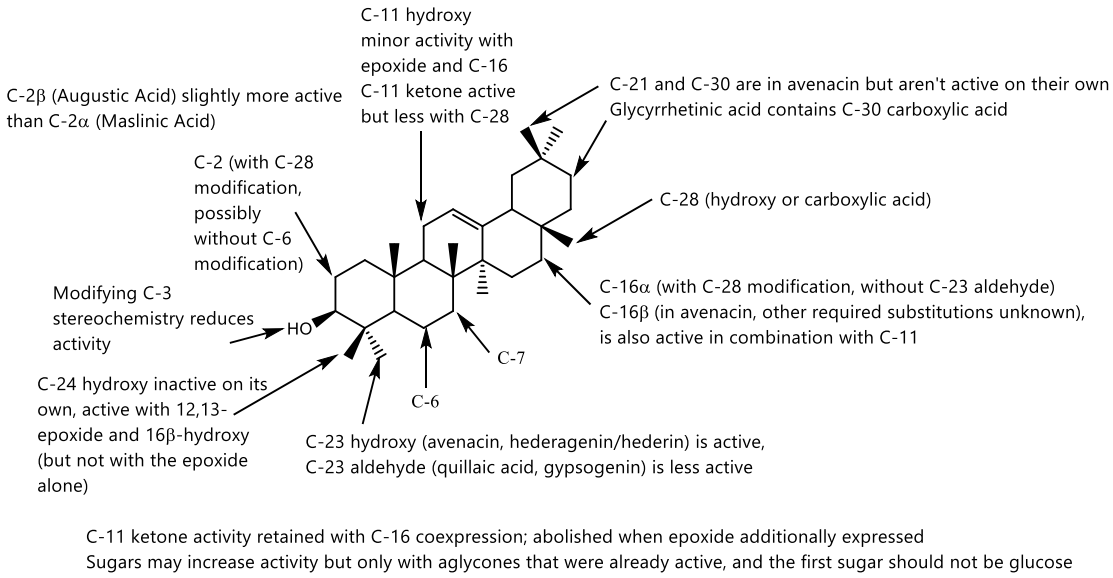


Figure 7-11: Summary of the features associated with the structure-activity relationships of triterpenes with the β-amyrin scaffold with respect to anti-proliferative effects of the HL-60 cell line.

Based on these findings, a number of potential triterpenes were designed with the aim of combining multiple scaffold oxidations important for bioactivity into molecules with greater bioactive potential. These molecules can be seen in **Figure 7-12**.

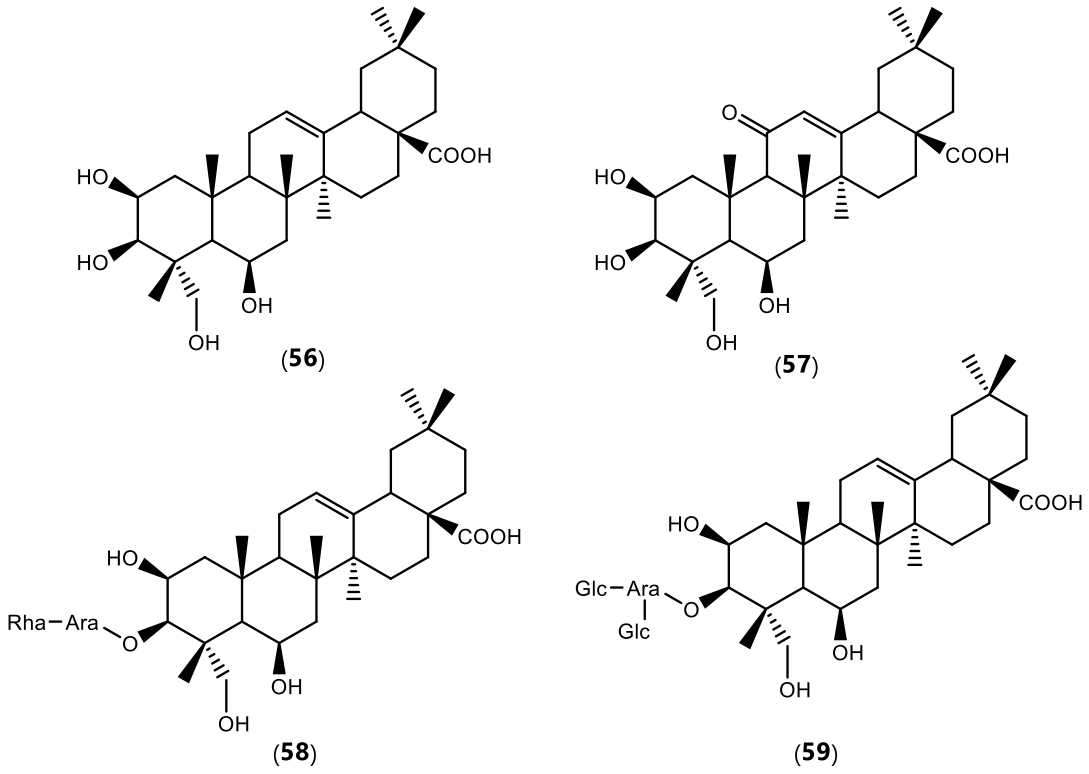


Figure 7-12: Structures of potentially promising triterpene bioactives designed using the insights gained from structure-activity relationship analysis. Rha = L-rhamnose, Ara = L-arabinose, Glc = D-glucose.

The above compounds contain a combination of the most clearly active positions on the β-amyrin scaffold: C-2, C-23, and C-6 hydroxylation combined with one or both of the 11-oxo or 28-carboxylic acid groups. The candidates have features in common with triterpenes that are known to inhibit cell proliferation such as hederagenin (C-28 carboxylic acid, C-23 alcohol) and asiatic acid (C-28 carboxylic acid, C-23 + C-2 alcohol). It is possible that the combination of the 11-oxo group with an oleanolic acid-derived compound could reduce activity, as seen for 11-oxo-oleanolic acid (**5**) itself (O'Connell lab data, see **Section A.5**). However, the 11-oxo group features prominently in bioactive triterpenes, including in this study (see **Chapter 6**), and in other previous studies [50], and therefore merits further investigation. The sugar chains of candidates (**58**) and (**59**) (**Figure 7-12**) are in common with those of the two β-amyrin derived triterpenes α-hederin (**53**) and the avenacins, which have the highest anti-proliferative activity observed so far (see **Section A.5**).

The aglycone candidates were chosen for bioproduction in *N. benthamiana* due to the focus of the current study being on the optimisation of β-amyrin-derived aglycones as potential bioactive molecules through the generation of structural variants of β-amyrin derivatives that differ in the number and positions of oxidation events around the scaffold.

The enzymes required for biosynthesis of the triterpene candidates (candidate (**56**) in **Figure 7-12**, hereafter called candidate triterpene 1, and candidate (**57**) in **Figure 7-12**, hereafter called candidate triterpene 2) were co-expressed in *N. benthamiana* as shown in **Figure 7-13** and leaf extracts were then analysed by LC-MS to detect the presence of the candidate triterpenes (see **Section 2.4.3**). The LC-MS spectra are shown in **Figure 7-14**.

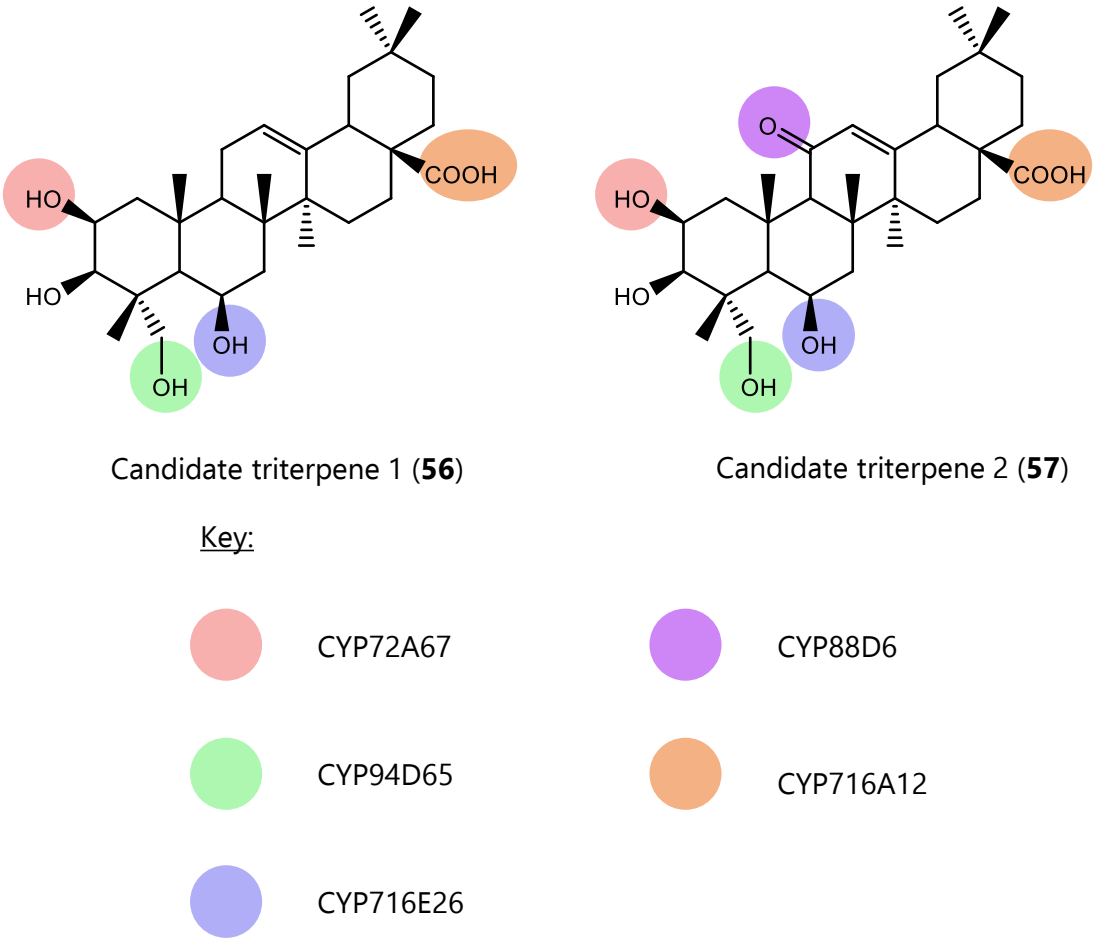


Figure 7-13: The two candidate triterpenes produced in *N. benthamiana* and the enzymes required for their biosynthesis.

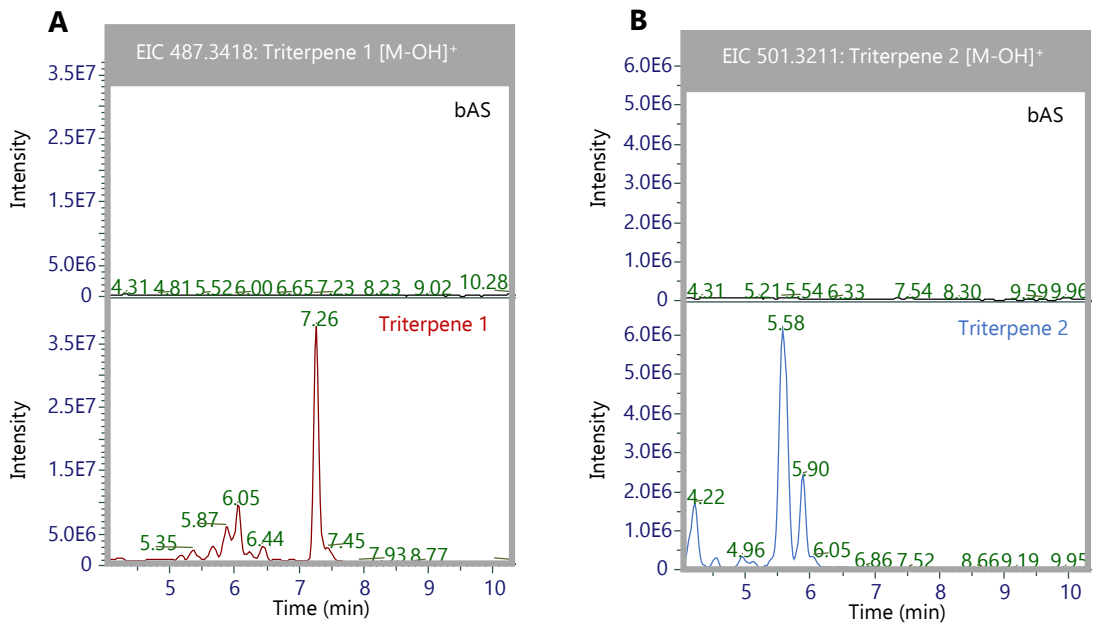


Figure 7-14: Testing of candidate triterpene expression via LC-MS. Extracted ion chromatograms (EICs) for A) *m/z* = 487.3418, corresponding to the predicted mass of candidate triterpene 1 $[M\text{-}OH]^+$ or B) *m/z* = 501.3211, corresponding to the predicted mass of candidate triterpene 2 $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1) as a control (black), and with addition of required CYPs to produce triterpene candidate 1 (red), or triterpene candidate 2 (blue).

These experiments demonstrate the successful production of the expected triterpene candidates at analytical scale. Future work will involve scale-up in order to generate preparative scale amounts of these compounds for investigations of bioactivity; also the generation and testing of other potential triterpene candidates. For example, although the C-2β hydroxyl group was selected for the current study, literature reports show C-2α hydroxyl triterpene derivatives to also be significantly bioactive and worthy of testing. These steps will lead to the optimisation of β-amyrin derived aglycones. A suite of aglycone molecules generated and optimised in this way can then be functionalised further through the use of decorating enzymes (such as glycosyl and acyl transferases) to produce additional triterpene diversity for future biological testing, using the range of oxidation events across the scaffold as handles for further modification.

7.3.3 Solid-Phase Extraction Improves the Throughput of Triterpene Testing

A significant bottleneck in the improvement and testing for triterpenes is the requirement for purified samples at milligram scale for use in bioassays. The

establishment of a method that would enable semi-crude *N. benthamiana* leaf extracts to be screened directly without the need to purify compounds at milligram scale would greatly accelerate the discovery of new bioactives. Recent investigations into the use of semi-purified extracts in biological assays [172] has raised the prospects of utilising high-throughput methods for faster screening of candidate molecules, such as through the use of solid-phase extraction for purification.

As mentioned in **Section 7.1**, SPE techniques can be used to improve the throughput of investigations into plant extract bioactivities. Samples loaded into an SPE cartridge are separated based on the relative affinity of the components towards either a stationary phase (which can constitute various matrices including normal-phase silica or hydrocarbon chain functionalised reverse-phase silica) or a mobile phase (a solvent passed through the cartridge) [170]. Unlike regular chromatography, which relies on the separation of compounds via the time taken for a compound to move through the column based on a continuous flow of either constant polarity or a gradient, SPE usually works through capturing a particular component of the mixture, preventing its movement through the column entirely. This can involve capturing impurities, leaving the compound of interest to flow through the column, or capturing the compound of interest, washing impurities from the column, and then later releasing the desired compound through application of an eluting solvent [170]. SPE is therefore not continuous in the flow of its mobile phase. This produces a series of discrete fractions which can be retained or discarded depending on the presence or absence of compounds of interest.

An investigation was therefore carried out into the feasibility of using solid-phase extraction methods on *N. benthamiana* extracts generated from plants transiently expressing different combinations of triterpene biosynthetic genes. The full method is given in (**Section 7.5.3**) but briefly, freeze-dried leaves (approximately 0.5 g) from *N. benthamiana* plants infiltrated with MMA buffer only, or with gene constructs for production of β-amyrin (**1**) and 11-oxo-β-amyrin (**2**), were ground with a pestle and mortar and extracted in parallel using a speed extractor (Buchi). (**Figure 7-15A**) The extracts were adsorbed onto cotton wool, which was then divided into three pieces,

each being loaded onto the top of a separate solid-phase extraction cartridge (**Figure 7-15B**). One piece was loaded directly; one was loaded onto a layer of Ambersep ion exchange resin; and one was loaded onto of a layer of charcoal. The purpose of including Ambersep and charcoal was to evaluate the ability of these materials to remove pigments such as brightly coloured chlorophylls and other compounds that may interfere with the colourimetric bioassays.

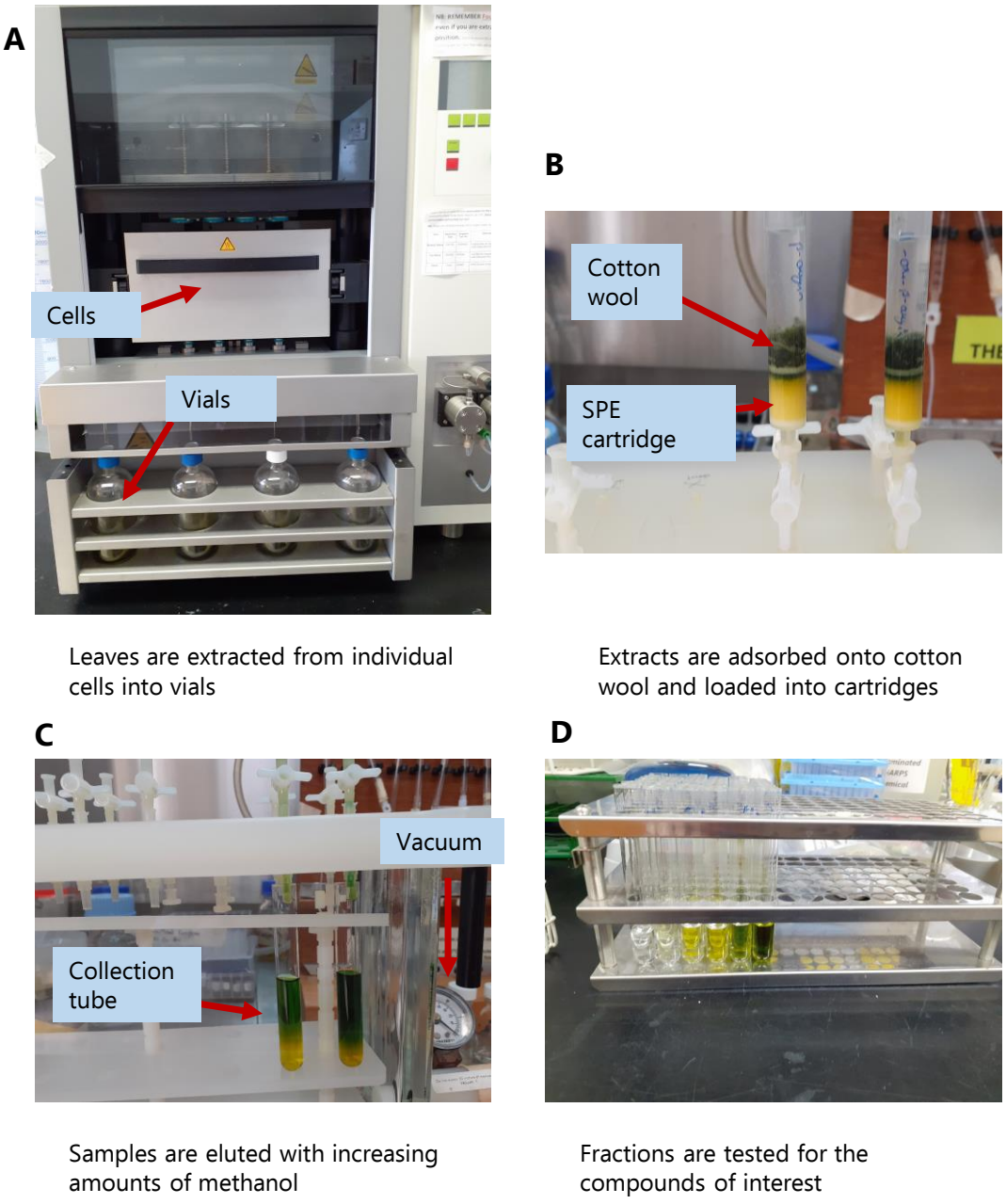


Figure 7-15: Fractionation of *N. benthamiana* extracts using solid-phase extraction for testing in bioactivity assays. A) Parallel extraction of ground freeze-dried leaves (loaded into the cells) into vials. B) Extracts are adsorbed onto cotton wool and loaded into SPE cartridges. C) Extracts are fractionated into collection tubes with increasing amounts of

methanol in water using a vacuum. D) Fractions are labelled and tested for compounds of interest.

Extracts were eluted from the solid-phase extraction cartridges using increasing proportions of methanol in water (**Figure 7-15C**), to give a series of fractions (labelled N, A, or C depending on whether they were not treated (N) , treated with Ambersep (A), or treated with charcoal (C), and also with the % of methanol used to elute them) (**Figure 7-15D**). These were analysed by GC-MS to identify those fractions containing triterpenes. Fractions of interest, and all of the fractions from the MMA extracts, were transferred into pre-weighed vials and dried and the dry mass of each fraction determined. Fractions were then dissolved in a sufficient amount of DMSO to make a 5 mg/ml solution. Fractions were then tested using the MTS assay (see **Section 7.5.4**).

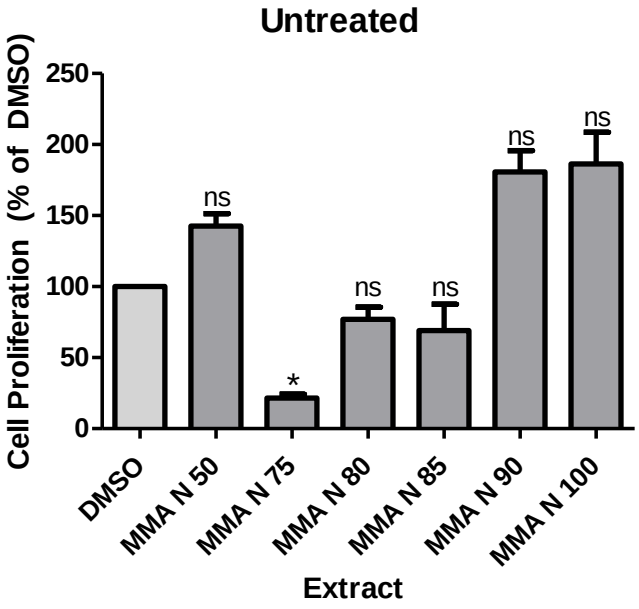


Figure 7-16: Proliferation of HL-60 cells exposed to fractions of untreated extracts from leaves infiltrated with MMA only. Cells were treated with 1 µl fraction at 5 mg/ml concentration (by total weight of fraction). Proliferation was measured by MTS assay after 72 hours and is shown relative to DMSO (vehicle) treated cells. Values represent the mean of 3 biological replicates (each the mean of three technical replicates) ± SD. Significance values (calculated by ANOVA) are denoted as follows: $P<0.05$ (*) or not significant (ns). In the names of the fractions given above, "N" signifies untreated fractions, and the number given for each fraction is the % of methanol required to elute that fraction.

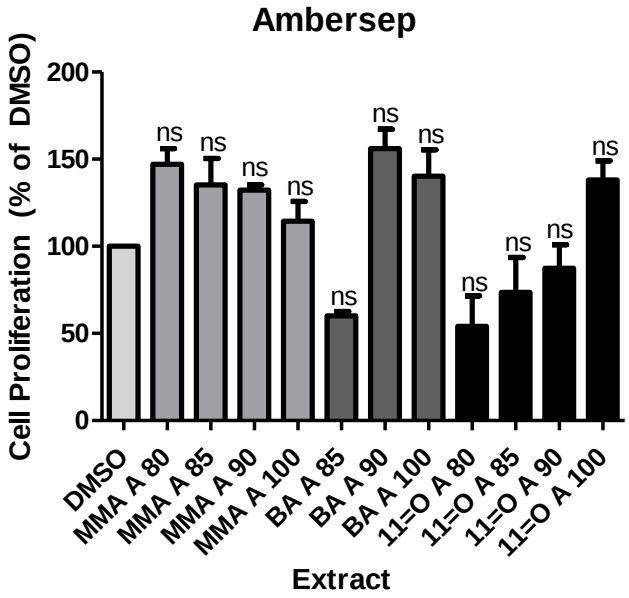


Figure 7-17: Proliferation of HL-60 cells exposed to fractions of Ambersep treated extracts from leaves infiltrated with MMA, β-amyrin synthesis constructs (BA), or 11-oxo-β-amyrin synthesis constructs (11=0). Cells were treated with 1 µl fraction at 5 mg/ml concentration (per total weight of fraction). Proliferation was measured by MTS assay after 72 hours and is shown relative to DMSO (vehicle) treated cells. Values represent the mean of 3 biological replicates (each the mean of three technical replicates) ± SD. Significance values (calculated by ANOVA) are denoted as follows: not significant (ns). In the names of the fractions given above, "A" signifies fractions treated with Ambersep, and the number given for each fraction is the % of methanol required to elute that fraction.

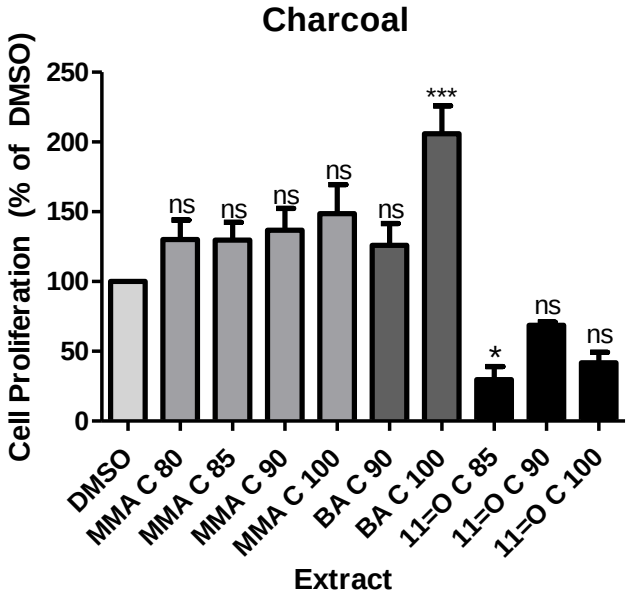


Figure 7-18: Proliferation of HL-60 cells exposed to fractions of charcoal treated extracts from leaves infiltrated with MMA, β-amyrin synthesis constructs (BA), or 11-oxo-β-amyrin synthesis constructs (11=0). Cells were treated with 1 µl fraction at 5 mg/ml concentration (per total weight of fraction). Proliferation was measured by MTS assay after 72 hours and is shown relative to DMSO (vehicle) treated cells. Values represent the mean of 3 biological replicates (each the mean of three technical replicates) ± SD. Significance values (calculated by ANOVA) are denoted as follows: $P<0.05$ (*), $P<0.001$(***), or not significant (ns). In the names of the fractions given above, "C" signifies fractions treated with charcoal, and the number given for each fraction is the % of methanol required to elute that fraction.

As can be seen in **Figure 7-16**, the untreated extracts contain constituents that interfere with the assay, as even leaf extracts from MMA (mock) infiltrations inhibit cell proliferation at 75 % MeOH and possibly at 80 % and 85 %. Ambersep treatment (**Figure 7-17**) does alleviate this background activity, however some inhibition can be seen for the β-amyrin (**1**) containing samples at 85 % MeOH, which is not expected as β-amyrin (**1**) itself is known not to inhibit cell proliferation in this particular assay. When the extracts were treated with charcoal (**Figure 7-18)**, only the samples containing the known anti-proliferative triterpene 11-oxo-β-amyrin (**2**) inhibit cell proliferation. The charcoal treatment therefore appears promising as a means of cleaning up the *N. benthamiana* leaf extracts sufficiently to be able to detect bioactivity.

These results are a proof of concept for the use of SPE techniques for improving the throughput of testing the bioactivity of triterpenes generated by transient

expression in *N. benthamiana*. This method holds considerable promise for rapid identification of new bioactive triterpenes without the need for purification in the first instance; under a week to produce fractions ready for testing from harvested leaves compared to the 3-4 weeks taken to purify most of the compounds mentioned in earlier chapters. This method of testing could be performed in parallel with multiple samples at a time, enabling large suites of molecules to be designed, generated by heterologous expression at small scale, and then screened for bioactivity at a rapid pace. Promising leads could then be prioritised for production at preparative scale and further in-depth evaluation of bioactivity following this initial triaging process. This overall approach will allow inactive molecules to be screened out before the stage of large-scale production, leaving only the most interesting compounds to go through to full purification and quantitative assays.

7.4 Conclusions

In this chapter, a comprehensive review of the structure-activity relationships of triterpenes with respect to anti-proliferative effects has been carried out. A focus on pentacyclic triterpenes, for which both the greatest amount of data and the greatest capacity for functionalisation exist, has resulted in the design of potentially more potent molecules, which have been produced in a small-scale in *N. benthamiana* as a demonstration of the viability of their biosynthesis. Additionally, the development of SPE techniques to enable semi-pure testing of triterpenes for biological activity represents an innovative step forward for the evaluation of the products of triterpene biosynthesis, dramatically reducing the time needed to assess novel triterpenes and allowing for the screening of entire libraries of compounds in parallel before focusing on the most interesting and valuable molecules for further study.

The future expansion of this work using the anti-inflammatory data and the insights gained therein will allow for greater insights into a variety of triterpene bioactivities. Additionally, the scope of this approach can be expanded beyond aglycones to investigate the potential of tailoring enzymes such as acyl and glucosyl transferases in further enhancing the bioactivity of triterpene molecules.

7.5 Materials and Methods for Chapter 7

7.5.1 Graphing Triterpene Cell Viability Data

Cell viability data over 10 cell lines obtained both within the O'Connell Lab and from literature sources was tabulated, and any value over 100 μM was set to 100 μM for ease of analysis. A total of 142 compounds had data assembled for analysis. Data was assigned categories based on a number of classifications (scaffold, type of compound, type of modifications present, number of modifications) and each category was assigned a number. Each compound therefore had an individual number for each potential classification. These category numbers were then plotted against IC_{50} values for each classification using excel (XY scatter plot). The full database of compounds showing their categorisation and IC_{50} values is given in (**Section A.5**).

7.5.2 Small-Scale Testing of Improved Triterpenes

A. tumefaciens strains containing tHMGR, SAD1, CYP716A12, CYP716E26, CYP72A67, CYP88D6, and CYP94D65 were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined into two combinations: a combination of tHMGR, SAD1, CYP716A12, CYP716E26, CYP72A67, and CYP94D65 to produce triterpene candidate 1, and a combination of all genes to produce triterpene candidate 2. The combined strains were used to infiltrate 1 leaf of 5-week-old *N. benthamania* plants using the manual infiltration method (see **Section 2.3.2**). The plants were grown for 5 days under greenhouse conditions before leaves were harvested and processed using the standard method for LC-MS analysis (see **Section 2.4.3**).

7.5.3 Solid-Phase Extraction Set-Up and Testing

A. tumefaciens strains containing tHMGR, SAD1, and CYP88D6 were cultured according to the standard procedure (see **Section 2.3.1**). Strains were combined into two combinations: a combination of tHMGR and SAD1 only and a combination of tHMGR, SAD1, and CYP88D6. Each combination of strains, along with an appropriate volume of MMA, was used to infiltrate 5 leaves per plant of 5-week-old *N. benthamiana* plants using the manual infiltration method (see **Section 2.3.2**).

Plants were grown for 5 days under greenhouse conditions before leaves were harvested.

Harvested leaves were frozen and freeze-dried for 24 h before being extracted using the speed extractor using the same programme as in **Section 2.4.6** but with the exception that each condition was loaded into a different cell of the instrument and extracted into separate vials instead of being pooled.

The extracts were then each adsorbed onto a piece of cotton wool, which was divided equally into 2 pieces (for the extracts expressing β-amyrin or 11-oxo-β-amyrin) or 3 pieces (for the MMA infiltrated extract). Each piece was loaded into the top of a solid-phase extraction cartridge (Hypersep C8 500 mg, Thermo). For each condition, one cartridge was used as-is, one was used with ~1 cm of Ambersep ion exchange resin (Sigma) between the top of the cartridge and the cotton wool, and one was used with ~1 cm of activated charcoal (Sigma) between the top of the cartridge and the cotton wool. Cartridges were equilibrated with 50 % methanol in water prior to the addition of the cotton wool.

The cartridges were attached to a vacuum manifold. Samples were eluted from the cartridge using a vacuum applied to the manifold. Elution was carried out using an increasing amount of methanol in water; 5 ml 50 % methanol, 5 ml 75 % methanol, 5 ml 80 % methanol, 5 ml 85 % methanol, 5 ml 90 % methanol, 5 ml 100 % methanol. Each of the elution volumes was collected separately in glass collection tubes and samples of each fraction sent for analysis using GC-MS using the 15-minute method (see **Section 2.4.1**). For the β-amyrin and 11-oxo-β-amyrin expressing samples, fractions were dried and retained for biological testing where GC-MS analysis showed them to contain significant amounts of the compound of interest. All of the fractions from the untreated MMA infiltrated sample were dried and retained for biological testing, along with the Ambersep and charcoal treated MMA infiltrated samples eluted at the same proportion of methanol as either β-amyrin or 11-oxo-β-amyrin containing fractions.

7.5.4 Qualitative MTS Antiproliferation Assay of Solid-Phase Extraction Fractions

MTS assays were carried out in 96 well plates under sterile conditions. The outermost wells contained 250 µl water to maintain humidity and minimise media evaporation. HL-60 cells were tested.

Each of the fractions from the solid-phase extraction procedure were dissolved in an appropriate amount of DMSO to make a 5 mg/ml stock solution prior to use.

For each plate, all of the fractions to be tested were assayed in triplicate. 1 µl of the stock concentration was used per 100 µl well to give a 100-fold dilution from the original stock. Untreated and vehicle (DMSO) treated controls were also included for reference. Three wells containing media only were included in order to subtract media absorbance during measurement.

Plates were returned to the incubator after the addition of compounds and maintained at 37 °C/5 % CO_2 for 3 days. The MTS reagent (Cell Titer 96 Aqueous non-radioactive assay, Promega) was prepared. 10 µl MTS was then added to each well in the dark (excluding the outermost wells). Plates were returned to the incubator for 4 hours, after which absorbance at 492 nm was measured using a plate reader (POLARstar OPTIMA, BMG Labtech) and the averaged media only reading subtracted from all wells.

The inhibition of each fraction was calculated and normalised to vehicle (DMSO) control before being plotted using the GraphPad Prism software (Version 5). Statistical analysis was performed via one-way ANOVA with post-hoc Tukey's multiple comparison test.

8 General Discussion

8.1 Introduction

Triterpenes are a diverse class of molecules found throughout the plant kingdom, and have found uses as part of formulations in traditional remedies for a range of different diseases (such as the use of ginseng to modulate the immune system and extracts of the frankincense tree to treat inflammation) [155,175]. Similarly, interest in triterpenes as modern therapeutic agents has led to the investigation of triterpenes and derivatives for diseases such as cancer and inflammation. An improved understanding of the properties of this major family of plant bioactives would be aided by taking a systematic approach towards investigating structure-activity relationships. However, such an approach depends on the ability to bioengineer suites of molecules with specific structural features, and on generating these molecules at preparative scale in order to evaluate bioactivity. Previous studies of the structure-activity relationships of triterpenes have been limited due to the low abundance of triterpenes in their natural sources, and their recalcitrance to chemical synthesis. The establishment of the transient expression platform in *N. benthamiana* and the creation of the triterpene toolkit (TTK) has allowed for the targeted engineering of triterpenes with desired structural features and the production at mg scale of a suite of functionalised β-amyrin derivatives as part of the present study. Investigation of the properties of these molecules and their mechanisms of action has provided fresh insights into the structural features that determine triterpene bioactivity.

8.2 Generation of Novel Structural Diversity Through Combinatorial Biosynthesis

Combinatorial biosynthesis allows for the significant expansion of triterpene structural diversity through the production of novel molecules. In the current study, pairwise combination of CYPs from the triterpene toolkit enabled the production of a substantial number of new products (66), 31 of which had putative structures not previously reported (**Chapter 3**). The success of this investigation proves the versatility of the combinatorial biosynthesis platform and lays the groundwork for

the use of the TTK to produce designer molecules with carefully chosen functionalisations.

The production of suites of specifically designed molecules relies upon the availability of a sufficient variety of biosynthetic enzymes in order to generate the required structural diversity. In **Chapter 4**, the feasibility of incorporating enzymes from beyond the Plant Kingdom into the TTK is investigated by evaluation of enzymes from bacterial and animal sources. The variety of products produced through combinatorial biosynthesis in which these enzymes are co-expressed with other plant enzymes from the TTK demonstrates their receptiveness to use in combinatorial experiments and the flexibility and promiscuity provided by them. These enzymes include CYP106A1 and CYP106A2 from *Bacillus megaterium*, which oxidise the previously unserved position 7β on the β-amyrin scaffold [141,142]. Additionally, some of the incorporated enzymes modify currently served positions in ways not previously accessible through the TTK, such as CYP716C11 from *Centella asiatica*, which oxidises the C-2 position in the α- rather than β-configuration [117].

Future approaches to diversifying the TTK could include harnessing an even greater diversity of enzymes from the Plant Kingdom and elsewhere, and also use of mutagenized enzymes with novel functionalities. For example, previous work in the Osbourn lab has led to the generation of mutant forms of the *A. strigosa* CYP51H10 enzyme that decouples the dual C-16β hydroxylation and C-12,13β epoxidation functionalities of this enzyme [120], to yield mutant forms that can perform each of the two types of oxidations independently (unpublished data). Such approaches have been used extensively for bacterial CYPs, which can be readily investigated due to their solubility and ease of purification. One particularly well-known example of this is BM3, otherwise known as CYP102A1 (from *Bacillus megaterium*), which has an especially high catalytic activity and has been modified extensively to accept a range of substrates, including drug-like molecules [51,138]. Mutagenesis of BM3 has led to the production of libraries of variants which are capable of a vast array of transformations [139]. The transient plant expression system has proved to be a powerful and rapid means of expressing plant CYPs and investigating their functions

[13,50,115,120,176], and so offers a route to screening for mutant variants of CYPs from plants and other organisms that could further expand the capability of the TTK.

The use of enzymes modified to have specific functions in this way would allow for the production of a far wider variety of molecules, truly allowing for the design and production of molecules with optimised bioactive properties. This could be further enabled through the advent of machine learning techniques and the use of molecular modelling to accurately predict the function of enzymes from their structure or even amino acid sequence [177]. This would enable the potential results of any mutagenesis study to be predicted *in silico* and so increase the throughput of these approaches. The predictions made through computer-based methods will still need to be evaluated at the laboratory bench, but here the throughput could be improved through the capture of reaction products for MS analysis using techniques such as matrix-assisted laser desorption ionisation (MALDI) [178,179]. These potential future directions open up the prospect of designing enzymes to perform the desired modifications to produce specifically designed molecules for medicinal and other applications.

8.3 Structural Elucidation of Triterpenes by NMR

The scale-up of triterpene production to the mg scale remains an important requirement for undisputed assignment of structure through NMR, as well as for the evaluation of triterpenes for biological activity. The putative structures for compounds identified in **Chapter 4** (predicted using observed molecular mass, GC-MS fragmentation patterns, and knowledge of enzyme activity) were confirmed following scale-up and purification in **Chapter 5**. The use of NMR techniques such as 2D NMR provides for the assignment of exact regiochemistry within a scaffold, something that can be inferred through GC-MS techniques but not definitively proven. Meanwhile, the stereochemistry of a compound can be proven through the use of nuclear-overhauser effect-related spectroscopy (NOESY and ROESY) to show correlations between protons in 3D space [180], and also through the comparison of ^{1}H-^{1}H coupling constants utilising the Karplus equation to determine bond angle, which is especially relevant to the triterpenes based on amyrin-type scaffolds

because they are polycyclic molecules formed from multiple fused cyclohexane rings [126]. The use of these techniques to ascertain stereochemistry can be critical, since while different stereoisomers may have different retention times by GC- and LC-analysis (as has been shown in this thesis), when working with only a single isomer in the absence of a verified standard, these methods are unable to distinguish the stereochemistry with certainty. Indeed, as can be seen with the example of the production of echinocystic acid through the combination of CYP716A12 and CYP716A141 (**Chapter 5**), the overreliance on assuming structure through previously observed activities of characterised enzymes can be misleading in cases where enzyme combinations result in unpredicted products. Therefore, structural characterisation of molecules by NMR is essential in order to accurately assign the structures of products generated by triterpene bioengineering.

8.4 SPE Methods Offer the Potential for Improved Throughput of Investigations of Triterpene Bioactivity

Chapter 5 also introduced the production of a suite of molecules for biological testing. The production of these molecules through standard scale-up and purification techniques was relatively slow, taking approximately 1-2 months per molecule. Since not all of these molecules showed bioactivity, this process represents a significant investment of time with limited returns. If the throughput for evaluating the biological activity of the products of co-expression could be accelerated this would represent a significant advance, especially when testing suites of molecules that have been deliberately designed to have improved biological effects. **Chapter 7** presents the use of SPE methods for qualitative evaluation of anti-proliferative effects of triterpenes. The advantage of this technique is its speed and throughput. The number of infiltrated plants required in order to generate sufficient quantity of compound for testing is reduced from approximately 100 down to one. The use of parallel processing techniques reduces the time required to prepare *N. benthamiana* extracts for biological testing from over one month per compound down to just over a week for a suite of up to 20 compounds. The results of these assays are qualitative but serve well to provide an indication of whether a

given compound produces a desired anti-proliferative effect or not. In the future, this would allow for inactive molecules to be screened out at an early stage, with only the desired, active molecules from a given library of compounds going on to further mg scale production, full purification, structural determination, and in-depth testing, so saving substantial amounts of time and speeding up the process of triterpene structure-function relationship investigation significantly.

8.5 The Potential of Triterpenes as Drug-Like Molecules

The results from evaluation of the biological activities of the triterpenes generated by *in planta* expression (**Chapter 6**), along with a range of standards, were consistent with previous reports for those triterpenes that had been tested previously [50]. They also revealed new positions on the β-amyrin scaffold associated with anti-proliferative and anti-inflammatory effects. Preliminary investigation into the mechanisms behind the anti-inflammatory effects of the triterpenes implicated the NfκB pathway. Future work into the mechanisms of action of these effects, such as quantitative PCR and reporter assays, will provide additional insights into the anti-inflammatory effects of the triterpenes. The mechanisms underlying observed anti-proliferative effects are yet to be established.

The expression of CYPs from human sources in *N. benthamiana* have a variety of uses aside from simple triterpene production. Although it was not the focus of the current study, human CYPs have the additional advantage of providing a means of predicting the fate of produced biomolecules within the human body. As the CYPs of the human liver are predominantly responsible for first-pass metabolism of ingested substances [181], products formed through combination of these CYPs with other TTK enzymes likely resemble those already frequently formed via regular human metabolism after ingestion of the substrate triterpenes [182,183]. The formation and testing of these products therefore provides further insight into potential metabolism and associated downstream effects of triterpenes in the body. This could have utility as a predictive tool for evaluating produced triterpenes for their likely effects post-metabolism, such as determining whether the metabolised products of known bioactive triterpenes retain any activity (or exhibit any unwanted

effects), or testing triterpenes known to be inactive for potential pro-drug properties (i.e., whether their metabolised products will begin to exhibit previously unknown bioactivities).

Moving forwards, promising triterpene candidates should be investigated to improve their "druggability" in order to make them more viable drug leads. In the case of most triterpenes, the major factor impacting their drug-like nature is their relative lack of solubility in water, which prevents them from being used in standard orally administered drug formulations. Glycyrrhetinic acid (**8**) (enoxolone) and its derivative carbenoxolone are formulated into sodium salts through their carboxylic acid groups to allow them to be administered orally [184]. This requires an available carboxylic acid group which is not the case for all of the triterpenes with potential bioactive properties (such as 11-oxo-β-amyrin (**2**)).

Consideration should therefore be given to the improvement of the drug-like nature of the triterpenes. A number of methods already exist within the drug discovery field to achieve this, and some degree of investigation into improving the water solubility (and therefore likely oral bioavailability) of triterpenes has already been carried out. The addition of simple groups such as esters and amines to an available hydroxyl or carboxyl group on a triterpene molecule, often at C-3 or C-28, has been investigated at length and received some success [28]. For example, investigation of a suite of derivatives of ursolic acid featuring a C-3 acetyl group and a variety of amines at C-28 found improvements in IC_{50} in a range of cell lines including MCF-7 and A549, as well as improved selectivity between malignant and non-malignant cells [185]. The addition of a greater number of acyltransferases to catalyse the formation of esters of triterpenes is therefore a valuable future direction, especially those that act on the C-3 hydroxyl group common to all triterpenes.

Glycosylation of triterpenes has also been investigated. Glycosyl groups are frequently cleaved in the human digestive tract, which may impact on their oral bioavailability [186,187]. However, some triterpene aglycones are themselves glycosylated by bacteria in the human gut, leading to increased bioavailability. The most well-known example of this is compound K (20-O-D-glucopyranosyl-20(S)-

protopanaxadiol), a known bioactive molecule formed in the human gut after ingestion of ginsenosides that accumulates within the blood plasma at a physiologically relevant concentration [186]. Glycosylation is therefore a strategy worth investigating, and expansion of the number of glycosyl transferases accessible in the toolkit is a worthwhile future direction.

Finally, the conjugation of polymers such as PEG groups has also been investigated as a means of creating pro-drugs. The conjugation of a multi-arm PEG group to betulinic acid led to improved solubility while maintaining an anti-tumour effect [188].

As well as the formulation of prodrugs, innovative drug-delivery systems can also improve the availability of triterpenes in the body. These could include the formation of micelles containing the compound of interest, the dispersal of molecule-containing droplets in an emulsion, or the loading of drugs into nanoparticles or liposomes. Ginsenoside compounds have previously been loaded into nanoparticles and liposomes with an associated improvement in bioavailability [74].

The handles added through typical combinatorial biosynthesis, such as hydroxyl and carboxylic acid groups, are available for further modification through chemical as well as enzymatic means. This raises the prospect of engaging in semi-synthesis of triterpene prodrugs, through the production of an optimised drug-like molecule through triterpene biosynthesis in sufficient yield, followed by the conjugation of functional moieties to transform the compound of interest into a prodrug carried out through conventional synthetic chemistry approaches. This coupling of biological and chemical techniques would allow for the maximum possible utilisation of available chemical space and subsequent diversity.

The biological investigations currently carried out relate to the potential treatment of diseases relating to cancer and inflammation. However, a much greater variety of human diseases exist for which triterpenes could theoretically be evaluated. Triterpenes have previously been evaluated for their potential effects as anti-viral

agents [167], in the treatment of parasitic disease such as leishmaniasis [189], as anti-bacterial agents [169], and in the treatment of diabetes [19]. Further development of triterpenes as therapeutics for this range of conditions would require the same understanding of structure-function relationships as has begun to be elucidated for the anti-proliferative effects of triterpenes. This would be greatly aided with the use of the same systematic approach to develop suites of compounds for biological testing. It is likely that different indications will be best treated by different carefully designed molecules. Some evidence of this can be seen in the fact that the structural positions responsible for anti-proliferative and anti-inflammatory activities seen in the present study do not entirely overlap (**Chapter 6**). The increased ability to produce designer molecules in high throughput raises the possibility of the creation of a library of compounds for testing against a library of different diseases, and the eventual elucidation of the separate structure-activity relationships underlying each different disease. Compounds could then be individually generated targeting each disease in turn.

8.6 Triterpene Structure-Activity Relationship Investigations: the Potential for Machine Learning

Preliminary investigations into triterpene structure-activity relationships (**Chapter 7**) have led to the initial suggestion of triterpene structural features associated with improved anti-proliferative effects, from which a set of optimised triterpenes was designed. The scale-up and testing of these compounds will allow for the validation of this approach. In the future, improved throughput techniques will allow for a greater number of candidates to be designed and screened using qualitative methods (such as through the SPE techniques outlined in **Chapter 7**) prior to scale-up for more investigative studies. The throughput of these investigations could be further enhanced through the use of machine learning. The ability to use existing triterpene bioactivity data, from published sources as well as databases such as ChEMBL [190,191], to predict the likely activity of designed molecules would provide an additional layer of *in silico* verification before the production of triterpenes at even a small scale for SPE-based qualitative testing. This would enable large libraries

of molecules to be systematically designed and screened *in silico* prior to small-scale qualitative evaluation of potentially hundreds or thousands of molecules in parallel using SPE techniques, after which compounds scoring a positive hit would be scaled up to the mg scale for full structural verification and in-depth studies into mechanism.

It is feasible to envision the coupling of the high-throughput and machine learning-based techniques discussed thus far into a pipeline for triterpene drug discovery, where potential bioactive molecules are predicted based on existing knowledge of structure-activity relationships. The biosynthetic pathways for these target molecules could then be assembled using existing toolkit enzymes, new enzymes garnered from nature, and/or novel mutant variants of enzymes designed through an understanding of enzyme structure-function relationships. This would lead to the generation of a library of compounds that could be processed through SPE in parallel and then tested qualitatively against a library of disease indications. Positive hits could then be scaled-up for further characterisation, quantitative biological activity determination and study of mechanisms of action, providing data to feed back into the understanding of structure-activity relationships while also driving forwards the development of triterpenes as drug lead molecules. Especially optimised molecules could be further investigated through chemical means to create prodrug molecules or encapsulation methods to move triterpenes forwards as a new source of novel pharmaceutical molecules.

8.7 Concluding Remarks

Natural products are an important source of compounds with either known or potential pharmaceutical applications. Triterpenes in particular show significant promise in a range of areas, particularly cancer and inflammation [100,192]. Despite this, triterpenes are often present at only low levels in the producing plants, often as components of complex mixtures of structurally related compounds. Furthermore, the plant species that make them are generally not cultivated, and in some cases may be endangered. These challenges have previously hampered structure-function investigations of triterpene bioactivity. Transient plant expression now offers

opportunities to produce entire libraries of molecules through synthetic biology methods. *Agrobacterium*-mediated expression in *N. benthamiana* enables rapid evaluation of the functions of individual enzymes and combinations of enzymes, and can be scaled up for purification and investigation of the triterpenes thus generated. The transient plant expression process is currently being used for commercial-scale production of protein therapeutics and antibodies, for example for the COVID-19 vaccine COVIFENZ®, which is produced by the Canadian company Medicago Ltd. [193]. To date, transient plant expression is not being used for the production of small molecules at commercial scale. However, research in the Osbourn lab has led to the elucidation of the complete 20-step biosynthetic pathway for the saponin vaccine adjuvant QS-21 and its reconstitution in *N. benthamiana* ([194] and Laetitia Martin, submitted). These examples, in combination with the results presented in this thesis, showcase the power of transient plant expression for bioproduction of a diverse range of pharmaceutically important molecules in a scalable fashion for applications as therapeutic agents.

References

1. Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C*., et al.* (2015) Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review. *Biotechnology Advances* 33, 1582-1614
2. Li, Y., Feng, L., Song, Z. F., Li, H. B., and Huai, Q. Y. (2016) Synthesis and Anticancer Activities of Glycyrrhetinic Acid Derivatives. *Molecules* 21
3. Chen, H. P., Zhao, Z. Z., Li, Z. H., Huang, Y., Zhang, S. B*., et al.* (2018) Anti-Proliferative and Anti-Inflammatory Lanostane Triterpenoids from the Polish Edible Mushroom *Macrolepiota procera*. *Journal of Agricultural and Food Chemistry* 66, 3146-3154
4. Gallego-Jara, J., Lozano-Terol, G., Sola-Martínez, R. A., Cánovas-Díaz, M., and de Diego Puente, T. (2020) A Compressive Review About Taxol(®): History and Future Challenges. *Molecules* 25
5. Muffler, K., Leipold, D., Scheller, M. C., Haas, C., Steingroewer, J*., et al.* (2011) Biotransformation of Triterpenes. *Process Biochemistry* 46, 1-15
6. Sawai, S., and Saito, K. (2011) Triterpenoid Biosynthesis and Engineering in Plants. *Frontiers in Plant Science* 2
7. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K*., et al.* (2013) High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin. *Nature* 496, 528-+
8. Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M*., et al.* (2014) Combinatorial Biosynthesis of Sapogenins and Saponins in *Saccharomyces cerevisiae* Using a C-16 Alpha Hydroxylase from *Bupleurum falcatum*. *Proceedings of the National Academy of Sciences of the United States of America* 111, 1634-1639
9. Hartmann, M.-A. (1998) Plant Sterols and the Membrane Environment. *Trends in Plant Science* 3, 170-175
10. 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, 2016-2031
11. Du, Y., Fu, X., Chu, Y., Wu, P., Liu, Y*., et al.* (2022) Biosynthesis and the Roles of Plant Sterols in Development and Stress Responses. *Int J Mol Sci* 23
12. Kemen, A. C., Honkanen, S., Melton, R. E., Findlay, K. C., Mugford, S. T*., et al.* (2014) Investigation of Triterpene Synthesis and Regulation in Oats Reveals a Role for β-Amyrin in Determining Root Epidermal Cell Patterning. *Proceedings of the National Academy of Sciences* 111, 8679-8684
13. Khakimov, B., Kuzina, V., Erthmann, P. O., Fukushima, E. O., Augustin, J. M*., et al.* (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 Journal* 84, 478-490
14. Inagaki, Y. S., Noutoshi, Y., Fujita, K., Imaoka, A., Arase, S*., et al.* (2013) Infection-Inhibition Activity of Avenacin Saponins against the Fungal Pathogens *Blumeria graminis* f. sp *hordei*, *Bipolaris oryzae*, and *Magnaporthe oryzae*. *Journal of General Plant Pathology* 79, 69-73
15. Pensec, F., Marmonier, A., Marchal, A., Gersch, S., Nassr, N*., et al.* (2013) *Gypsophila paniculata* Root Saponins as an Environmentally Safe Treatment against Two Nematodes, Natural Vectors of Grapevine Fanleaf Degeneration. *Australian Journal of Grape and Wine Research* 19, 439-445

16. Mugford, S. T., Louveau, T., Melton, R., Qi, X. Q., Bakht, S.*, et al.* (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, 1078-1092
17. Kao, T. C., Shyu, M. H., and Yen, G. C. (2010) Glycyrrhizic Acid and 18 Beta-Glycyrrhetinic Acid Inhibit Inflammation Via PI3K/AKT/GSK3 Beta Signaling and Glucocorticoid Receptor Activation. *Journal of Agricultural and Food Chemistry* 58, 8623-8629
18. Syrovets, T., Büchele, B., Krauss, C., Laumonnier, Y., and Simmet, T. (2005) Acetyl-Boswellic Acids Inhibit Lipopolysaccharide-Mediated TNF-α Induction in Monocytes by Direct Interaction with IκB Kinases. *The Journal of Immunology* 174, 498-506
19. Shah, B. A., Qazi, G. N., and Taneja, S. C. (2009) Boswellic Acids: A Group of Medicinally Important Compounds. *Natural product reports* 26, 72-89
20. Datoo, M. S., Natama, H. M., Somé, A., Bellamy, D., Traoré, O.*, et al.* (2022) Efficacy and Immunogenicity of R21/Matrix-M Vaccine against Clinical Malaria after 2 Years' Follow-up in Children in Burkina Faso: A Phase 1/2b Randomised Controlled Trial. *The Lancet Infectious Diseases* 22, 1728-1736
21. Maltz, F., and Fidler, B. (2019) Shingrix: A New Herpes Zoster Vaccine. *P t* 44, 406-433
22. Novavax. (2022) Novavax Investigational Covid-19 Vaccine (Nvx-Cov2373).
23. Lacaille-Dubois, M. A., and Wagner, H. (2017) New Perspectives for Natural Triterpene Glycosides as Potential Adjuvants. *Phytomedicine*
24. Kim, D. H., Shin, E. K., Kim, Y. H., Lee, B. W., Jun, J.-G.*, et al.* (2009) Suppression of Inflammatory Responses by Celastrol, a Quinone Methide Triterpenoid Isolated from *Celastrus regelii*. *European Journal of Clinical Investigation* 39, 819-827
25. Corey, E. J., and Lee, J. (1993) Enantioselective Total Synthesis of Oleanolic Acid, Erythrodiol, β-Amyrin, and Other Pentacyclic Triterpenes from a Common Intermediate. *Journal of the American Chemical Society* 115, 8873-8874
26. Huang, A. X., Xiong, Z. M., and Corey, E. J. (1999) An Exceptionally Short and Simple Enantioselective Total Synthesis of Pentacyclic Triterpenes of the β-Amyrin Family. *Journal of the American Chemical Society* 121, 9999-10003
27. Wang, P., Kim, Y.-J., Navarro-Villalobos, M., Rohde, B. D., and Gin, D. Y. (2005) Synthesis of the Potent Immunostimulatory Adjuvant Qs-21A. *Journal of the American Chemical Society* 127, 3256-3257
28. Zhou, M., Zhang, R. H., Wang, M., Xu, G. B., and Liao, S. G. (2017) Prodrugs of Triterpenoids and Their Derivatives. *European Journal of Medicinal Chemistry* 131, 222-236
29. Sporn, M. B., Liby, K. T., Yore, M. M., Fu, L. F., Lopchuk, J. M.*, et al.* (2011) New Synthetic Triterpenoids: Potent Agents for Prevention and Treatment of Tissue Injury Caused by Inflammatory and Oxidative Stress. *Journal of Natural Products* 74, 537-545
30. 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
31. Kuzuyama, T., and Seto, H. (2012) Two Distinct Pathways for Essential Metabolic Precursors for Isoprenoid Biosynthesis. *Proc Jpn Acad Ser B Phys Biol Sci* 88, 41-52
32. Zhao, L., Chang, W.-c., Xiao, Y., Liu, H.-w., and Liu, P. (2013) Methylerythritol Phosphate Pathway of Isoprenoid Biosynthesis. *Annual Review of Biochemistry* 82, 497-530
33. Reed, J., and Osbourn, A. (2018) Engineering Terpenoid Production through Transient Expression in *Nicotiana benthamiana*. *Plant Cell Reports* 37, 1431-1441

34. Vranova, E., Coman, D., and Gruissem, W. (2013) Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis. in *Annual Review of Plant Biology, Vol 64* (Merchant, S. S. ed.), Annual Reviews, Palo Alto. pp 665-700
35. Hemmerlin, A., Harwood, J. L., and Bach, T. J. (2012) A Raison D'être for Two Distinct Pathways in the Early Steps of Plant Isoprenoid Biosynthesis? *Progress in Lipid Research* 51, 95-148
36. Poulter, C. D. (1990) Biosynthesis of Non-Head-to-Tail Terpenes. Formation of 1'-1 and 1'-3 Linkages. *Accounts of Chemical Research* 23, 70-77
37. Hoshino, T., Nakano, S.-i., Kondo, T., Sato, T., and Miyoshi, A. (2004) Squalene–Hopene Cyclase: Final Deprotonation Reaction, Conformational Analysis for the Cyclization of (3R,S)-2,3-Oxidosqualene and Further Evidence for the Requirement of an Isopropylidene Moiety Both for Initiation of the Polycyclization Cascade and for the Formation of the 5-Membered E-Ring. *Organic & Biomolecular Chemistry* 2, 1456-1470
38. Nowosielski, M., Hoffmann, M., Wyrwicz, L. S., Stepniak, P., Plewczynski, D. M.*, et al.* (2011) Detailed Mechanism of Squalene Epoxidase Inhibition by Terbinafine. *Journal of Chemical Information and Modeling* 51, 455-462
39. Abe, I. (2007) Enzymatic Synthesis of Cyclic Triterpenes. *Natural product reports* 24, 1311-1331
40. Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., and Osbourn, A. (2014) Triterpene Biosynthesis in Plants. in *Annual Review of Plant Biology, Vol 65* (Merchant, S. S. ed.), Annual Reviews, Palo Alto. pp 225-257
41. Wendt, K. U., Schulz, G. E., Corey, E. J., and Liu, D. R. (2000) Enzyme Mechanisms for Polycyclic Triterpene Formation. *Angewandte Chemie-International Edition* 39, 2812-+
42. Xu, R., Fazio, G. C., and Matsuda, S. P. T. (2004) On the Origins of Triterpenoid Skeletal Diversity. *Phytochemistry* 65, 261-291
43. Segura, M. J. R., Meyer, M. M., and Matsuda, S. P. T. (2000) Arabidopsis Thaliana LUP1 Converts Oxidosqualene to Multiple Triterpene Alcohols and a Triterpene Diol. *Organic Letters* 2, 2257-2259
44. Phillips, D. R., Rasbery, J. M., Bartel, B., and Matsuda, S. P. T. (2006) Biosynthetic Diversity in Plant Triterpene Cyclization. *Current Opinion in Plant Biology* 9, 305-314
45. Stephenson, M. J., Field, R. A., and Osbourn, A. (2019) The Protosteryl and Dammarenyl Cation Dichotomy in Polycyclic Triterpene Biosynthesis Revisited: Has This 'Rule' Finally Been Broken? *Natural product reports*
46. Ghosh, S. (2017) Triterpene Structural Diversification by Plant Cytochrome P450 Enzymes. *Frontiers in Plant Science* 8
47. Cramer, J., Sager, C. P., and Ernst, B. (2019) Hydroxyl Groups in Synthetic and Natural-Product-Derived Therapeutics: A Perspective on a Common Functional Group. *Journal of Medicinal Chemistry*
48. McLean, K. J., Sabri, M., Marshall, K. R., Lawson, R. J., Lewis, D. G.*, et al.* (2005) Biodiversity of Cytochrome P450 Redox Systems. *Biochemical Society Transactions* 33, 796-801
49. Mizutani, M., and Sato, F. (2011) Unusual P450 Reactions in Plant Secondary Metabolism. *Archives of Biochemistry and Biophysics* 507, 194-203
50. Reed, J., Stephenson, M. J., Miettinen, K., Brouwer, B., Leveau, A.*, et al.* (2017) A Translational Synthetic Biology Platform for Rapid Access to Gram-Scale Quantities of Novel Drug-Like Molecules. *Metabolic Engineering* 42, 185-193
51. Munro, A. W., Girvan, H. M., Mason, A. E., Dunford, A. J., and McLean, K. J. (2013) What Makes a P450 Tick? *Trends in Biochemical Sciences* 38, 140-150

52. Renault, H., Bassard, J. E., Hamberger, B., and Werck-Reichhart, D. (2014) Cytochrome P450-Mediated Metabolic Engineering: Current Progress and Future Challenges. *Current Opinion in Plant Biology* 19, 27-34
53. Owatworakit, A., Townsend, B., Louveau, T., Jenner, H., Rejzek, M.*, et al.* (2013) Glycosyltransferases from Oat (*Avena*) Implicated in the Acylation of Avenacins. *Journal of Biological Chemistry* 288, 3696-3704
54. Louveau, T., and Osbourn, A. (2019) The Sweet Side of Plant-Specialized Metabolism. *Cold Spring Harbor perspectives in biology*
55. Orme, A., Louveau, T., Stephenson, M. J., Appelhagen, I., Melton, R.*, et al.* (2019) A Noncanonical Vacuolar Sugar Transferase Required for Biosynthesis of Antimicrobial Defense Compounds in Oat. *Proceedings of the National Academy of Sciences of the United States of America* 116, 27105-27114
56. Chung, S. Y., Seki, H., Fujisawa, Y., Shimoda, Y., Hiraga, S.*, et al.* (2020) A Cellulose Synthase-Derived Enzyme Catalyses 3-O-Glucuronosylation in Saponin Biosynthesis. *Nature Communications* 11, 5664
57. D'Auria, J. C. (2006) Acyltransferases in Plants: A Good Time to Be BAHD. *Current Opinion in Plant Biology* 9, 331-340
58. Mugford, S. T., and Milkowski, C. (2012) Chapter Fourteen - Serine Carboxypeptidase-Like Acyltransferases from Plants. in *Methods in Enzymology* (Hopwood, D. A. ed.), Academic Press. pp 279-297
59. Mugford, S. T., Qi, X., Bakht, S., Hill, L., Wegel, E.*, et al.* (2009) A Serine Carboxypeptidase-Like Acyltransferase Is Required for Synthesis of Antimicrobial Compounds and Disease Resistance in Oats. *Plant Cell* 21, 2473-2484
60. 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*, 246-248
61. Bally, J., Jung, H., Mortimer, C., Naim, F., Philips, J. G.*, et al.* (2018) The Rise and Rise of *Nicotiana benthamiana*: A Plant for All Reasons. *Annual Review of Phytopathology* 56, 405-426
62. Gelvin, S. B. (2003) *Agrobacterium*-Mediated Plant Transformation: The Biology Behind the "Gene-Jockeying" Tool. *Microbiology and Molecular Biology Reviews* 67, 16-37
63. 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 Biotechnology Journal* 7, 682-693
64. 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, 11
65. Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S.*, et al.* (2014) A Golden Gate Modular Cloning Toolbox for Plants. *Acs Synthetic Biology* 3, 839-843
66. Harker, M., Holmberg, N., Clayton, J. C., Gibbard, C. L., Wallace, A. D.*, et al.* (2003) Enhancement of Seed Phytosterol Levels by Expression of an N-Terminal Truncated *Hevea brasiliensis* (Rubber Tree) 3-Hydroxy-3-Methylglutaryl-Coa Reductase. *Plant Biotechnology Journal* 1, 113-121
67. Holmberg, N., Harker, M., Wallace, A. D., Clayton, J. C., Gibbard, C. L.*, et al.* (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. *The Plant Journal* 36, 12-20
68. Stephenson, M. J., Reed, J., Brouwer, B., and Osbourn, A. (2018) Transient Expression in *Nicotiana benthamiana* Leaves for Triterpene Production at a Preparative Scale. *Jove-Journal of Visualized Experiments*, 11

69. Reyes-Zurita, F. J., Pachón-Peña, G., Lizárraga, D., Rufino-Palomares, E. E., Cascante, M*., et al.* (2011) The Natural Triterpene Maslinic Acid Induces Apoptosis in HT29 Colon Cancer Cells by a JNK-p53-Dependent Mechanism. *BMC Cancer* 11, 154
70. Akihisa, T., Kamo, S., Uchiyama, T., Akazawa, H., Banno, N*., et al.* (2006) Cytotoxic Activity of *Perilla frutescens* Var. Japonica Leaf Extract Is Due to High Concentrations of Oleanolic and Ursolic Acids. *Journal of Natural Medicines* 60, 331-333
71. 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 (Basel)* 2, 2428-66
72. Adrian, T. E., and Collin, P. (2018) The Anti-Cancer Effects of Frondoside A. *Marine Drugs* 16, 16
73. Chen, M. C., Hsu, H. H., Chu, Y. Y., Cheng, S. F., Shen, C. Y*., et al.* (2018) Lupeol Alters ER Stress-Signaling Pathway by Downregulating ABCG2 Expression to Induce Oxaliplatin-Resistant LoVo Colorectal Cancer Cell Apoptosis. *Environmental Toxicology* 33, 587-593
74. Pan, W. L., Xue, B. L., Yang, C. L., Miao, L. L., Zhou, L. L*., et al.* (2018) Biopharmaceutical Characters and Bioavailability Improving Strategies of Ginsenosides. *Fitoterapia* 129, 272-282
75. World Health Organisation. (2020) The Top 10 Causes of Death.
76. Yang, W.-H., Xu, J., Mu, J.-B., and Xie, J. (2017) Revision of the Concept of Anti-Angiogenesis and Its Applications in Tumor Treatment. *Chronic Diseases and Translational Medicine* 3, 33-40
77. Wiman, K. G., and Zhivotovsky, B. (2017) Understanding Cell Cycle and Cell Death Regulation Provides Novel Weapons against Human Diseases. *Journal of Internal Medicine* 281, 483-495
78. Kumari, R., and Jat, P. (2021) Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. *Frontiers in Cell and Developmental Biology* 9
79. Dasari, S., and Bernard Tchounwou, P. (2014) Cisplatin in Cancer Therapy: Molecular Mechanisms of Action. *European Journal of Pharmacology* 740, 364-378
80. Lui, V. W. Y., Lau, C. P. Y., Cheung, C. S. F., Ho, K., Ng, M. H. L*., et al.* (2010) An RNA-Directed Nucleoside Anti-Metabolite, 1-(3-C-Ethynyl-Beta-D-Ribo-Pentofuranosyl)Cytosine (ECyd), Elicits Antitumor Effect Via TP53-Induced Glycolysis and Apoptosis Regulator (TIGAR) Downregulation. *Biochemical Pharmacology* 79, 1772-1780
81. Yang, Y., Wang, Y. W., Zhao, M., Jia, H. B., Li, B*., et al.* (2018) Tormentic Acid Inhibits IL-1β-Induced Chondrocyte Apoptosis by Activating the PI3K/Akt Signaling Pathway. *Molecular Medicine Reports* 17, 4753-4758
82. Matson, D. R., and Stukenberg, P. T. (2011) Spindle Poisons and Cell Fate: A Tale of Two Pathways. *Mol Interv* 11, 141-50
83. Spradlin, J. N., Hu, X., Ward, C. C., Brittain, S. M., Jones, M. D*., et al.* (2019) Harnessing the Anti-Cancer Natural Product Nimbolide for Targeted Protein Degradation. *Nature Chemical Biology* 15, 747-755
84. Reyes-Zurita, F. J., Rufino-Palomares, E. E., Lupiáñez, J. A., and Cascante, M. (2009) Maslinic Acid, a Natural Triterpene from *Olea europaea L.*, Induces Apoptosis in HT29 Human Colon-Cancer Cells Via the Mitochondrial Apoptotic Pathway. *Cancer Letters* 273, 44-54
85. Siddiqui, M. Z. (2011) *Boswellia serrata* , a Potential Antiinflammatory Agent: An Overview. *Indian Journal of Pharmaceutical Sciences* 73, 255-261

86. Sharma, S., Gupta, S., Khajuria, V., Bhagat, A., Ahmed, Z*., et al.* (2016) Analogues of Boswellic Acids as Inhibitors of Pro-Inflammatory Cytokines TNF-α and IL-6. *Bioorganic & Medicinal Chemistry Letters* 26, 695-698
87. Chen, L., Deng, H., Cui, H., Fang, J., Zuo, Z*., et al.* (2018) Inflammatory Responses and Inflammation-Associated Diseases in Organs. *Oncotarget* 9, 7204-7218
88. Cavaillon, J. M., and Adib-Conquy, M. (2002) The Pro-Inflammatory Cytokine Cascade. in *Immune Response in the Critically Ill* (Marshall, J. C., and Cohen, J. eds.), Springer Berlin Heidelberg, Berlin, Heidelberg. pp 37-66
89. Liu, T., Zhang, L., Joo, D., and Sun, S.-C. (2017) NF-κB Signaling in Inflammation. *Signal Transduction and Targeted Therapy* 2, 17023
90. Ahmed, S. M. U., Luo, L., Namani, A., Wang, X. J., and Tang, X. (2017) Nrf2 Signaling Pathway: Pivotal Roles in Inflammation. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1863, 585-597
91. O'Connell, M. A., and Hayes, J. D. (2015) The Keap1/Nrf2 Pathway in Health and Disease: From the Bench to the Clinic. *Biochemical Society Transactions* 43, 687-689
92. Sun, S.-C., and Liu, Z.-G. (2011) A Special Issue on NF-κB Signaling and Function. *Cell Research* 21, 1-2
93. Christian, F., Smith, E. L., and Carmody, R. J. (2016) The Regulation of NF-κB Subunits by Phosphorylation. *Cells* 5
94. Tak, P. P., and Firestein, G. S. (2001) NF-κB: A Key Role in Inflammatory Diseases. *Journal of Clinical Investigation* 107, 7-11
95. Oh, H., and Ghosh, S. (2013) NF-κB: Roles and Regulation in Different CD4(+) T-Cell Subsets. *Immunological Reviews* 252, 41-51
96. Saha, S., Buttari, B., Panieri, E., Profumo, E., and Saso, L. (2020) An Overview of Nrf2 Signaling Pathway and Its Role in Inflammation. *Molecules* 25
97. Liu, G.-H., Qu, J., and Shen, X. (2008) NF-κB/P65 Antagonizes Nrf2-ARE Pathway by Depriving CBP from Nrf2 and Facilitating Recruitment of HDAC3 to MafK. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1783, 713-727
98. Liu, X., Zhang, X., Ding, Y., Zhou, W., Tao, L*., et al.* (2017) Nuclear Factor E2-Related Factor-2 Negatively Regulates NLRP3 Inflammasome Activity by Inhibiting Reactive Oxygen Species-Induced NLRP3 Priming. *Antioxidants and Redox Signaling* 26, 28-43
99. Lee, J.-H., Koo, T. H., Yoon, H., Jung, H. S., Jin, H. Z*., et al.* (2006) Inhibition of NF-κB Activation through Targeting IκB Kinase by Celastrol, a Quinone Methide Triterpenoid. *Biochemical Pharmacology* 72, 1311-1321
100. Hussain, H., Al-Harrasi, A., Csuk, R., Shamraiz, U., Green, I. R*., et al.* (2017) Therapeutic Potential of Boswellic Acids: A Patent Review (1990-2015). *Expert Opin Ther Pat* 27, 81-90
101. Taherzadeh, D., Baradaran Rahimi, V., Amiri, H., Ehtiati, S., Yahyazadeh, R*., et al.* (2022) Acetyl-11-Keto-*β*-Boswellic Acid (AKBA) Prevents Lipopolysaccharide-Induced Inflammation and Cytotoxicity on H9C2 Cells. *Evidence-Based Complementary and Alternative Medicine* 2022, 2620710
102. 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. *Journal of Agricultural and Food Chemistry* 58, 8623-8629
103. Classen-Houben, D., Schuster, D., Da Cunha, T., Odermatt, A., Wolber, G*., et al.* (2009) Selective Inhibition of 11β-Hydroxysteroid Dehydrogenase 1 by 18α-Glycyrrhetinic Acid but Not 18β-Glycyrrhetinic Acid. *The Journal of Steroid Biochemistry and Molecular Biology* 113, 248-252

104. Zhang, L., Fan, Z. R., Wang, L., Liu, L. Q., Li, X. Z.*, et al.* (2020) Carbenoxolone Decreases Monocrotaline-Induced Pulmonary Inflammation and Pulmonary Arteriolar Remodeling in Rats by Decreasing the Expression of Connexins in T Lymphocytes. *International Journal of Molecular Medicine* 45, 81-92
105. He, F., Cheng, Q., Li, N., and Shang, Y. (2022) Carbenoxolone Ameliorates Allergic Airway Inflammation through NF-κB/NLRP3 Pathway in Mice. *Biological and Pharmaceutical Bulletin* 45, 743-750
106. Ahmad, R., Raina, D., Meyer, C., Kharbanda, S., and Kufe, D. (2006) Triterpenoid CDDO-Me Blocks the NF-Kappa B Pathway by Direct Inhibition of IKK Beta on Cys-179. *Journal of Biological Chemistry* 281, 35764-35769
107. Dinkova-Kostova, A. T., Liby, K. T., Stephenson, K. K., Holtzclaw, W. D., Gao, X.*, et al.* (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, 4584-9
108. Neymotin, A., Calingasan, N. Y., Wille, E., Naseri, N., Petri, S.*, et al.* (2011) Neuroprotective Effect of Nrf2/ARE Activators, CDDO Ethylamide and CDDO Trifluoroethylamide, in a Mouse Model of Amyotrophic Lateral Sclerosis. *Free Radic Biol Med* 51, 88-96
109. Tamura, K., Teranishi, Y., Ueda, S., Suzuki, H., Kawano, N.*, et al.* (2017) Cytochrome P450 Monooxygenase CYP716A141 Is a Unique β-Amyrin C-16β Oxidase Involved in Triterpenoid Saponin Biosynthesis in *Platycodon grandiflorus*. *Plant and Cell Physiology* 58, 874-884
110. Hodgson, H., De La Peña, R., Stephenson, M. J., Thimmappa, R., Vincent, J. L.*, et al.* (2019) Identification of Key Enzymes Responsible for Protolimonoid Biosynthesis in Plants: Opening the Door to Azadirachtin Production. *Proceedings of the National Academy of Sciences* 116, 17096
111. 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, 2063-2066
112. Sparkman, O. D., Zelda, P., and Fulton, G. K. (2011) *Gas Chromatography and Mass Spectrometry: A Practical Guide*, Academic Press, Burlington, MA
113. Uddin, J., Muhsinah, A. B., Imran, M., Khan, M. N., and Musharraf, S. G. (2022) Structure–Fragmentation Study of Pentacyclic Triterpenoids Using Electrospray Ionization Quadrupole Time-of-Flight Tandem Mass Spectrometry (ESI-QTOFMS/MS). *Rapid Communications in Mass Spectrometry* 36, e9243
114. Fukushima, E. O., Seki, H., Ohyama, K., Ono, E., Umemoto, N.*, et al.* (2011) CYP716A Subfamily Members Are Multifunctional Oxidases in Triterpenoid Biosynthesis. *Plant and Cell Physiology* 52, 2050-2061
115. Moses, T., Pollier, J., Shen, Q., Soetaert, S., Reed, J.*, et al.* (2015) OSC2 and CYP716A14v2 Catalyze the Biosynthesis of Triterpenoids for the Cuticle of Aerial Organs of *Artemisia annua*. *Plant Cell* 27, 286-301
116. Yasumoto, S., Seki, H., Shimizu, Y., Fukushima, E. O., and Muranaka, T. (2017) Functional Characterization of CYP716 Family P450 Enzymes in Triterpenoid Biosynthesis in Tomato. *Frontiers in Plant Science* 8
117. Miettinen, K., Pollier, J., Buyst, D., Arendt, P., Csuk, R.*, et al.* (2017) The Ancient CYP716 Family Is a Major Contributor to the Diversification of Eudicot Triterpenoid Biosynthesis. *Nature Communications* 8, 14153
118. Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T.*, et al.* (2011) Triterpene Functional Genomics in Licorice for Identification of CYP72A154 Involved in the Biosynthesis of Glycyrrhizin. *Plant Cell* 23, 4112-4123

119. Yano, R., Takagi, K., Takada, Y., Mukaiyama, K., Tsukamoto, C.*, et al.* (2017) Metabolic Switching of Astringent and Beneficial Triterpenoid Saponins in Soybean Is Achieved by a Loss-of-Function Mutation in Cytochrome P450 72A69. *The Plant Journal* 89, 527-539
120. Geisler, K., Hughes, R. K., Sainsbury, F., Lomonossoff, G. P., Rejzek, M.*, et al.* (2013) Biochemical Analysis of a Multifunctional Cytochrome P450 (CYP51) Enzyme Required for Synthesis of Antimicrobial Triterpenes in Plants. *Proceedings of the National Academy of Sciences* 110, E3360-E3367
121. Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T.*, et al.* (2008) Licorice β-Amyrin 11-Oxidase, a Cytochrome P450 with a Key Role in the Biosynthesis of the Triterpene Sweetener Glycyrrhizin. *Proceedings of the National Academy of Sciences of the United States of America* 105, 14204-14209
122. Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T.*, et al.* (2006) Identification of β-Amyrin and Sophoradiol 24-Hydroxylase by Expressed Sequence Tag Mining and Functional Expression Assay. *Febs Journal* 273, 948-959
123. Li, Y., Leveau, A., Zhao, Q., Feng, Q., Lu, H.*, et al.* (2021) Subtelomeric Assembly of a Multi-Gene Pathway for Antimicrobial Defense Compounds in Cereals. *Nature Communications* 12, 2563
124. Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S.*, et al.* (2001) A New Class of Oxidosqualene Cyclases Directs Synthesis of Antimicrobial Phytoprotectants in Monocots. *Proceedings of the National Academy of Sciences of the United States of America* 98, 13431-13436
125. Zhang, K., and Zuo, Y. (2005) Pitfalls and Solution for Simultaneous Determination of Estrone and 17α-Ethinylestradiol by Gas Chromatography–Mass Spectrometry after Derivatization with N,O-Bis(Trimethylsilyl)Trifluoroacetamide. *Analytica Chimica Acta* 554, 190-196
126. Clayden, J., Greeves, N., Warren, S., and Wothers, P. (2001) *Organic Chemistry*, 1st ed., Oxford University Press
127. 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
128. Floss, H. G. (2006) Combinatorial Biosynthesis—Potential and Problems. *Journal of Biotechnology* 124, 242-257
129. Malhotra, K., and Franke, J. (2022) Cytochrome P450 Monooxygenase-Mediated Tailoring of Triterpenoids and Steroids in Plants. *Beilstein Journal of Organic Chemistry* 18, 1289-1310
130. Anonymous. (2017) TriForC Database.
131. Miettinen, K., Iñigo, S., Kreft, L., Pollier, J., De Bo, C.*, et al.* (2018) The TriForC Database: A Comprehensive up-to-Date Resource of Plant Triterpene Biosynthesis. *Nucleic Acids Research* 46, D586-D594
132. Kushiro, T., Shibuya, M., and Ebizuka, Y. (1998) β-Amyrin Synthase - Cloning of Oxidosqualene Cyclase That Catalyzes the Formation of the Most Popular Triterpene among Higher Plants. *European Journal of Biochemistry* 256, 238-244
133. Poulous, T. L., and Johnson, E. F. (2015) Structures of Cytochrome P450 Enzymes. in *Cytochrome P450 Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed.), 4th Edition Ed., Springer International Publishing. pp 3-32
134. Putkaradze, N., Litzenburger, M., Abdulmughni, A., Milhim, M., Brill, E.*, et al.* (2017) CYP109E1 Is a Novel Versatile Statin and Terpene Oxidase from *Bacillus megaterium*. *Applied Microbiology and Biotechnology* 101, 8379-8393

135. Diczfalusy, U., and Björkhem, I. (2011) Still Another Activity by the Highly Promiscuous Enzyme CYP3A4: 25-Hydroxylation of Cholesterol. *Journal of Lipid Research* 52, 1447-1449
136. Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I.*, et al.* (2015) CYP72A67 Catalyzes a Key Oxidative Step in *Medicago truncatula* Hemolytic Saponin Biosynthesis. *Molecular Plant* 8, 1493-1506
137. Kelly, S. L., and Kelly, D. E. (2013) Microbial Cytochromes P450: Biodiversity and Biotechnology. Where Do Cytochromes P450 Come from, What Do They Do and What Can They Do for Us? *Philos Trans R Soc Lond B Biol Sci* 368, 20120476
138. Jeffreys, L. N., Poddar, H., Golovanova, M., Levy, C. W., Girvan, H. M.*, et al.* (2019) Novel Insights into P450 BM3 Interactions with FDA-Approved Antifungal Azole Drugs. *Scientific Reports* 9, 1577
139. Thistlethwaite, S., Jeffreys, L. N., Girvan, H. M., McLean, K. J., and Munro, A. W. (2021) A Promiscuous Bacterial P450: The Unparalleled Diversity of BM3 in Pharmaceutical Metabolism. *International journal of molecular sciences* 22, 11380
140. Schmitz, D., Zapp, J., and Bernhardt, R. (2012) Hydroxylation of the Triterpenoid Dipterocarpol with CYP106A2 from *Bacillus megaterium*. *FEBS Journal* 279, 1663-74
141. Brill, E., Hannemann, F., Zapp, J., Bruning, G., Jauch, J.*, et al.* (2014) A New Cytochrome P450 System from *Bacillus megaterium* DSM319 for the Hydroxylation of 11-Keto-β-Boswellic Acid (KBA). *Applied Microbiology and Biotechnology* 98, 1701-17
142. Kiss, F. M., Schmitz, D., Zapp, J., Dier, T. K., Volmer, D. A.*, et al.* (2015) Comparison of CYP106A1 and CYP106A2 from *Bacillus megaterium* - Identification of a Novel 11-Oxidase Activity. *Applied Microbiology and Biotechnology* 99, 8495-514
143. Bleif, S. (2012) *Von der identifizierung neuer substrate Der 15ß-Hydroxylase (CYP106A2) Zur cytochrom P450-Ganzzellkatalyse In Bacillus megaterium* PhD Doctoral thesis, Saarland University
144. O'Keefe, D. P., Tepperman, J. M., Dean, C., Leto, K. J., Erbes, D. L.*, et al.* (1994) Plant Expression of a Bacterial Cytochrome P450 That Catalyzes Activation of a Sulfonylurea Pro-Herbicide. *Plant Physiology* 105, 473-482
145. Van Aken, B. (2009) Transgenic Plants for Enhanced Phytoremediation of Toxic Explosives. *Current Opinion in Biotechnology* 20, 231-6
146. Zhao, M., Ma, J., Li, M., Zhang, Y., Jiang, B.*, et al.* (2021) Cytochrome P450 Enzymes and Drug Metabolism in Humans. *International journal of molecular sciences* 22, 12808
147. Shet, M. S., Fisher, C. W., Holmans, P. L., and Estabrook, R. W. (1993) Human Cytochrome-P450-3A4 - Enzymatic-Properties of a Purified Recombinant Fusion Protein Containing NADPH-P450 Reductase. *Proceedings of the National Academy of Sciences of the United States of America* 90, 11748-11752
148. Shiota, N., Nagasawa, A., Sakaki, T., Yabusaki, Y., and Ohkawa, H. (1994) Herbicide-Resistant Tobacco Plants Expressing the Fused Enzyme between Rat Cytochrome P4501A1 (CYP1A1) and Yeast NADPH-Cytochrome P450 Oxidoreductase. *Plant Physiology* 106, 17-23
149. Shiota, N., Kodama, S., Inui, H., and Ohkawa, H. (2000) Expression of Human Cytochromes P450 1A1 and P450 1A2 as Fused Enzymes with Yeast NADPH-Cytochrome P450 Oxidoreductase in Transgenic Tobacco Plants. *Bioscience Biotechnology and Biochemistry* 64, 2025-2033
150. Sheludko, Y. V., Gerasymenko, I. M., and Warzecha, H. (2018) Transient Expression of Human Cytochrome P450s 2D6 and 3A4 in *Nicotiana benthamiana* Provides a Possibility for Rapid Substrate Testing and Production of Novel Compounds. *Biotechnology Journal* 13, 1700696

151. Huang, A. C., Jiang, T., Liu, Y. X., Bai, Y. C., Reed, J.*, et al.* (2019) A Specialized Metabolic Network Selectively Modulates *Arabidopsis* Root Microbiota. *Science* 364
152. Schweizer, S., Eichele, K., Ammon, H. P. T., and Safayhi, H. (2000) 3-Acetoxy Group of Genuine AKBA (Acetyl-11-Keto-β-Boswellic Acid) Is α-Configurated. *Planta Medica* 66, 781-782
153. Reed, J. (2017) *Transient Expression for Engineering Triterpenoid Diversity in Plants.* PhD Doctoral Thesis, University of East Anglia
154. Park, K. J., Subedi, L., Kim, S. Y., Choi, S. U., and Lee, K. R. (2018) Bioactive Triterpenoids from Twigs of *Betula schmidtii*. *Bioorganic Chemistry* 77, 527-533
155. Biswas, T., Mathur, A. K., and Mathur, A. (2017) A Literature Update Elucidating Production of *Panax* Ginsenosides with a Special Focus on Strategies Enriching the Anti-Neoplastic Minor Ginsenosides in Ginseng Preparations. *Applied Microbiology and Biotechnology* 101, 4009-4032
156. Bian, Z., Xu, F., Liu, H., and Du, Y. (2022) Ursolic Acid Ameliorates the Injury of H9c2 Cells Caused by Hypoxia and Reoxygenation through Mediating CXCL2/NF-κB Pathway. *International Heart Journal* 63, 755-762
157. Fujiwara, Y., Komohara, Y., Ikeda, T., and Takeya, M. (2011) Corosolic Acid Inhibits Glioblastoma Cell Proliferation by Suppressing the Activation of Signal Transducer and Activator of Transcription-3 and Nuclear Factor-Kappa B in Tumor Cells and Tumor-Associated Macrophages. *Cancer Science* 102, 206-211
158. Cascão, R., Fonseca, J. E., and Moita, L. F. (2017) Celastrol: A Spectrum of Treatment Opportunities in Chronic Diseases. *Frontiers in Medicine* 4
159. Buttke, T. M., McCubrey, J. A., and Owen, T. C. (1993) Use of an Aqueous Soluble Tetrazolium/Formazan Assay to Measure Viability and Proliferation of Lymphokine-Dependent Cell Lines. *Journal of Immunological Methods* 157, 233-240
160. Cory, A. H., Owen, T. C., Barltrop, J. A., and Cory, J. G. (1991) Use of an Aqueous Soluble Tetrazolium/Formazan Assay for Cell Growth Assays in Culture. *Cancer Commun* 3, 207-12
161. Reddy, S. D., Siva, B., Phani Babu, V. S., Vijaya, M., Nayak, V. L.*, et al.* (2017) New Cycloartane Type-Triterpenoids from the Areal Parts of *Caragana sukiensis* and Their Biological Activities. *European Journal of Medicinal Chemistry* 136, 74-84
162. Sporn, M. B., Liby, K. T., Yore, M. M., Fu, L., Lopchuk, J. M.*, et al.* (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
163. Chanput, W., Mes, J. J., and Wichers, H. J. (2014) THP-1 Cell Line: An in Vitro Cell Model for Immune Modulation Approach. *International Immunopharmacology* 23, 37-45
164. Rushworth, S. A., MacEwan, D. J., and O'Connell, M. A. (2008) Lipopolysaccharide-Induced Expression of NAD(P)H:Quinone Oxidoreductase 1 and Heme Oxygenase-1 Protects against Excessive Inflammatory Responses in Human Monocytes. *Journal of Immunology* 181, 6730-6737
165. Lee, J., Rhee, M. H., Kim, E., and Cho, J. Y. (2012) BAY 11-7082 Is a Broad-Spectrum Inhibitor with Anti-Inflammatory Activity against Multiple Targets. *Mediators Inflamm* 2012, 416036
166. Vo, N. N. Q., Fukushima, E. O., and Muranaka, T. (2017) Structure and Hemolytic Activity Relationships of Triterpenoid Saponins and Sapogenins. *Journal of Natural Medicines* 71, 50-58
167. Song, G., Shen, X., Li, S., Li, Y., Si, H.*, et al.* (2016) Structure-Activity Relationships of 3-O-β-Chacotriosyl Oleanane-Type Triterpenoids as Potential H5N1 Entry Inhibitors. *European Journal of Medicinal Chemistry* 119, 109-121

168. Sui, J.-J., Zhou, W.-H., Liu, D.-Y., Li, M.-Q., and Sun, J.-S. (2017) Highly Efficient Synthesis of Bioactive Oleanane-Type Saponins. *Carbohydrate Research* 452, 43-46
169. Yang, Y.-H., Dai, S.-Y., Deng, F.-H., Peng, L.-H., Li, C.*, et al.* (2022) Recent Advances in Medicinal Chemistry of Oleanolic Acid Derivatives. *Phytochemistry* 203, 113397
170. Wells, D. A. (2003) Chapter 11 Solid-Phase Extraction: High Throughput Techniques. in *Progress in Pharmaceutical and Biomedical Analysis* (Wells, D. A. ed.), Elsevier. pp 361-432
171. Thornburg, C. C., Britt, J. R., Evans, J. R., Akee, R. K., Whitt, J. A.*, et al.* (2018) NCI Program for Natural Product Discovery: A Publicly-Accessible Library of Natural Product Fractions for High-Throughput Screening. *Acs Chemical Biology* 13, 2484-2497
172. Grkovic, T., Akee, R. K., Thornburg, C. C., Trinh, S. K., Britt, J. R.*, et al.* (2020) National Cancer Institute (NCI) Program for Natural Products Discovery: Rapid Isolation and Identification of Biologically Active Natural Products from the NCI Prefractionated Library. *Acs Chemical Biology* 15, 1104-1114
173. National Cancer Institute. (2019) The NCI Program for Natural Product Discovery (NPNPD) Prefractionated Library.
174. Li, H. M., He, N., Li, X. Y., Zhou, L., Zhao, M. Y.*, et al.* (2013) Oleanolic Acid Inhibits Proliferation and Induces Apoptosis in NB4 Cells by Targeting PML/RAR α. *Oncology Letters* 6, 885-890
175. Abdel-Tawab, M., Werz, O., and Schubert-Zsilavecz, M. (2011) *Boswellia serrata* an Overall Assessment of in Vitro, Preclinical, Pharmacokinetic and Clinical Data. *Clinical Pharmacokinetics* 50, 349-369
176. Brendolise, C., Yauk, Y. K., Eberhard, E. D., Wang, M., Chagne, D.*, et al.* (2011) An Unusual Plant Triterpene Synthase with Predominant α-Amyrin-Producing Activity Identified by Characterizing Oxidosqualene Cyclases from *Malus x domestica*. *Febs Journal* 278, 2485-2499
177. Zou, Z., Tian, S., Gao, X., and Li, Y. (2019) mlDEEPre: Multi-Functional Enzyme Function Prediction with Hierarchical Multi-Label Deep Learning. *Frontiers in Genetics* 9
178. Fu, L., Zhang, J., and Si, T. (2020) Recent Advances in High-Throughput Mass Spectrometry That Accelerates Enzyme Engineering for Biofuel Research. *BMC Energy* 2, 1
179. Pluchinsky, A. J., Wackelin, D. J., Huang, X., Arnold, F. H., and Mrksich, M. (2020) High Throughput Screening with SAMDI Mass Spectrometry for Directed Evolution. *Journal of the American Chemical Society* 142, 19804-19808
180. Jacobsen, N. E. (2007) *NMR Spectroscopy Explained : Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*, John Wiley & Sons, Incorporated, Hoboken, UNITED STATES
181. McDonnell, A. M., and Dang, C. H. (2013) Basic Review of the Cytochrome P450 System. *J Adv Pract Oncol* 4, 263-8
182. Zanger, U. M., and Schwab, M. (2013) Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation. *Pharmacology & Therapeutics* 138, 103-141
183. Basheer, L., and Kerem, Z. (2015) Interactions between CYP3A4 and Dietary Polyphenols. *Oxid Med Cell Longev* 2015, 854015
184. National Center for Biotechnology Information. (2022) Carbenoxolone Sodium.
185. Wang, J., Jiang, Z., Xiang, L., Li, Y., Ou, M.*, et al.* (2014) Synergism of Ursolic Acid Derivative US597 with 2-Deoxy-D-Glucose to Preferentially Induce Tumor Cell Death by Dual-Targeting of Apoptosis and Glycolysis. *Scientific Reports* 4, 5006

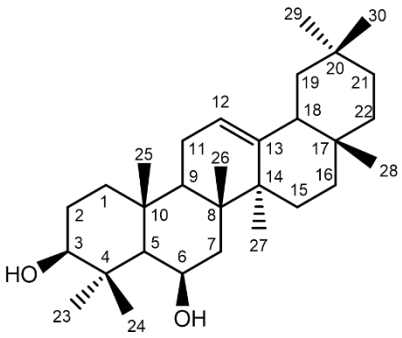
186. Christensen, L. P. (2009) Ginsenosides: Chemistry, Biosynthesis, Analysis, and Potential Health Effects. in *Advances in Food and Nutrition Research* (Taylor, S. L. ed.), 1st Edition Ed., Academic Press. pp
187. Omar, H. R., Komarova, I., El-Ghonemi, M., Fathy, A., Rashad, R.*, et al.* (2012) Licorice Abuse: Time to Send a Warning Message. *Therapeutic advances in endocrinology and metabolism* 3, 125-38
188. Dai, L., Li, D., Cheng, J., Liu, J., Deng, L.-H.*, et al.* (2014) Water Soluble Multiarm-Polyethylene Glycol–Betulinic Acid Prodrugs: Design, Synthesis, and in Vivo Effectiveness. *Polymer Chemistry* 5, 5775-5783
189. de Jesus, J. A., Laurenti, M. D., Antonangelo, L., Faria, C. S., Lago, J. H. G.*, et al.* (2021) Related Pentacyclic Triterpenes Have Immunomodulatory Activity in Chronic Experimental Visceral Leishmaniasis. *Journal of Immunology Research* 2021
190. Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M.*, et al.* (2011) ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Research* 40, D1100-D1107
191. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T.*, et al.* (2019) The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Research* 47, W636-W641
192. Ren, Y., and Kinghorn, A. D. (2019) Natural Product Triterpenoids and Their Semi-Synthetic Derivatives with Potential Anticancer Activity. *Planta Med* 85, 802-814
193. GSK. (2022) Medicago and GSK Announce the Approval by Health Canada of COVIFENZ®, an Adjuvanted Plant-Based COVID-19 Vaccine.
194. Reed, J., Orme, A., El-Demerdash, A., Owen, C., Martin, L. B. B.*, et al.* (2023) Elucidation of the Pathway for Biosynthesis of Saponin Adjuvants from the Soapbark Tree. *Science* 379, 1252-1264
195. Kim, D. K., Nam, I. Y., Kim, J. W., Shin, T. Y., and Lim, J. P. (2002) Pentacyclic Triterpenoids from *Ilex macropoda*. *Arch Pharm Res* 25, 617-620
196. Feng, X., Zou, Z., Fu, S., Sun, L., Su, Z.*, et al.* (2010) Microbial Oxidation and Glucosidation of Echinocystic Acid by *Nocardia corallina*. *Journal of Molecular Catalysis B: Enzymatic* 66, 219-223
197. Leveau, A., Reed, J., Qiao, X., Stephenson, M. J., Mugford, S. T.*, et al.* (2019) Towards Take-All Control: A C-21β Oxidase Required for Acylation of Triterpene Defence Compounds in Oat. *The New phytologist* 221, 1544-1555
198. Fukushima, E. O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K.*, et al.* (2013) Combinatorial Biosynthesis of Legume Natural and Rare Triterpenoids in Engineered Yeast. *Plant and Cell Physiology* 54, 740-749
199. Moses, T., Pollier, J., Faizal, A., Apers, S., Pieters, L.*, et al.* (2015) Unraveling the Triterpenoid Saponin Biosynthesis of the African Shrub *Maesa lanceolata. Molecular Plant* 8, 122-135
200. Sheludko, Y. V., Gerasymenko, I. M., Herrmann, F. J., and Warzecha, H. (2022) Evaluation of Biotransformation Capacity of Transplastomic Plants and Hairy Roots of *Nicotiana tabacum* Expressing Human Cytochrome P450 2D6. *Transgenic Research* 31, 351-368
201. Kahnt, M., Heller, L., Grabandt, P., Al-Harrasi, A., and Csuk, R. (2018) Platanic Acid: A New Scaffold for the Synthesis of Cytotoxic Agents. *European Journal of Medicinal Chemistry* 143, 259-265
202. Kiem, P. V., Dung, D. T., Yen, P. H., Nhiem, N. X., Quang, T. H.*, et al.* (2018) New Isomalabaricane Analogues from the Sponge *Rhabdastrella providentiae* and Their Cytotoxic Activities. *Phytochemistry Letters* 26, 199-204
203. Rozimamat, R., Hu, R., and Aisa, H. A. (2018) New Isopimarane Diterpenes and Nortriterpene with Cytotoxic Activity from *Ephorbia alatavica* Boiss. *Fitoterapia* 127, 328-333

204. Teng, Y., Zhang, H. Q., Zhou, J. F., Zhan, G. Q., and Yao, G. M. (2018) Hebecarposides A-K, Antiproliferative Lanostane-Type Triterpene Glycosides from the Leaves of *Lyonia ovalifolia* Var. *hebecarpa*. *Phytochemistry* 151, 32-41
205. Xu, F., Huang, X. L., Wu, H. M., and Wang, X. P. (2018) Beneficial Health Effects of Lupenone Triterpene: A Review. *Biomedicine & Pharmacotherapy* 103, 198-203
206. Zhang, S. B., Li, Z. H., Stadler, M., Chen, H. P., Huang, Y.*, et al.* (2018) Lanostane Triterpenoids from *Tricholoma pardinum* with NO Production Inhibitory and Cytotoxic Activities. *Phytochemistry* 152, 105-112
207. Zhu, N. L., Sun, Z. H., Hu, M. G., Li, Y. D., Zhang, D. W.*, et al.* (2018) Cucurbitane-Type Triterpenes from the Tubers of *Hemsleya penxianensis* and Their Bioactive Activity. *Phytochemistry* 147, 49-56

A. Supplemental Information

A.1 NMR Spectra

Carbon numbering scheme



Selected 2D NMR

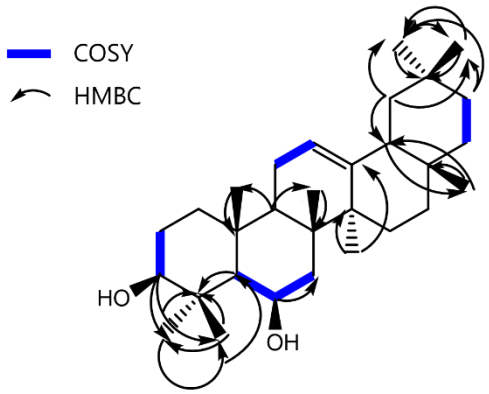


Table A-1: NMR assignments for 6β-hydroxy-β-amyrin (400 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
13	144.32	C	N/A	N/A
12	122.06	CH	5.24	t, J = 3.4 Hz
3	79.11	CH	3.20-3.12	m
6	68.73	CH	4.58	s (br)
5	55.59	CH	0.77-0.74	m
9	47.94	CH	1.71-1.57	m
18	47.25	CH	2.08-1.90	m
19	46.85	CH_2	1.71-1.57, 1.07-1.03	m
14	42.34	C	N/A	N/A
7	40.77	CH_2	1.90-1.74, 1.52-1.46	m
1	40.73	CH_2	1.71-1.57	m
4	39.57	C	N/A	N/A
8	38.91	C	N/A	N/A
22	37.13	CH_2	1.26-1.19	m
10	36.42	C	N/A	N/A
21	34.74	CH_2	1.14-1.07	m
29	33.33	CH_3	0.88	s
17	32.50	C	N/A	N/A
20	31.10	C	N/A	N/A
28	28.36	CH_3	0.84	s
23	27.92	CH_3	1.08	s
2	27.42	CH_2	1.71-1.57	m
16	26.95	CH_2	2.08-1.90, 0.82-0.78	m
15	26.12	CH_2	1.90-1.74, 0.98-0.90	m
27	26.03	CH_3	1.11	s
30	23.69	CH_3	0.88	s
11	23.43	CH_2	2.08-1.90	m
26	18.38	CH_3	1.27	s
24	17.04	CH_3	1.18	s
25	17.04	CH_3	1.33	s

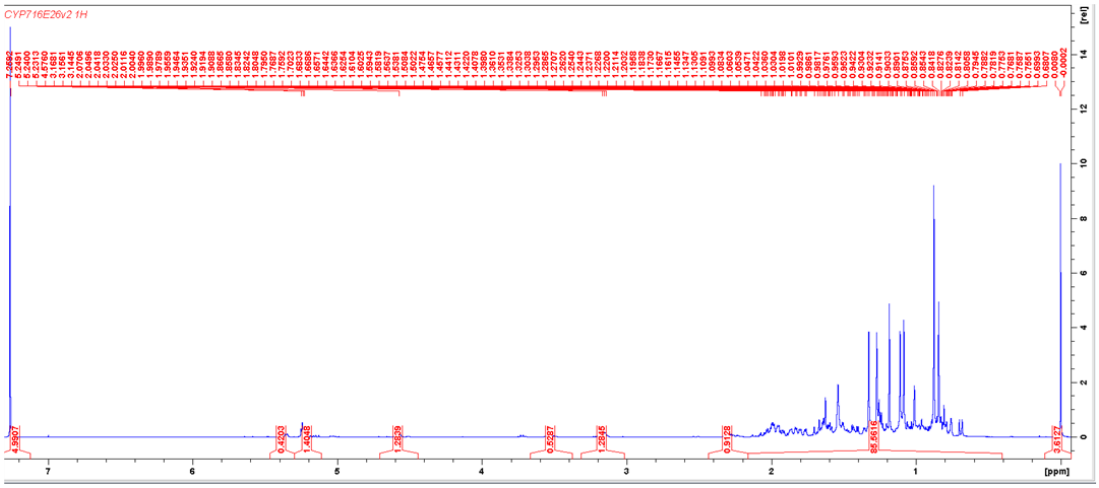


Figure A-1: ^{1}H NMR spectrum for 6β-hydroxy-β-amyrin (400 MHz)

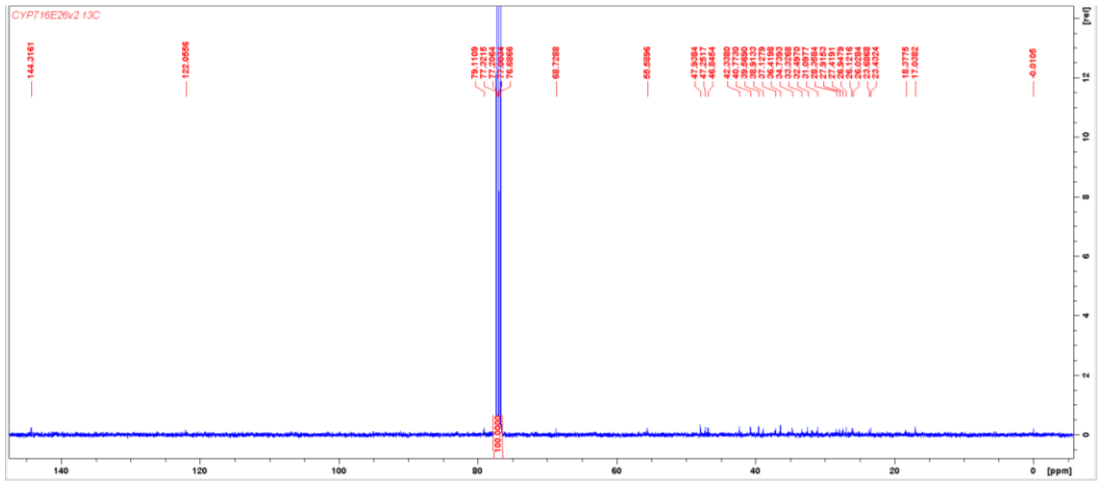


Figure A-2: ^{13}C NMR spectrum for 6β-hydroxy-β-amyrin (400 MHz)

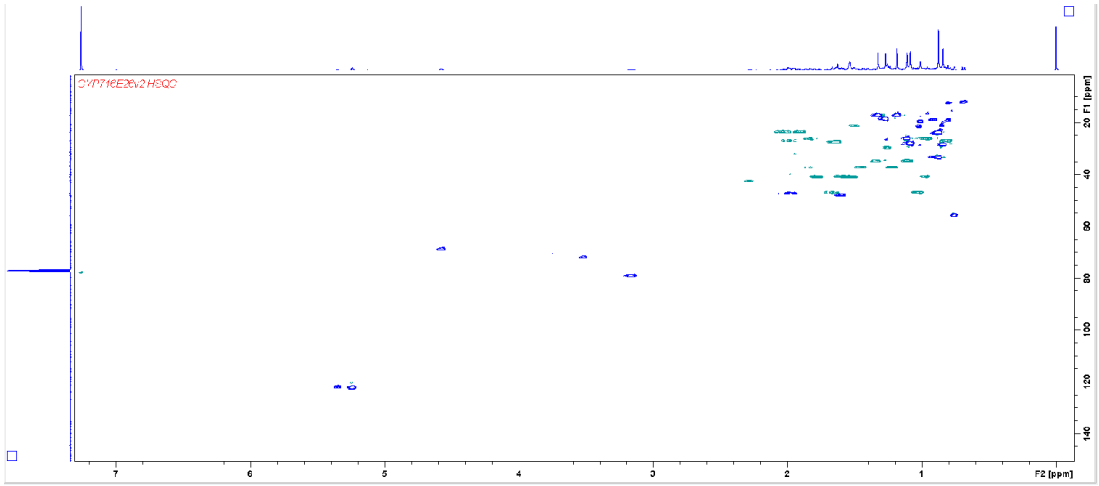


Figure A-3: HSQC NMR spectrum for 6β-hydroxy-β-amyrin (400 MHz)

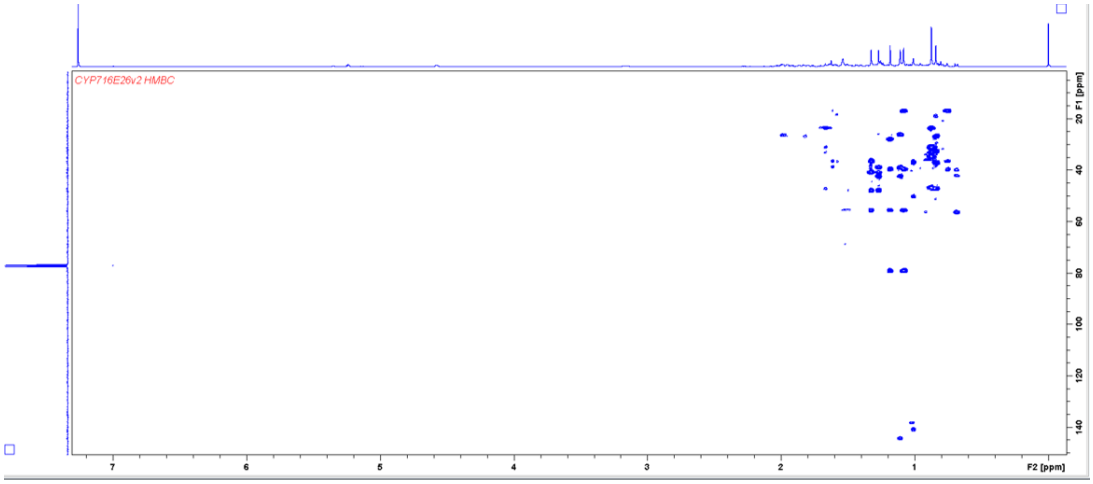


Figure A-4: HMBC NMR spectrum for 6β-hydroxy-β-amyrin (400 MHz)

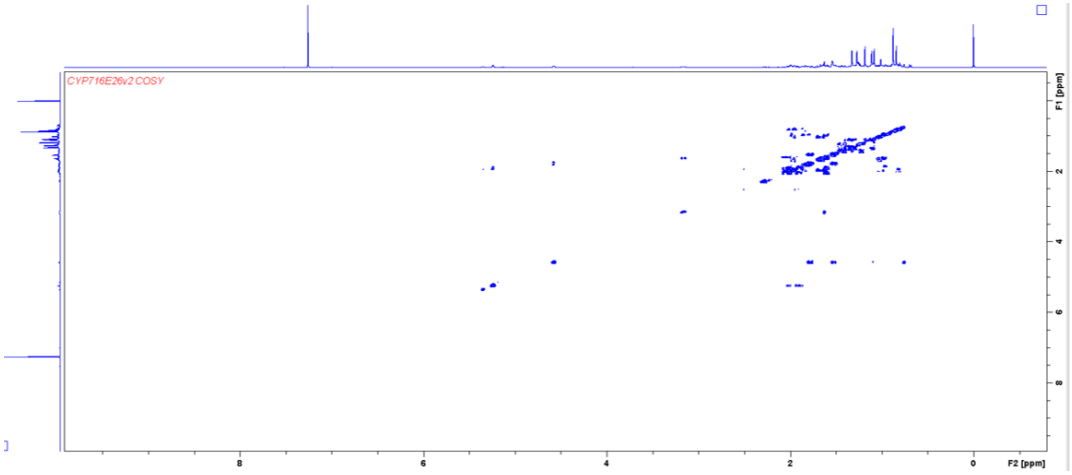
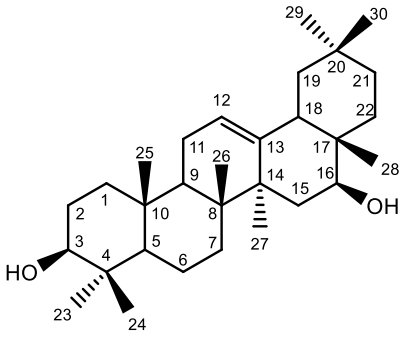


Figure A-5: COSY NMR spectrum for 6β-hydroxy-β-amyrin (400 MHz)

Carbon numbering scheme



Selected 2D NMR

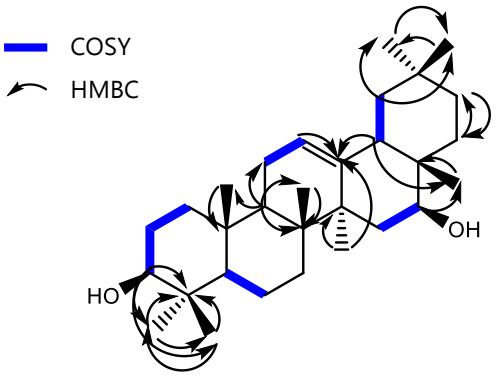


Table A-2: NMR assignments for 16β-hydroxy-β-amyrin (600 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
13	143.53	C	N/A	N/A
12	122.33	CH	5.25	t, J = 3.5 Hz
3	78.99	CH	3.22	d, J = 11.4 Hz
16	66.03	CH	4.20	d, J = 11.0 Hz
5	55.19	CH	0.73	dd, J = 11.8, 1.6 Hz
18	49.11	CH	2.15	dd, J = 14.0, 4.6 Hz
9	46.85	CH	1.49-1.54	m
19	46.56	CH_2	1.57-1.72, 1.03-1.09	m
14	43.81	C	N/A	N/A
8	39.89	C	N/A	N/A
4	38.79	C	N/A	N/A
1	38.60	CH_2	1.57-1.72, 0.95-0.98	m
17	37.34	C	N/A	N/A
10	36.89	C	N/A	N/A
15	35.61	CH_2	1.57-1.72, 1.26-1.46	m
21	34.18	CH_2	1.26-1.46, 1.10-1.20	m
29	33.26	CH_3	0.89	s
7	32.66	CH_2	1.49-1.54, 1.26-1.46	m
20	30.93	C	N/A	N/A
22	30.57	CH_2	1.80-1.95, 1.10-1.20	m
23	28.10	CH_3	1.00	s
2	27.22	CH_2	1.57-1.72	m
27	27.14	CH_3	1.22	s
30	23.97	CH_3	0.91	s
11	23.54	CH_2	1.80-1.95	m
28	21.48	CH_3	0.80	s
6	18.35	CH_2	1.57, 1.26-1.46	m
26	16.85	CH_3	0.99	s
24	15.58	CH_3	0.79	s
25	15.53	CH_3	0.94	s

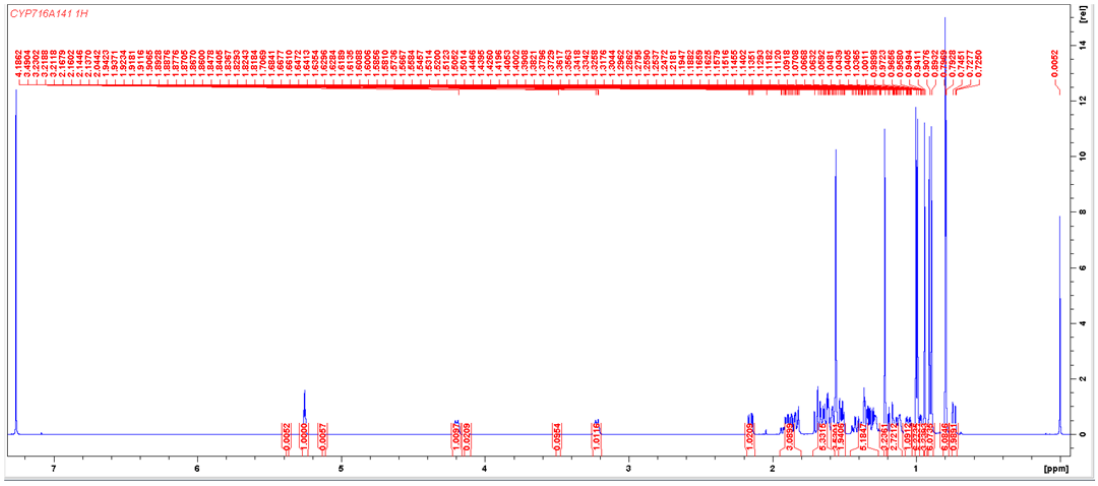


Figure A-6: ^{1}H NMR spectrum for 16β-hydroxy-β-amyrin (600 MHz)

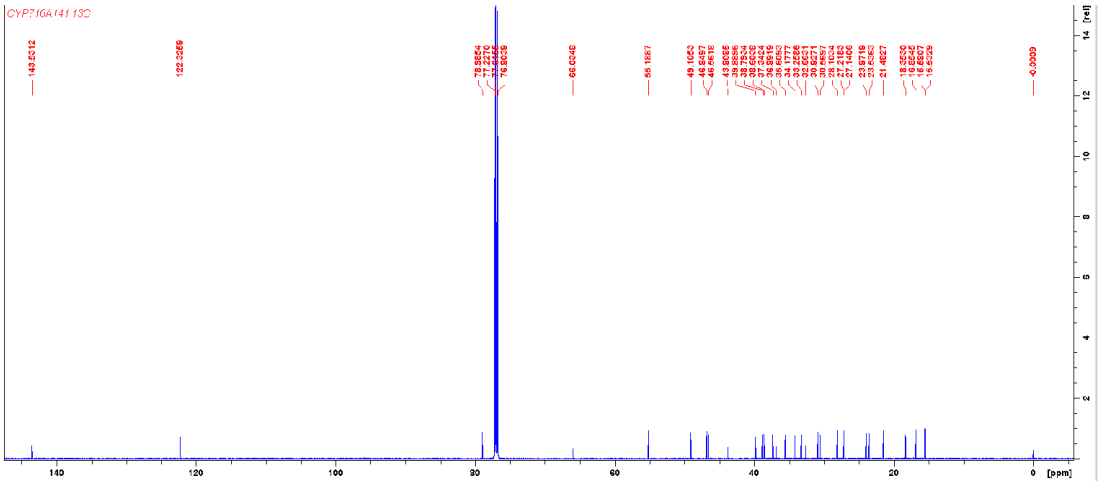


Figure A-7: ^{13}C NMR spectrum for 16β-hydroxy-β-amyrin (600 MHz)

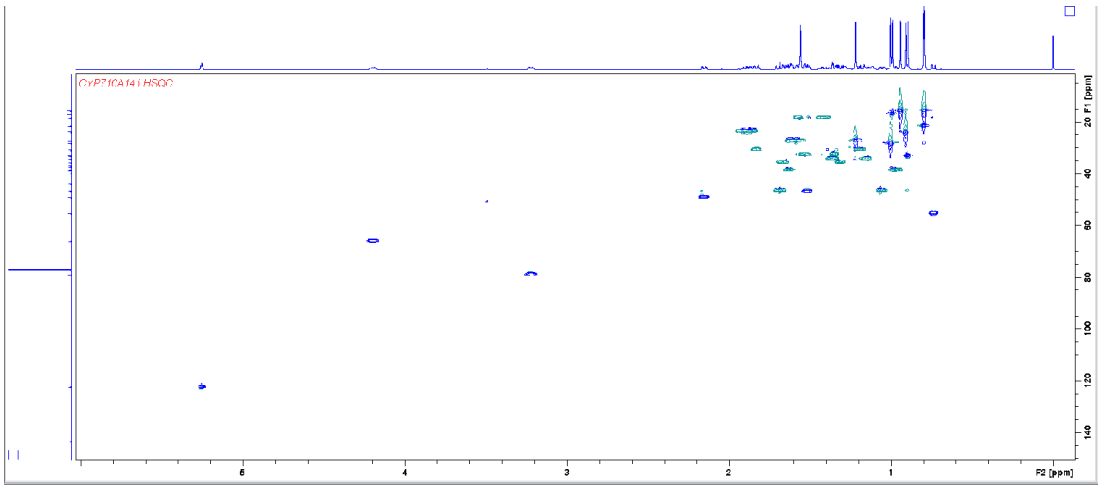


Figure A-8: HSQC NMR spectrum for 16β-hydroxy-β-amyrin (600 MHz)

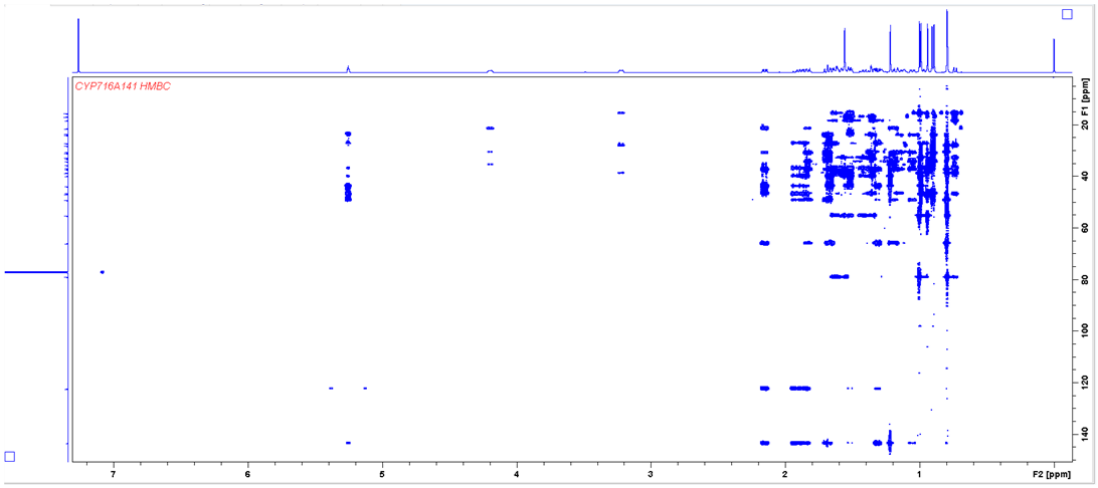


Figure A-9: HMBC NMR spectrum for 16β-hydroxy-β-amyrin (600 MHz)

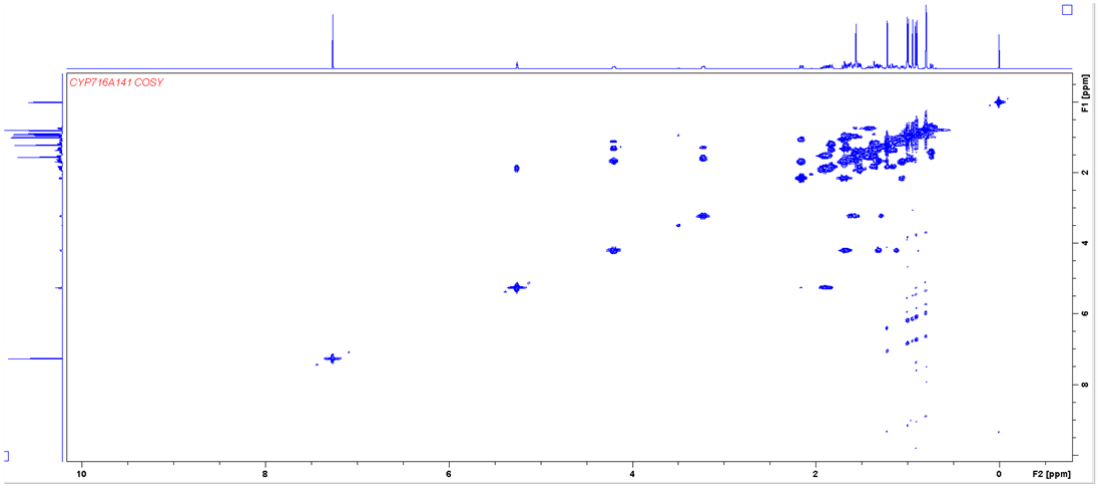


Figure A-10: COSY NMR spectrum for 16β-hydroxy-β-amyrin (600 MHz)

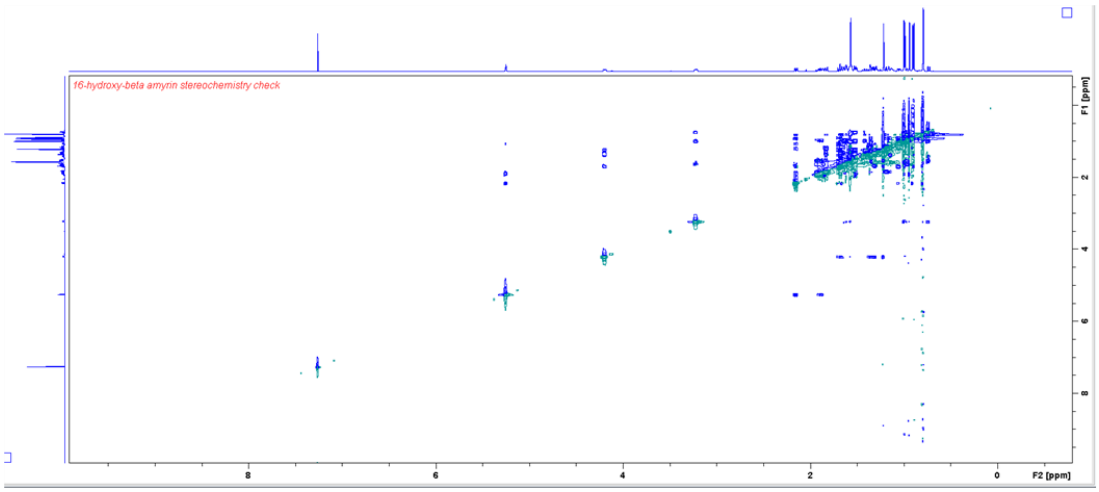
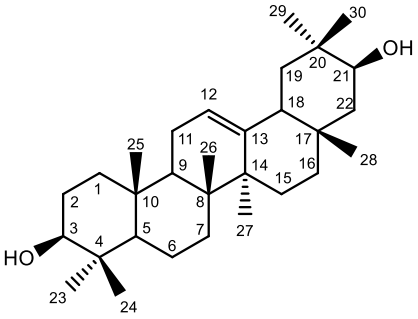


Figure A-11: ROESY NMR spectrum for 16β-hydroxy-β-amyrin (600 MHz)

Carbon numbering scheme



Selected 2D NMR

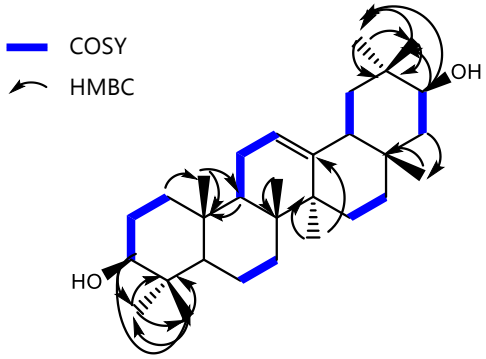


Table A-3: NMR assignments for 21β-hydroxy-β-amyrin (400 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
13	143.75	C	N/A	N/A
12	122.60	CH	5.22	t, J = 3.5 Hz
21	79.02	CH	3.22	dt, J = 11.4, 5.3 Hz
3	74.05	CH	3.55-3.50	dt, J = 12.0, 4.7 Hz
18	55.19	CH	0.74	d, J = 11.7 Hz
5	47.62	CH	1.66-1.52	m
2	47.03	CH_2	1.80-1.73, 1.16-1.13	m
9	46.66	CH	2.02-1.92	m
1	45.43	CH_2	1.52-1.47, 1.39-1.31	m
14	41.68	C	N/A	N/A
8	39.77	C	N/A	N/A
20	38.78	C	N/A	N/A
22	38.60	CH_2	1.66-1.55, 0.99-0.97	m
17	36.96	C	N/A	N/A
4	36.31	C	N/A	N/A
10	35.02	C	N/A	N/A
19	32.67	CH_2	1.52-1.47, 1.39-1.31	m
23	29.07	CH_3	0.96	s
25	28.30	CH_3	0.87	s
6	28.25	CH_2	2.02-1.92	m
29	28.11	CH_3	1.00	s
7	27.23	CH_2	1.66-1.55	m
15	26.09	CH_2	1.80-1.73, 0.99-0.97	m
27	25.91	CH_3	1.13	s
11	23.56	CH_2	1.92-1.82	m
16	18.37	CH_2	1.44-1.39, 1.66-1.55	m
24	16.93	CH_3	0.86	s
26	16.81	CH_3	0.95	s
30	15.59	CH_3	0.79	s
28	15.51	CH_3	0.94	s

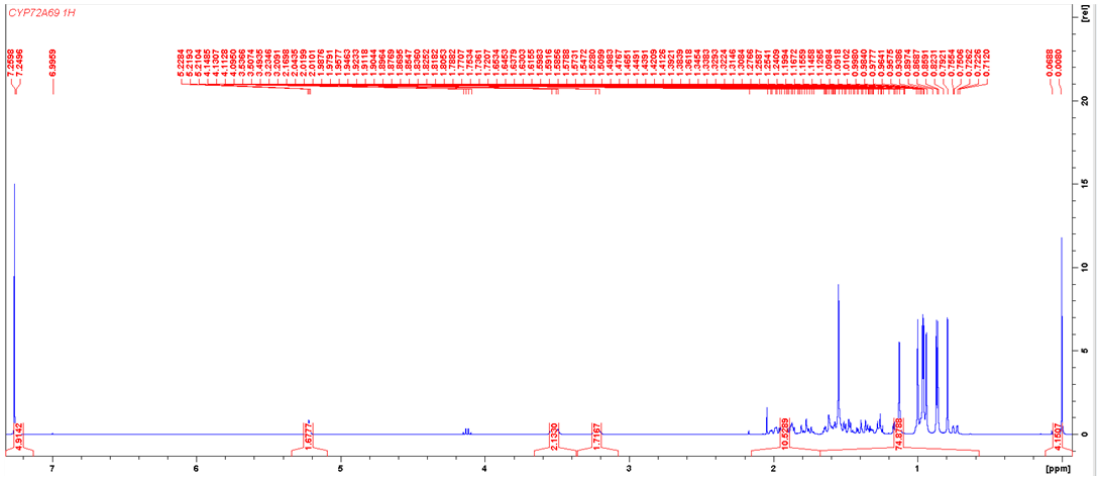


Figure A-12: ^{1}H NMR spectrum for 21β-hydroxy-β-amyrin (400 MHz)

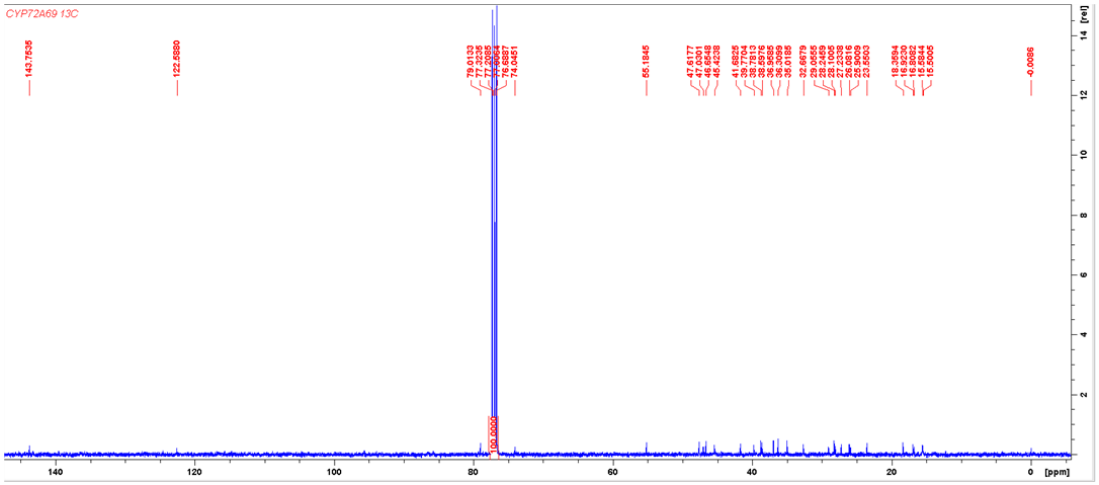


Figure A-13: ^{13}C NMR spectrum for 21β-hydroxy-β-amyrin (400 MHz)

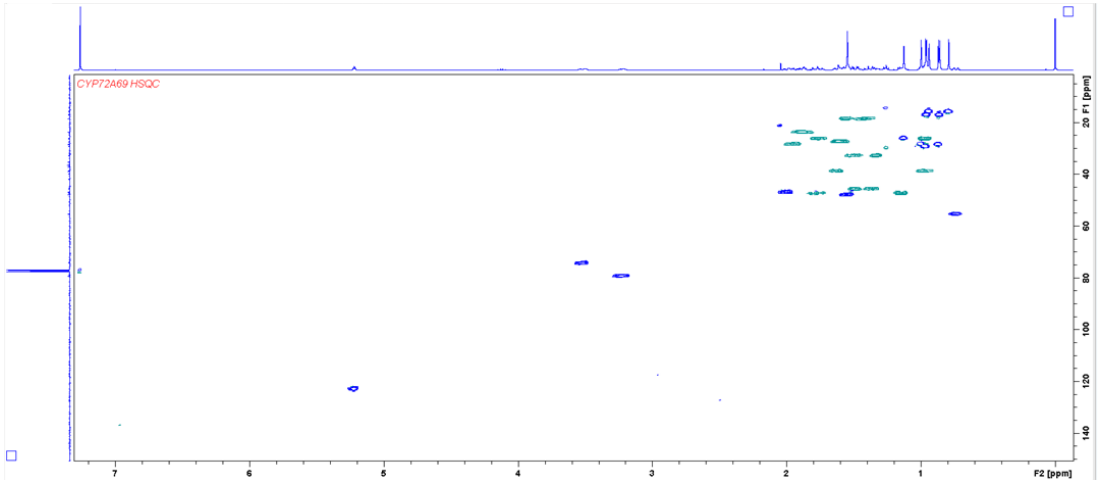


Figure A-14: HSQC NMR spectrum for 21β-hydroxy-β-amyrin (400 MHz)

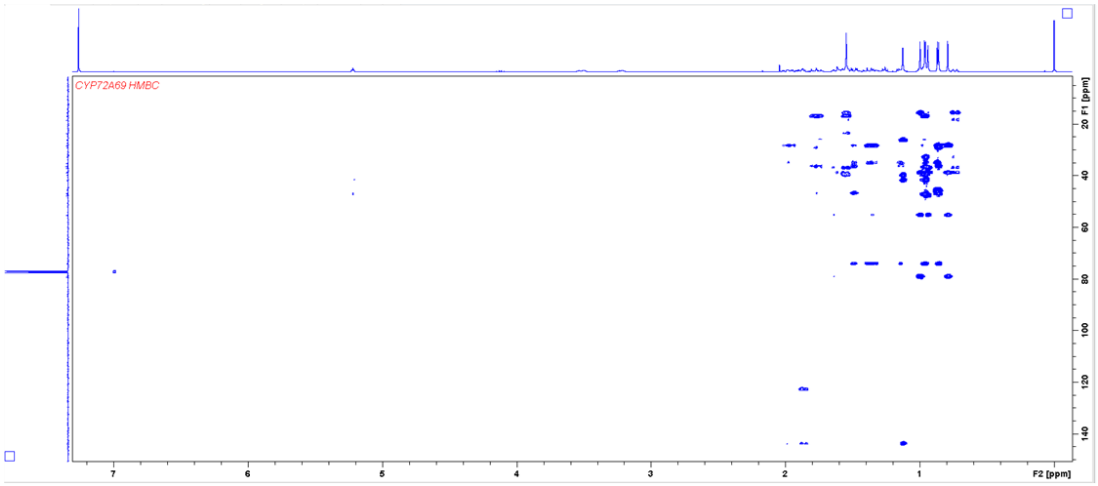


Figure A-15: HMBC NMR spectrum for 21β-hydroxy-β-amyrin (400 MHz)

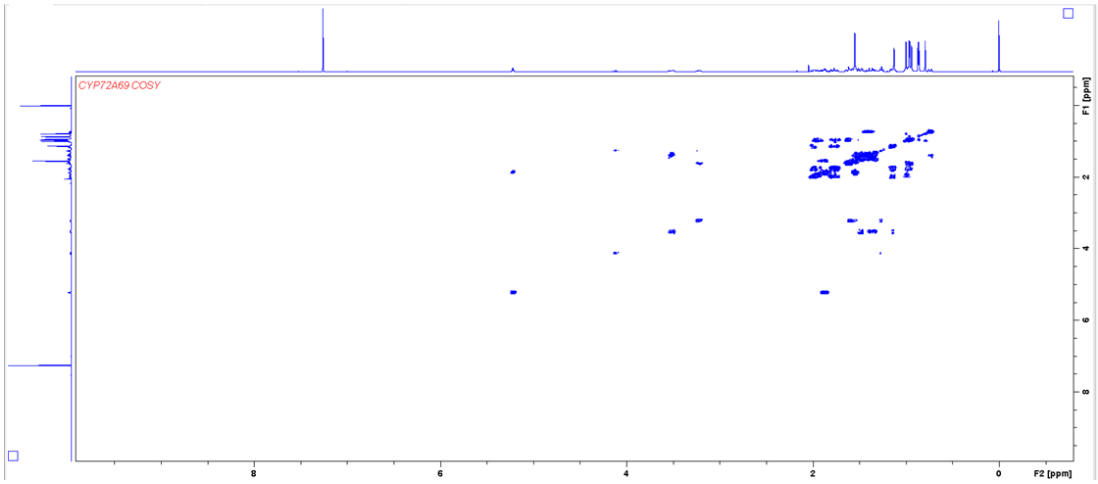
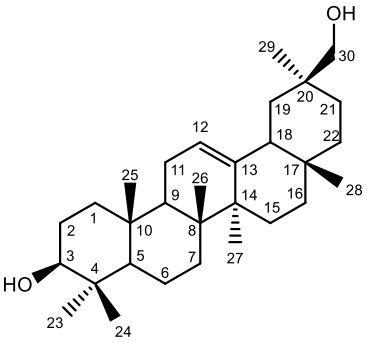


Figure A-16: COSY NMR spectrum for 21β-hydroxy-β-amyrin (400 MHz)

Carbon numbering scheme



Selected 2D NMR

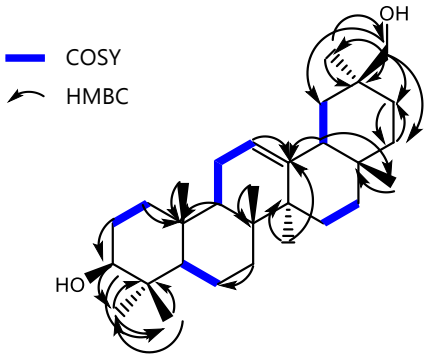


Table A-4: NMR assignments for 30-hydroxy-β-amyrin (600 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
13	144.58	C	N/A	N/A
12	122.30	CH	5.19	t, J = 3.5 Hz
3	79.02	CH	3.22	dd, J = 11.3, 4.3 Hz
30	66.76	CH_2	3.52	dd, J = 44.3, 10.8 Hz
5	55.18	CH	0.74	dd, J = 11.8, 1.2 Hz
9	47.63	CH	1.67-1.53	m
18	46.74	CH	1.94-1.81	m
19	41.95	CH_2	1.67-1.53, 1.45-1.21	m
14	41.74	C	N/A	N/A
8	39.79	C	N/A	N/A
4	38.78	C	N/A	N/A
1	38.59	CH_2	1.67-1.53, 0.97	m
10	36.95	C	N/A	N/A
22	36.50	CH_2	1.45-1.21	m
20	35.53	C	N/A	N/A
7	32.65	CH_2	1.53-1.48, 1.45-1.21	m
17	32.44	C	N/A	N/A
21	29.64	CH_2	1.45-1.21	m
28	28.23	CH_3	0.83	s
23	28.10	CH_3	1.00	s
29	27.48	CH_3	0.90	s
16	27.24	CH_2	0.88-0.85, 2.04-1.98	m
2	27.23	CH_2	1.67-1.53	m
15	26.09	CH_2	1.77, 1.00-0.98	td, J = 13.9, 4.6 Hz
27	25.99	CH_3	1.15	s
11	23.52	CH_2	1.94-1.81	m
6	18.37	CH_2	1.67-1.53, 1.45-1.21	m
26	16.81	CH_3	0.96	s
24	15.59	CH_3	0.79	s
25	15.51	CH_3	0.94	s

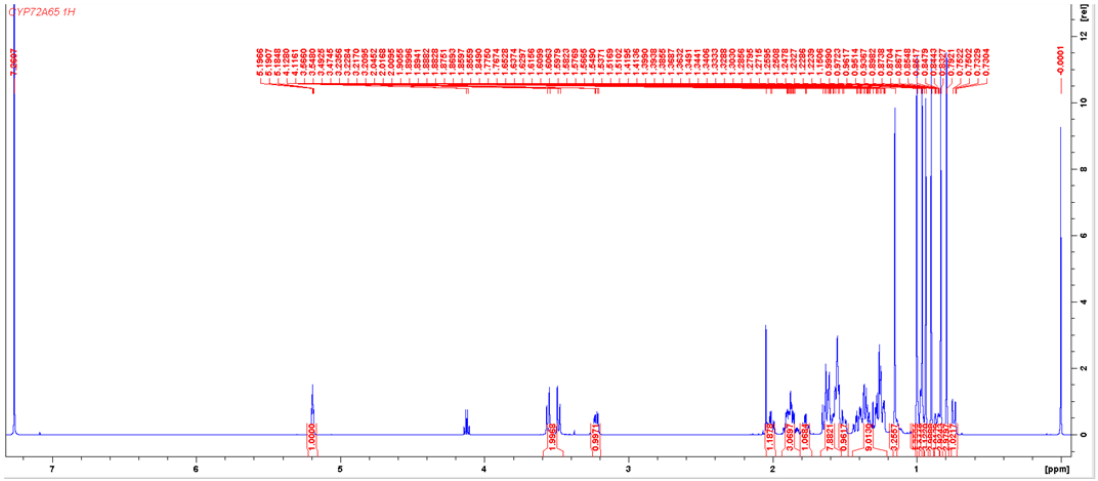


Figure A-17: ^{1}H NMR spectrum for 30-hydroxy-β-amyrin (600MHz)

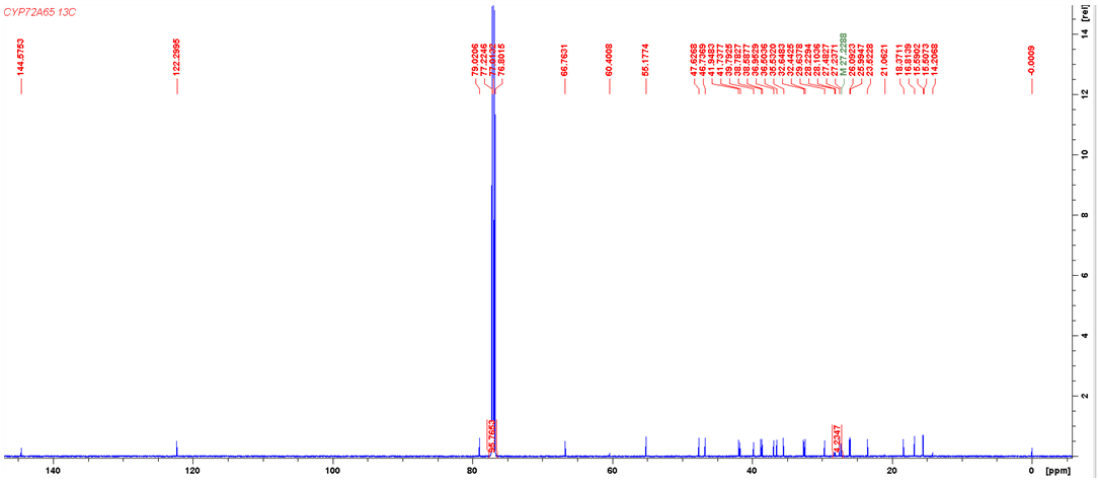


Figure A-18: ^{13}C NMR spectrum for 30-hydroxy-β-amyrin (600MHz)

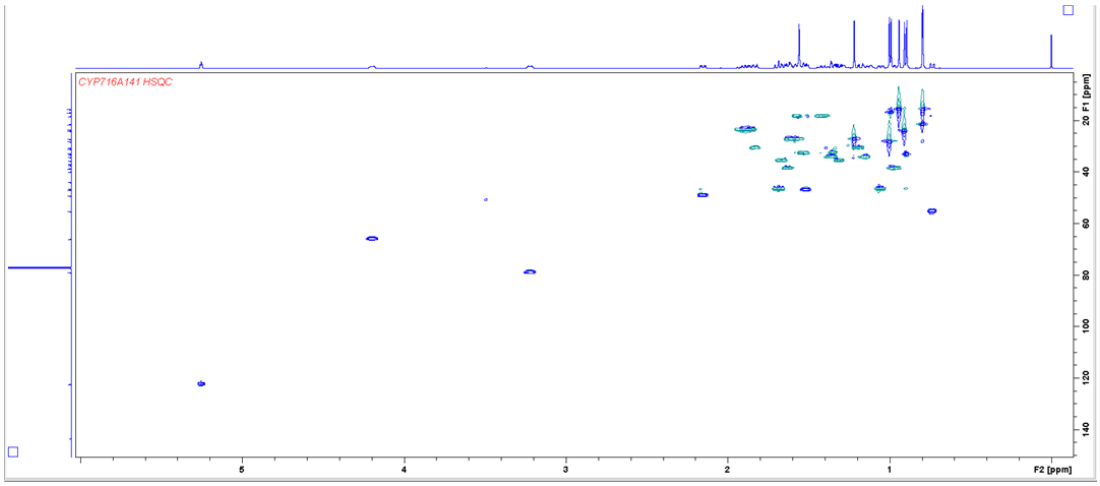


Figure A-19: HSQC NMR spectrum for 30-hydroxy-β-amyrin (600MHz)

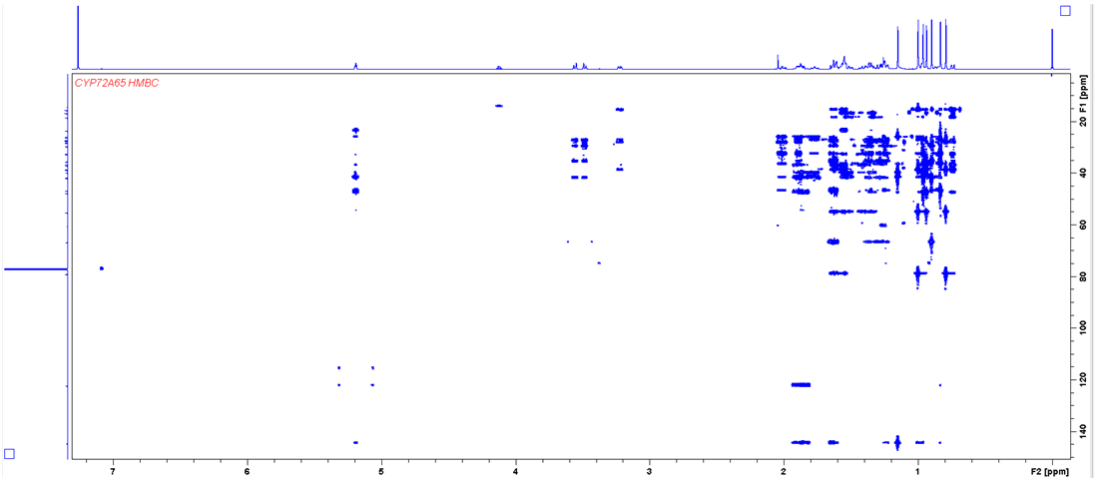


Figure A-20: HMBC NMR spectrum for 30-hydroxy-β-amyrin (600MHz)

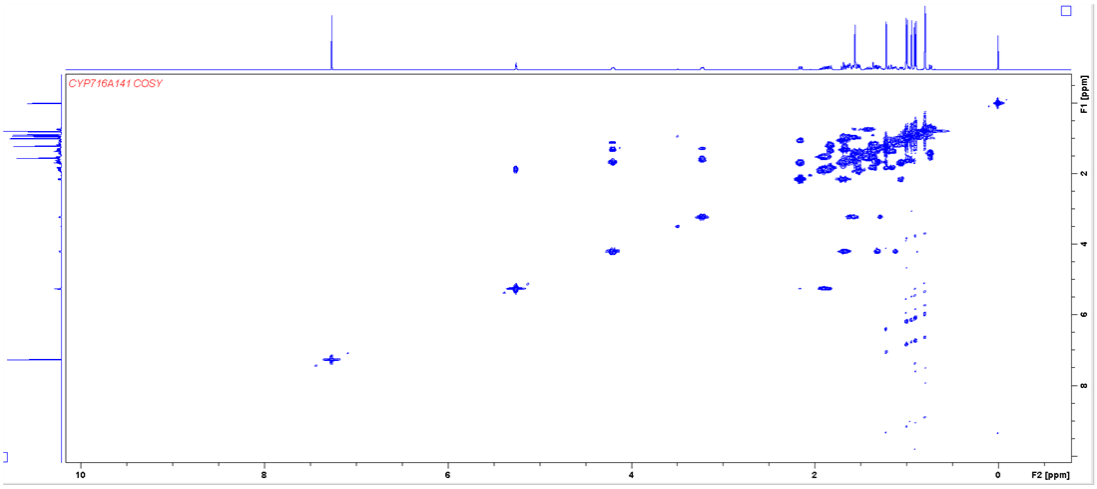
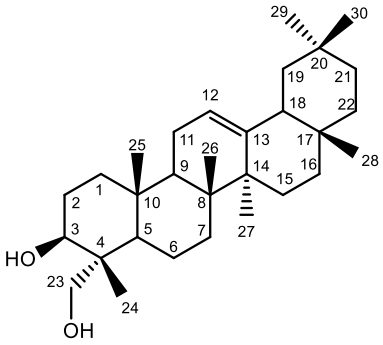


Figure A-21: COSY NMR spectrum for 30-hydroxy-β-amyrin (600MHz)

Carbon numbering scheme



Selected 2D NMR

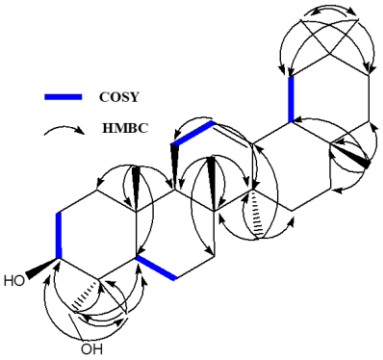


Table A-5: NMR assignments for 23-hydroxy-β-amyrin (600 MHz) (NMR performed by Michael Stephenson)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
13	145.17	C	N/A	N/A
12	121.64	CH	5.18	t, *J*= 3.5
3	77.01	CH	3.64	t, *J*= 7.5
23	72.29	CH_2	3.74, 3.45	d, *J*= 10.3, d *J*= 10.5
5	49.82	CH	0.87	m
9	47.62	CH	1.58	m
18	47.21	CH	1.95	m
19	46.82	CH_2	1.66, 1.01	m
4	41.83	C	N/A	N/A
14	41.74	C	N/A	N/A
8	39.79	C	N/A	N/A
1	38.25	CH_2	1.64, 0.98	m
22	37.13	CH_2	1.42, 1.22	m
10	36.84	C	N/A	N/A
21	34.73	CH_2	1.32, 1.10	m
29	33.34	CH_3	0.87	s
17	32.49	C	N/A	N/A
7	32.44	CH_2	1.50, 1.32	m
20	31.09	C	N/A	N/A
28	28.40	CH_3	0.83	s
2	26.92	CH_2	1.99, 0.81	m
16	26.83	CH_2	1.62	m
15	26.14	CH_2	1.76, 0.96	m
27	26.01	CH_3	1.13	s
30	23.69	CH_3	0.87	s
11	23.51	CH_2	1.87	m
6	18.55	CH_2	1.46, 1.37	m
26	16.82	CH_3	0.97	s
25	15.87	CH_3	0.99	s
24	11.36	CH_3	0.91	s

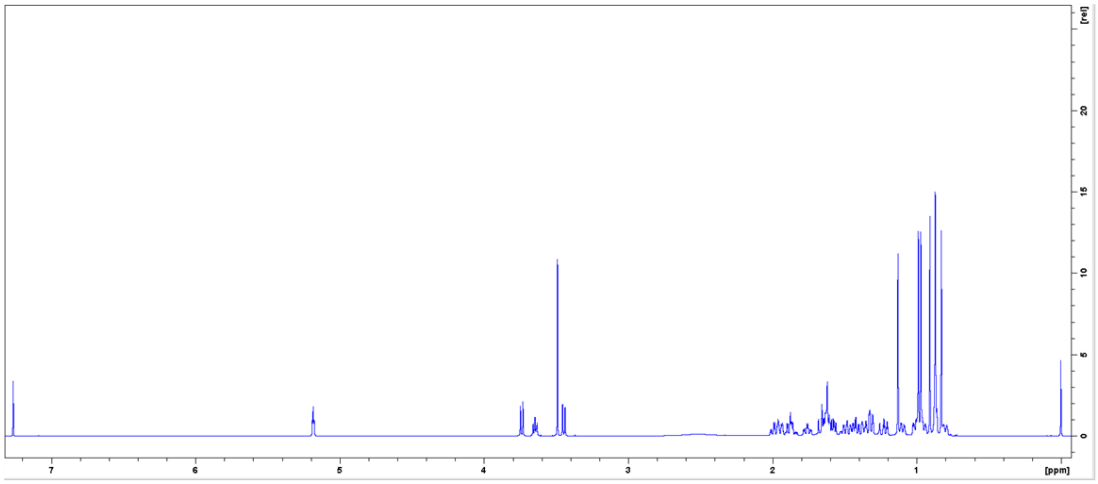


Figure A-22: ^{1}H NMR spectrum for 23-hydroxy-β-amyrin (600 MHz)

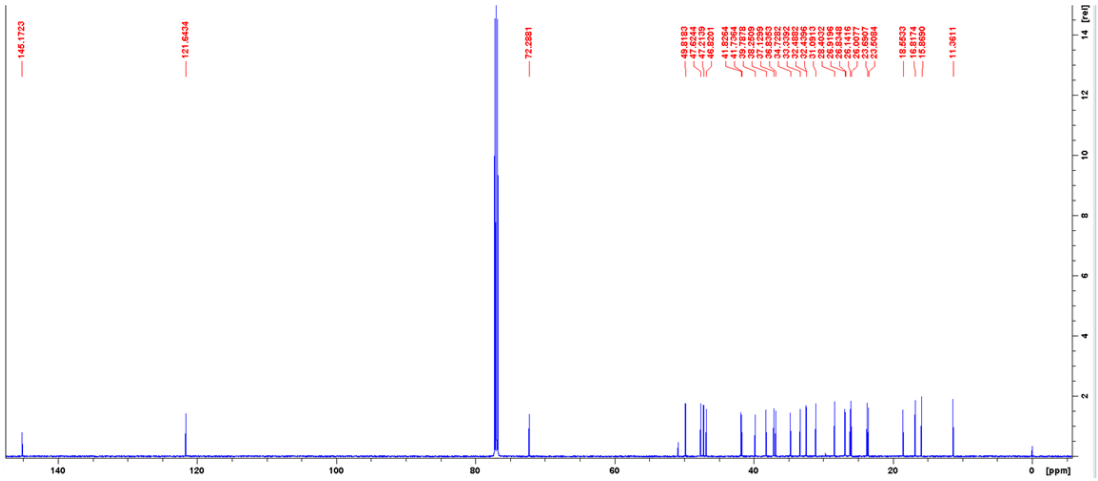


Figure A-23: ^{13}C NMR spectrum for 23-hydroxy-β-amyrin (600 MHz)

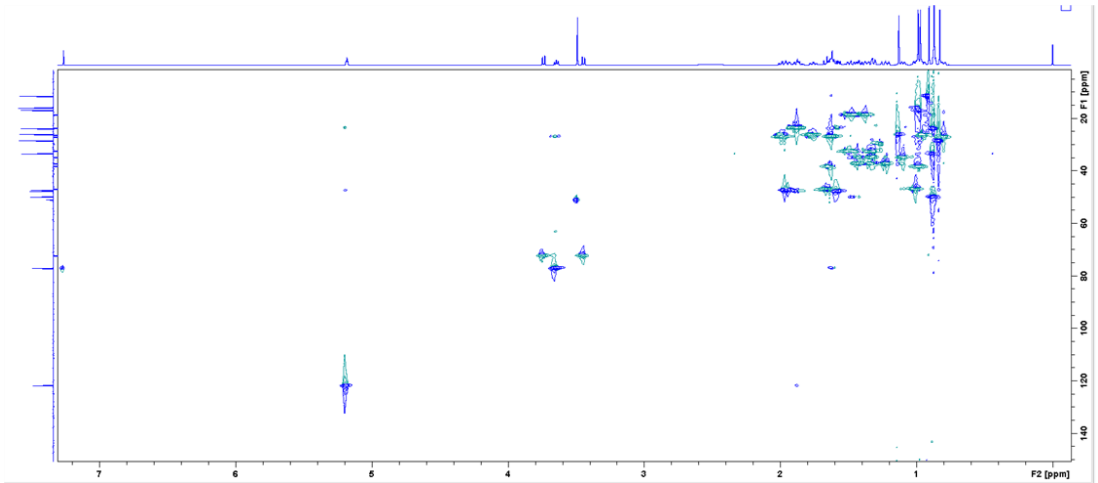


Figure A-24: HSQC NMR spectrum for 23-hydroxy-β-amyrin (600 MHz)

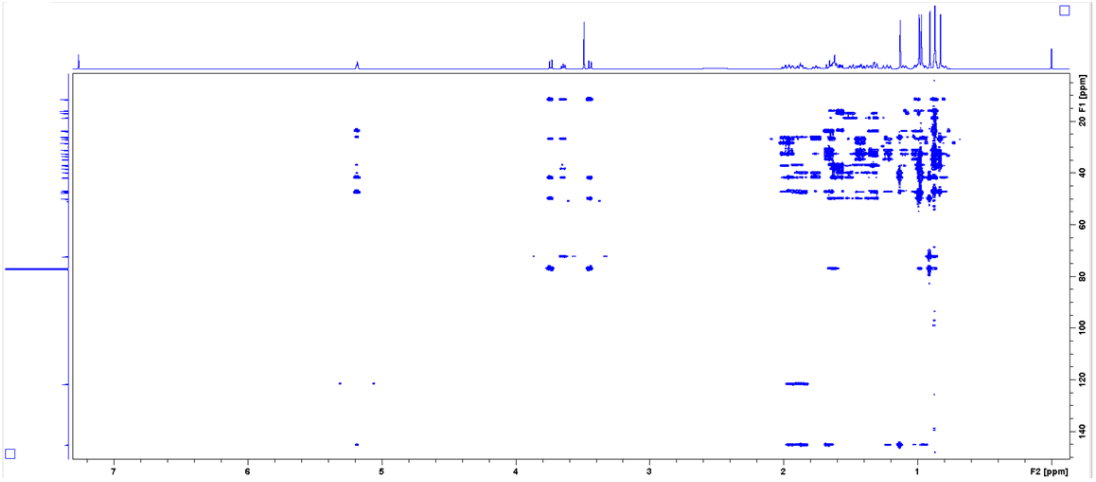


Figure A-25: HMBC NMR spectrum for 23-hydroxy-β-amyrin (600 MHz)

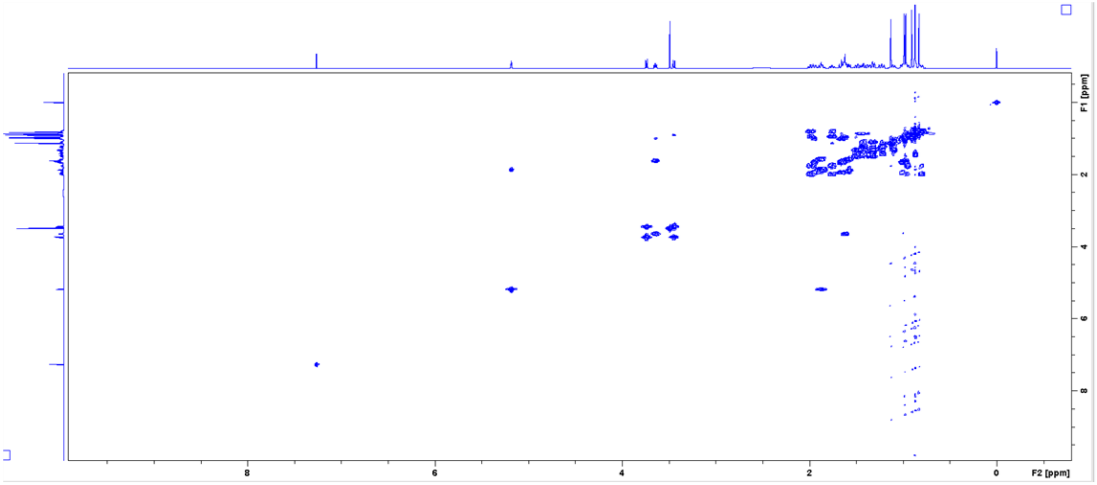
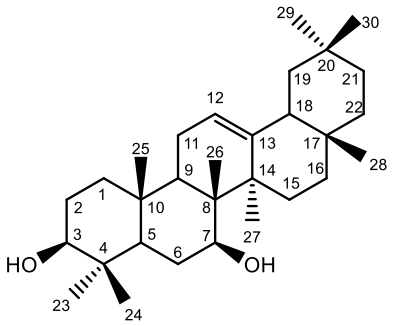


Figure A-26: COSY NMR spectrum for 23-hydroxy-β-amyrin (600 MHz)

Carbon numbering scheme



Selected 2D NMR

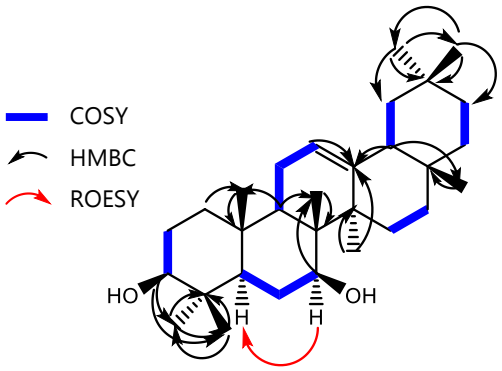


Table A-6: NMR assignments for 7β-hydroxy-β-amyrin (600 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	38.38	CH_2	1.67-1.60, 0.96-0.88	m
2	27.22	CH_2	2.07-1.97, 1.66-1.52	m
3	78.76	CH	3.22	d, J = 11.2 Hz
4	38.55	C	N/A	N/A
5	52.23	CH	0.82	d, J = 14 Hz
6	29.28	CH_2	1.74, 1.54-1.45	m
7	73.29	CH	3.95	m
8	45.38	C	N/A	N/A
9	47.66	CH	1.48	m
10	37.06	C	N/A	N/A
11	23.65	CH_2	2.01-1.84	m
12	122.13	CH	5.23	t, J = 3.6 Hz
13	144.70	C	N/A	N/A
14	43.15	C	N/A	N/A
15	29.94	CH_2	2.06-1.99, 1.30-1.25	m
16	27.10	CH_2	2.07-1.97, 0.88-0.82	m
17	32.32	C	N/A	N/A
18	47.78	CH	1.97	m
19	46.86	CH_2	1.72-1.64, 1.06-0.99	m
20	31.12	C	N/A	N/A
21	34.67	CH_2	1.33, 1.10	td, J = 13.2, 3.9 Hz
22	36.94	CH_2	1.42, 1.25-1.20	td, J = 13.7, 3.9 Hz
23	28.04	CH_3	1.01	s
24	15.56	CH_3	0.80	s
25	15.33	CH_3	0.94	s
26	9.85	CH_3	0.99	s
27	26.26	CH_3	1.23	s
28	28.31	CH_3	0.84	s
29	33.33	CH_3	0.88	s
30	23.66	CH_3	0.87	s

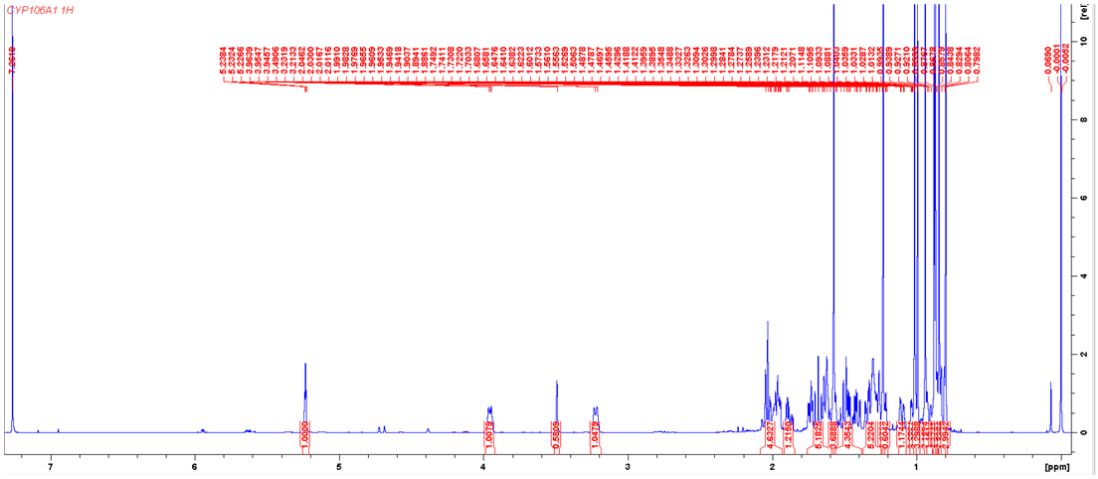


Figure A-27: ^{1}H NMR spectrum for 7β-hydroxy-β-amyrin (600 MHz)

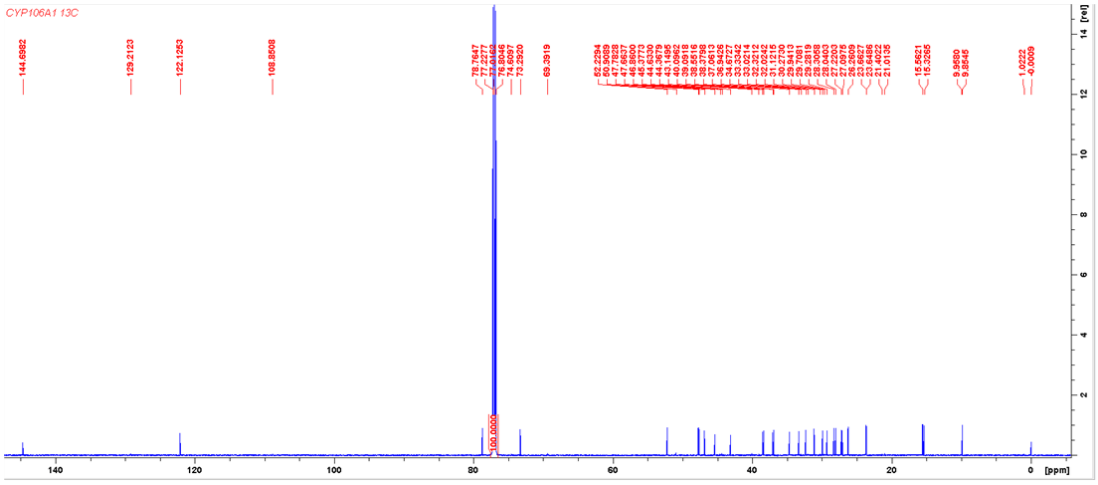


Figure A-28: ^{13}C NMR spectrum for 7β-hydroxy-β-amyrin (600 MHz)

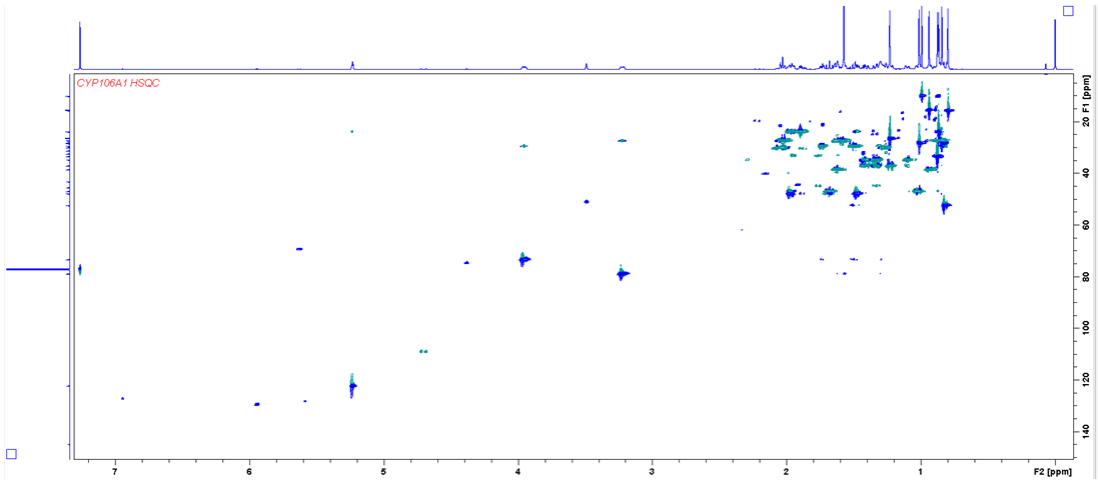


Figure A-29: HSQC NMR spectrum for 7β-hydroxy-β-amyrin (600 MHz)

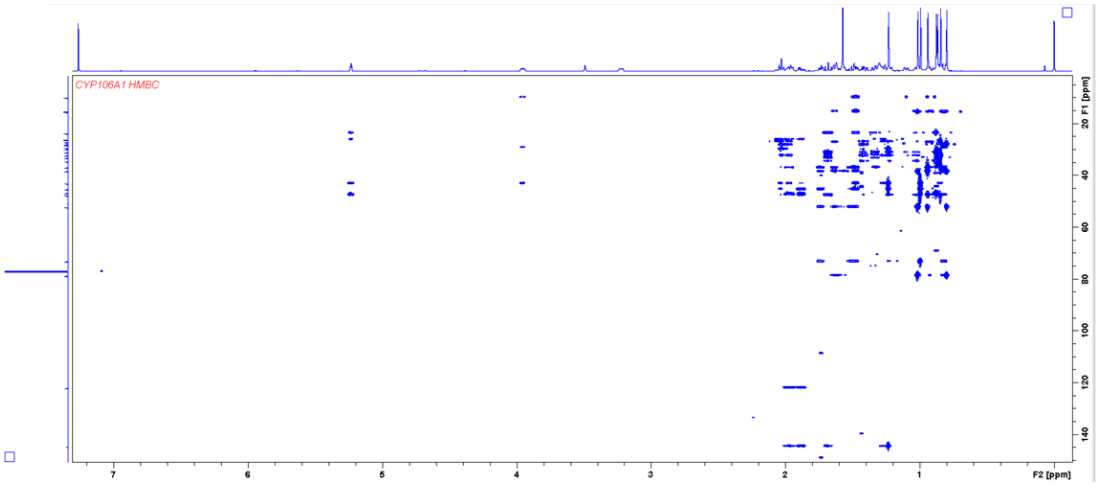


Figure A-30: HMBC NMR spectrum for 7β-hydroxy-β-amyrin (600 MHz)

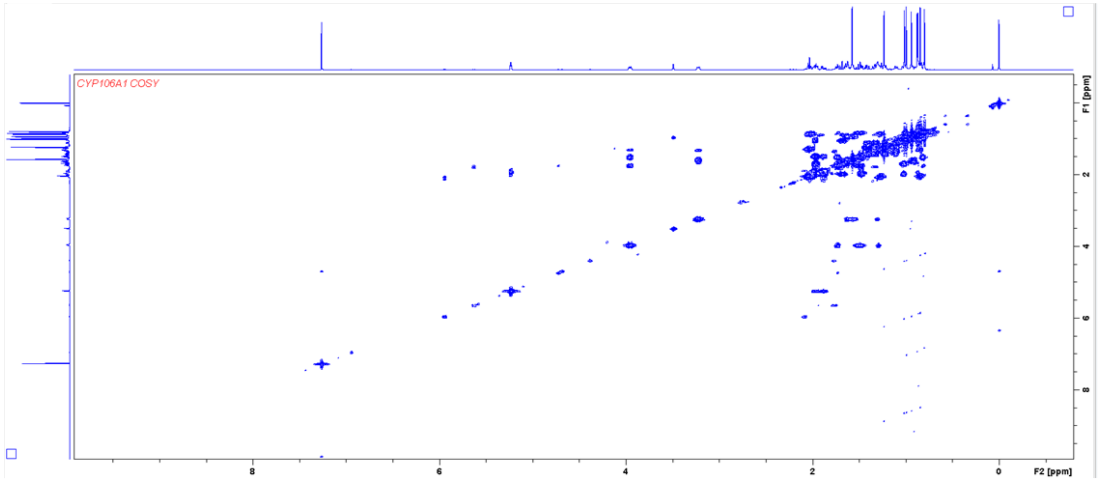


Figure A-31: COSY NMR spectrum for 7β-hydroxy-β-amyrin (600 MHz)

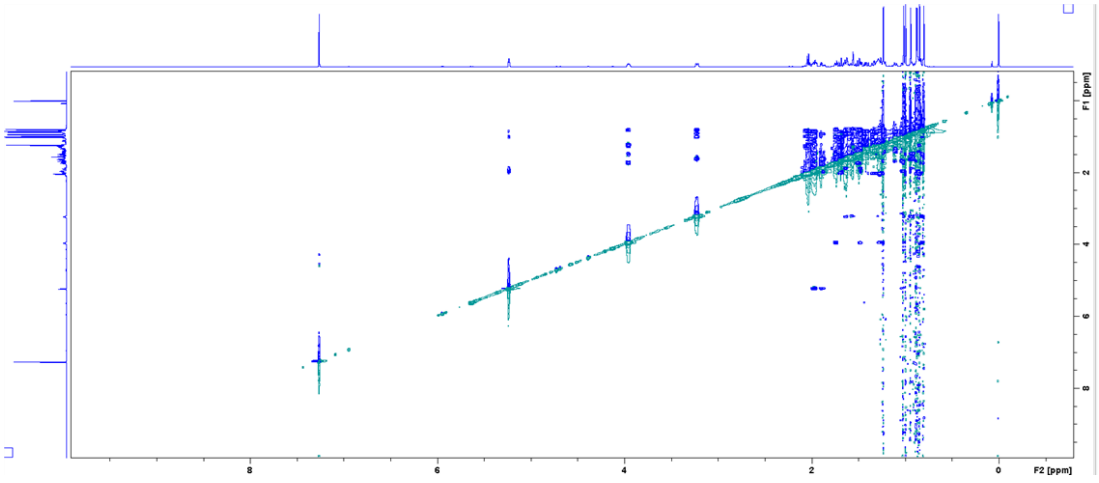
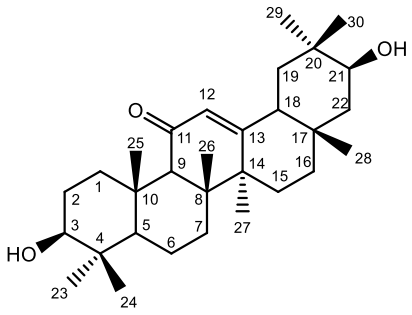


Figure A-32: ROESY NMR spectrum for 7β-hydroxy-β-amyrin (600 MHz)

Carbon numbering scheme



Selected 2D NMR

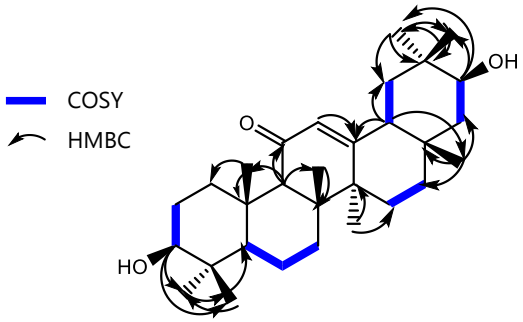


Table A-7: NMR assignments for 21β-hydroxy-11-oxo-β-amyrin (600 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	39.14	CH_2	2.79, 0.97	dt, J = 13.5, 3.5 Hz
2	27.31	CH_2	1.68-1.59	m
3	78.78	CH	3.23	dd, J = 11, 4.8 Hz
4	31.93	C	N/A	N/A
5	54.95	CH	0.70	d, J = 11.5 Hz
6	17.49	CH_2	1.63-1.57, 1.46-1.42	m
7	32.78	CH_2	1.68-1.61, 1.43-1.38	m
8	45.36	C	N/A	N/A
9	61.80	CH	2.33	s
10	37.11	C	N/A	N/A
11	200.08	C	N/A	N/A
12	128.62	CH	5.60	s
13	168.91	C	N/A	N/A
14	43.35	C	N/A	N/A
15	26.39	CH_2	1.81, 1.19-1.14	td, J = 13.7, 4.4 Hz
16	27.71	CH_2	2.02, 1.16-1.11	td, J = 13.7, 4.4 Hz
17	34.85	C	N/A	N/A
18	46.92	CH	2.17	dd, J = 14, 4 Hz
19	45.44	CH_2	1.77, 1.20	dd J = 14, 4.4 Hz
20	36.37	C	N/A	N/A
21	73.61	CH	3.54	dd, J = 12, 4.8 Hz
22	44.74	CH_2	1.58-1.53, 1.44-1.38	m
23	28.10	CH_3	1.00	s
24	15.58	CH_3	0.81	s
25	16.37	CH_3	1.13	s
26	18.73	CH_3	1.12	s
27	23.35	CH_3	1.351	s
28	28.66	CH_3	0.90	s
29	28.83	CH_3	0.99	s
30	16.82	CH_3	0.88	s

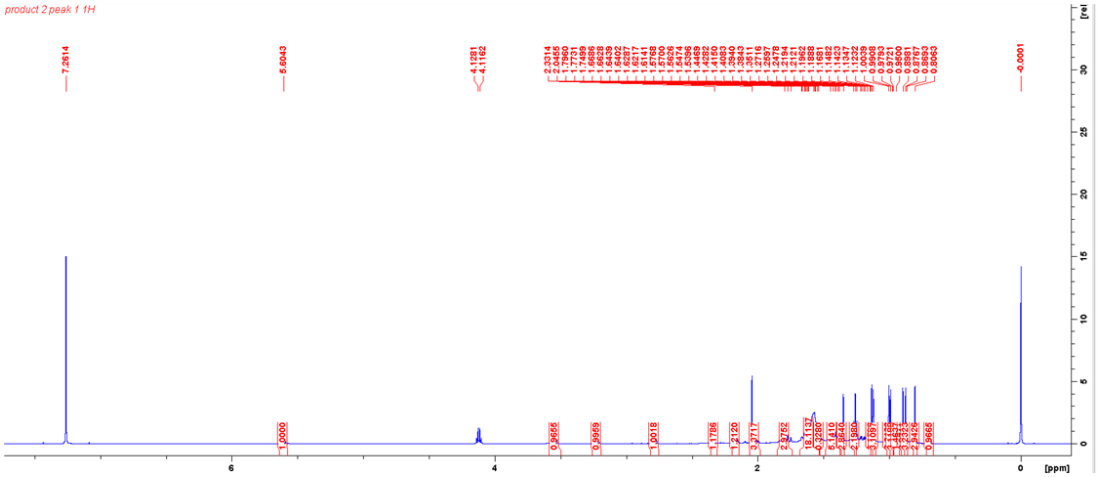


Figure A-33: ^{1}H NMR spectrum for 21β-hydroxy-11-oxo-β-amyrin (600 MHz)

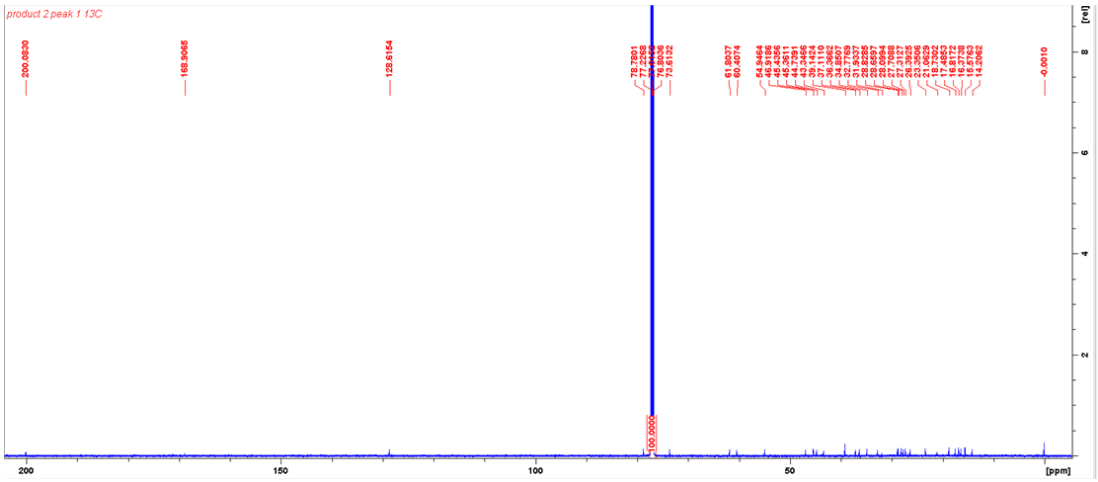


Figure A-34: ^{13}C NMR spectrum for 21β-hydroxy-11-oxo-β-amyrin (600 MHz)

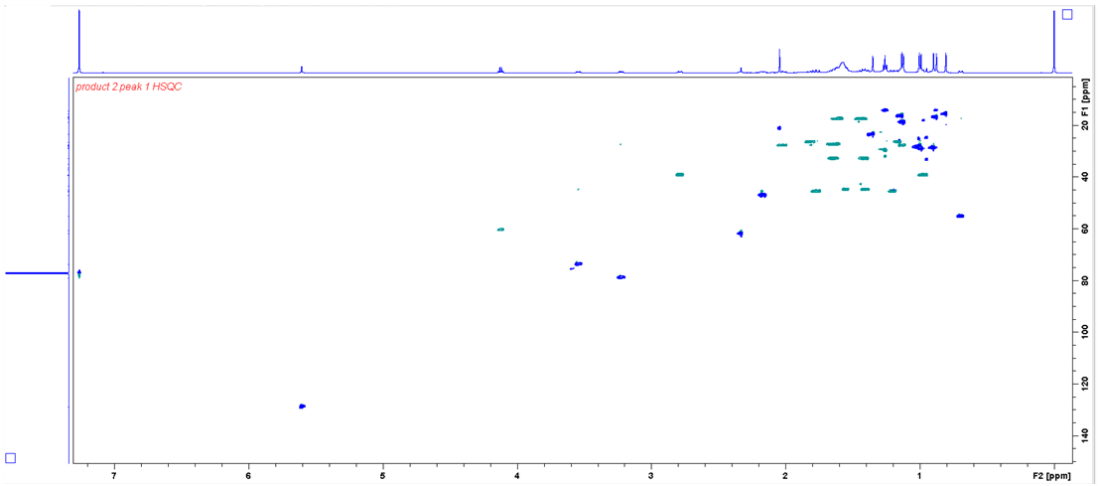


Figure A-35: HSQC NMR spectrum for 21β-hydroxy-11-oxo-β-amyrin (600 MHz)

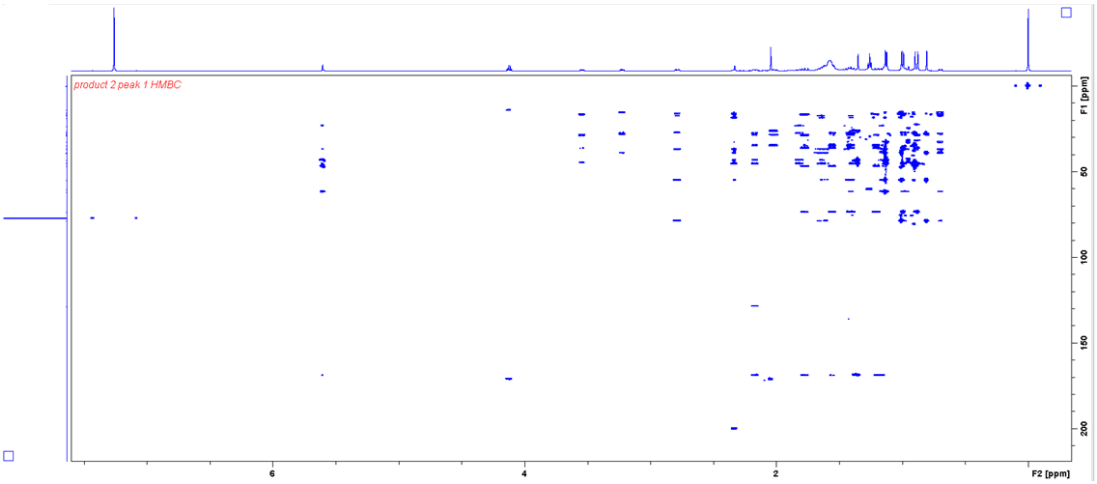


Figure A-36: HMBC NMR spectrum for 21β-hydroxy-11-oxo-β-amyrin (600 MHz)

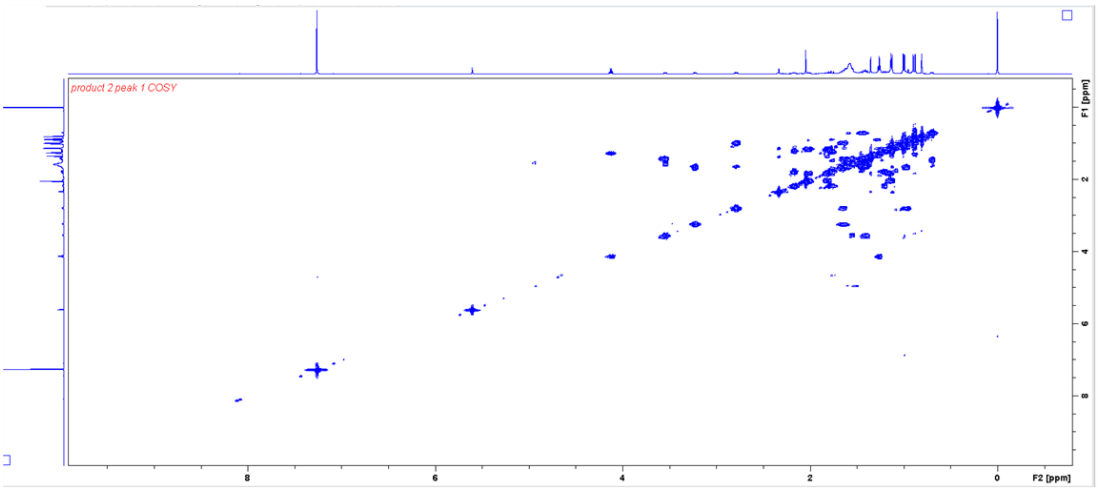
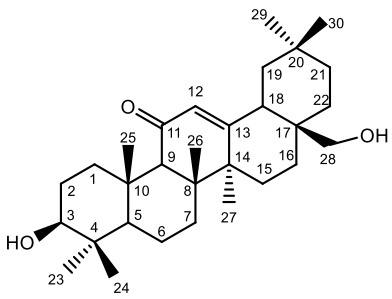


Figure A-37: COSY NMR spectrum for 21β-hydroxy-11-oxo-β-amyrin (600 MHz)

Carbon numbering scheme



Selected 2D NMR

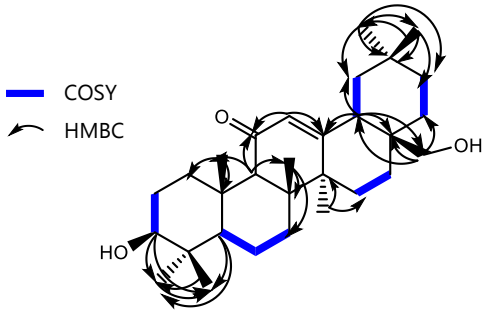


Table A-8: NMR assignments for 11-oxo-erythrodiol (600 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	39.14	CH_2	2.78, 1.02-0.94	dt, J = 13.5, 3.5 Hz
2	27.31	CH_2	1.67-1.59	m
3	78.79	CH	3.23	m
4	29.67	C	N/A	N/A
5	54.94	CH	0.70	d, J = 11 Hz
6	17.50	CH_2	1.64-1.57, 1.46-1.41	m
7	32.70	CH_2	1.69-1.62, 1.44-1.38	m
8	45.41	C	N/A	N/A
9	61.78	CH	2.34	s
10	37.088	C	N/A	N/A
11	200.11	C	N/A	N/A
12	128.28	CH	5.57	s
13	169.41	C	N/A	N/A
14	43.42	C	N/A	N/A
15	25.83	CH_2	1.77, 1.18	td, J = 14.0, 4.5 Hz
16	21.51	CH_2	1.94, 1.38-1.33	td, J = 14.0, 4.5 Hz
17	36.97	C	N/A	N/A
18	42.66	CH	2.15	dd, J = 13.6, 4.5 Hz
19	44.90	CH_2	1.77-1.67, 1.16-1.10	m
20	31.06	C	N/A	N/A
21	33.85	CH_2	1.36-1.29, 1.27-1.22	m
22	30.64	CH_2	1.61-1.53, 1.47-1.42	m
23	28.09	CH_3	1.00	s
24	15.57	CH_3	0.81	s
25	16.40	CH_3	1.13	s
26	18.60	CH_3	1.11	s
27	23.40	CH_3	1.39	s
28	69.67	CH_2	3.48, 3.22	d, J = 10.9 Hz
29	32.94	CH_3	0.92	s
30	23.41	CH_3	0.89	s

NMR is in accordance with the literature [195].

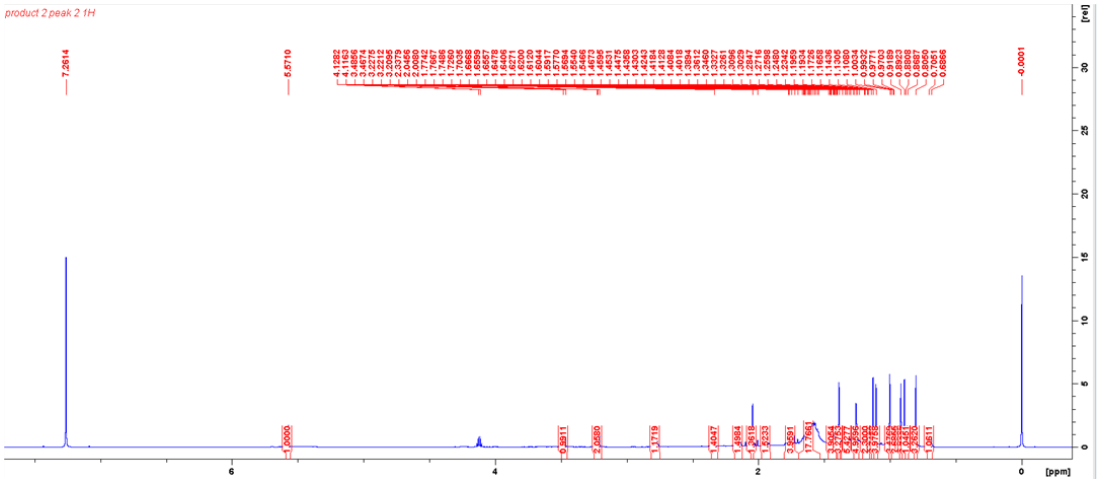


Figure A-38: ^{1}H NMR spectrum for 11-oxo-erythrodiol (600 MHz)

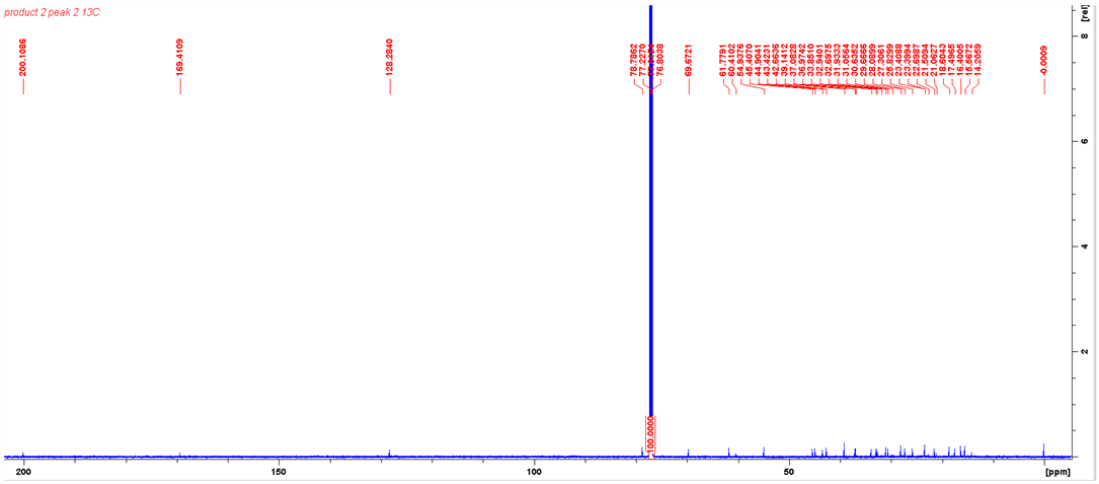


Figure A-39: ^{13}C NMR spectrum for 11-oxo-erythrodiol (600 MHz)

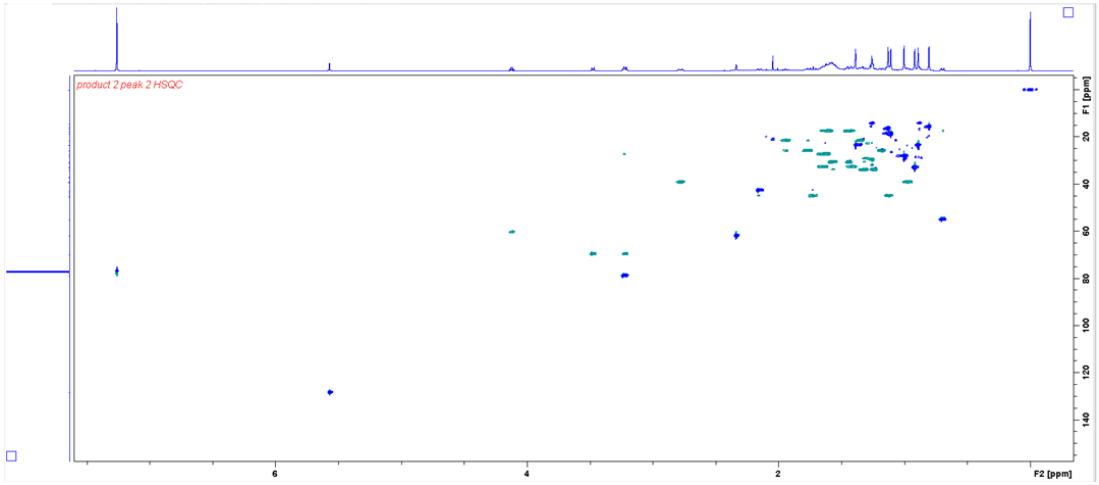


Figure A-40: HSQC NMR spectrum for 11-oxo-erythrodiol (600 MHz)

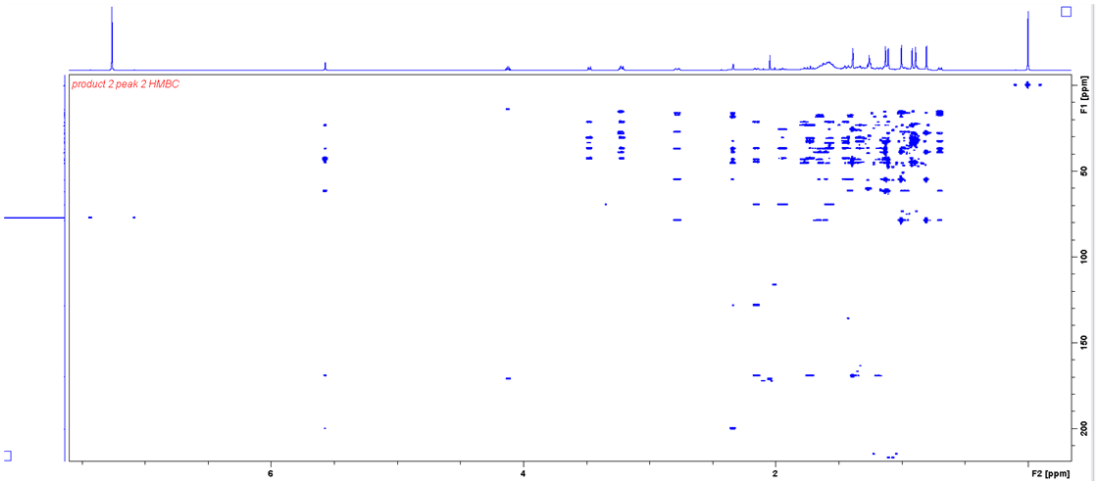


Figure A-41: HMBC NMR spectrum for 11-oxo-erythrodiol (600 MHz)

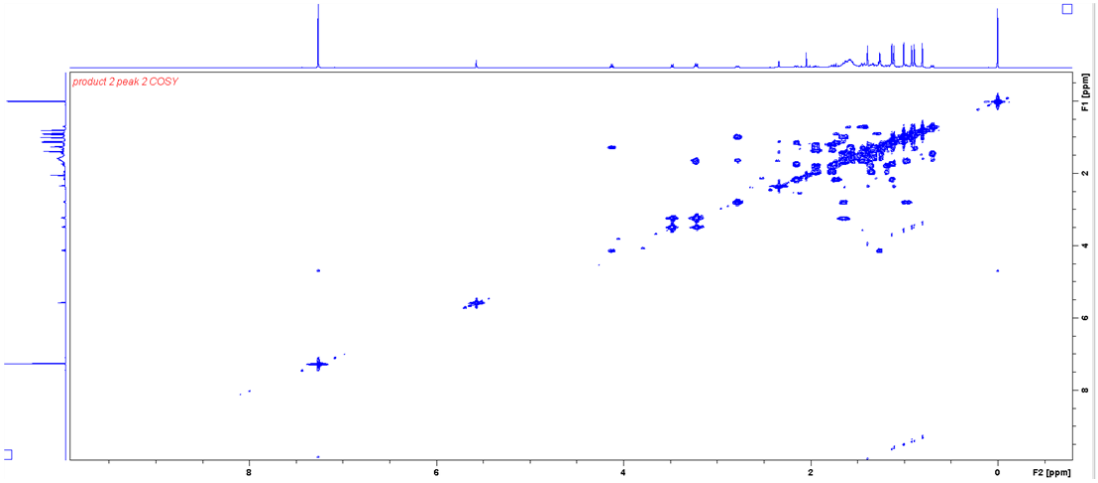


Figure A-42: COSY NMR spectrum for 11-oxo-erythrodiol (600 MHz)

Carbon numbering scheme **Selected 2D NMR**

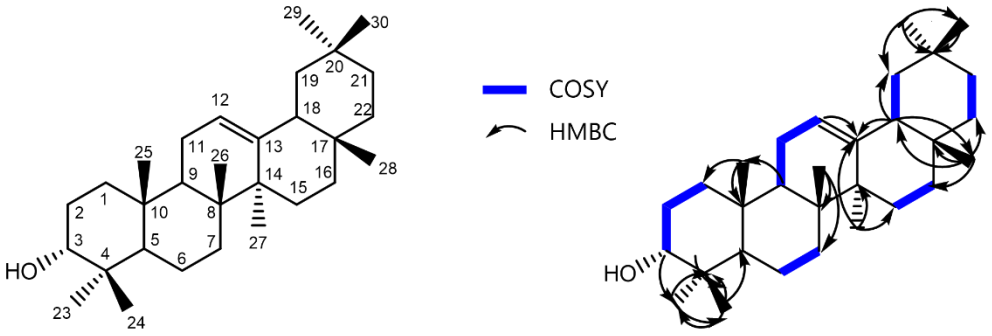


Table A-9: NMR assignments for epi-β-amyrin (600 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	33.02	CH_2	1.37-1.26	m
2	25.23	CH_2	1.98-1.90, 1.58-1.50	m
3	76.15	CH	3.41	t, J = 2.5 Hz
4	37.35	C	N/A	N/A
5	48.92	CH	1.28-1.19	m
6	18.31	CH_2	1.58-1.53, 1.48-1.37	m
7	32.52	CH2	1.59-1.51	m
8	39.95	C	N/A	N/A
9	47.39	CH	1.70-1.63	m
10	37.01	C	N/A	N/A
11	23.43	CH_2	1.91-1.86, 1.70-1.66	m
12	121.77	CH	5.18	t, J = 3.7 Hz
13	145.17	C	N/A	N/A
14	41.77	C	N/A	N/A
15	26.09	CH_2	1.76, 1.01-0.93	td, J = 13.7, 4.7 Hz
16	26.93	CH_2	1.99, 0.82-0.78	td, J = 13.5, 4.5 Hz
17	32.47	C	N/A	N/A
18	47.19	CH	1.97-1.91	m
19	46.81	CH_2	1.70-1.63, 1.05-0.98	m
20	31.08	C	N/A	N/A
21	34.74	CH_2	1.37-1.29, 1.12-1.06	m
22	37.17	CH_2	1.46-1.39, 1.24-1.18	m
23	28.28	CH_3	0.96	s
24	22.34	CH_3	0.85	s
25	15.28	CH_3	0.95	s
26	16.81	CH_3	0.97	s
27	26.14	CH_3	1.15	s
28	28.41	CH_3	0.83	s
29	33.35	CH_3	0.87	s
30	23.70	CH_3	0.87	s

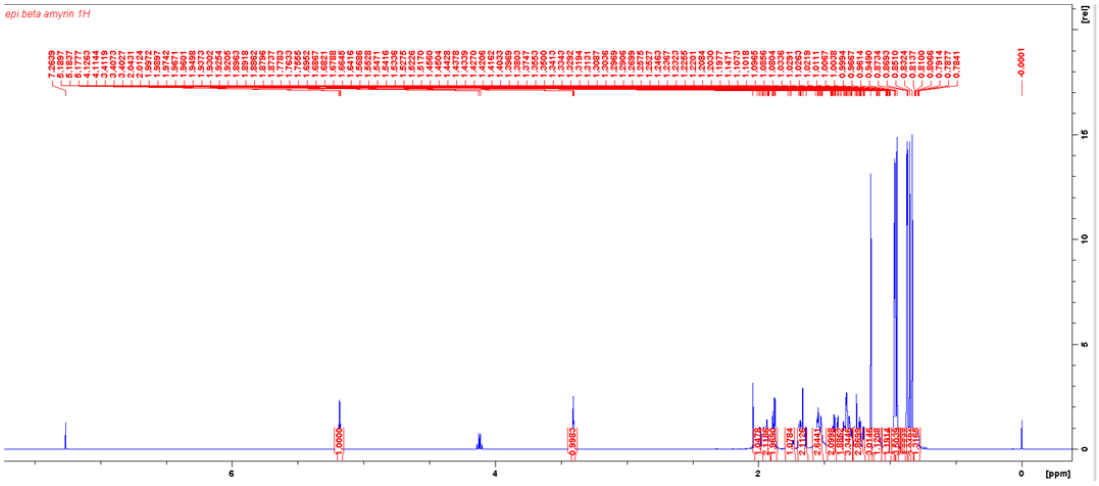


Figure A-43: ^{1}H NMR spectrum for epi-β-amyrin (600 MHz)

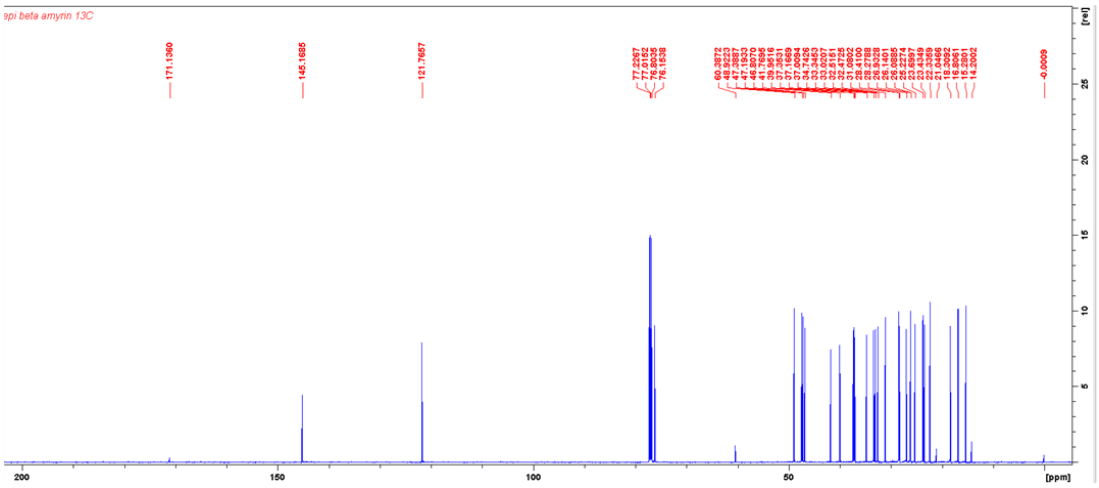


Figure A-44: ^{13}C NMR spectrum for epi-β-amyrin (600 MHz)

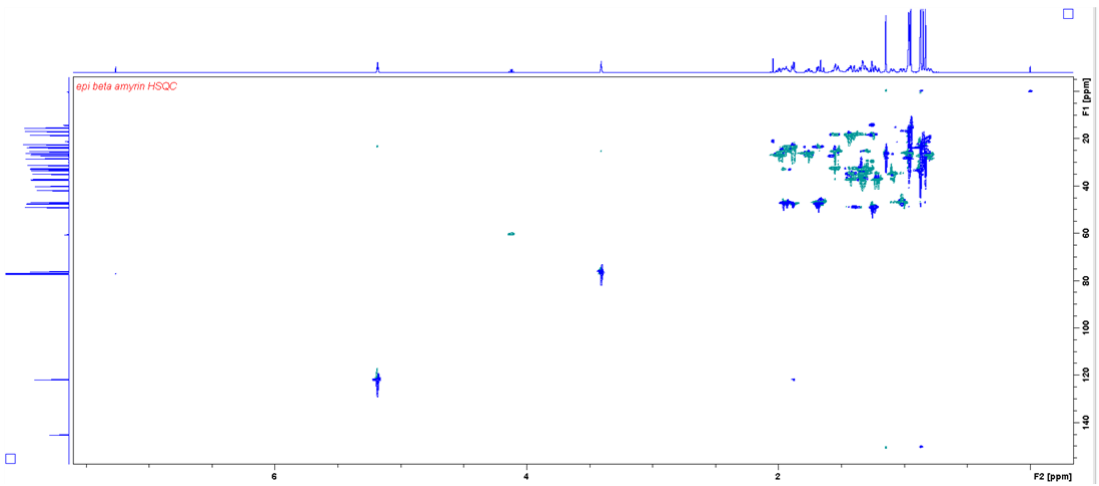


Figure A-45: HSQC NMR spectrum for epi-β-amyrin (600 MHz)

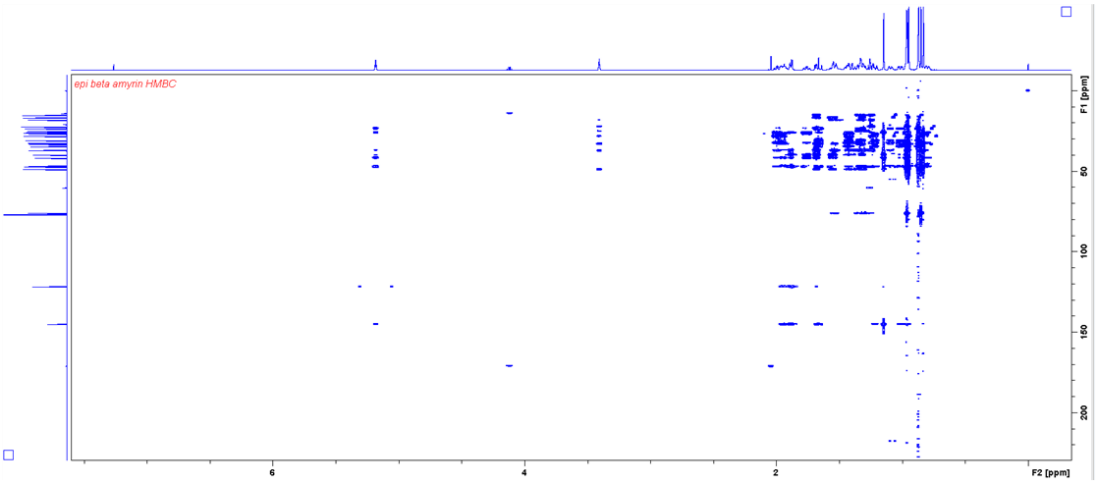


Figure A-46: HMBC NMR spectrum for epi-β-amyrin (600 MHz)

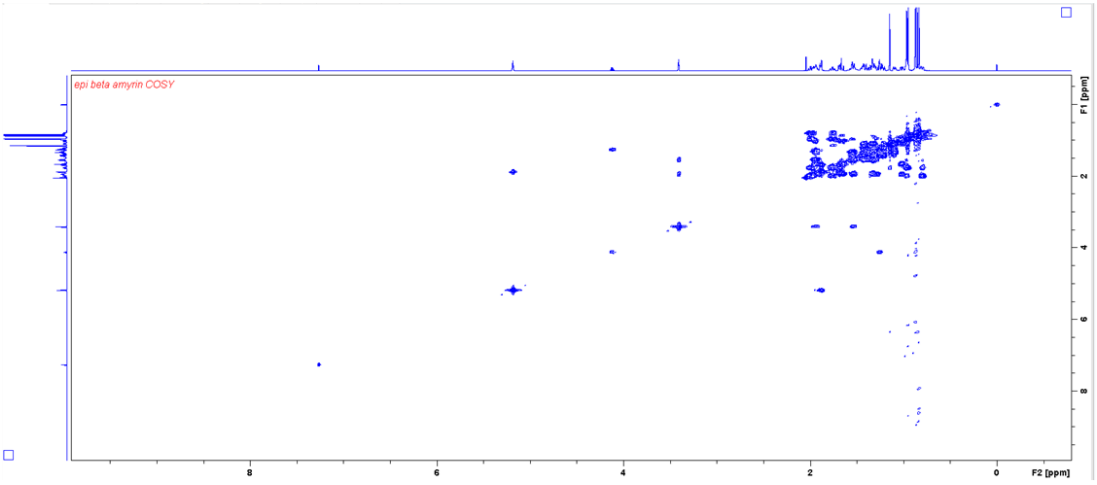
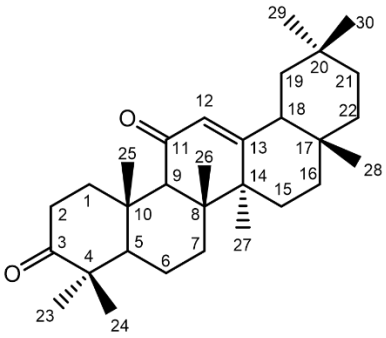


Figure A-47: COSY NMR spectrum for epi-β-amyrin (600 MHz)

Carbon numbering scheme



Selected 2D NMR

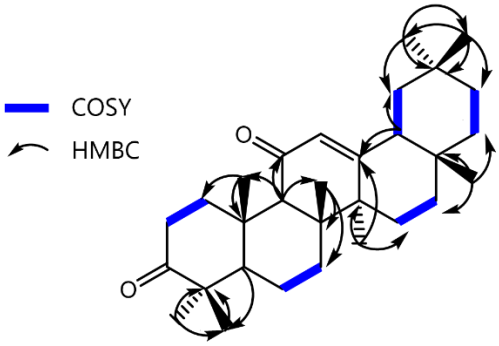


Table A-10: NMR assignments for 11-oxo-β-amyrone (400 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	39.78	CH_2	2.94-2.85, 1.37-1.32	qd, J = 7.1, 4.1 Hz
2	34.24	CH_2	2.61-2.51, 2.34-2.25	m
3	171.11	C	N/A	N/A
4	47.78	C	N/A	N/A
5	55.44	CH	1.27-1.21	m
6	18.82	CH_2	1.54-1.43	m
7	32.13	CH_2	1.65-1.60, 1.43-1.36	m
8	45.26	C	N/A	N/A
9	61.03	CH	2.37	s
10	36.71	C	N/A	m
11	199.53	C	N/A	N/A
12	128.02	CH	5.56	s
13	171.11	C	N/A	N/A
14	43.52	C	N/A	N/A
15	26.52	CH_2	1.83-1.72, 1.17-1.08	td, J = 13.5, 4.2 Hz
16	26.39	CH_2	2.05-1.95, 0.93-0.87	td, J = 13.5, 4.4 Hz
17	32.39	C	N/A	N/A
18	47.66	CH	2.12-2.05	dd, J = 13.9, 3.5 Hz
19	45.26	CH_2	1.60, 1.06-0.98	m
20	31.07	C	N/A	N/A
21	34.44	CH_2	1.30-1.28, 1.14-1.07	m
22	36.50	CH_2	1.43-1.40, 1.27-1.20	m
23	26.42	CH_3	1.04	s
24	21.42	CH_3	1.00	s
25	15.68	CH_3	1.20	s
26	18.56	CH_3	1.11	s
27	23.37	CH_3	1.30	s
28	28.81	CH_3	0.81	s
29	33.05	CH_3	0.84	s
30	23.49	CH_3	0.82	s

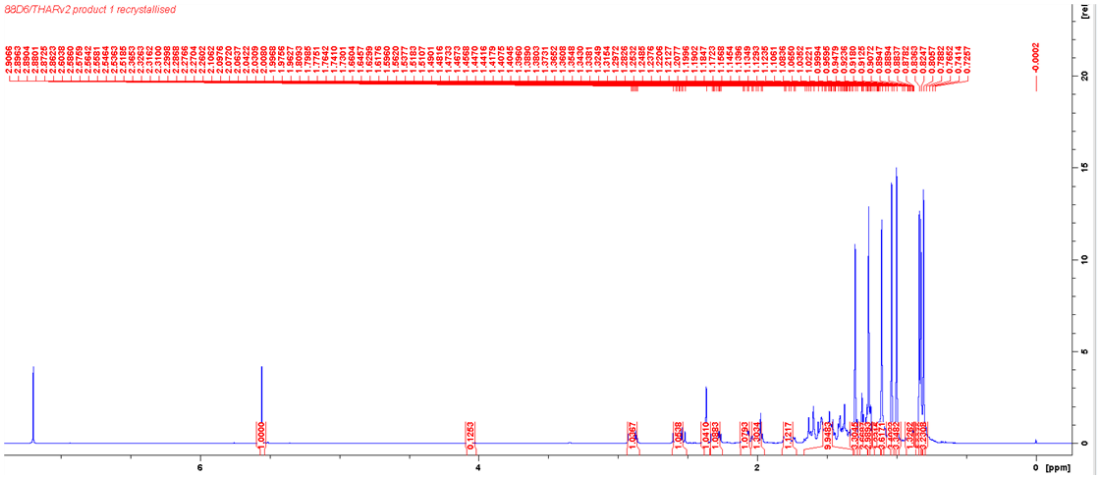


Figure A-48: ^{1}H NMR spectrum for 11-oxo-β-amyrone (400 MHz)

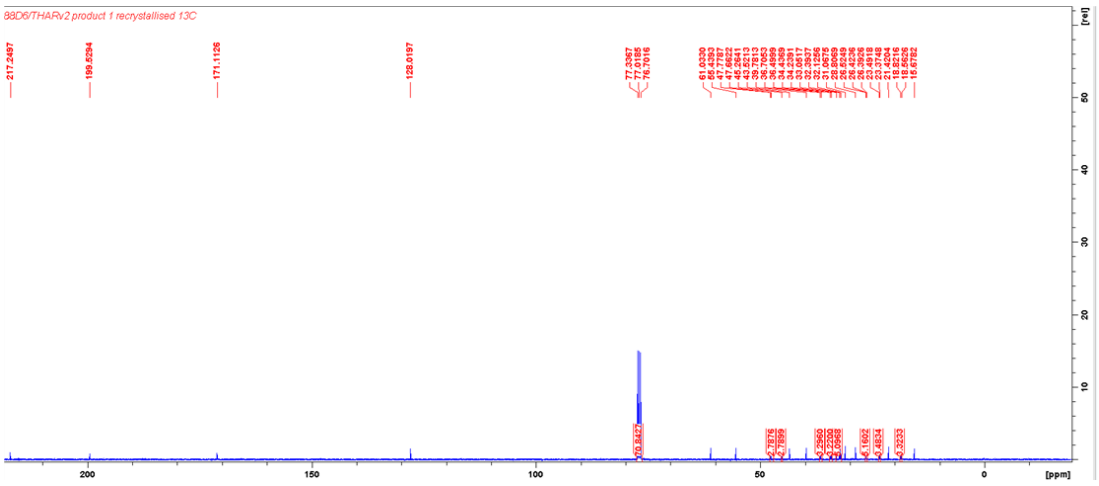


Figure A-49: ^{13}C NMR spectrum for 11-oxo-β-amyrone (400 MHz)

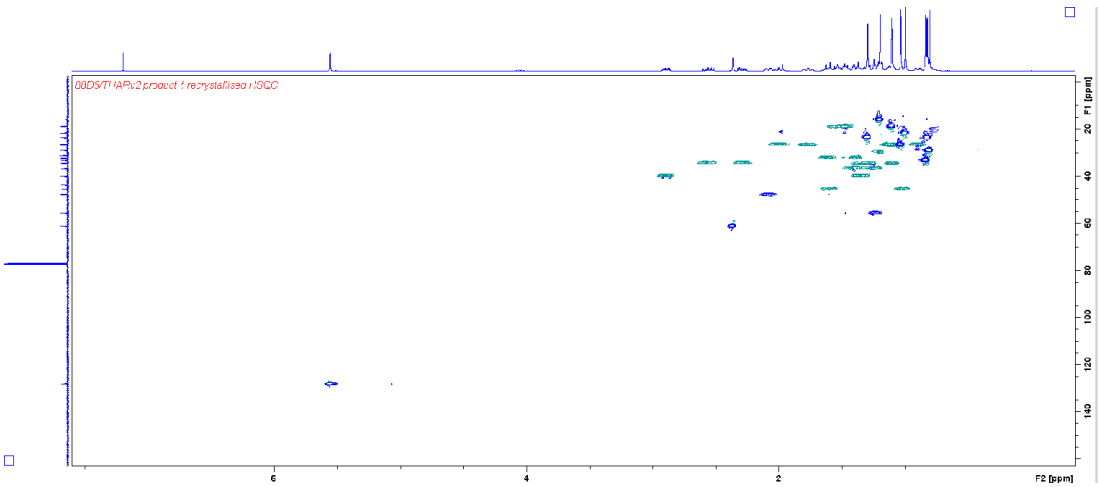


Figure A-50: HSQC NMR spectrum for 11-oxo-β-amyrone (400 MHz)

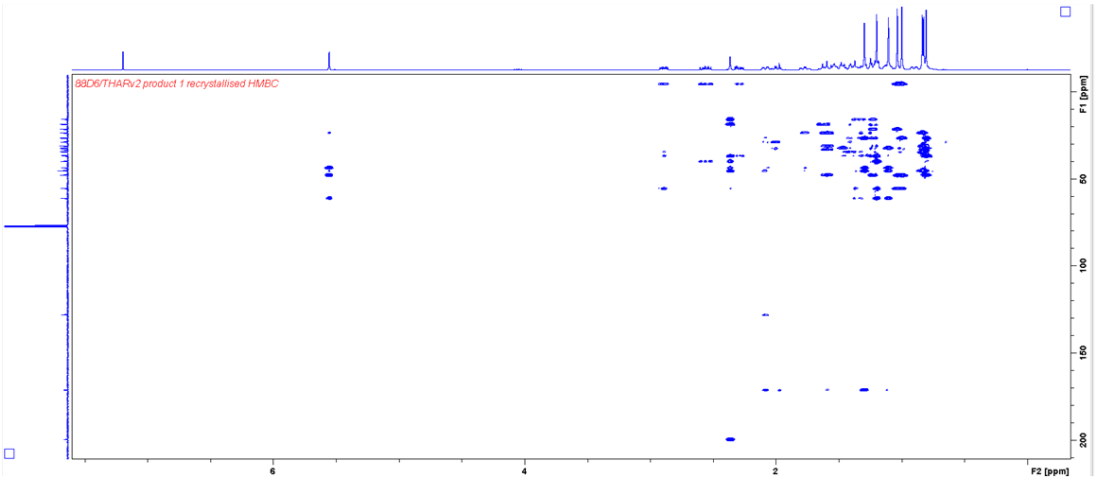


Figure A-51: HMBC NMR spectrum for 11-oxo-β-amyrone (400 MHz)

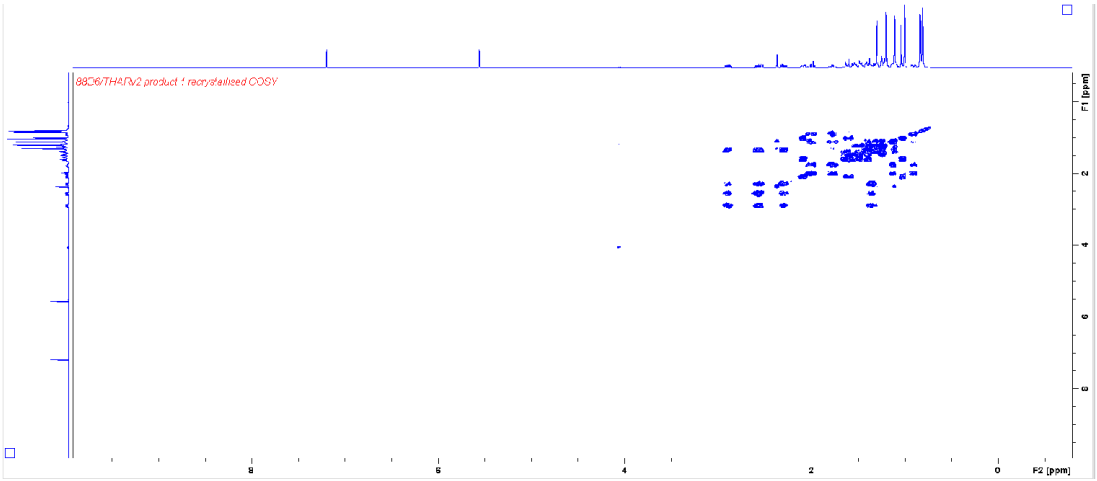


Figure A-52: COSY NMR spectrum for 11-oxo-β-amyrone (400 MHz)

Carbon numbering scheme **Selected 2D NMR**

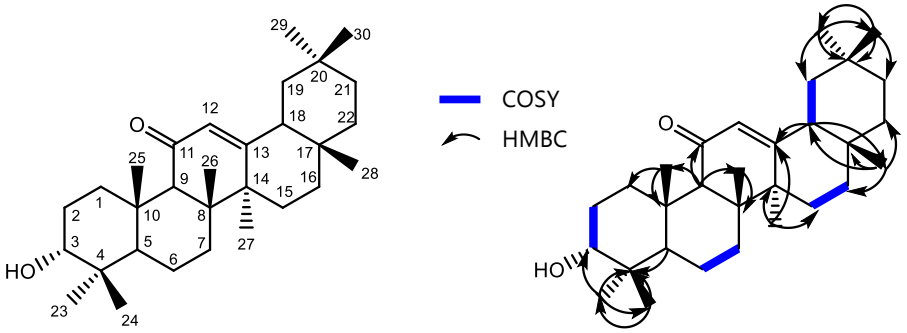


Table A-11: NMR assignments for 11-oxo-epi-β-amyrin (400 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	33.45	CH_2	2.44, 1.29-1.09	m
2	25.41	CH_2	2.03-1.90, 1.47-1.332.50-	m
3	75.86	CH	3.35	t, J = 2.7 Hz
4	37.54	C	N/A	N/A
5	48.41	CH	1.29-1.09	m
6	32.65	CH_2	1.68-1.56, 1.47-1.32	m
7	17.41	CH_2	1.47-1.32	m
8	45.64	C	N/A	N/A
9	61.67	CH	2.39	s
10	37.17	C	N/A	N/A
11	200.58	C	N/A	N/A
12	128.12	CH	5.51	s
13	170.65	C	N/A	N/A
14	43.47	C	N/A	N/A
15	26.39	CH_2	1.74, 0.88-0.85	td, J = 13.8, 4.4 Hz
16	26.41	CH_2	2.03-1.90, 1.29-1.09	m
17	32.37	C	N/A	N/A
18	47.60	CH	2.10-2.03	dd, J= 12.9, 3.4 Hz
19	45.18	CH_2	1.68-1.56, 1.04-0.97	m
20	31.05	C	N/A	N/A
21	34.47	CH_2	1.47-1.32, 1.29-1.09	m
22	36.54	CH_2	1.47-1.32, 1.29-1.09	m
23	28.46	CH_3	0.89	s
24	22.36	CH_3	0.80	s
25	16.29	CH_3	1.08	s
26	18.72	CH_3	1.07	s
27	23.64	CH_3	1.30	s
28	28.77	CH_3	0.79	s
29	33.08	CH_3	0.84	s
30	23.51	CH_3	0.82	s

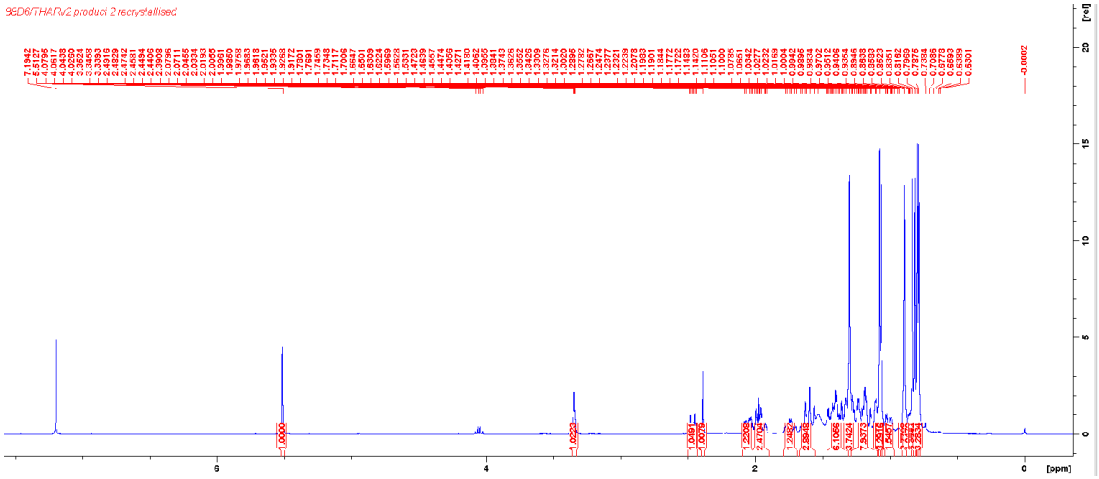


Figure A-53: ^{1}H NMR spectrum for 11-oxo-epi-β-amyrin (400 MHz)

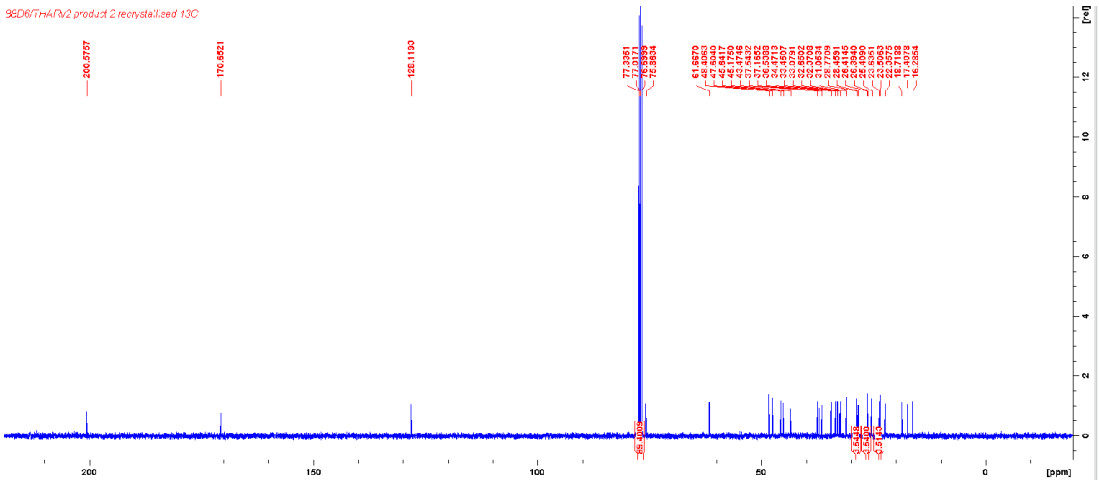


Figure A-54: ^{13}C NMR spectrum for 11-oxo-epi-β-amyrin (400 MHz)

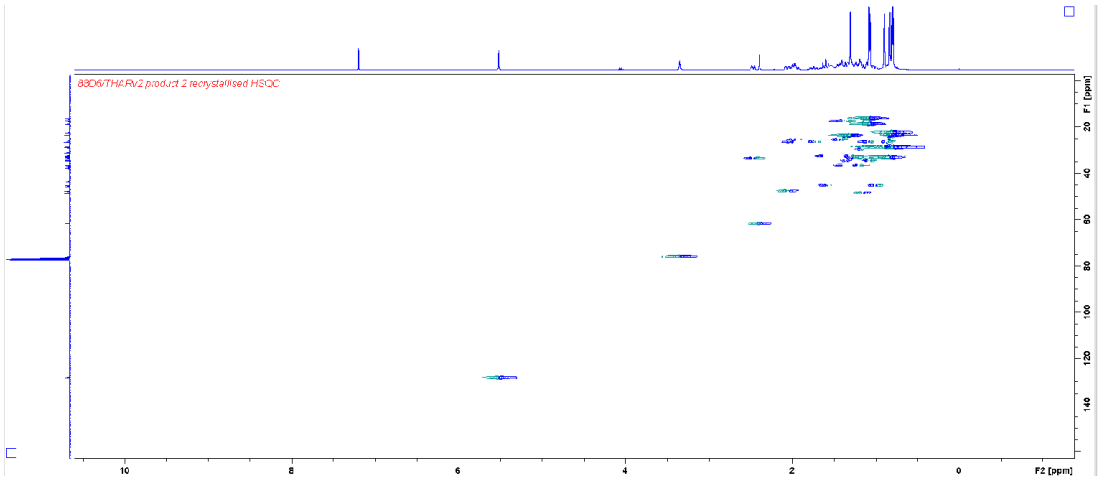


Figure A-55: HSQC NMR spectrum for 11-oxo-epi-β-amyrin (400 MHz)

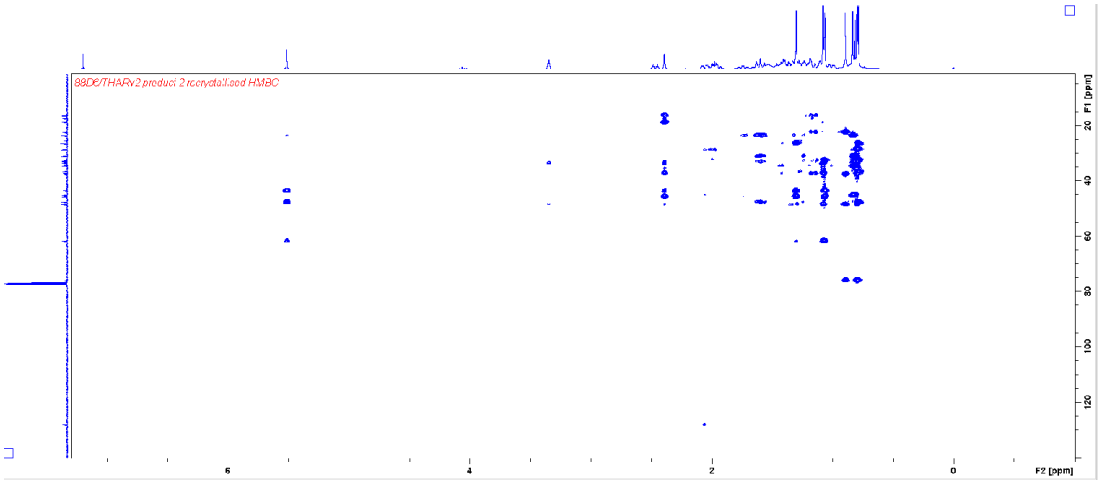


Figure A-56: HMBC NMR spectrum for 11-oxo-epi-β-amyrin (400 MHz)

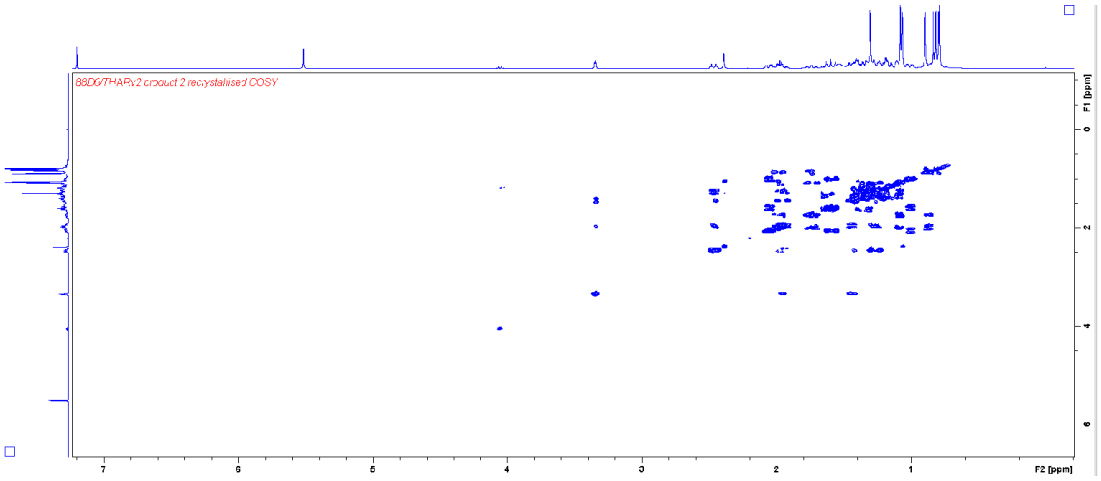


Figure A-57: COSY NMR spectrum for 11-oxo-epi-β-amyrin (400 MHz)

Carbon numbering scheme **Selected 2D NMR**

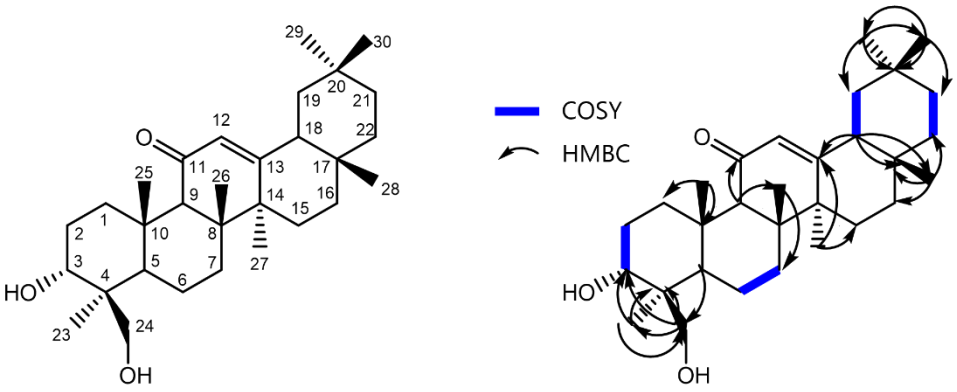


Table A-12: NMR assignments for 24-hydroxy-11-oxo-epi-β-amyrin (400 MHz)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	34.01	CH_2	2.53, 1.46-1.33	dt, J = 13.6, 4.0 Hz
2	25.47	CH_2	2.01-1.91, 1.60-1.50	m
3	70.52	CH	3.86	t, J = 3.1 Hz
4	42.91	C	N/A	N/A
5	49.12	CH	1.35	m
6	17.82	CH_2	1.61-1.52, 1.47-1.33	m
7	32.94	CH_2	1.71-1.62, 1.46-1.33	m
8	45.59	C	N/A	N/A
9	61.81	CH	2.47	s
10	36.89	C	N/A	N/A
11	200.34	C	N/A	N/A
12	128.03	CH	5.59	s
13	170.88	C	N/A	N/A
14	43.45	C	N/A	N/A
15	26.38	CH_2	2.06, 1.80/1.16, 0.95	td, J = 13.6, 4.3 Hz
16	26.38	CH_2	2.06, 1.80/1.16, 0.95	td, J = 13.6, 4.3 Hz
17	32.36	C	N/A	N/A
18	47.62	CH	2.13	dd, J = 13.8, 3.4 Hz
19	45.18	CH	1.71-1.62, 1.09-1.03	m
20	31.05	C	N/A	N/A
21	34.45	CH_2	1.33-1.24, 1.20-1.14	m
22	36.51	CH_2	1.52-1.41, 1.33-1.24	m
23	21.74	CH_3	1.10	s
24	66.58	CH_2	3.64	dd, J = 73.2, 11.0 Hz
25	17.09	CH_3	1.13	s
26	18.52	CH_3	1.11	s
27	23.59	CH_3	1.37	s
28	28.77	CH_3	0.85	s
29	33.07	CH_3	0.90	s
30	23.50	CH_3	0.88	s

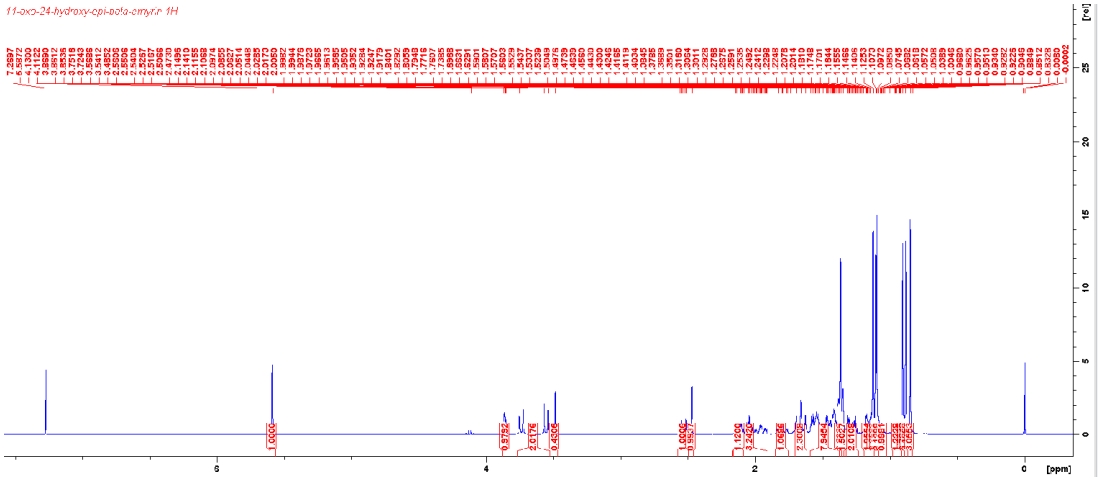


Figure A-58: ^{1}H NMR spectrum for 24-hydroxy-11-oxo-epi-β-amyrin (400 MHz)

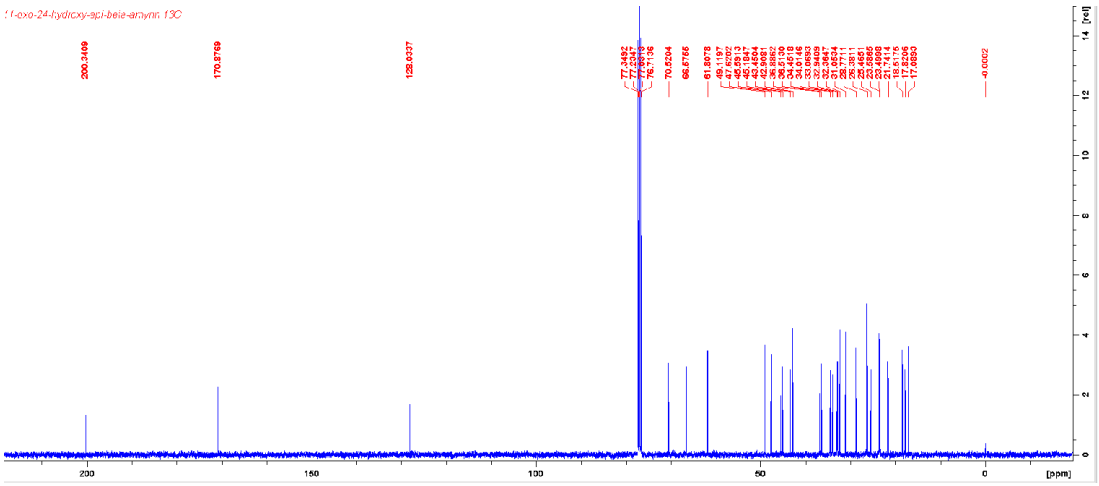


Figure A-59: ^{13}C NMR spectrum for 24-hydroxy-11-oxo-epi-β-amyrin (400 MHz)

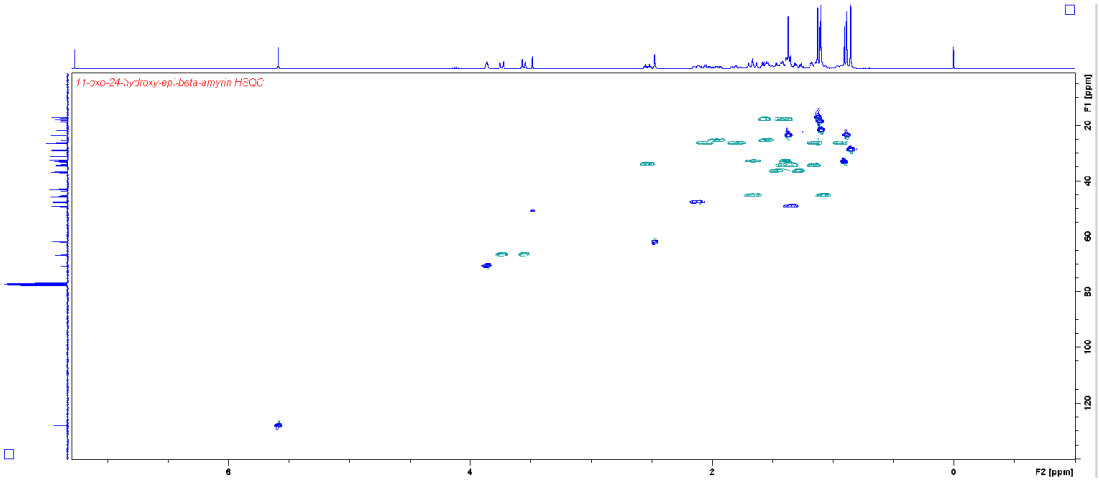


Figure A-60: HSQC NMR spectrum for 24-hydroxy-11-oxo-epi-β-amyrin (400 MHz)

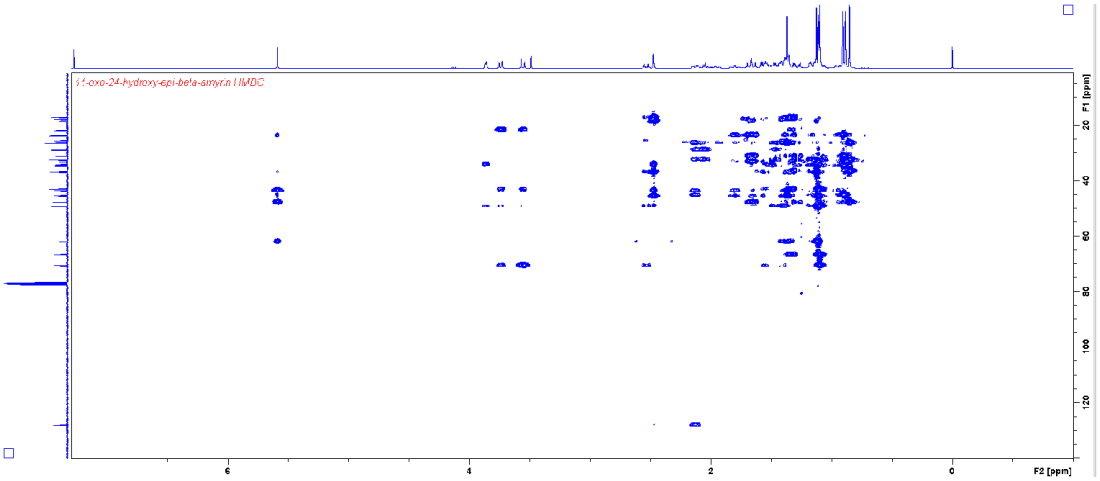


Figure A-61: HMBC NMR spectrum for 24-hydroxy-11-oxo-epi-β-amyrin (400 MHz)

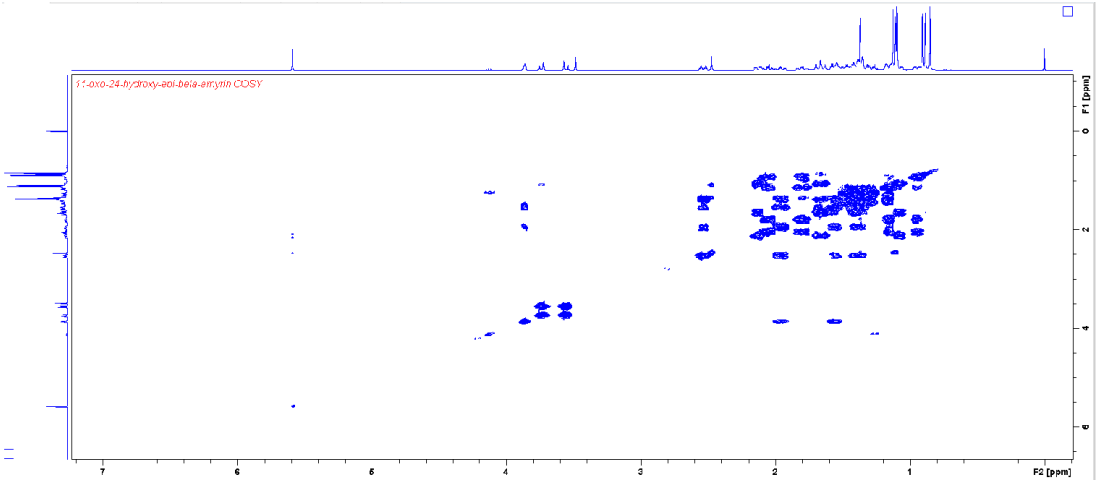


Figure A-62: COSY NMR spectrum for 24-hydroxy-11-oxo-epi-β-amyrin (400 MHz)

Carbon numbering scheme **Selected 2D NMR**

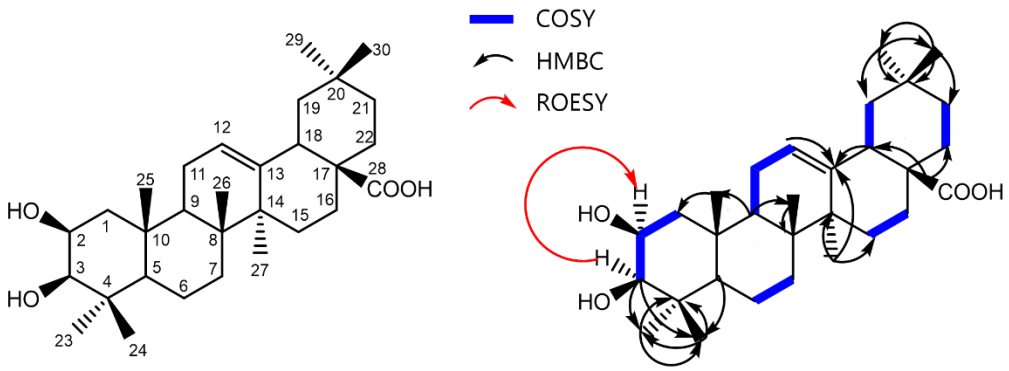


Table A-13: NMR assignments for augustic acid (600 MHz, pyridine)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	44.71	CH_2	2.30, 1.25-1.16	dd, J = 14.3, 2.6 Hz
2	71.21	CH	4.39	m
3	78.11	CH	3.42	d, J = 3.7 Hz
4	38.58	C	N/A	N/A
5	55.74	CH	1.01	m
6	18.41	CH_2	1.65-1.59, 1.59-1.51	m
7	33.11	CH_2	1.59-1.51, 1.38-1.33	m
8	39.68	C	N/A	N/A
9	48.30	CH	1.67	dd, J = 11.2, 6.7 Hz
10	37.15	C	N/A	N/A
11	23.73	CH_2	2.15-2.07, 2.08-1.99	m
12	122.47	CH	5.50	t, J = 3.1 Hz
13	144.64	C	N/A	N/A
14	42.11	C	N/A	N/A
15	28.03	CH_2	2.20, 1.25-1.16	td, J = 13.6, 3.9 Hz
16	23.49	CH_2	2.15-2.07, 1.99-1.94	m
17	46.46	C	N/A	N/A
18	41.82	CH	3.31	dd, J = 14.1, 4.4 Hz
19	46.26	CH_2	1.81, 1.33-1.25	m
20	30.74	C	N/A	N/A
21	32.99	CH_2	2.08-1.99, 1.86-1.79	m
22	34.00	CH_2	1.44, 1.25-1.16	td, J = 13.8, 3.6 Hz
23	30.03	CH_3	1.26	s
24	17.94	CH_3	1.36	s
25	16.39	CH_3	1.51	s
26	17.26	CH_3	1.07	s
27	26.01	CH_3	1.30	s
28	179.98	C	N/A	N/A
29	33.04	CH_3	0.93	s
30	23.53	CH_3	0.99	s

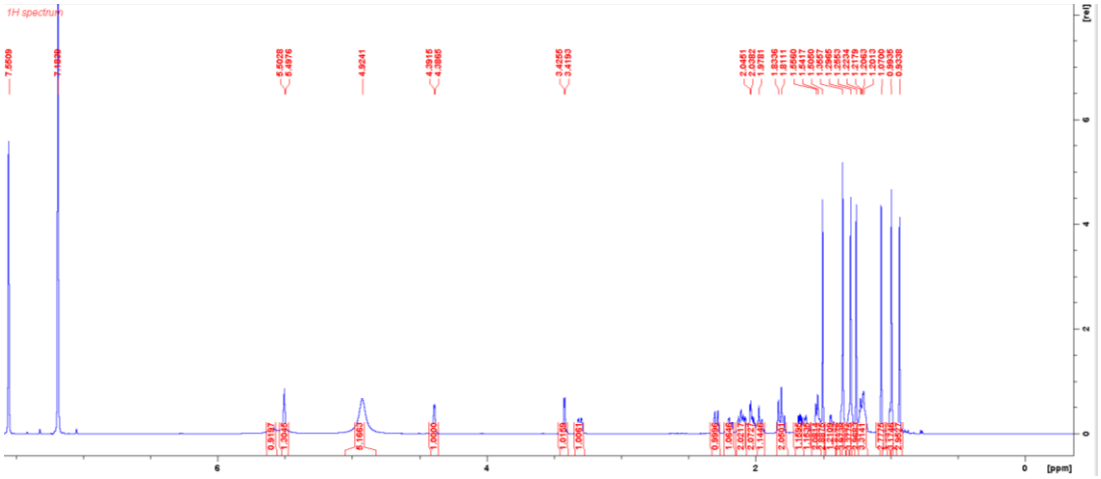


Figure A-63: ^{1}H NMR spectrum for augustic acid (600 MHz, pyridine)

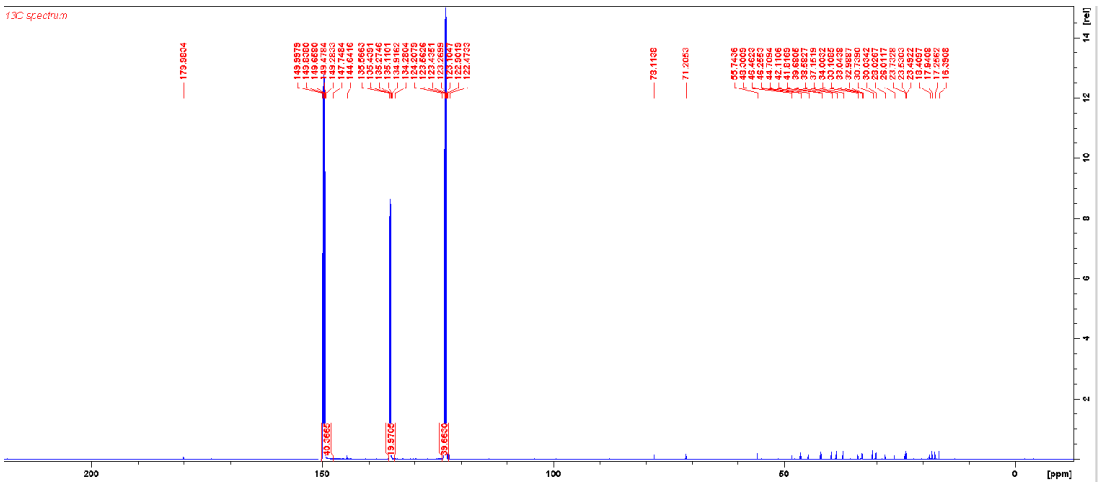


Figure A-64: ^{13}C NMR spectrum for augustic acid (600 MHz, pyridine)

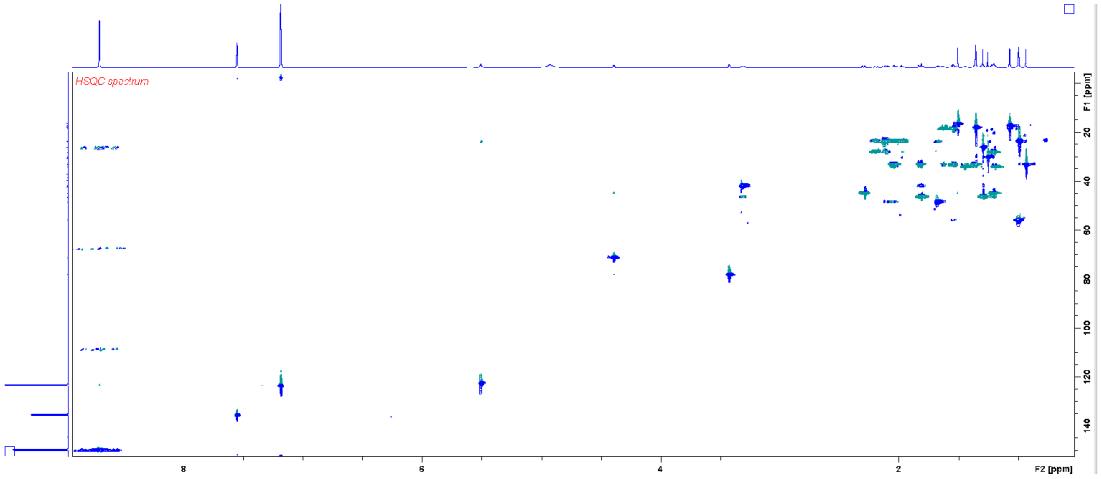


Figure A-65: HSQC NMR spectrum for augustic acid (600 MHz, pyridine)

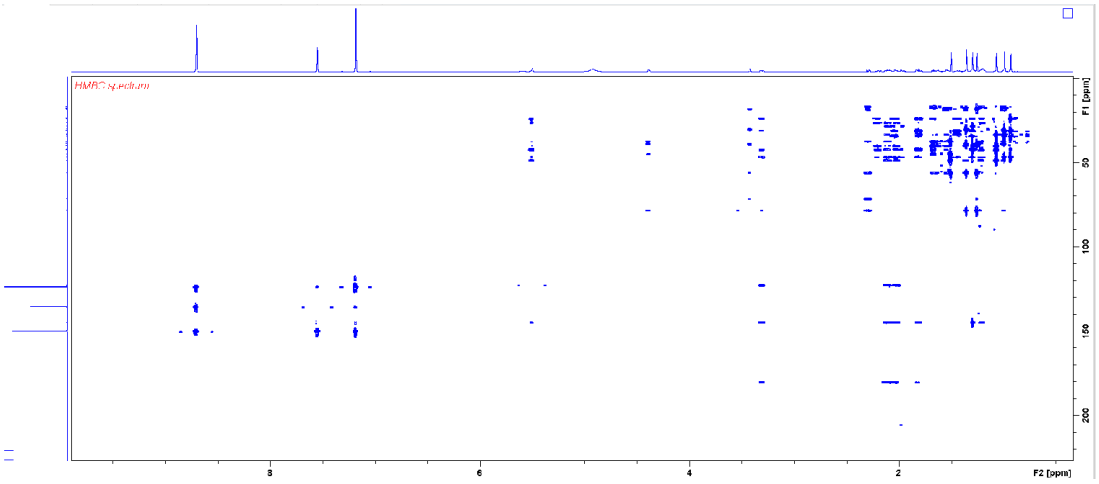


Figure A-66: HMBC NMR spectrum for augustic acid (600 MHz, pyridine)

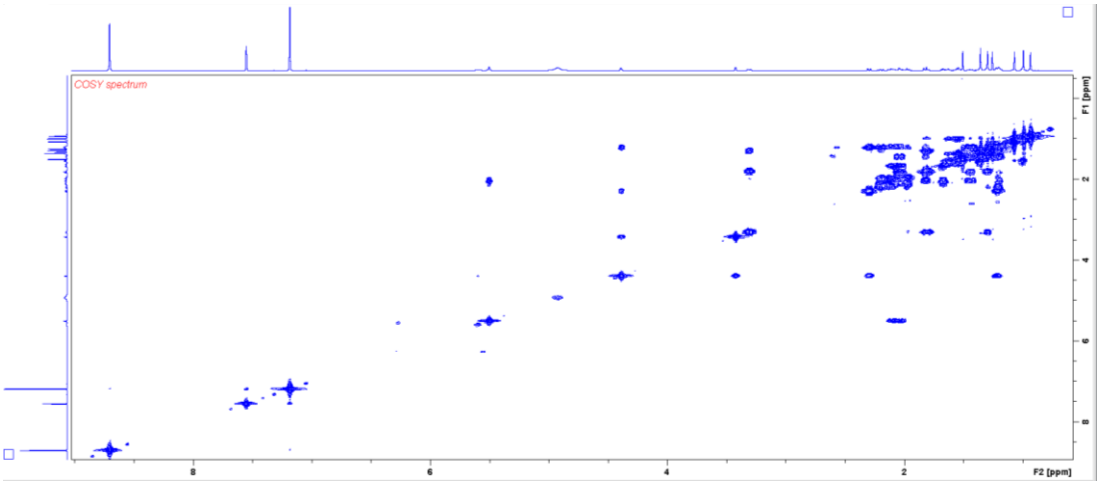


Figure A-67: COSY NMR spectrum for augustic acid (600 MHz, pyridine)

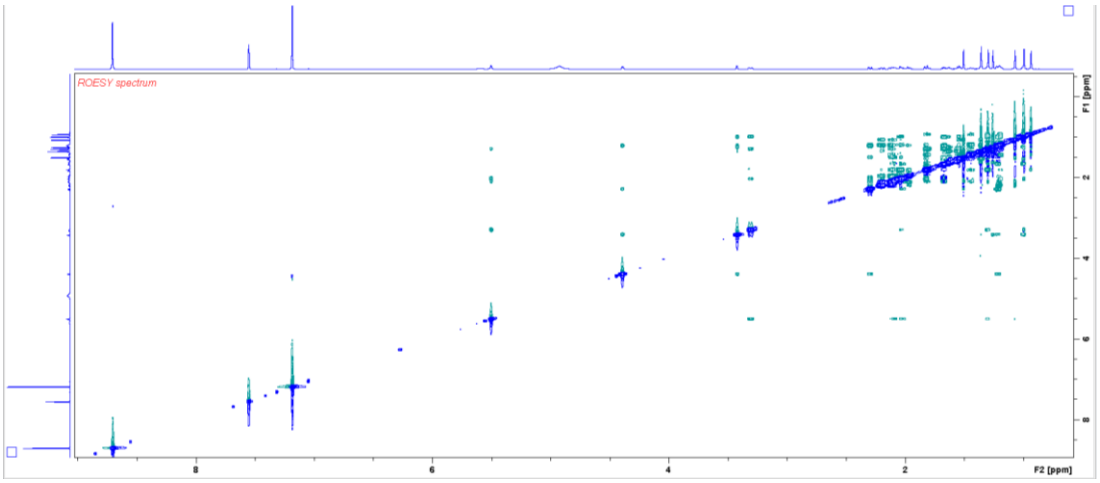


Figure A-68: ROESY NMR spectrum for augustic acid (600 MHz, pyridine)

Carbon numbering scheme **Selected 2D NMR**

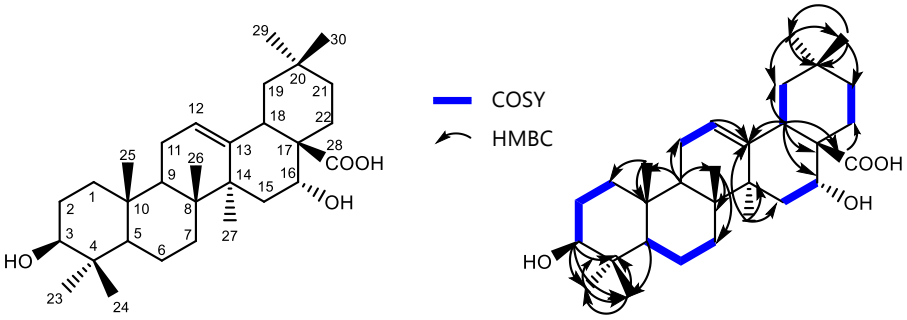


Table A-14: NMR assignments for echinocystic acid (600 MHz, pyridine)

Atom Number	^{13}C δ (ppm)	^{13}C Type	^{1}H δ (ppm)	^{1}H Multiplicity
1	38.80	CH_2	1.61-1.55, 1.08-1.01	m
2	27.94	CH_2	1.87-1.80	m
3	77.87	CH	3.45	dd, J = 11.1, 5.1 Hz
4	39.18	C	N/A	N/A
5	55.69	CH	0.90	d, J = 11.1 Hz
6	18.61	CH_2	1.58-1.53, 1.41-1.33	m
7	33.37	CH_2	1.64, 1.42-1.37	td J = 13.6, 3.8 Hz
8	39.70	C	N/A	N/A
9	47.07	CH	1.87-1.81	m
10	37.21	C	N/A	N/A
11	23.64	CH_2	2.02	m
12	122.21	CH	5.66	t, J = 3.2 Hz
13	144.91	C	N/A	N/A
14	41.92	C	N/A	N/A
15	35.94	CH_2	2.43-2.37, 1.76	dd, J = 14.5, 2.4 Hz
16	74.54	CH	5.26	m
17	48.67	C	N/A	N/A
18	41.25	CH	3.63	dd, J = 14.3, 4.2 Hz
19	47.09	CH_2	2.84, 1.44-1.38	m
20	30.84	C	N/A	N/A
21	36.00	CH_2	2.51, 1.39-1.33	td J = 12.9, 4.4 Hz
22	32.67	CH_2	2.46-2.39, 2.26	td, J = 13.7, 5 Hz
23	28.55	CH_3	1.22	s
24	16.35	CH_3	1.02	s
25	15.46	CH_3	0.93	s
26	17.32	CH_3	1.06	s
27	27.02	CH_3	1.84	s
28	179.76	C	N/A	N/A
29	33.13	CH_3	1.05	s
30	24.50	CH_3	1.17	s

NMR is in accordance with the literature [196].

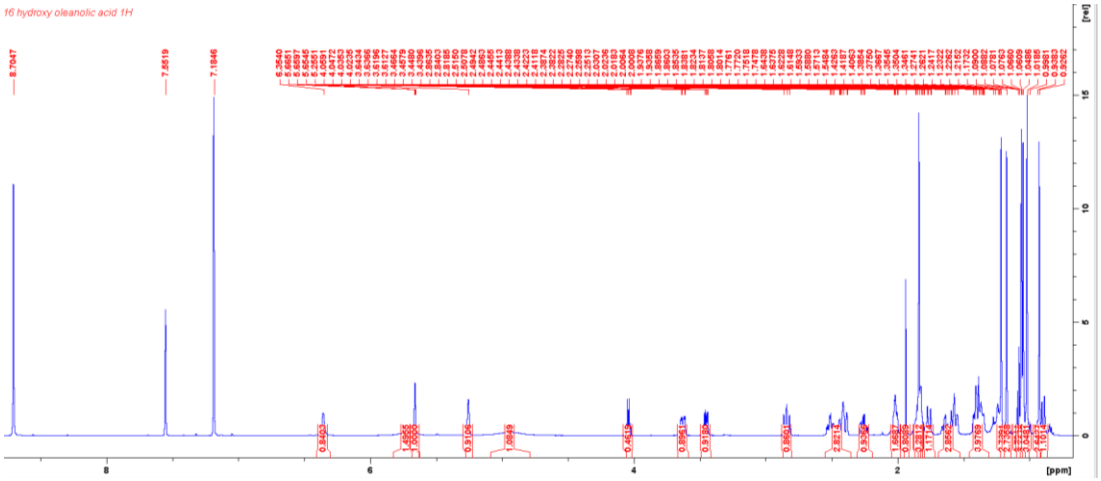


Figure A-69: ^{1}H NMR spectrum for echinocystic acid (600 MHz, pyridine)

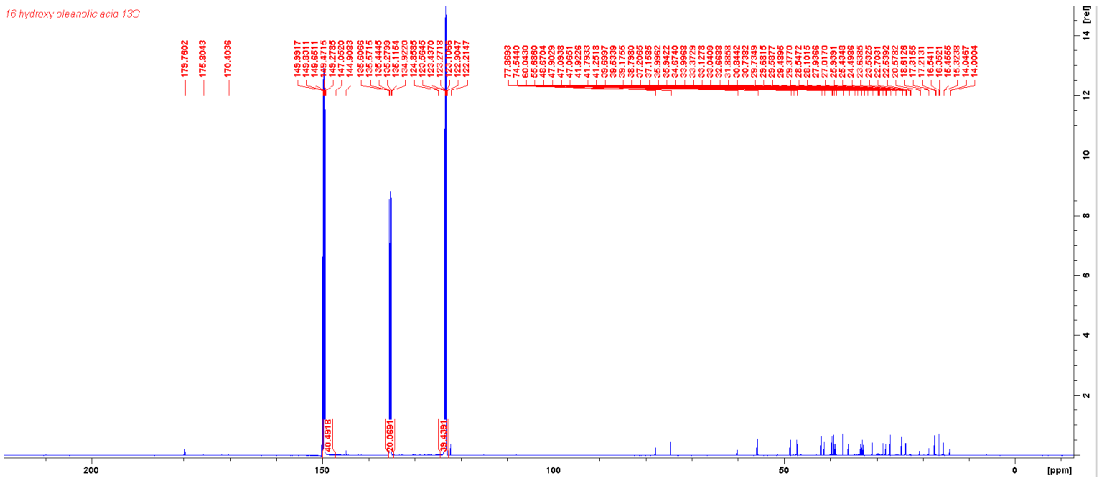


Figure A-70: ^{13}C NMR spectrum for echinocystic acid (600 MHz, pyridine)

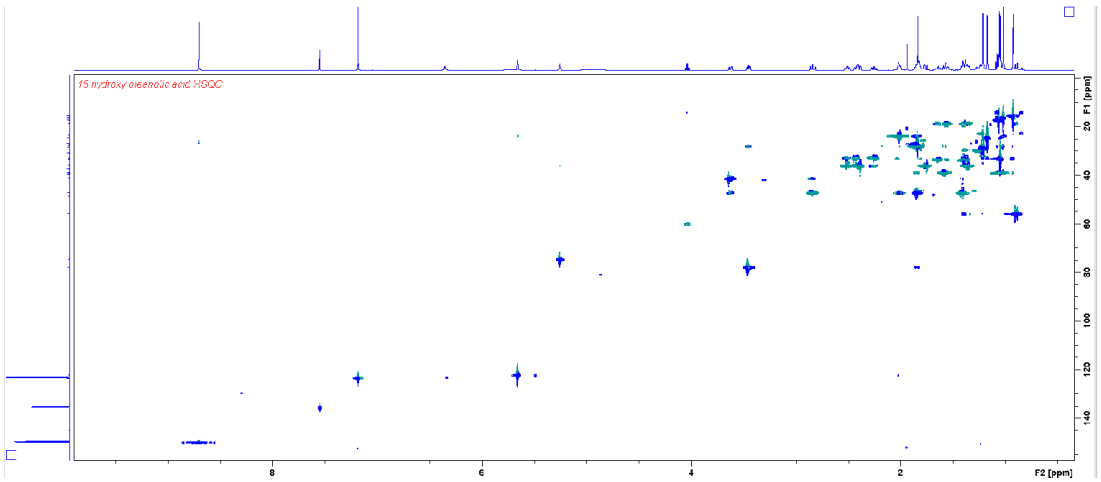


Figure A-71: HSQC NMR spectrum for echinocystic acid (600 MHz, pyridine)

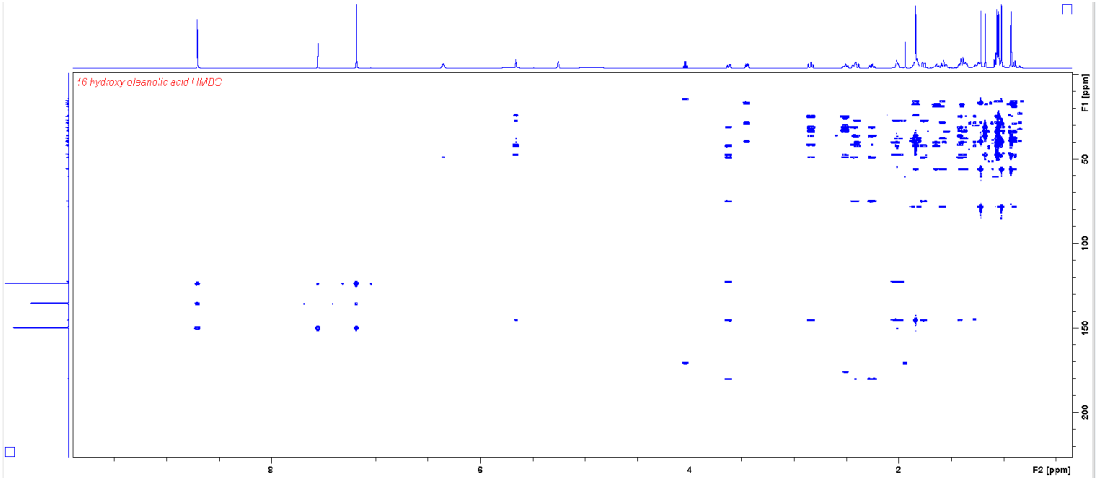


Figure A-72: HMBC NMR spectrum for echinocystic acid (600 MHz, pyridine)

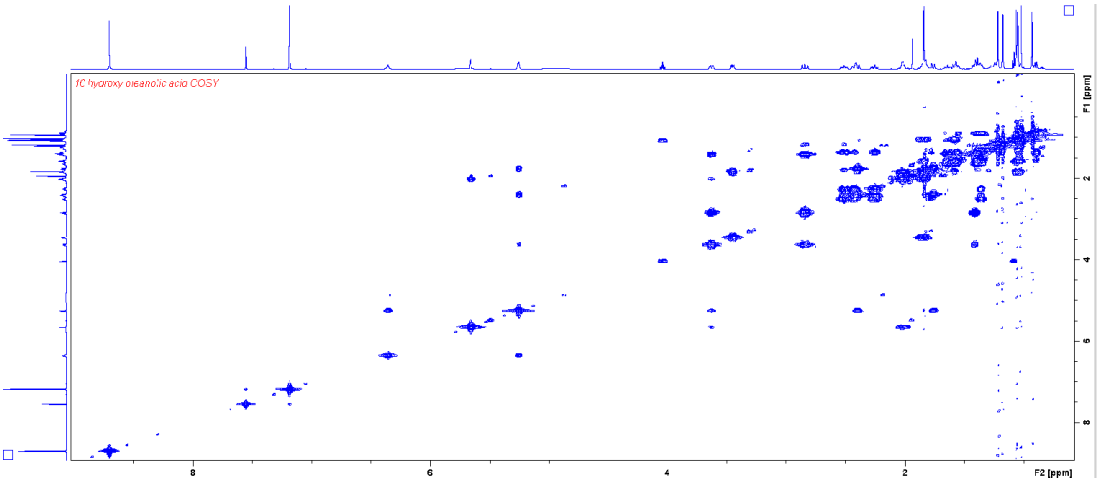


Figure A-73: COSY NMR spectrum for echinocystic acid (600 MHz, pyridine)

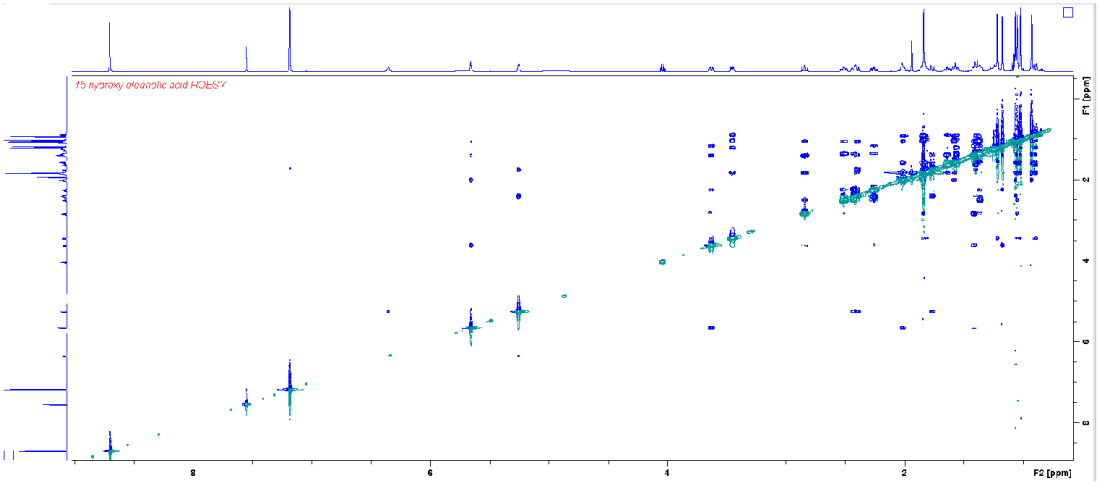


Figure A-74: ROESY NMR spectrum for echinocystic acid (600 MHz, pyridine)

A.2 Additional GC-MS Spectra

A.2.1 Chapter 3

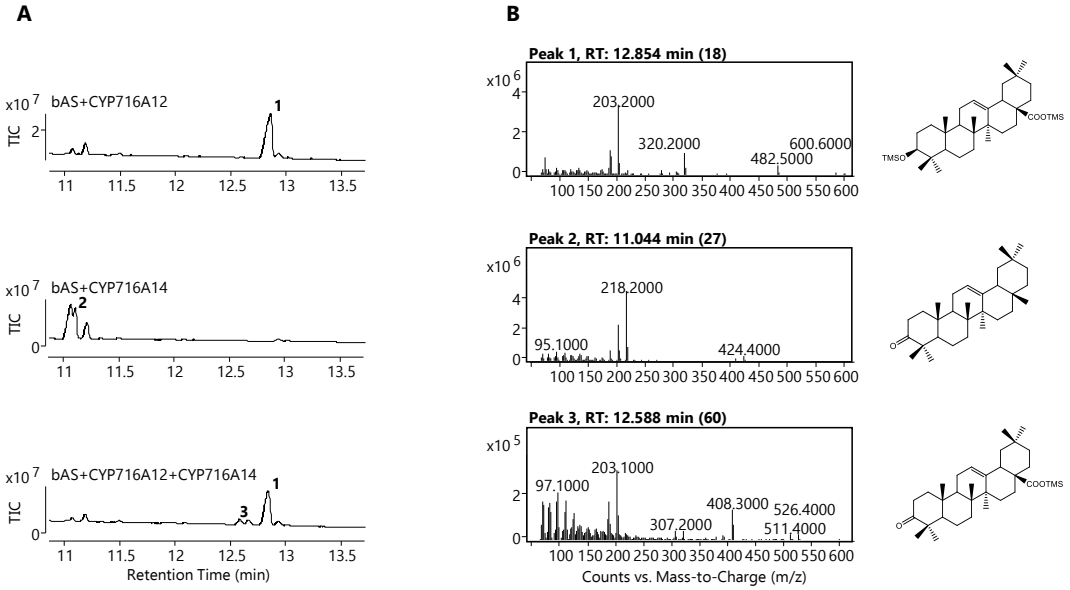


Figure A-75: CYP716A12+CYP716A14 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP716A14 (middle); and CYP716A12+CYP716A14 (bottom). B) Mass spectra for the peaks at 12.9 (peak 1), 11.0 (peak 2), and 12.6 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

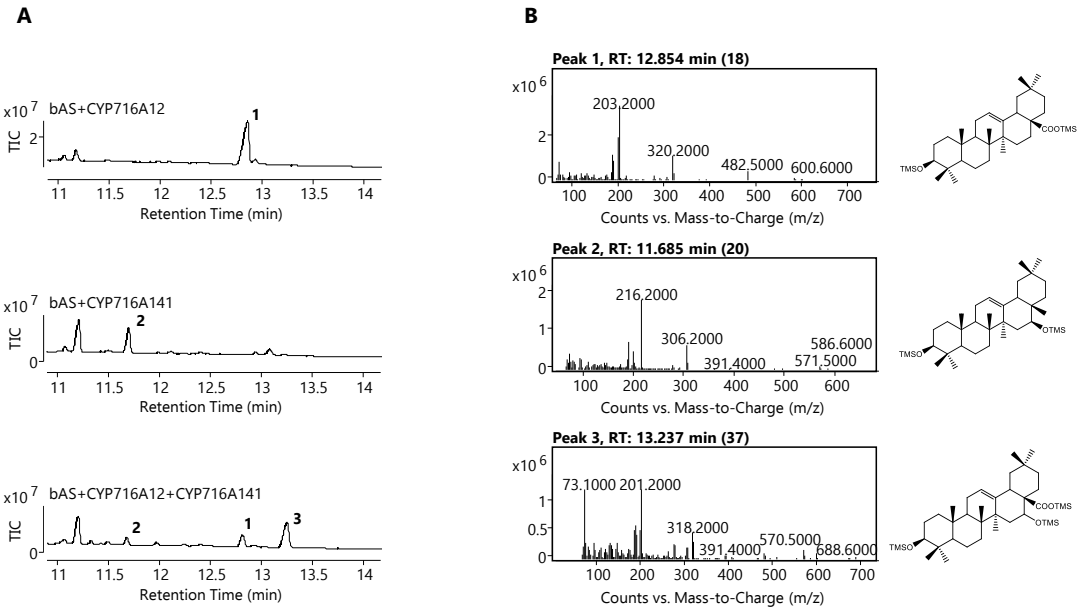


Figure A-76: CYP716A12+CYP716A141 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP716A141 (middle); and CYP716A12+CYP716A141 (bottom). B) Mass spectra for the peaks at 12.9 (peak 1), 11.7 (peak 2), and 13.2 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

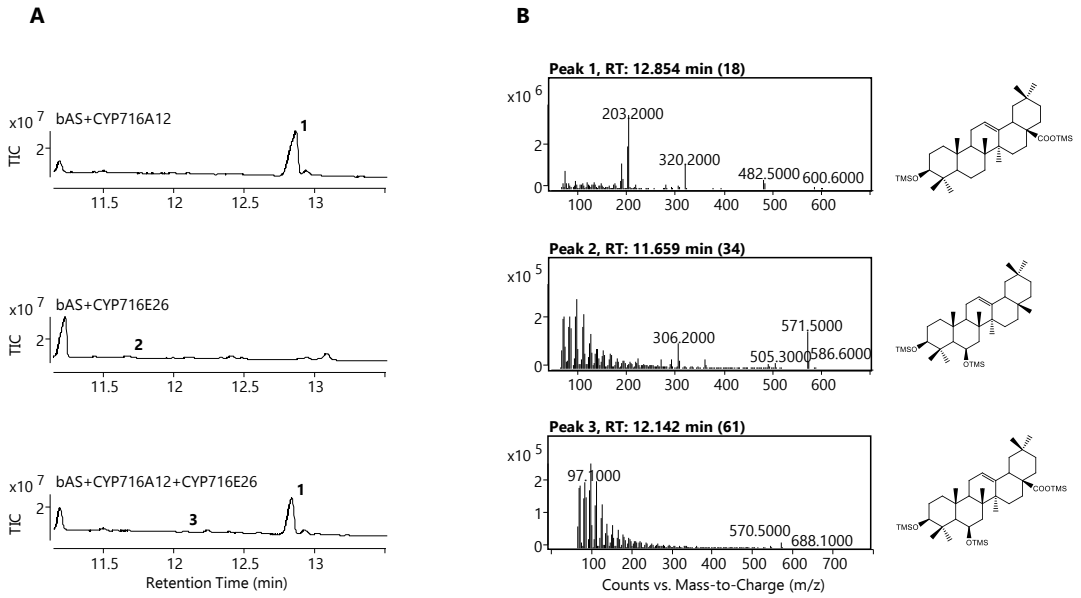


Figure A-77: CYP716A12+CYP716E26 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP716E26 (middle); and CYP716A12+CYP716E26 (bottom). B) Mass spectra for the peaks at 12.9 (peak 1), 11.7 (peak 2), and 12.1 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

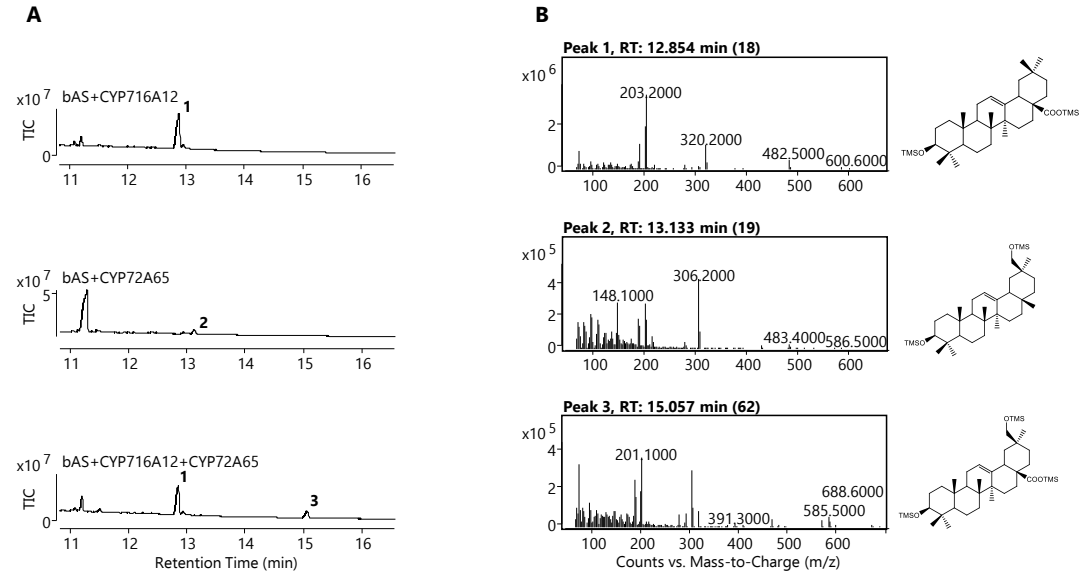


Figure A-78: CYP716A12+CYP72A65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP72A65 (middle); and CYP716A12+CYP72A65 (bottom). B) Mass spectra for the peaks at 12.9 (peak 1), 13.1 (peak 2), and 15.1 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

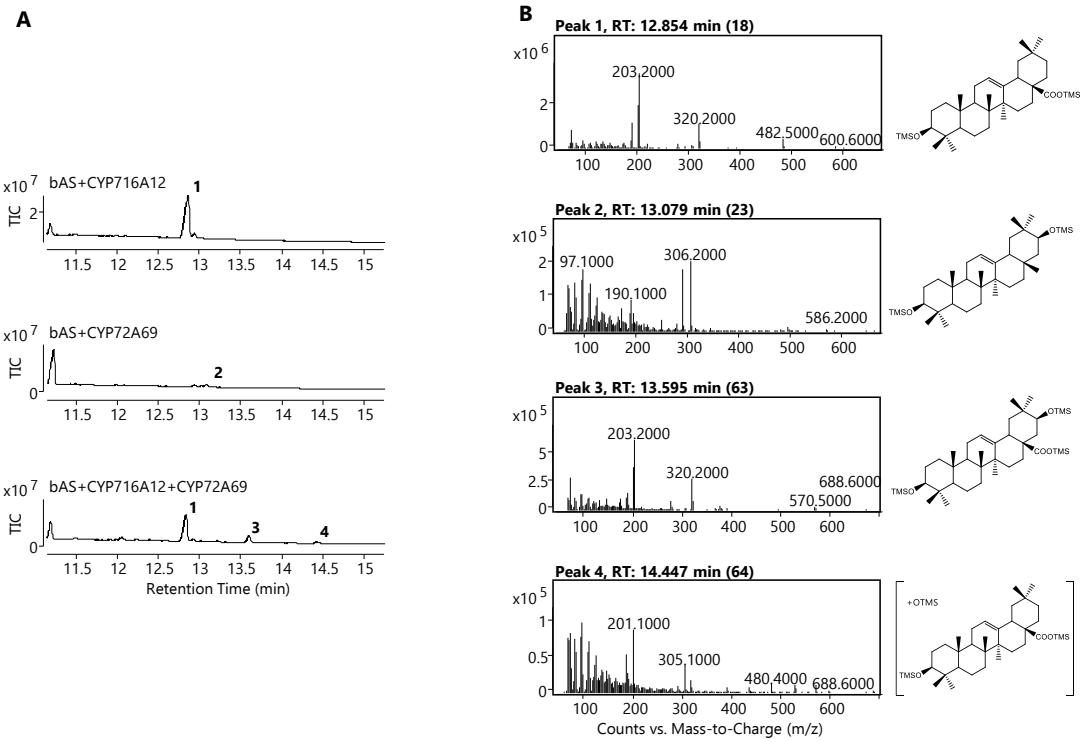


Figure A-79: CYP716A12+CYP72A69 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP72A69 (middle); and CYP716A12+CYP72A69 (bottom). B) Mass spectra for the peaks at 12.9 (peak 1), 13.1 (peak 2), 13.6 (peak 3), and 14.4 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

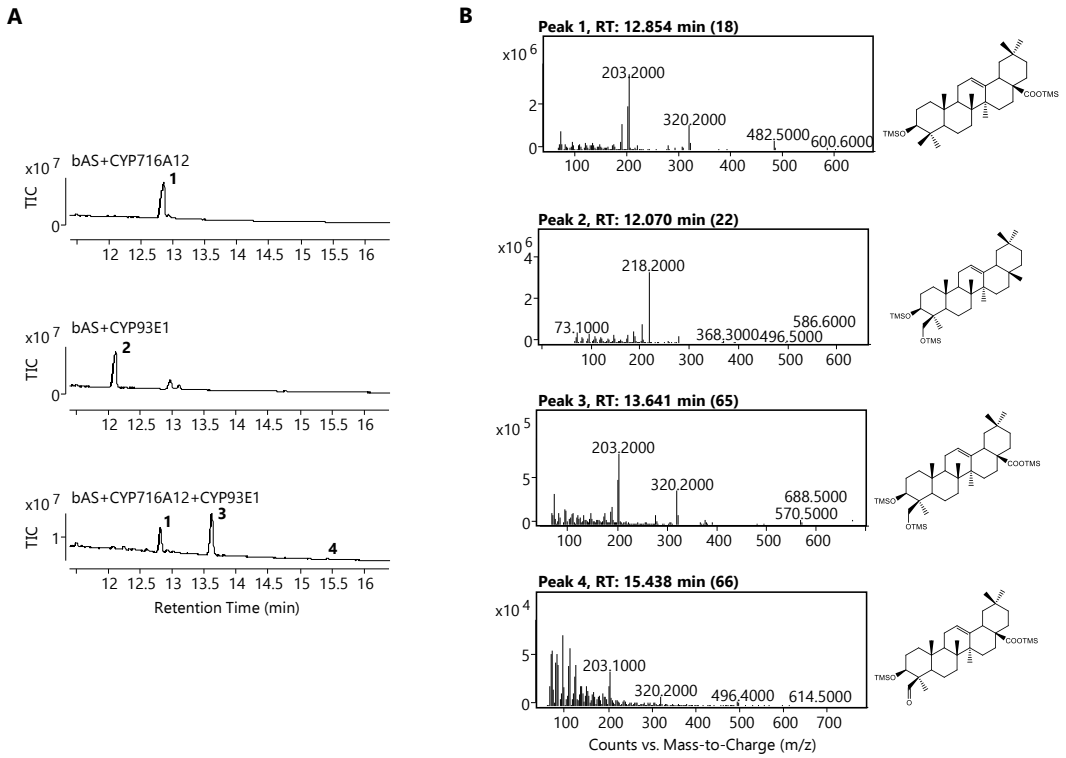


Figure A-80: CYP716A12+CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP93E1 (middle); and CYP716A12+CYP93E1 (bottom). B) Mass spectra for the peaks at 12.9 (peak 1), 12.0 (peak 2), 13.6 (peak 3), and 15.4 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

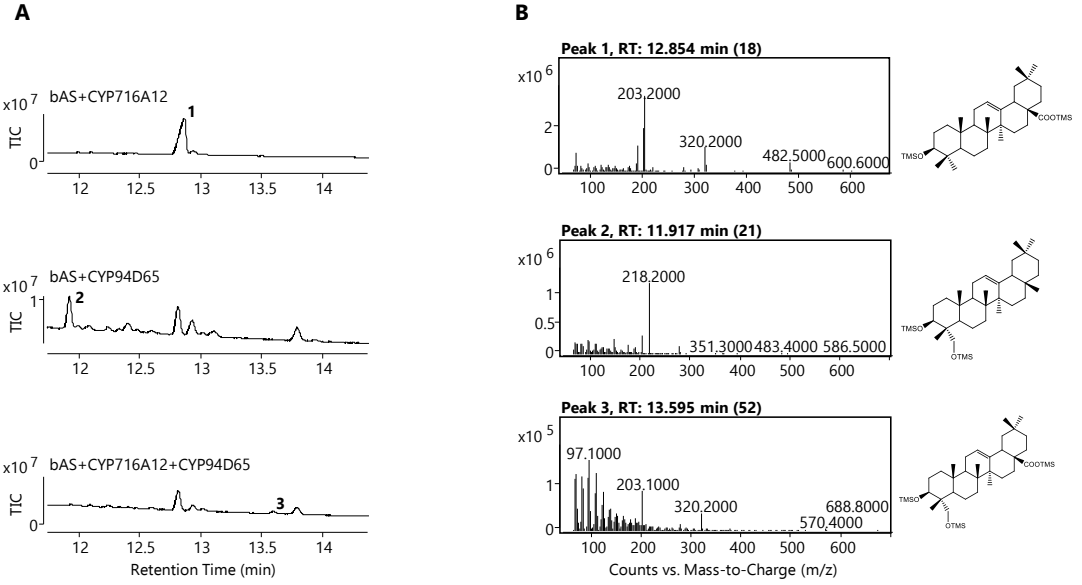


Figure A-81: CYP716A12+CYP94D65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A12 only (top); CYP94D65 (middle); and CYP716A12+CYP94D65 (bottom). B) Mass spectra for the peaks at 12.9 (peak 1), 11.9 (peak 2), and 13.6 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

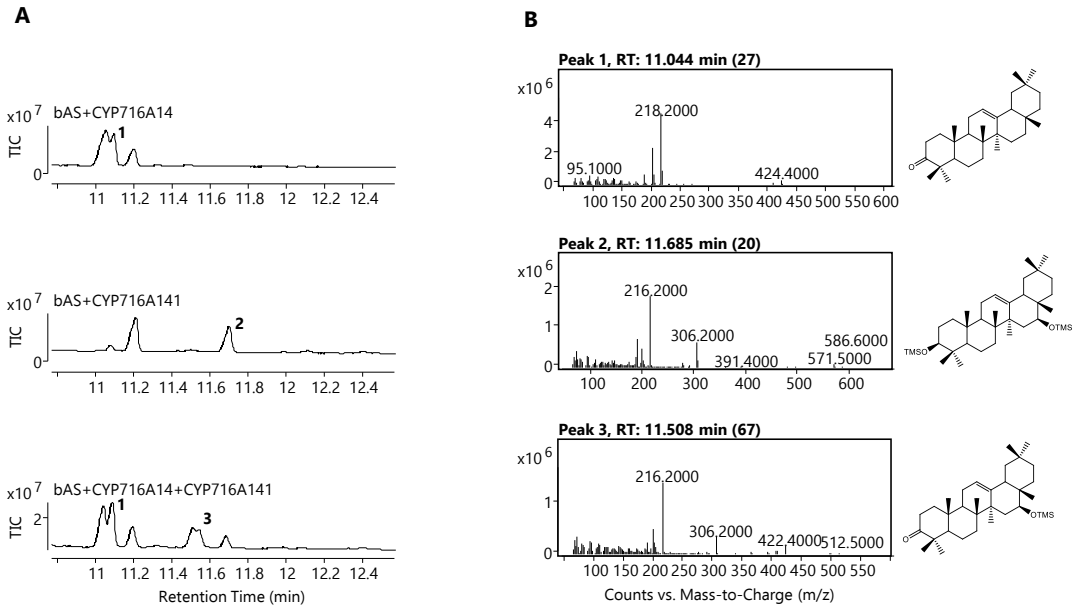


Figure A-82: CYP716A14+CYP716A141 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A14 only (top); CYP716A141 (middle); and CYP716A14+CYP716A141 (bottom). B) Mass spectra for the peaks at 11.0 (peak 1), 11.7 (peak 2), and 11.5 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

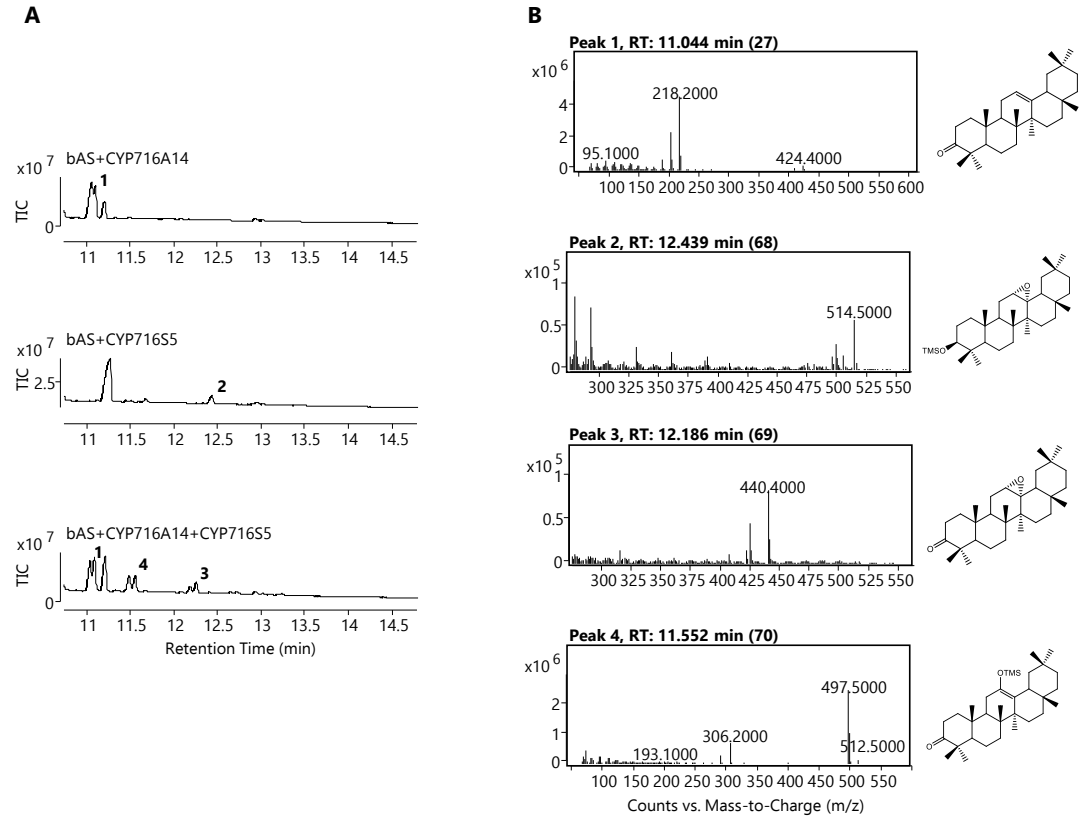


Figure A-83: CYP716A14+CYP716S5 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A14 only (top); CYP716S5 (middle); and CYP716A14+CYP716S5 (bottom). B) Mass spectra for the peaks at 11.0 (peak

1), 12.4 (peak 2), 12.1 (peak 3), and 11.5 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

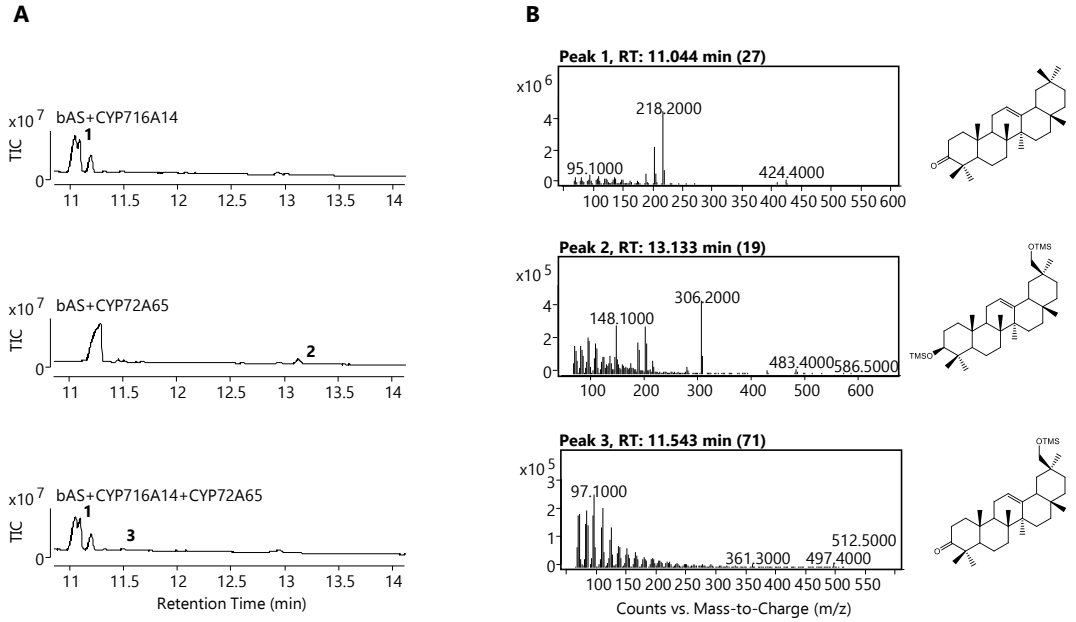


Figure A-84: CYP716A14+CYP72A65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A14 only (top); CYP72A65 (middle); and CYP716A14+CYP72A65 (bottom). B) Mass spectra for the peaks at 11.0 (peak 1), 13.1 (peak 2), and 11.5 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

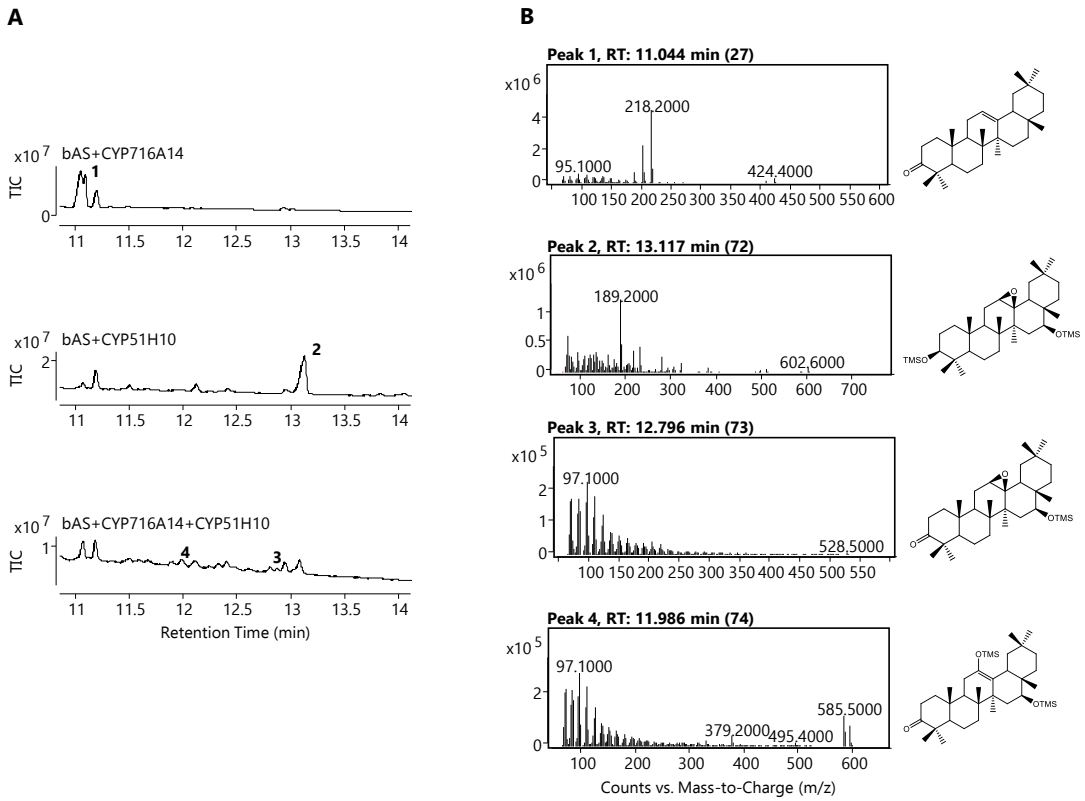


Figure A-85: CYP716A14+CYP51H10 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A14 only (top); CYP51H10 (middle); and CYP716A14+CYP51H10 (bottom). B) Mass spectra for the peaks at 11.0 (peak 1), 12.1 (peak 2), 12.8 (peak 3), and 12.0 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

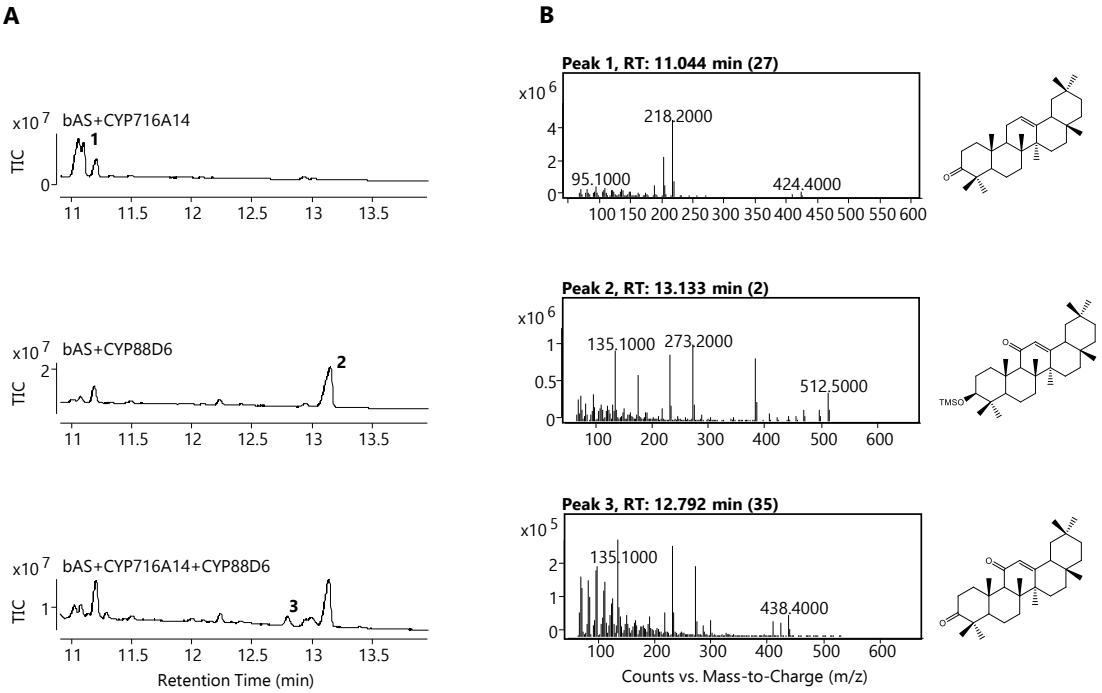


Figure A-86: CYP716A14+CYP88D6 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A14 only (top); CYP88D6 (middle); and CYP716A14+CYP88D6 (bottom). B) Mass spectra for the peaks at 11.0 (peak 1), 13.1 (peak 2), and 12.8 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

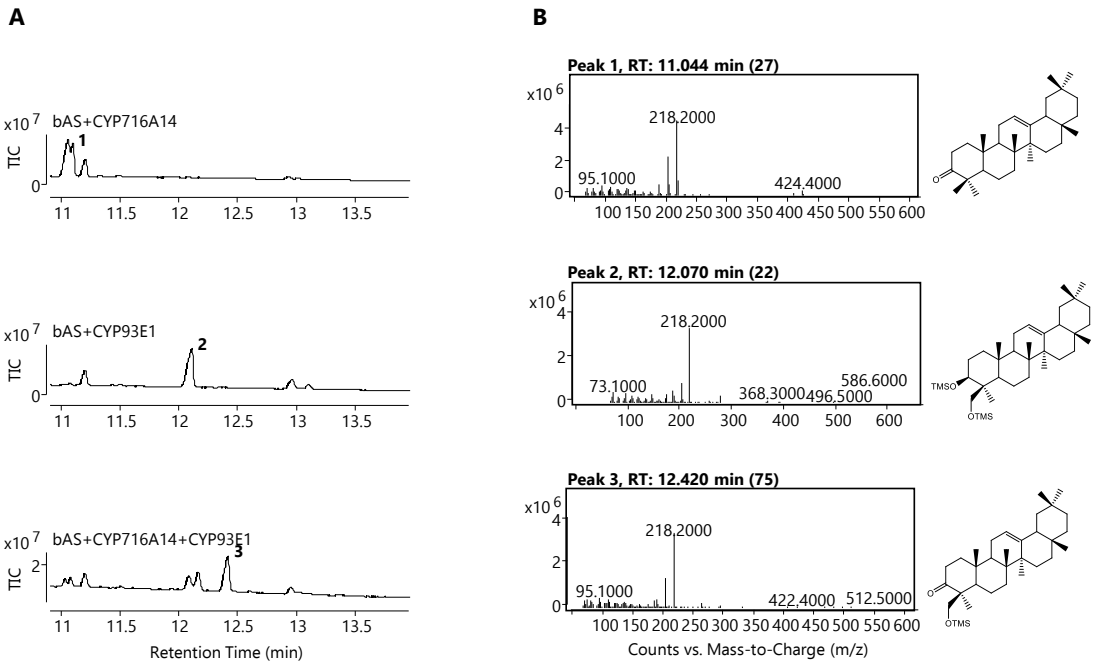


Figure A-87: CYP716A14+CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A14 only (top); CYP93E1 (middle); and CYP716A14+CYP93E1 (bottom). B) Mass spectra for the peaks at 11.0 (peak 1), 12.0 (peak 2),

and 12.4 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

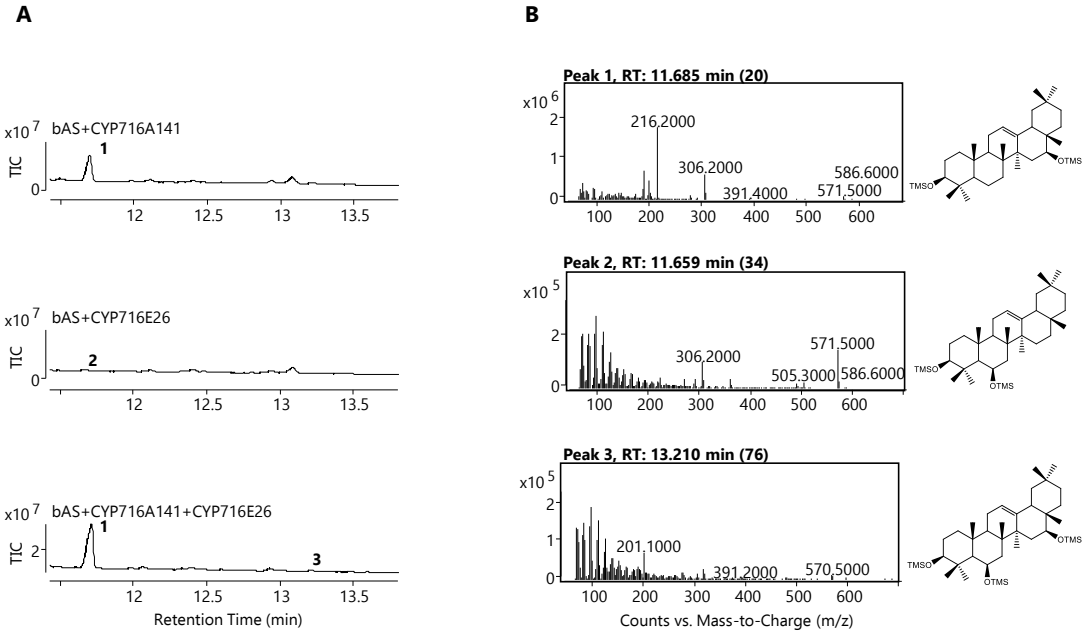


Figure A-88: CYP716A141+CYP716E26 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A141 only (top); CYP716E26 (middle); and CYP716A141+CYP716E26 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1), 11.7 (peak 2), and 13.2 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

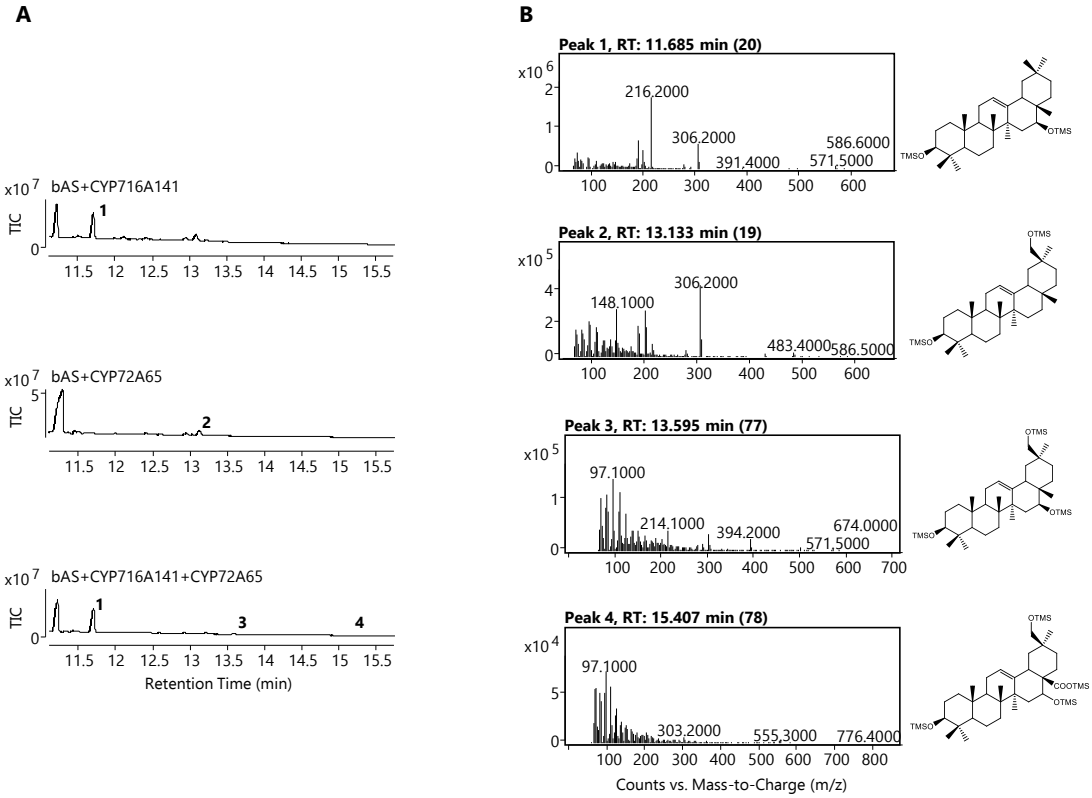


Figure A-89: CYP716A141+CYP72A65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A141 only (top); CYP72A65 (middle); and CYP716A141+CYP72A65 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1), 13.1 (peak 2), 13.6 (peak 3), and 15.4 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

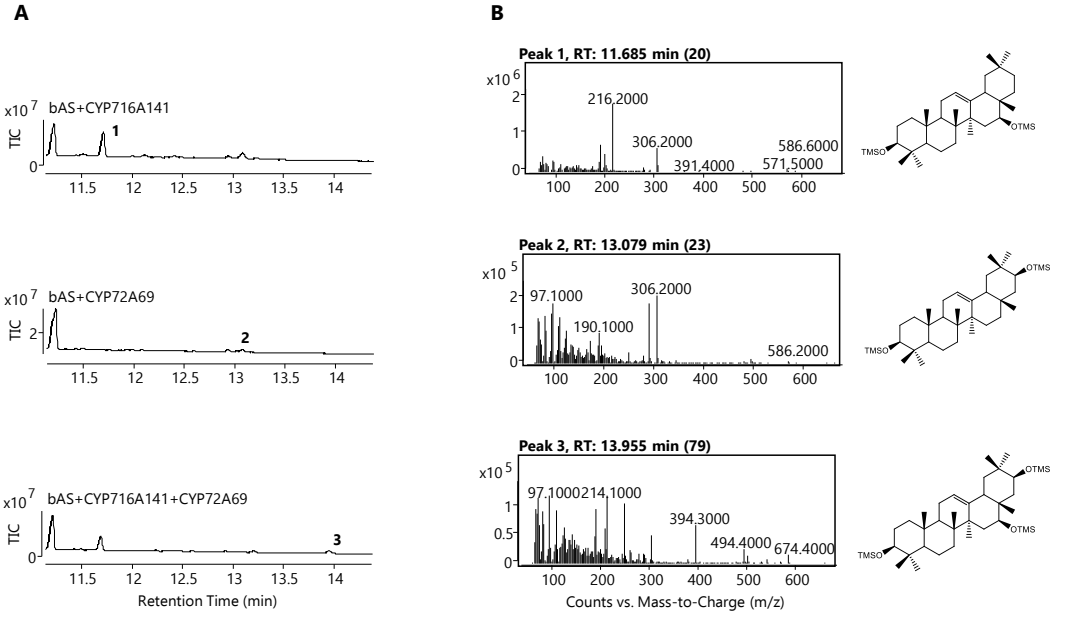


Figure A-90: CYP716A141+CYP72A69 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A141 only (top); CYP72A69 (middle); and CYP716A141+CYP72A69 (bottom). B) Mass spectra for the peaks at 11.7 (peak

1), 13.1 (peak 2), and 14.0 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

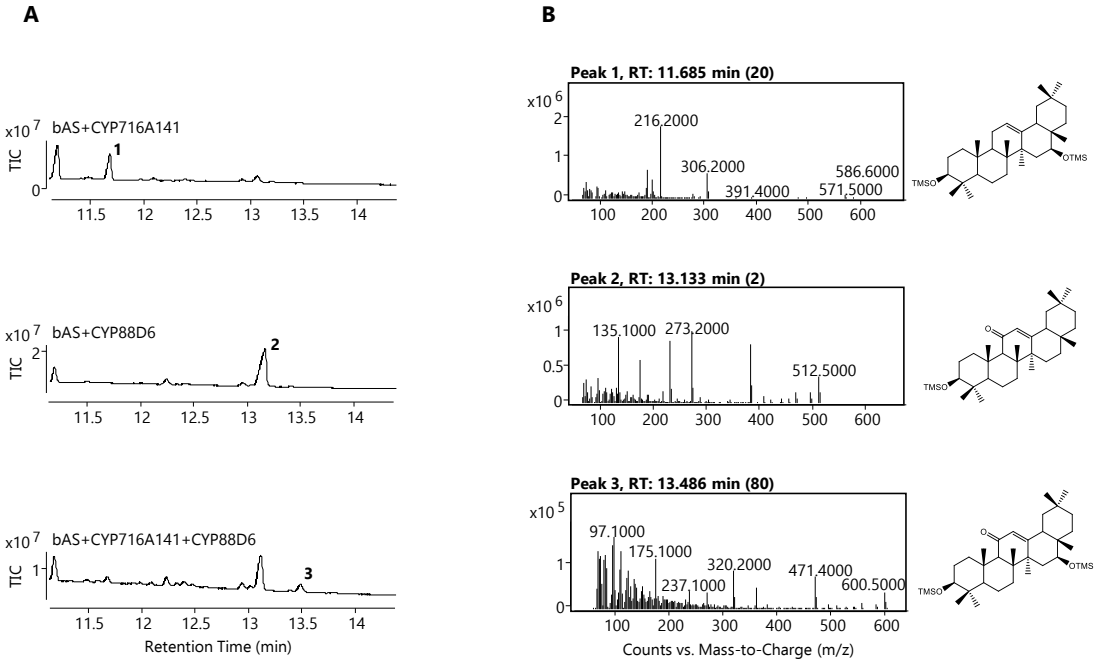


Figure A-91: CYP716A141+CYP88D6 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A141 only (top); CYP88D6 (middle); and CYP716A141+CYP88D6 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1), 13.1 (peak 2), and 13.5 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

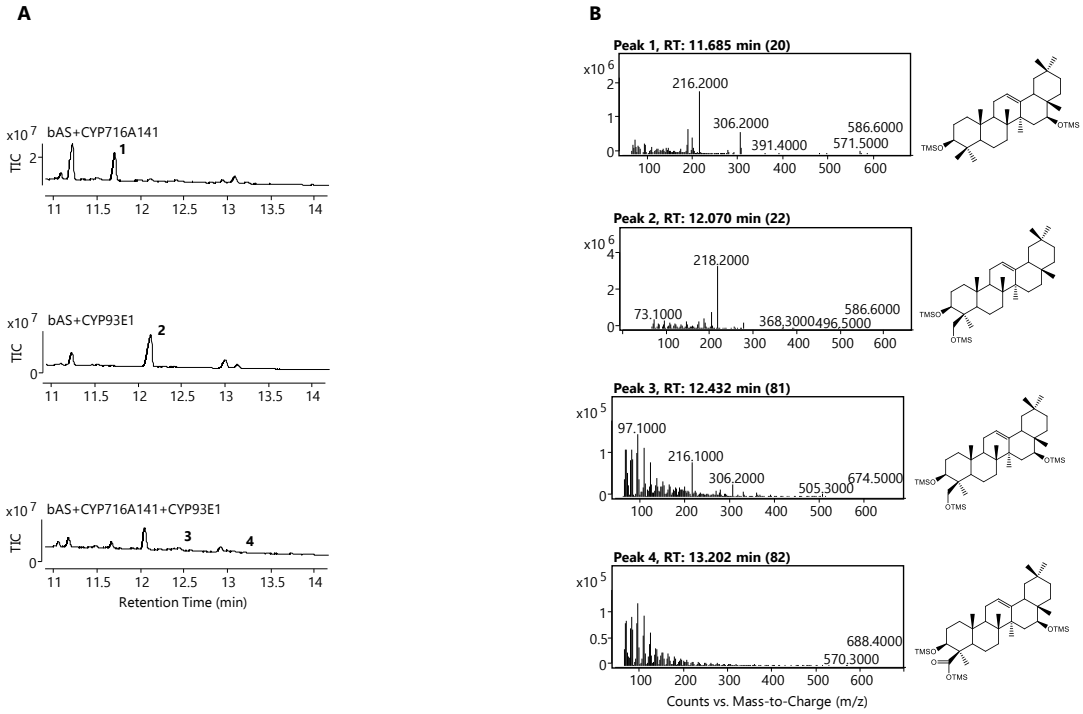


Figure A-92: CYP716A141+CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716A141 only (top); CYP93E1 (middle); and CYP716A141+CYP93E1 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1), 12.0 (peak 2), 12.4 (peak 3), and 13.2 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

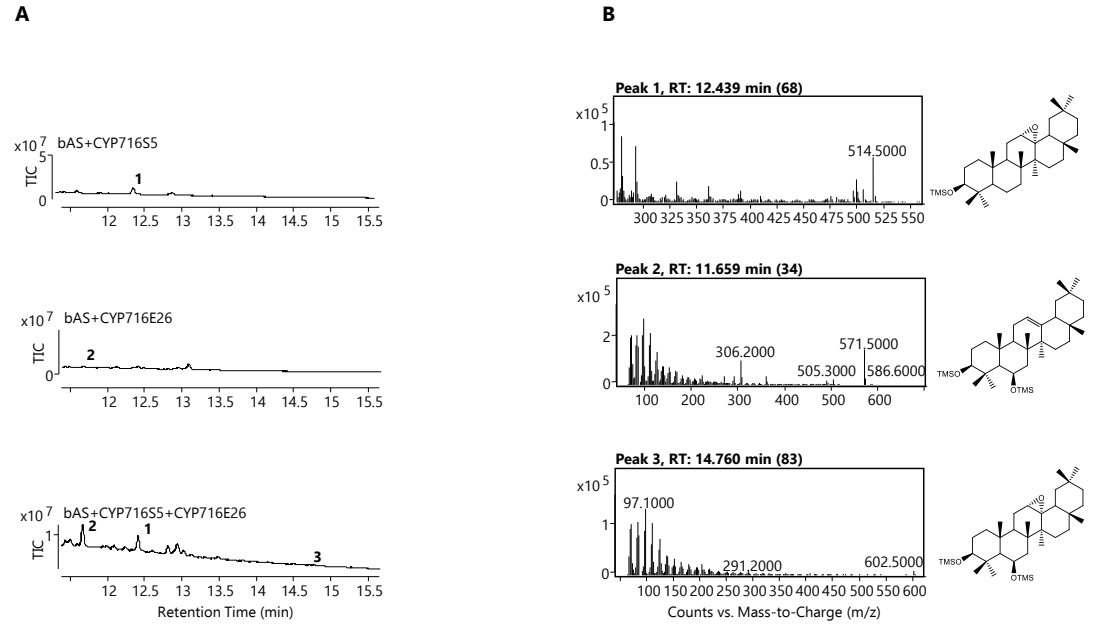


Figure A-93: CYP716S5+CYP716E26 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716S5 only (top); CYP716E26 (middle); and CYP716S5+CYP716E26 (bottom). B) Mass spectra for the peaks at 12.4 (peak 1), 11.7 (peak 2), and 14.8 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

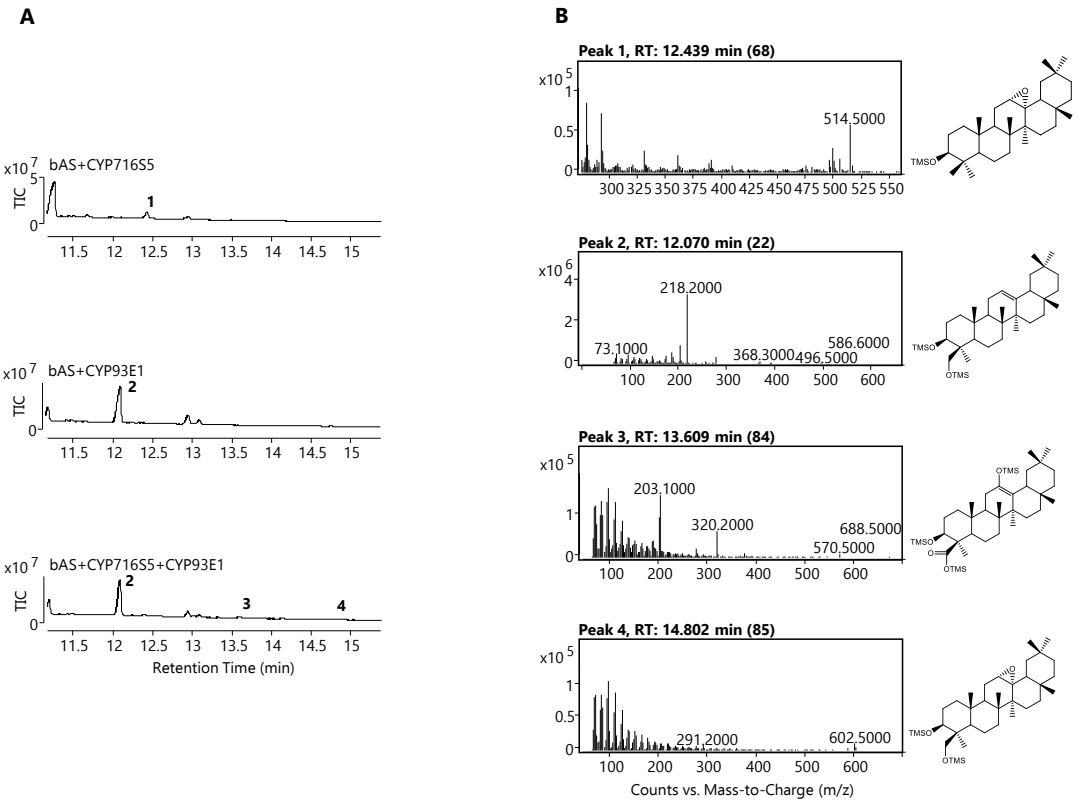


Figure A-94: CYP716S5+CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716S5 only (top); CYP93E1 (middle); and CYP716S5+CYP93E1 (bottom). B) Mass spectra for the peaks at 12.4 (peak 1), 13.6 (peak 2), and 14.8 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

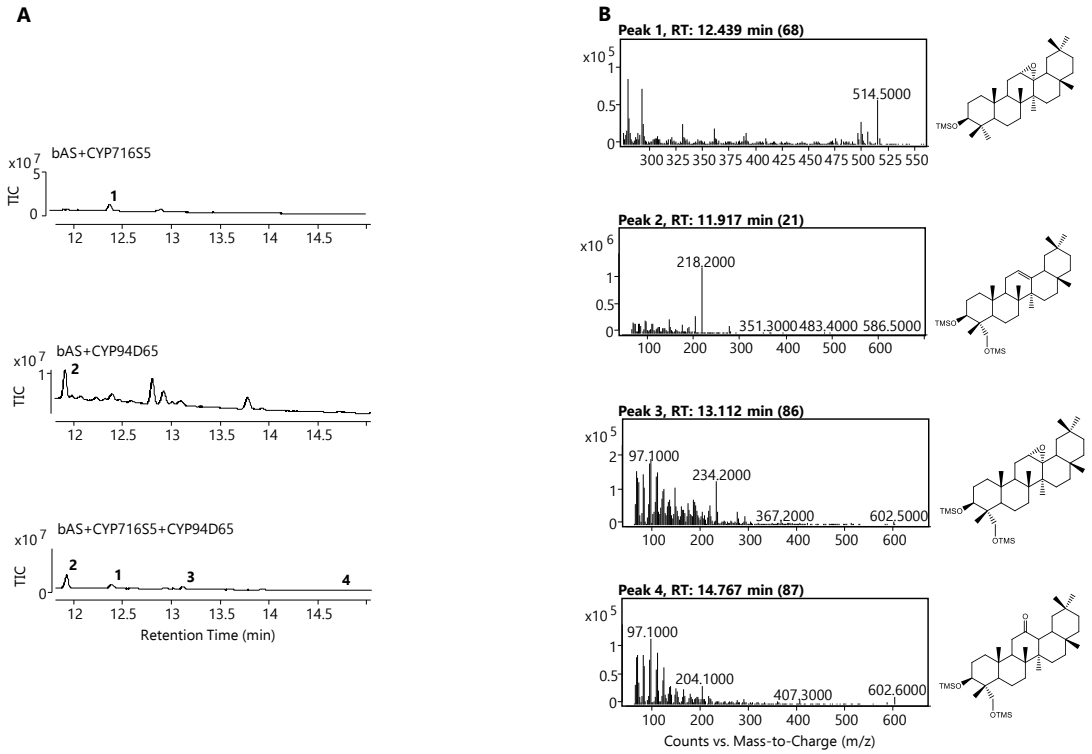


Figure A-95: CYP716S5+CYP94D65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716S5 only (top); CYP94D65 (middle); and CYP716S5+CYP94D65 (bottom). B) Mass spectra for the peaks at 12.4 (peak 1), 11.9 (peak 2), 13.1 (peak 3), and 14.8 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

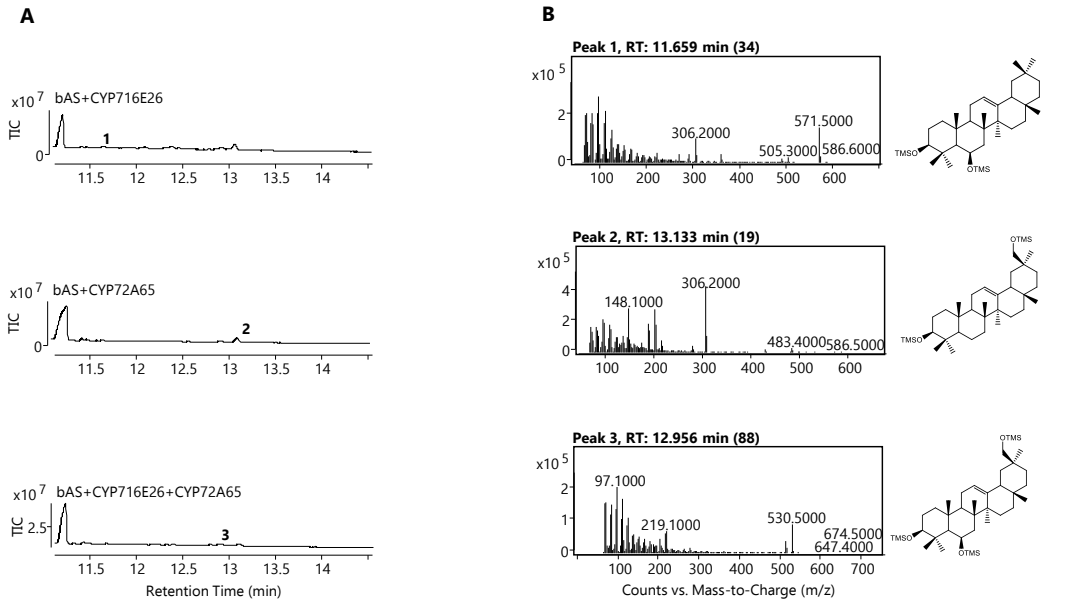


Figure A-96: CYP716E26+CYP72A65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716E26 only (top); CYP72A65 (middle); and CYP716E26+CYP72A65 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1), 13.1 (peak 2), and 13.0 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

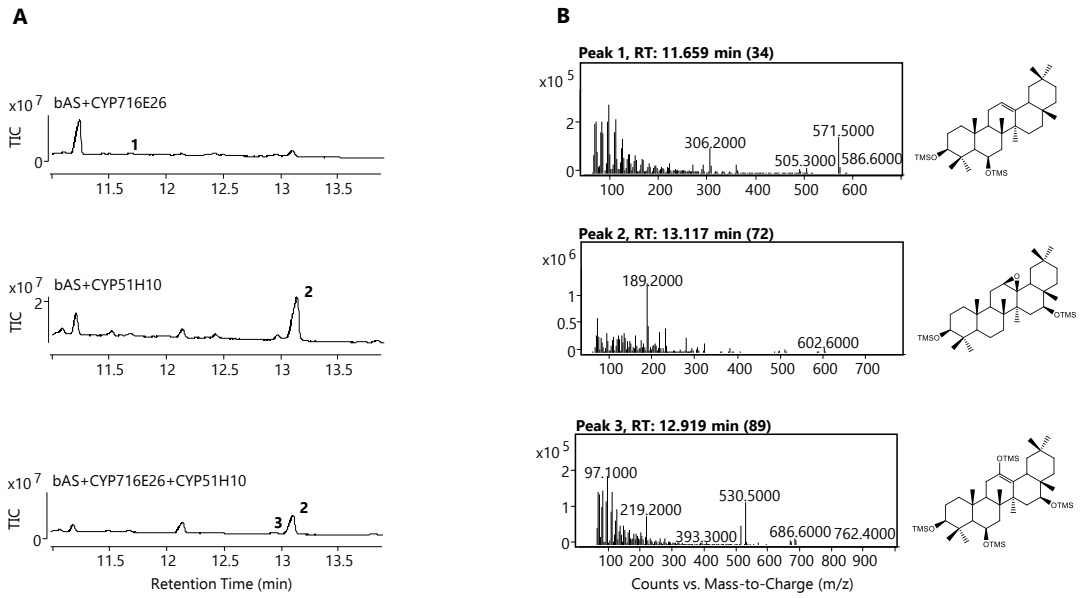


Figure A-97: CYP716E26+CYP51H10 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716E26 only (top); CYP51H10 (middle); and CYP716E26+CYP51H10 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1), 13.1 (peak 2), and 12.9 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

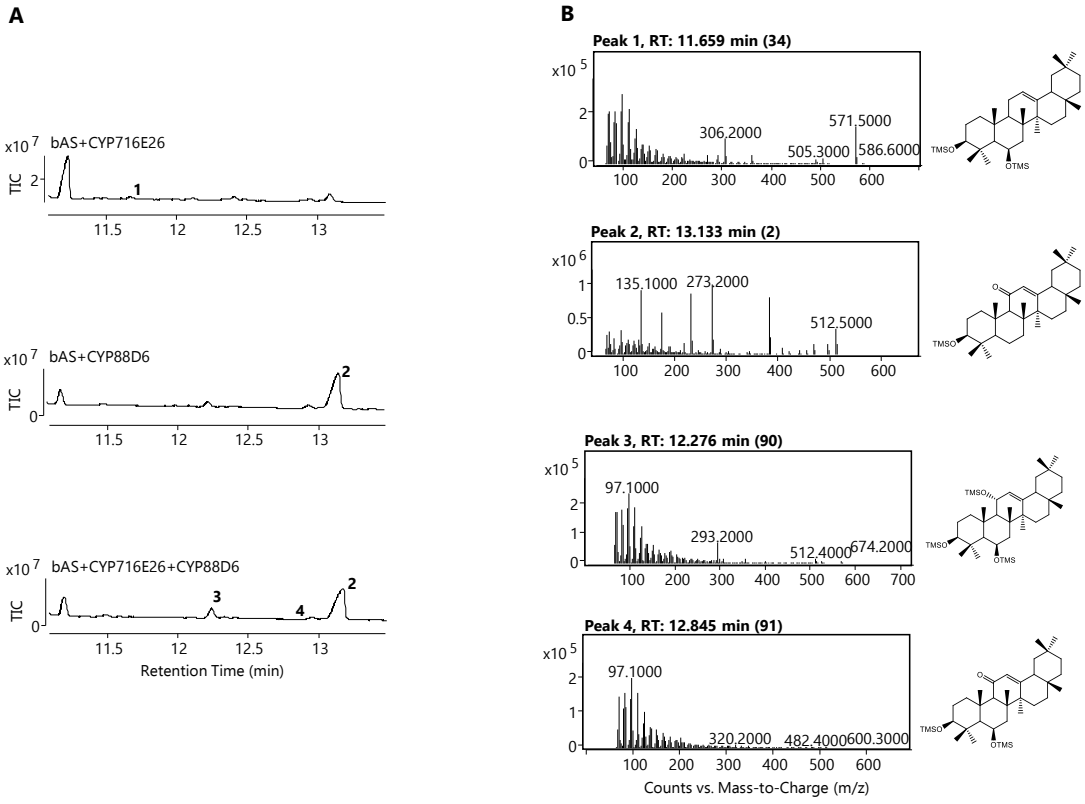


Figure A-98: CYP716E26+CYP88D6 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716E26 only (top); CYP88D6 (middle); and CYP716E26+CYP88D6 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1),

13.1 (peak 2), 12.3 (peak 3), and 12.8 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

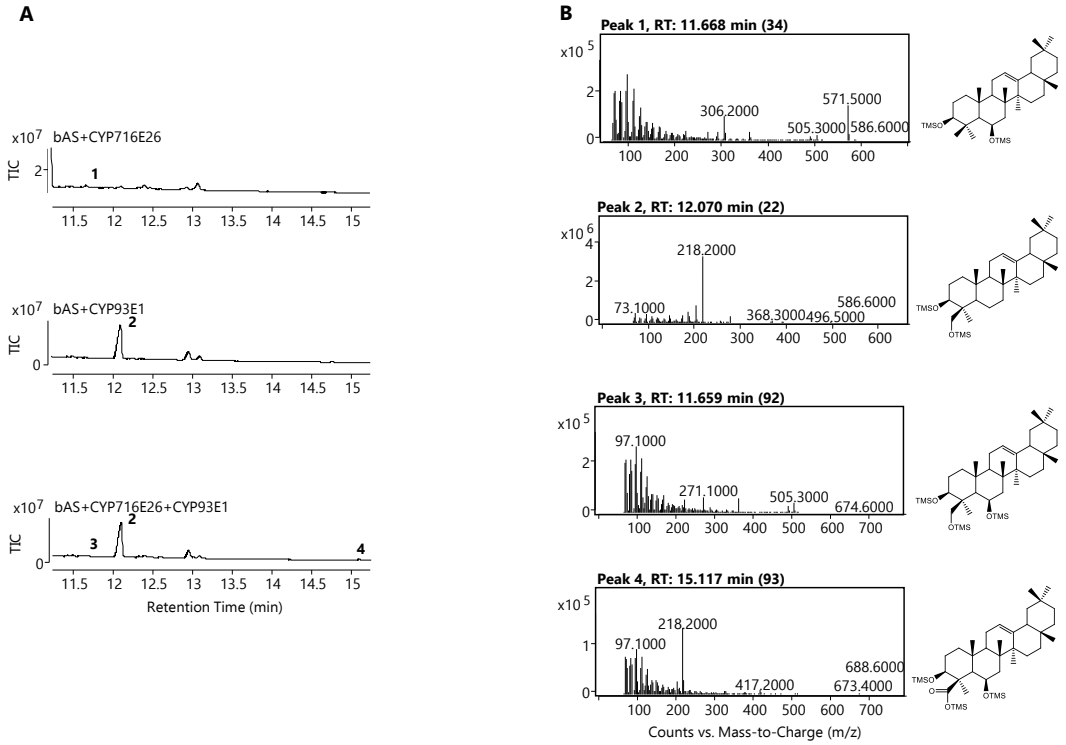


Figure A-99: CYP716E26+CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP716E26 only (top); CYP93E1 (middle); and CYP716E26+CYP93E1 (bottom). B) Mass spectra for the peaks at 11.7 (peak 1), 12.1 (peak 2), 11.7 (peak 3), and 15.1 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

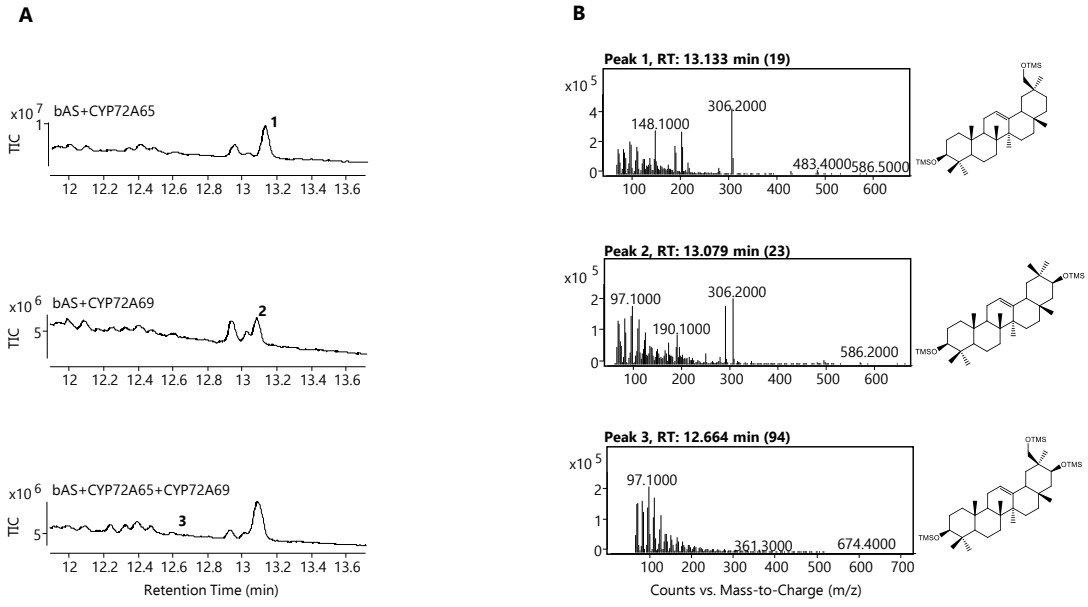


Figure A-100: CYP72A65 and CYP72A69 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP72A65 only (top); CYP72A69

(middle); and CYP72A65+CYP72A69 (bottom). B) Mass spectra for the peaks at 13.1 (peak 1), 13.1 (peak 2), and 12.6 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

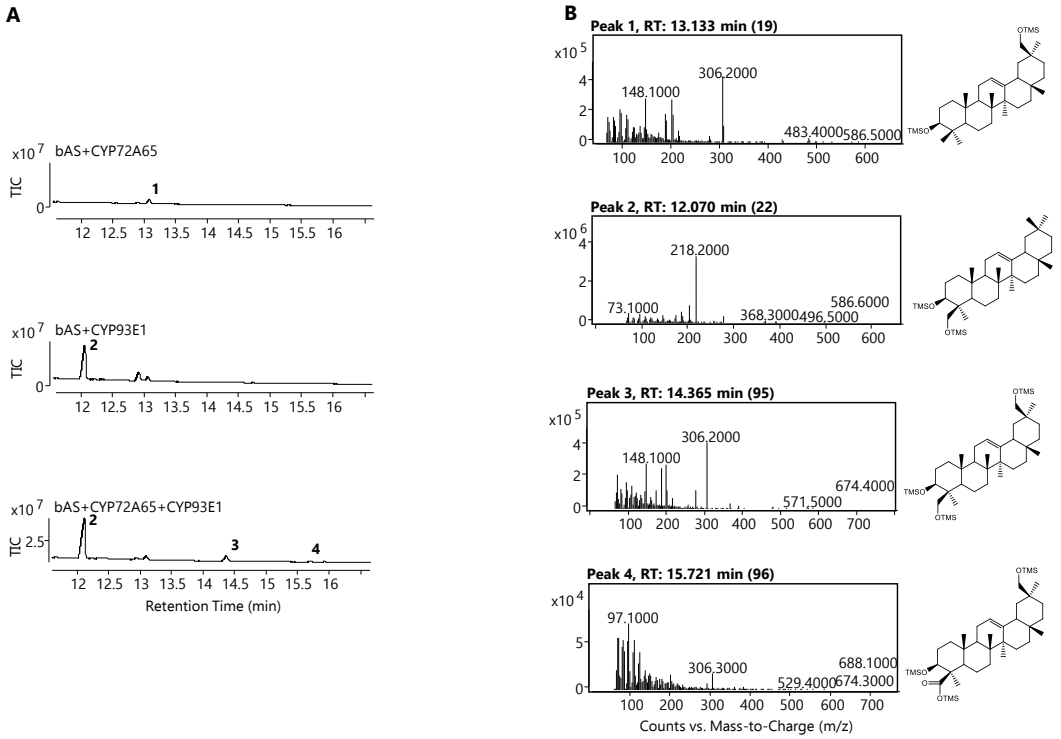


Figure A-101: CYP72A65 and CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP72A65 only (top); CYP93E1 (middle); and CYP72A65+CYP93E1 (bottom). B) Mass spectra for the peaks at 13.1 (peak 1), 12.1 (peak 2), 14.4 (peak 3), and 15.7 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

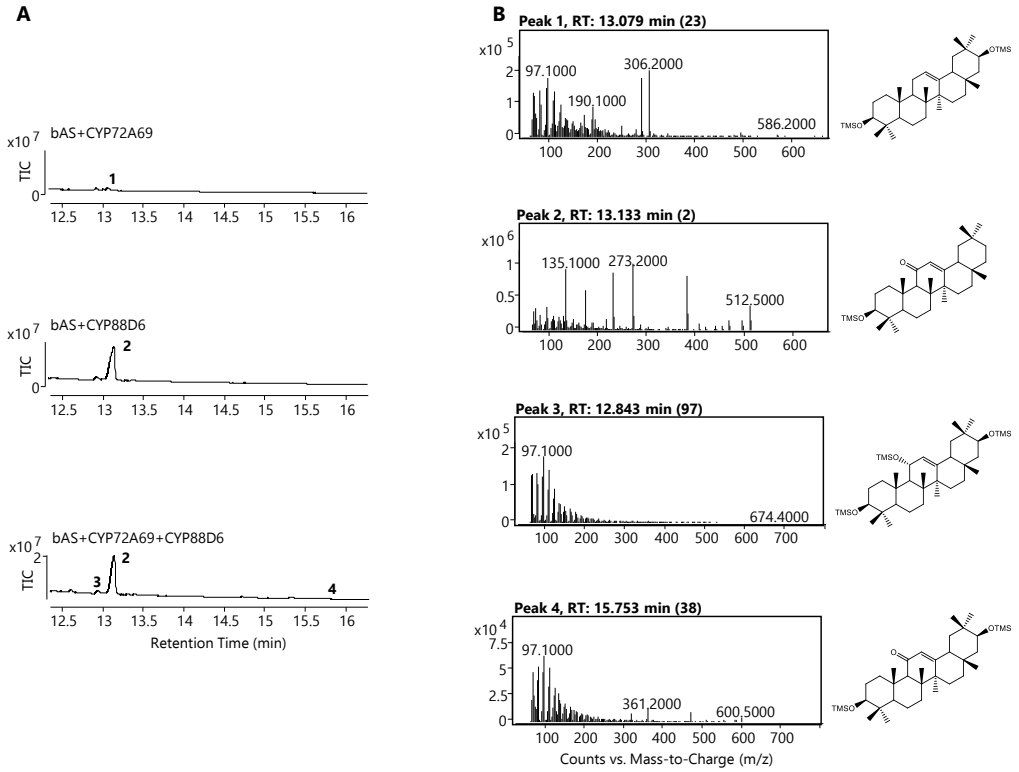


Figure A-102: CYP72A69 and CYP88D6 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP72A69 only (top); CYP88D6 (middle); and CYP72A69+CYP88D6 (bottom). B) Mass spectra for the peaks at 13.1 (peak 1), 13.1 (peak 2), 12.8 (peak 3), and 15.8 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

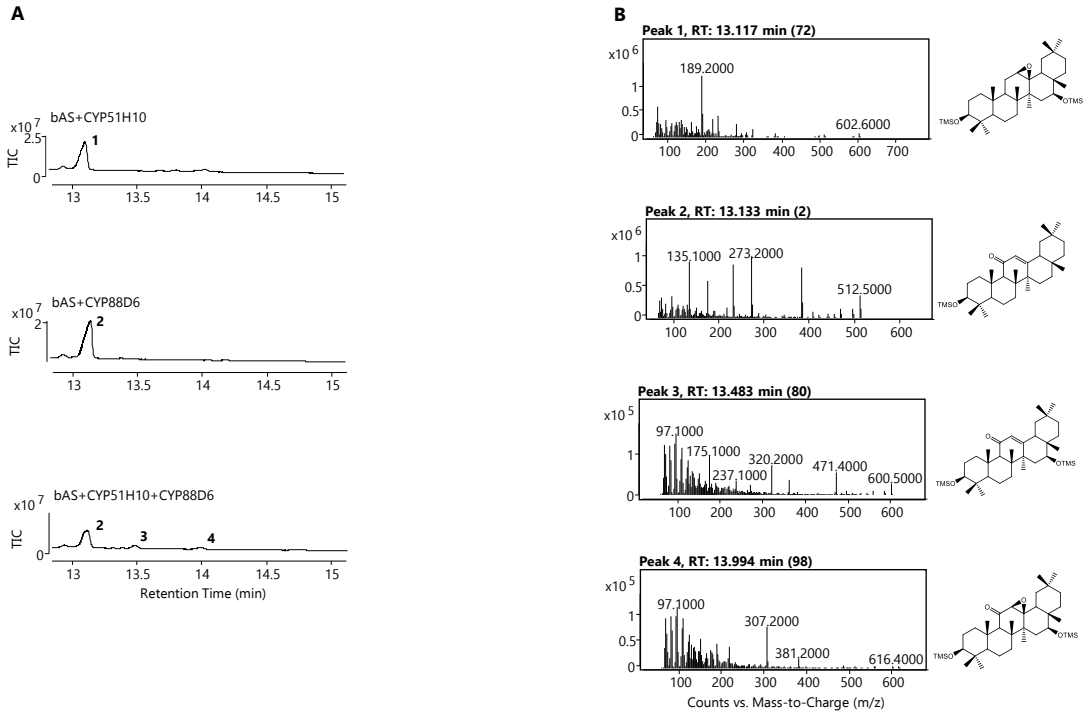


Figure A-103: CYP51H10 and CYP88D6 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP51H10 only (top); CYP88D6 (middle); and CYP51H10+CYP88D6 (bottom). B) Mass spectra for the peaks at 13.1 (peak 1), 13.1 (peak 2), 13.5 (peak 3), and 14.0 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

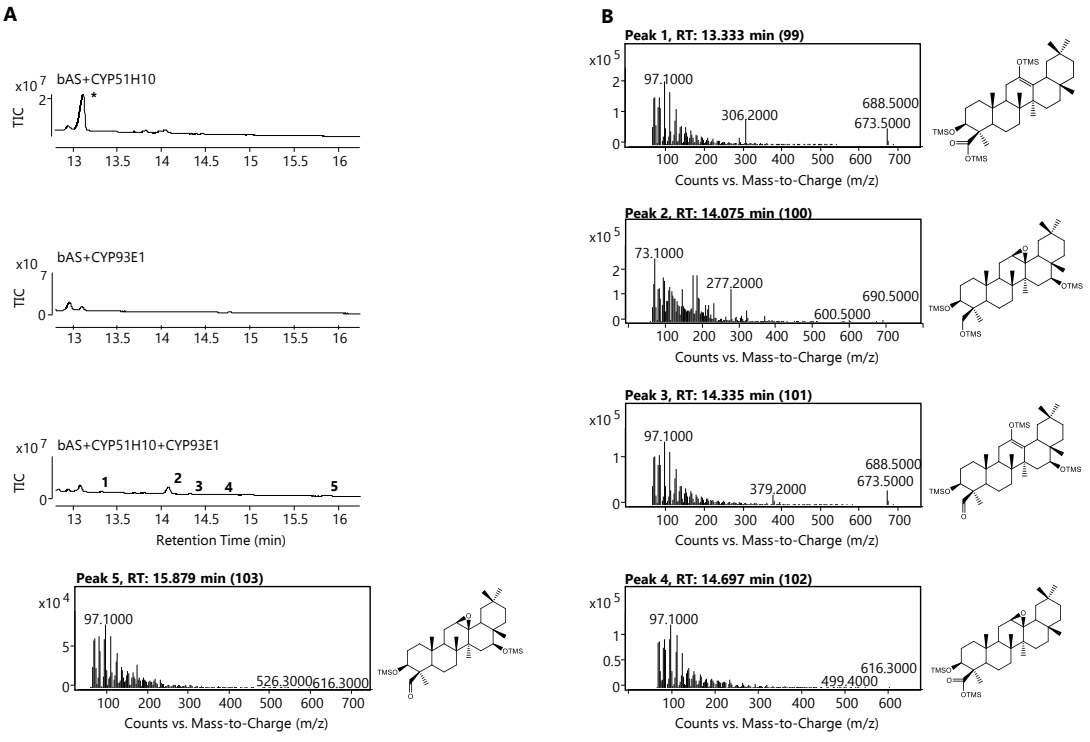


Figure A-104: CYP51H10 and CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP51H10 only (top); CYP93E1

(middle); and CYP51H10+CYP93E1 (bottom). B) Mass spectra for the peaks at 13.3 (peak 1), 14.1 (peak 2), 14.3 (peak 3), 14.7 (peak 4), and 15.9 (peak 5) minutes with predicted structures alongside. TIC = total ion chromatogram. * = 12,13-epoxy-16β-hydroxy-β-amyrin.

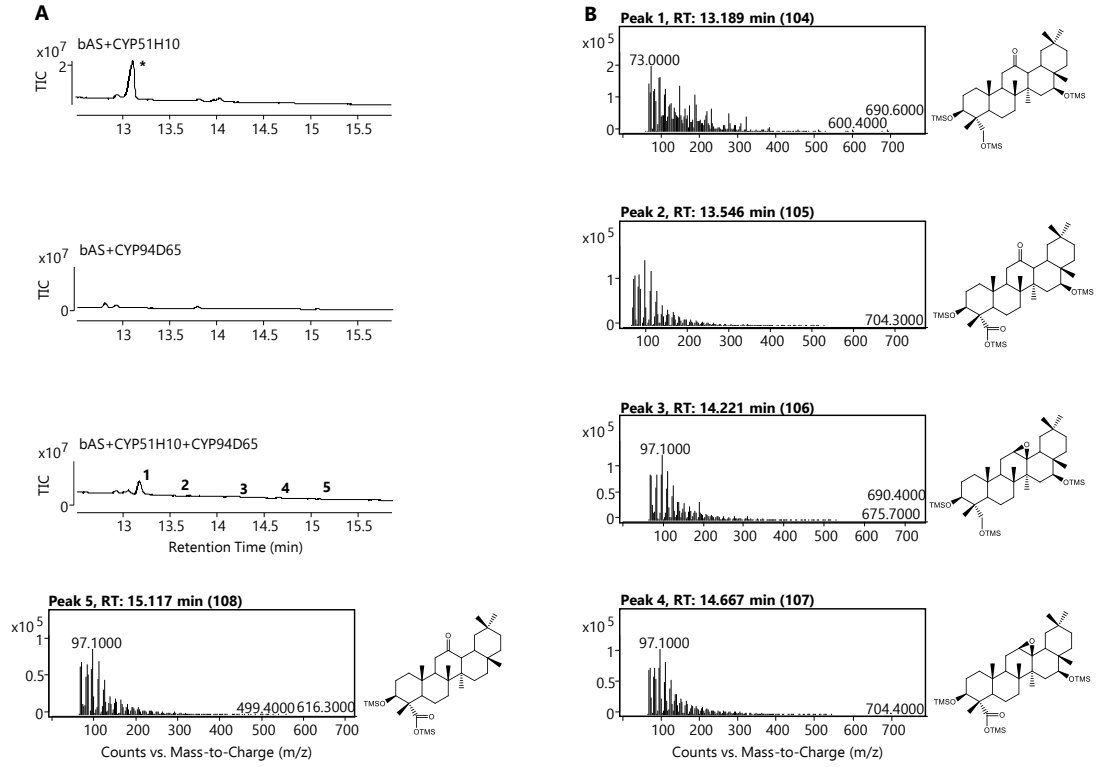


Figure A-105: CYP51H10 and CYP94D65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP51H10 only (top); CYP94D65 (middle); and CYP51H10+CYP94D65 (bottom). B) Mass spectra for the peaks at 13.2 (peak 1), 13.5 (peak 2), 14.2 (peak 3), 14.7 (peak 4), and 15.1 (peak 5) minutes with predicted structures alongside. TIC = total ion chromatogram. * = 12,13-epoxy-16β-hydroxy-β-amyrin.

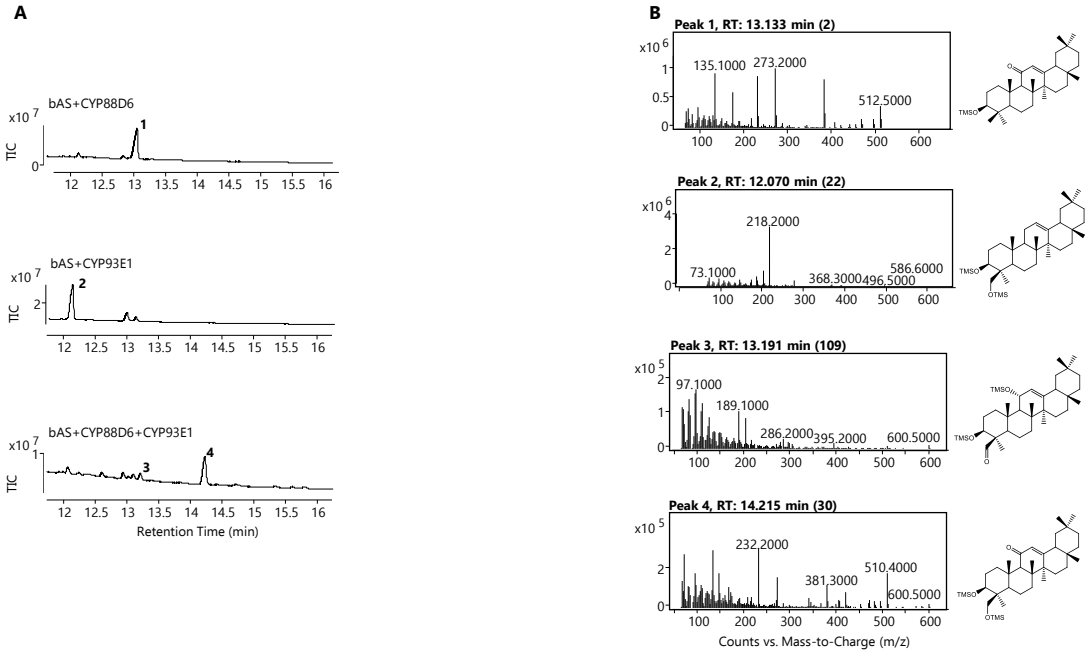


Figure A-106: CYP88D6 and CYP93E1 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP88D6 only (top); CYP93E1 (middle); and CYP88D6+CYP93E1 (bottom). B) Mass spectra for the peaks at 13.1 (peak 1), 12.1 (peak 2), 13.2 (peak 3), and 14.2 (peak 4) minutes with predicted structures alongside. TIC = total ion chromatogram.

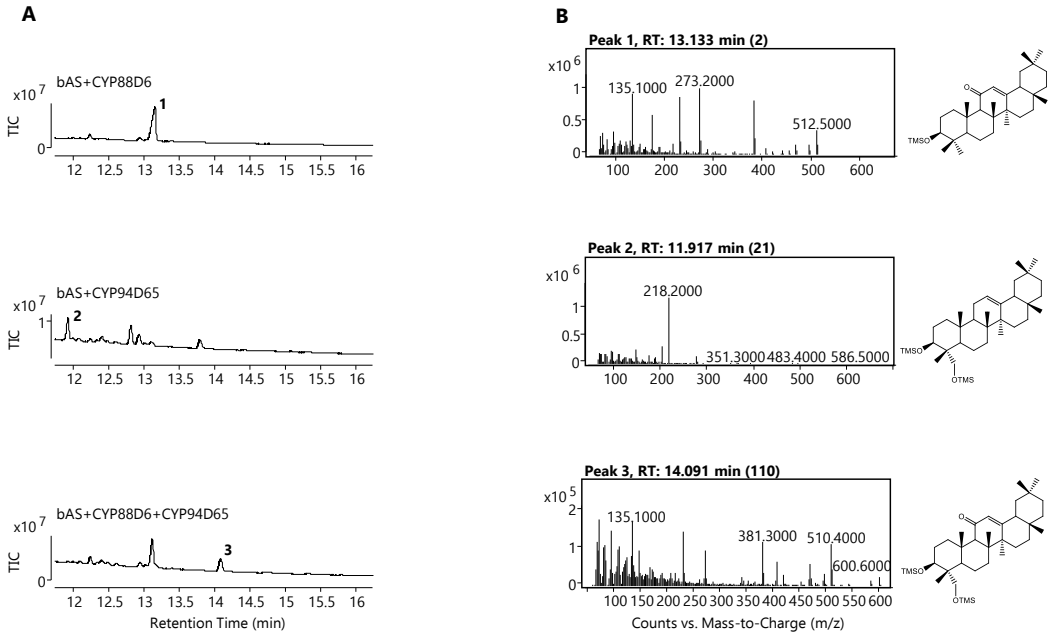


Figure A-107: CYP88D6 and CYP94D65 products. A) Products generated following co-expression of tHMGR, the β-amyrin synthase SAD1, and: CYP88D6 only (top); CYP94D65 (middle); and CYP88D6+CYP94D65 (bottom). B) Mass spectra for the peaks at 13.1 (peak 1), 11.9 (peak 2), and 14.1 (peak 3) minutes with predicted structures alongside. TIC = total ion chromatogram.

A.2.2 Chapter 4

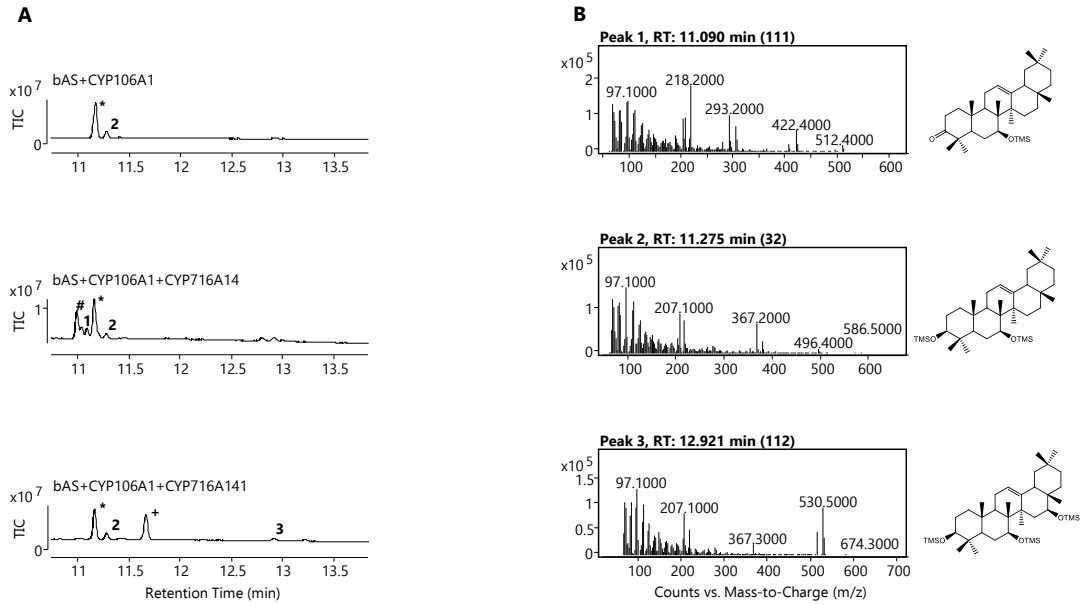


Figure A-108: CYP106A1 produces new products in co-expression with CYP716A14 or CYP716A141. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP106A1 (top); tHMGR, SAD1, CYP106A1, and CYP716A14 (middle); and tHMGR, SAD1, CYP106A1, and CYP716A141 (bottom). B) Mass spectra for the observed products at 11.1 (peak 1), 12.3 (peak 2), and 12.9 (peak 3) minutes with predicted products alongside. TIC = total ion chromatogram. * = β-amyrin, + = 16β-hydroxy-β-amyrin, # = β-amyrone.

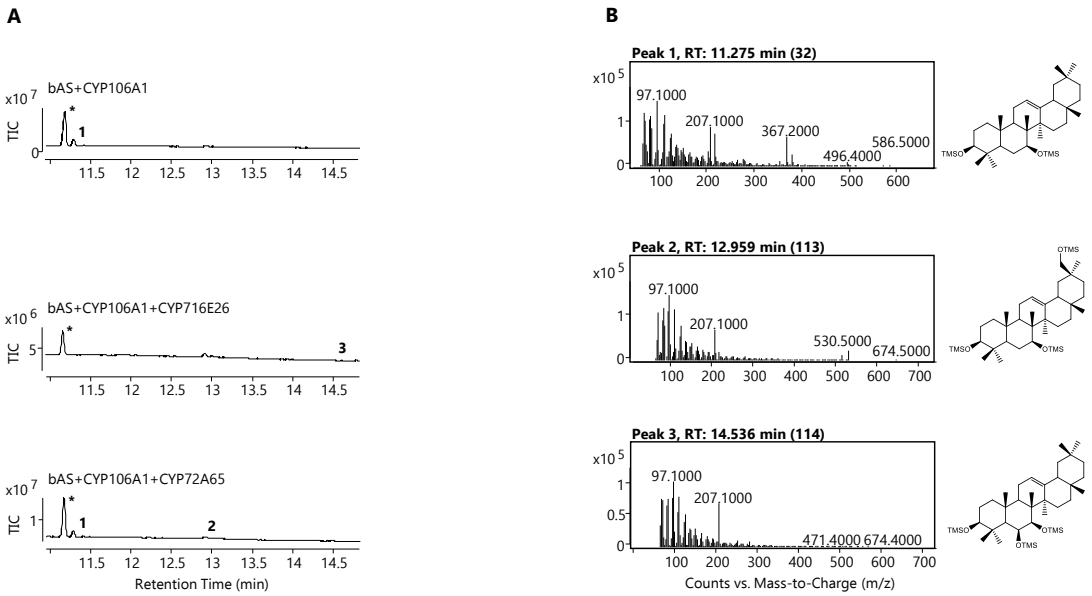


Figure A-109: CYP106A1 produces new products in co-expression with CYP716E26 or CYP72A65. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP106A1 (top); tHMGR, SAD1, CYP106A1, and CYP76E26 (middle); and tHMGR, SAD1, CYP106A1, and CYP72A65 (bottom). B) Mass spectra for the observed products at 11.3 (peak 1), 13.0 (peak 2), and 14.5 (peak 3) minutes with predicted products alongside. TIC = total ion chromatogram. * = β-amyrin.

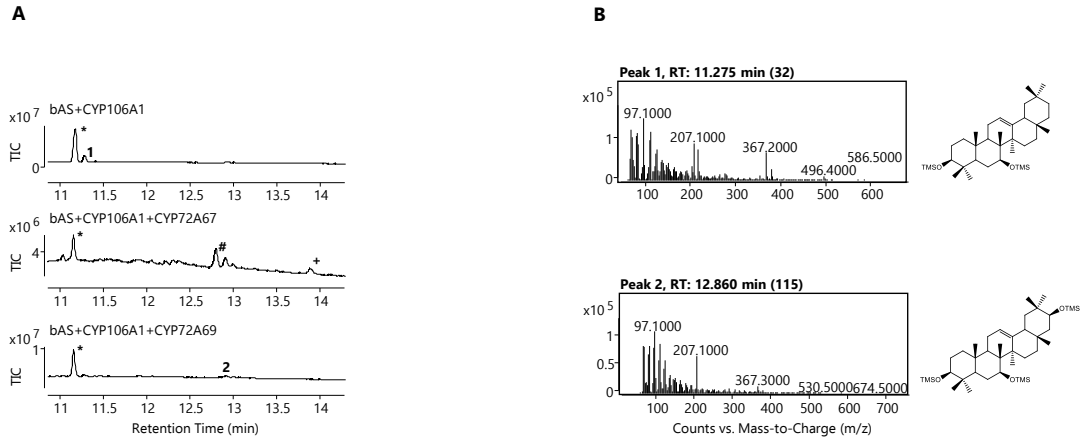


Figure A-110: CYP106A1 produces new products in co-expression with CYP716A12 and CYP72A67, or CYP72A69. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP106A1 (top); tHMGR, SAD1, CYP106A1, CYP716A12, and CYP72A67 (middle); and tHMGR, SAD1, CYP106A1, and CYP72A69 (bottom). B) Mass spectra for the observed products at 11.3 (peak 1), and 12.9 (peak 2) minutes with predicted products alongside. TIC = total ion chromatogram. * = β-amyrin, + = augustic acid, # = oleanolic acid.

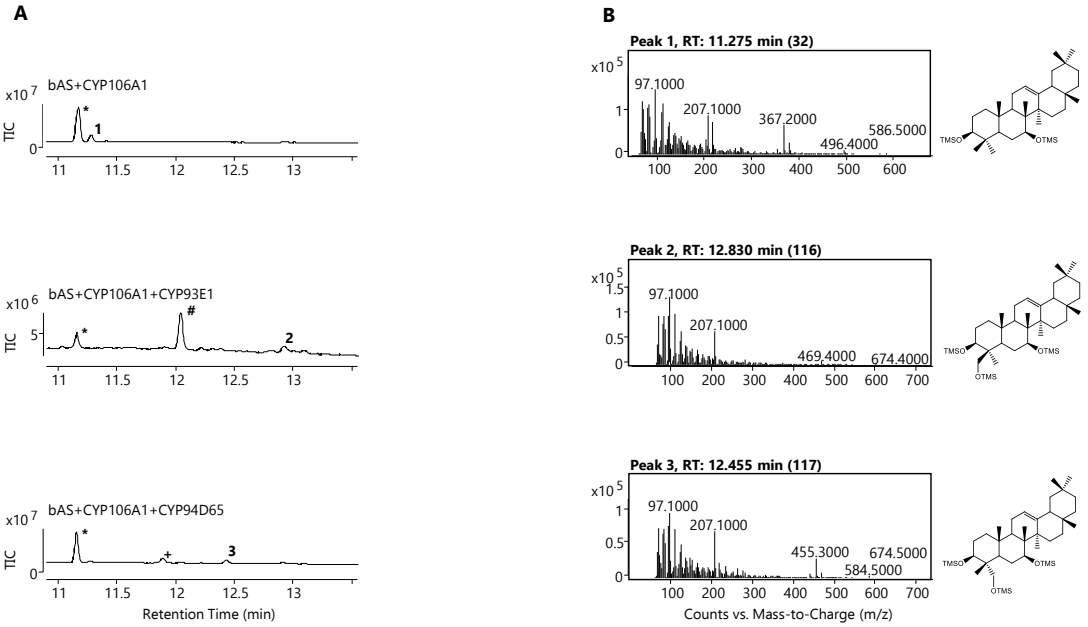


Figure A-111: CYP106A1 produces new products in co-expression with CYP93E1 or CYP94D65. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP106A1 (top); tHMGR, SAD1, CYP106A1, and CYP93E1 (middle); and tHMGR, SAD1, CYP106A1, and CYP94D65 (bottom). B) Mass spectra for the observed products at 11.3 (peak 1), 12.8 (peak 2), and 12.5 (peak 3) minutes with predicted products alongside. TIC = total ion chromatogram. * = β-amyrin, + = 23-hydroxy-β-amyrin, # = 24-hydroxy-β-amyrin.

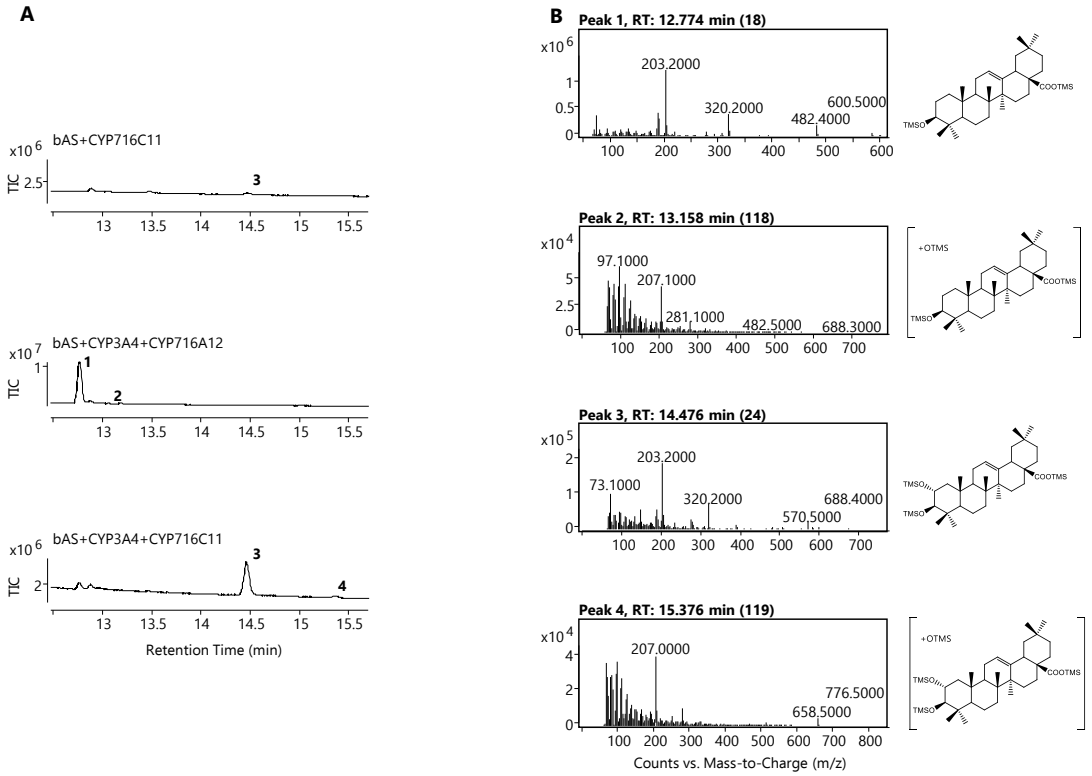


Figure A-112: CYP3A4 produces new products in co-expression with CYP716A12 and CYP716C11. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), CYP716A12, and CYP716C11 (top); tHMGR, SAD1, CYP716A12, and CYP3A4 (middle); and tHMGR, SAD1, CYP716A12, CYP716C11, and CYP3A4 (bottom). B) Mass spectra for the observed products at 12.8 (peak 1), 13.2 (peak 2), 14.5 (peak 3), and 15.4 (peak 4) minutes with predicted products alongside. TIC = total ion chromatogram.

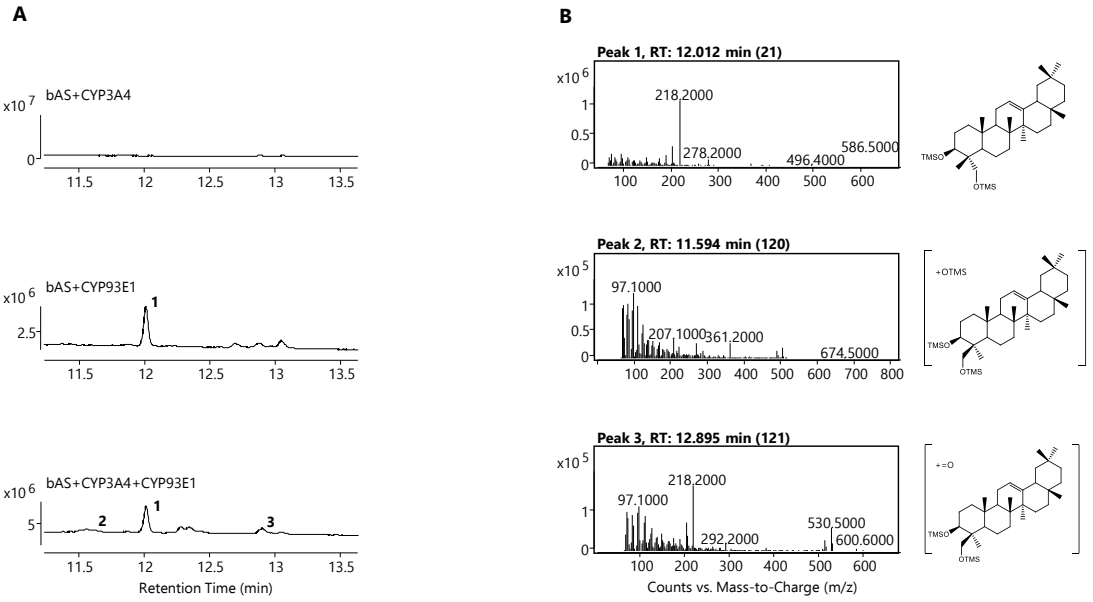


Figure A-113: CYP3A4 produces new products in co-expression with CYP93E1. A) GC-MS traces showing combinations of tHMGR, β-amyrin synthase (SAD1), and CYP3A4 (top); tHMGR, SAD1, and CYP93E1 (middle); and tHMGR, SAD1, CYP93E1, and CYP3A4 (bottom). B)

Mass spectra for the observed products at 12.0 (peak 1), 11.6 (peak 2), and 12.9 (peak 3) minutes with predicted products alongside. TIC = total ion chromatogram.

A.3 Additional LC-MS Spectra

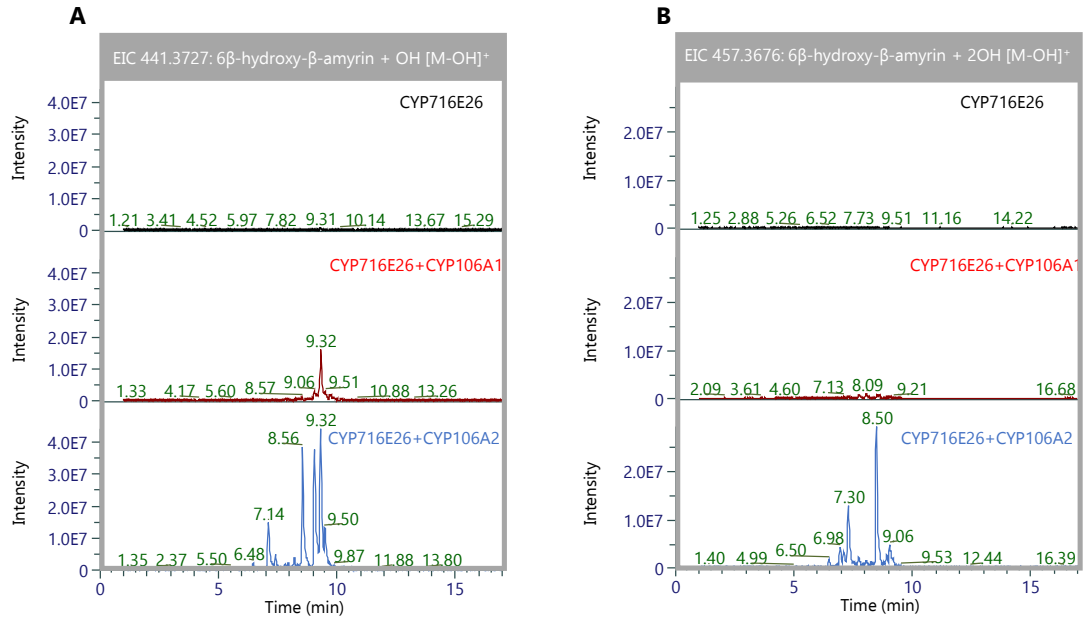


Figure A-114: CYP106A1 and CYP716E26 or CYP106A2 and CYP716E26 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 6β-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 457.3676, corresponding to 6β-hydroxy-β-amyrin plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP716E26 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

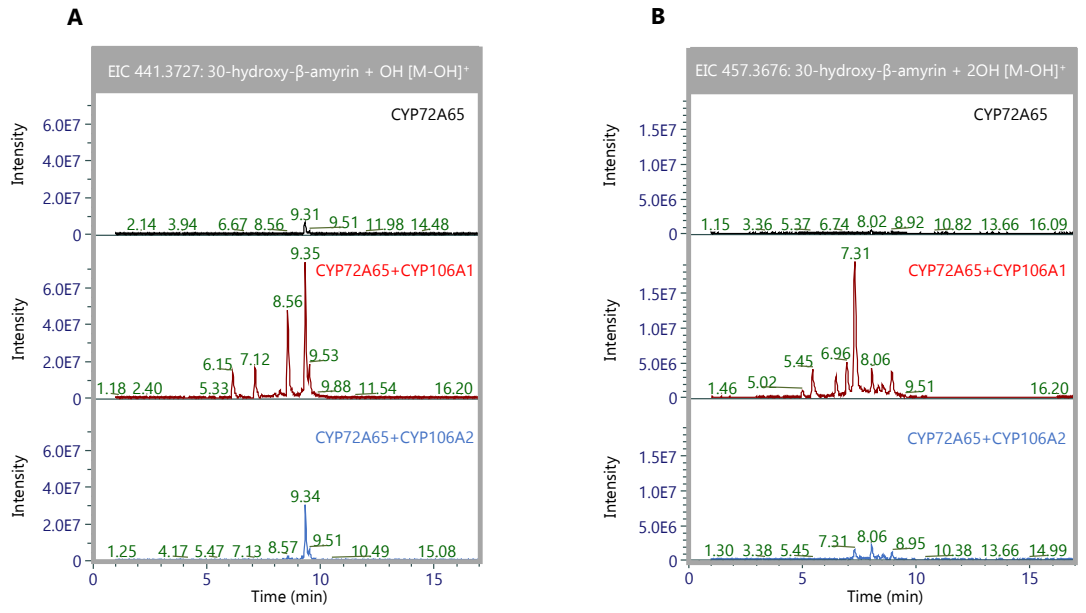


Figure A-115: CYP106A1 and CYP72A65 or CYP106A2 and CYP72A65 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 30-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 457.3676, corresponding to 30-hydroxy-β-amyrin plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts

expressing tHMGR, β-amyrin synthase (SAD1), and CYP72A65 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

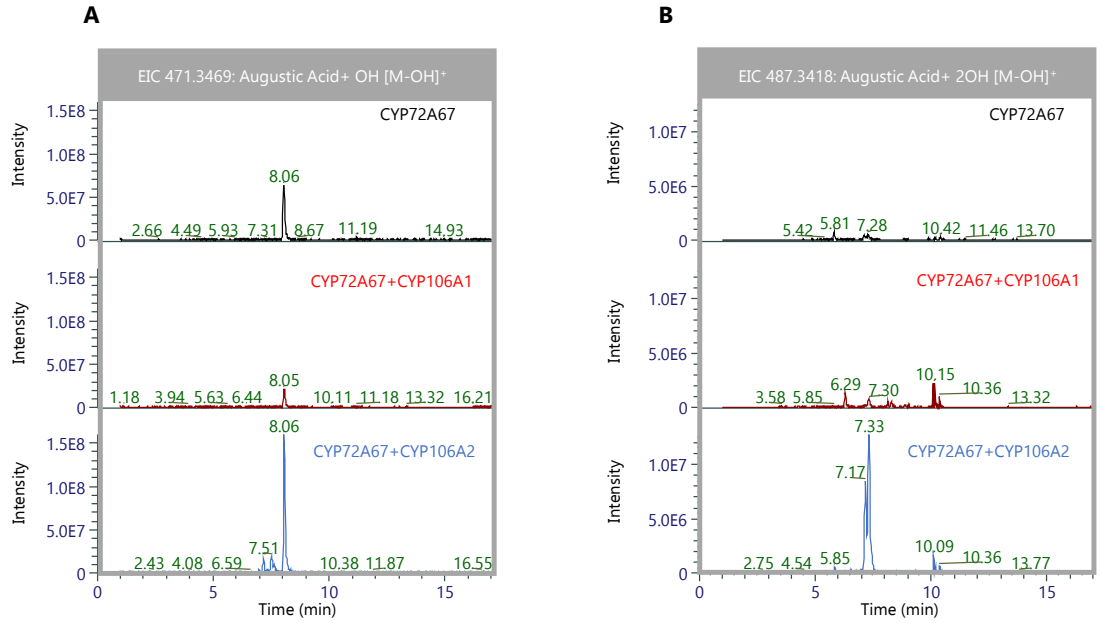


Figure A-116: CYP106A1, CYP716A12, and CYP72A67 or CYP106A2, CYP716A12, and CYP72A67 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 471.3469, corresponding to hydroxylated augustic acid $[M\text{-}OH]^+$, or B) *m/z* = 487.3418, corresponding to augustic acid plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), CYP716A12, and CYP72A67 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

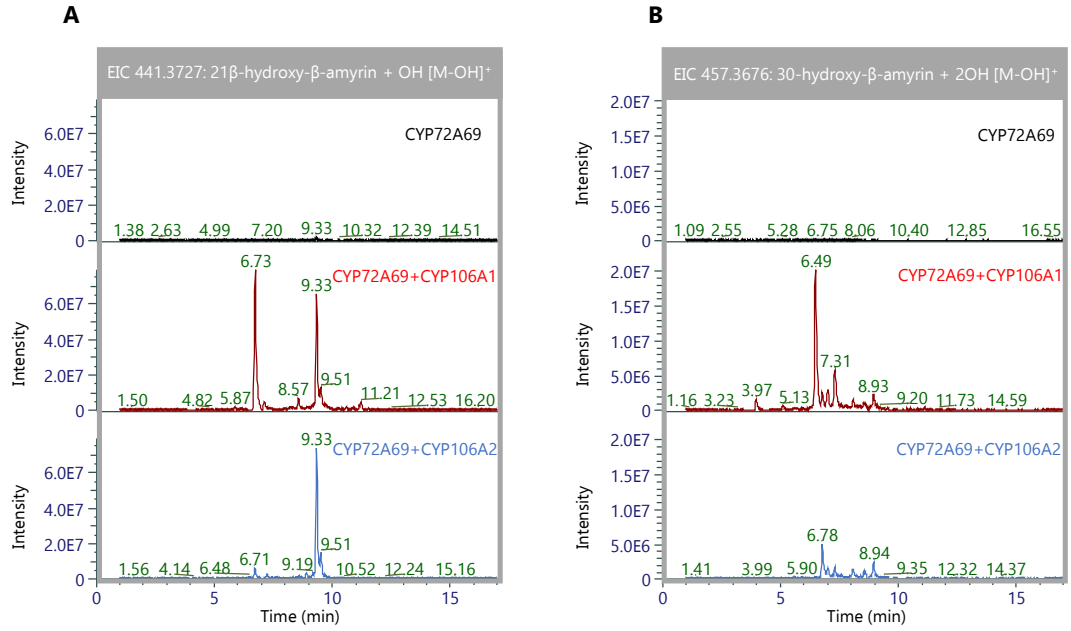


Figure A-117: CYP106A1 and CYP72A69 or CYP106A2 and CYP72A69 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding

to hydroxylated 21β-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 457.3676, corresponding to 21β-hydroxy-β-amyrin plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP72A69 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

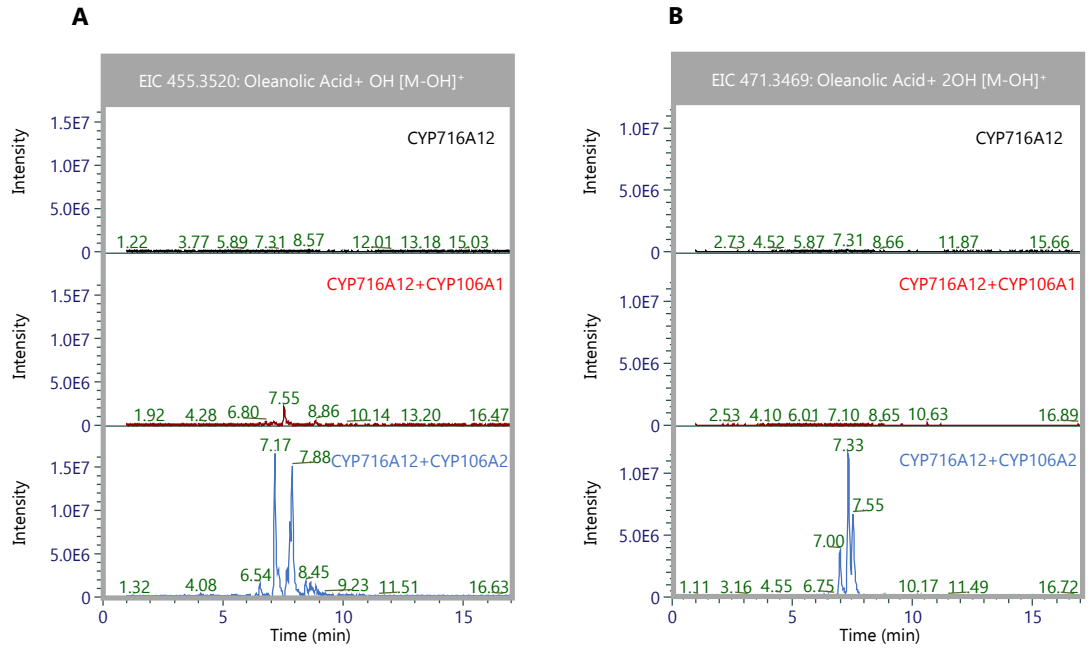


Figure A-118: CYP106A1 and CYP716A12 or CYP106A2 and CYP716A12 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 455.3520, corresponding to hydroxylated oleanolic acid $[M\text{-}OH]^+$, or B) *m/z* = 471.3469, corresponding to oleanolic acid plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP716A12 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

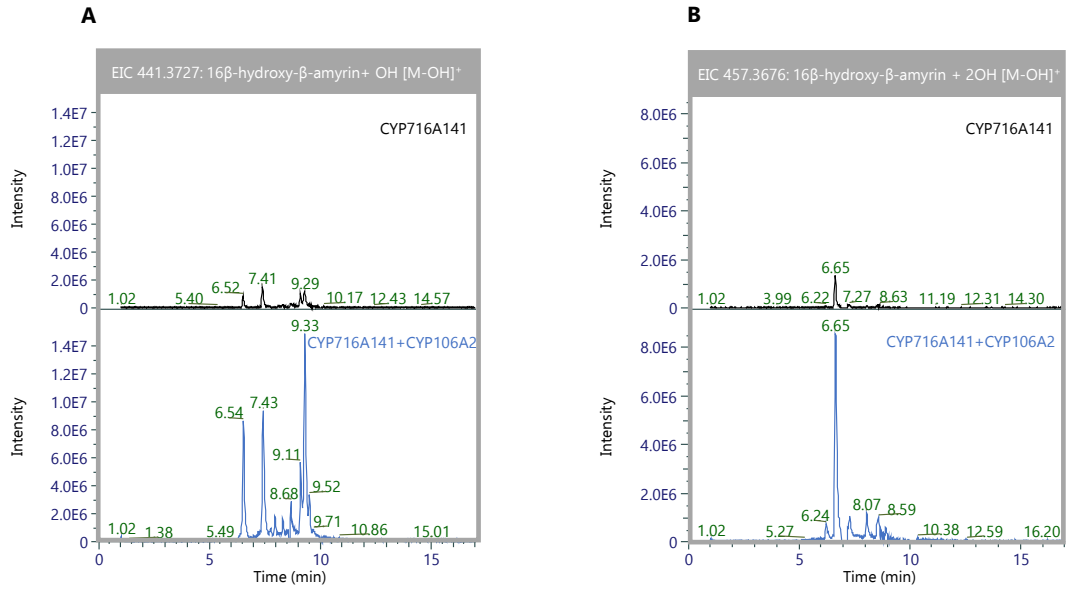


Figure A-119: CYP106A2 and CYP716A141 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated

16β-hydroxy-β-amyrin [M-OH]$^+$, or B) *m/z* = 457.3676, corresponding to 16β-hydroxy-β-amyrin plus two hydroxyl groups [M-OH]$^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP716A141 as a control (black), and with addition of CYP106A2 (blue).

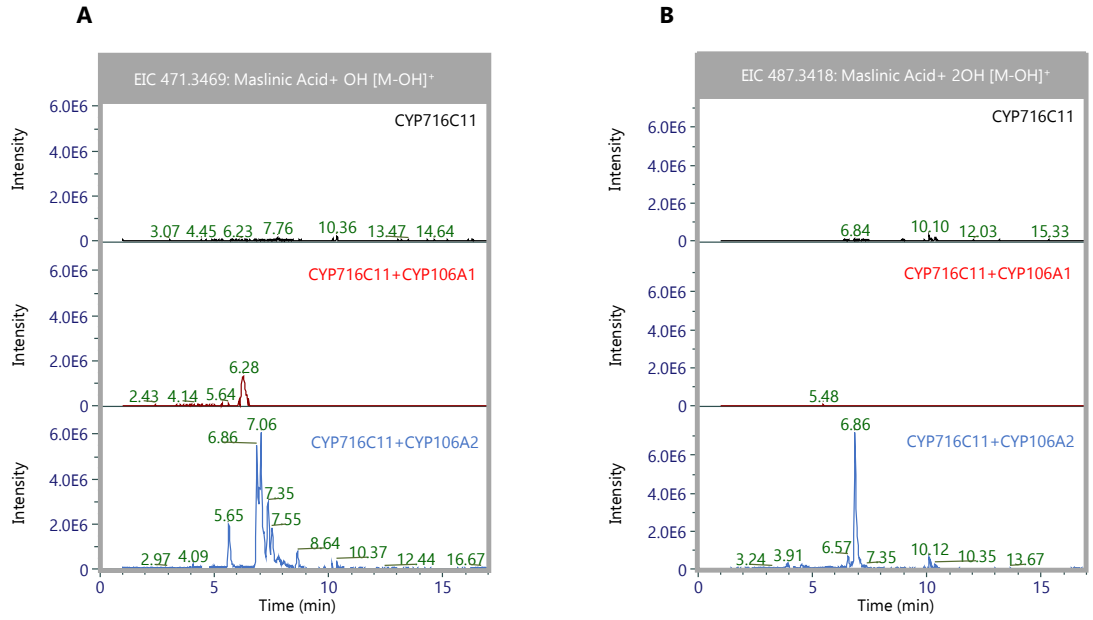


Figure A-120: CYP106A1, CYP716A12, and CYP716C11 or CYP106A2, CYP716A12, and CYP716C11 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 471.3469, corresponding to hydroxylated maslinic acid [M-OH]$^+$, or B) *m/z* = 487.3418, corresponding to maslinic acid plus two hydroxyl groups [M-OH]$^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), CYP716A12, and CYP716C11 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

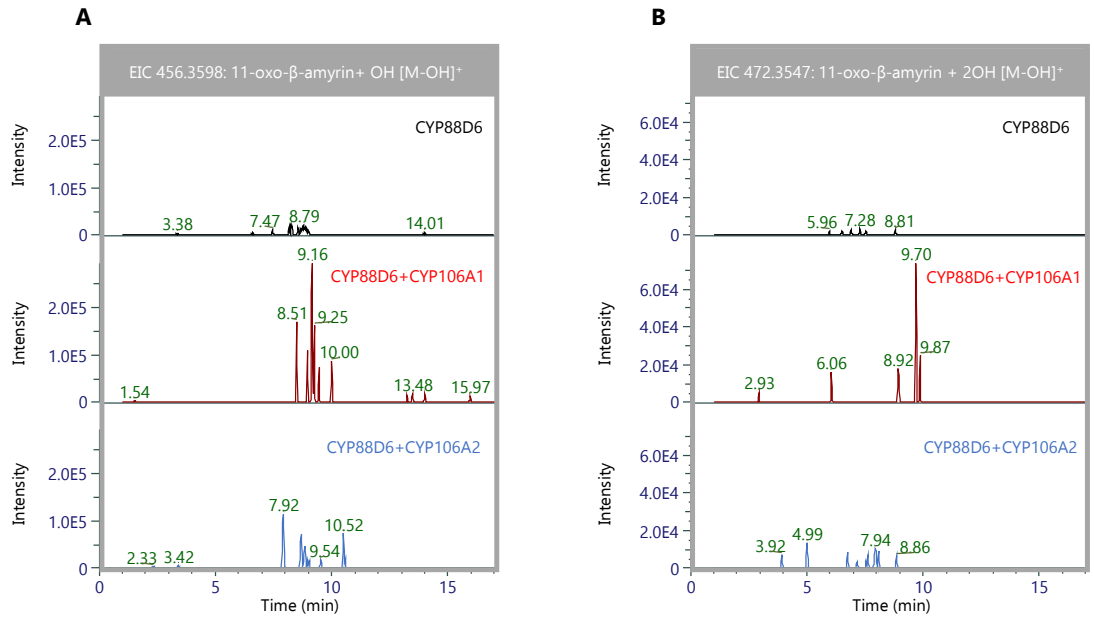


Figure A-121: CYP106A1 and CYP88D6 or CYP106A2 and CYP88D6 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 456.3598, corresponding to hydroxylated 11-oxo-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 472.3547, corresponding to 11-oxo-β-amyrin plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), CYP716A12, and CYP88D6 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

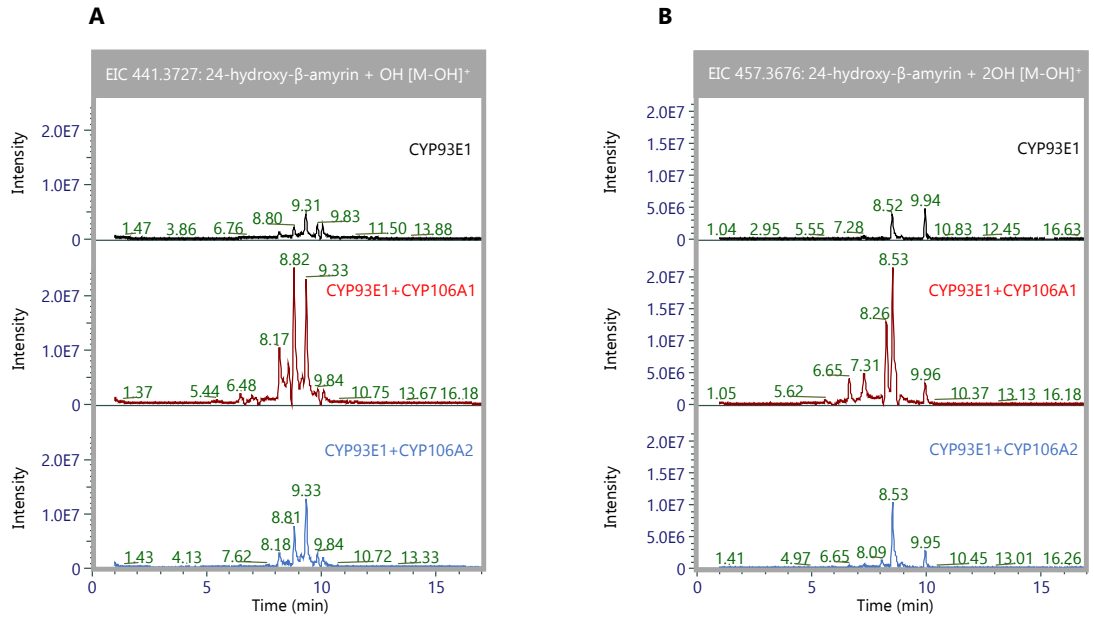


Figure A-122: CYP106A1 and CYP93E1 or CYP106A2 and CYP93E1 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 24-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 457.3676, corresponding to 24-hydroxy-β-amyrin plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP93E1 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

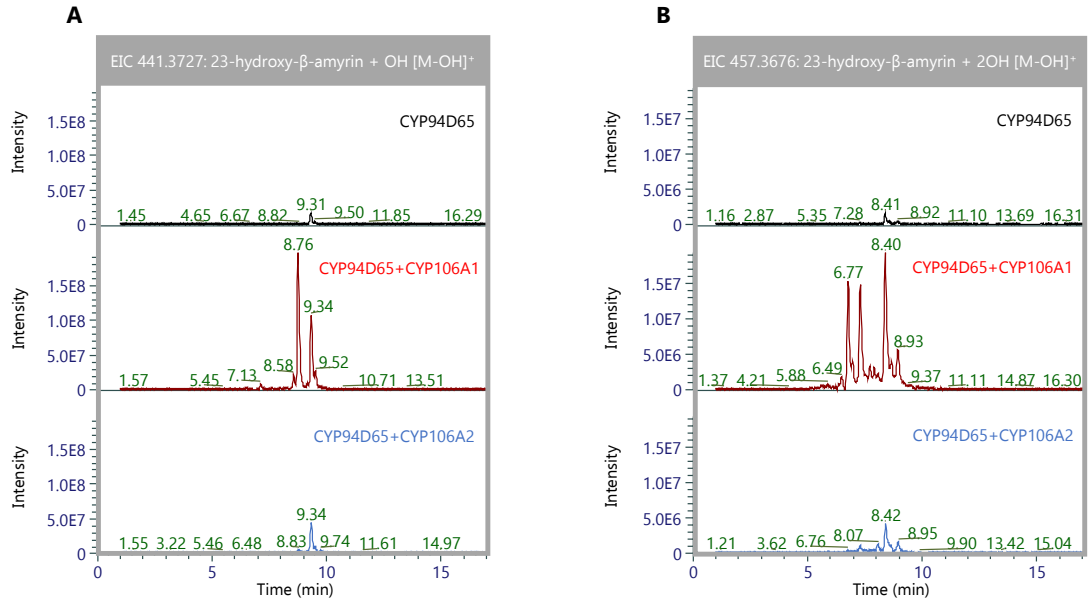


Figure A-123: CYP106A1 and CYP94D65 or CYP106A2 and CYP94D65 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 23-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 457.3676, corresponding to 23-hydroxy-β-amyrin plus two hydroxyl groups $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP94D65 as a control (black), and with addition of either CYP106A1 (red) or CYP106A2 (blue).

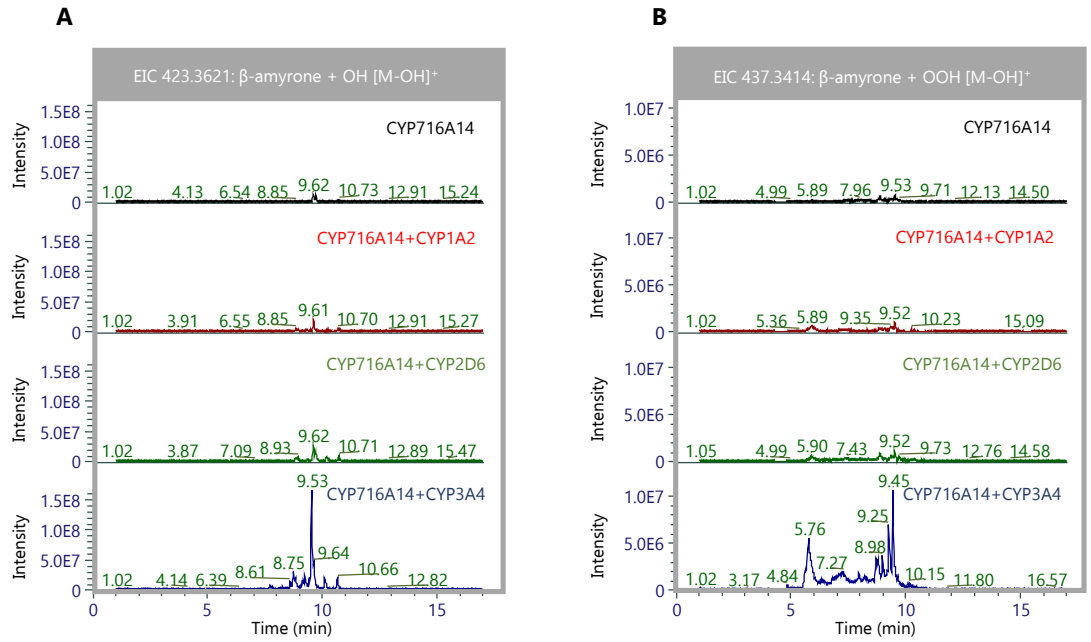


Figure A-124: CYP1A2 and CYP716A14, CYP2D6 and CYP716A14, or CYP3A4 and CYP716A14 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 423.3621, corresponding to hydroxylated β-amyrone $[M\text{-}OH]^+$, or B) *m/z* = 437.3414, corresponding to β-amyrone plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP716A14 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

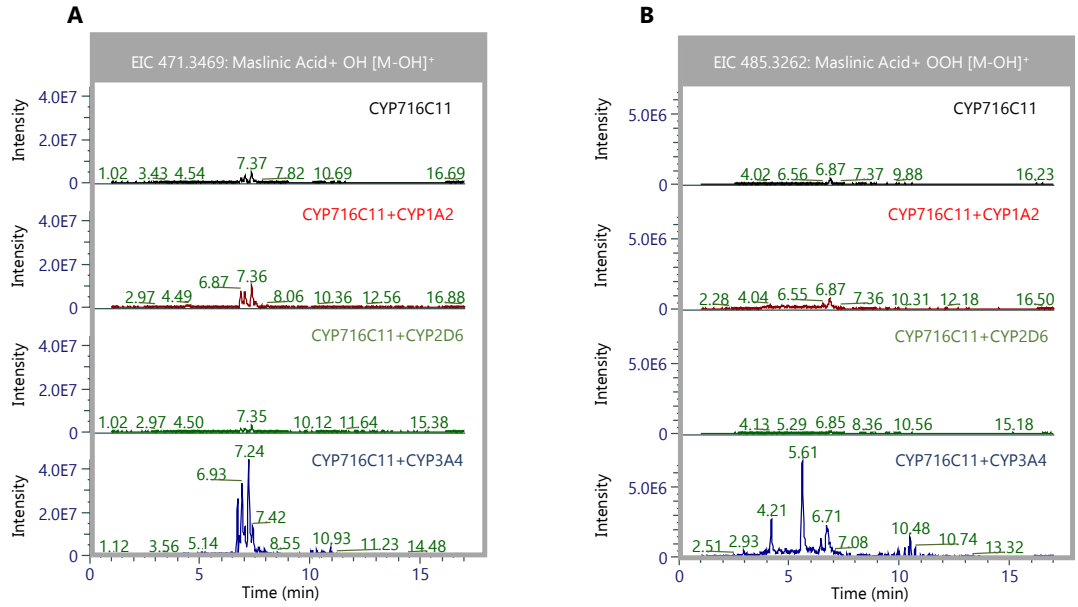


Figure A-125: CYP1A2, CYP716A12, and CYP716C11, CYP2D6, CYP716A12, and CYP716C11, or CYP3A4, CYP716A12, and CYP716C11 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 471.3469, corresponding to hydroxylated maslinic acid $[M\text{-}OH]^+$, or B) *m/z* = 485.3262, corresponding to maslinic acid plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), CYP716A12, and CYP716C11 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

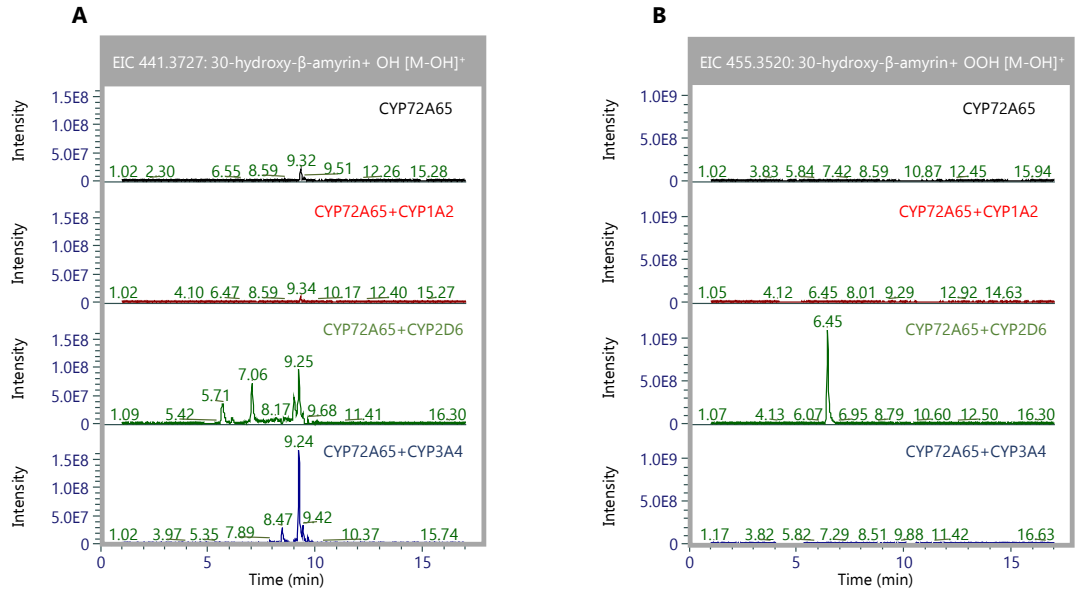


Figure A-126: CYP1A2 and CYP72A65, CYP2D6 and CYP72A65, or CYP3A4 and CYP72A65 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 30-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 455.3520, corresponding to 30-hydroxy-β-amyrin plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP72A65 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

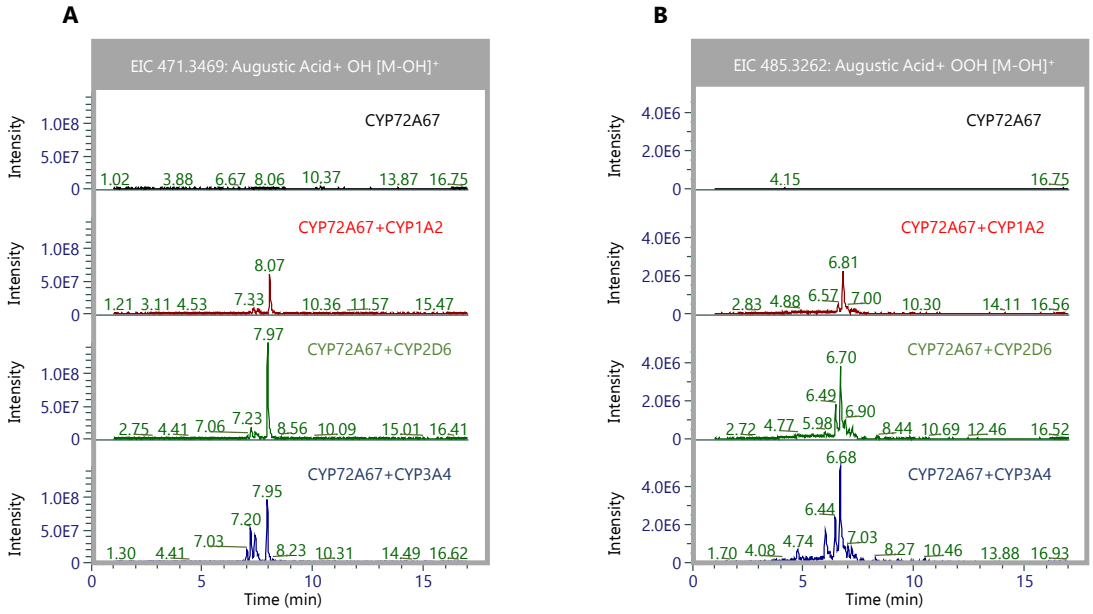


Figure A-127: CYP1A2, CYP716A12, and CYP72A67, CYP2D6, CYP716A12, and CYP72A67, or CYP3A4, CYP716A12, and CYP72A67 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 471.3469, corresponding to hydroxylated augustic acid $[M\text{-}OH]^+$, or B) *m/z* = 485.3262, corresponding to augustic acid plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), CYP716A12, and CYP72A67 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

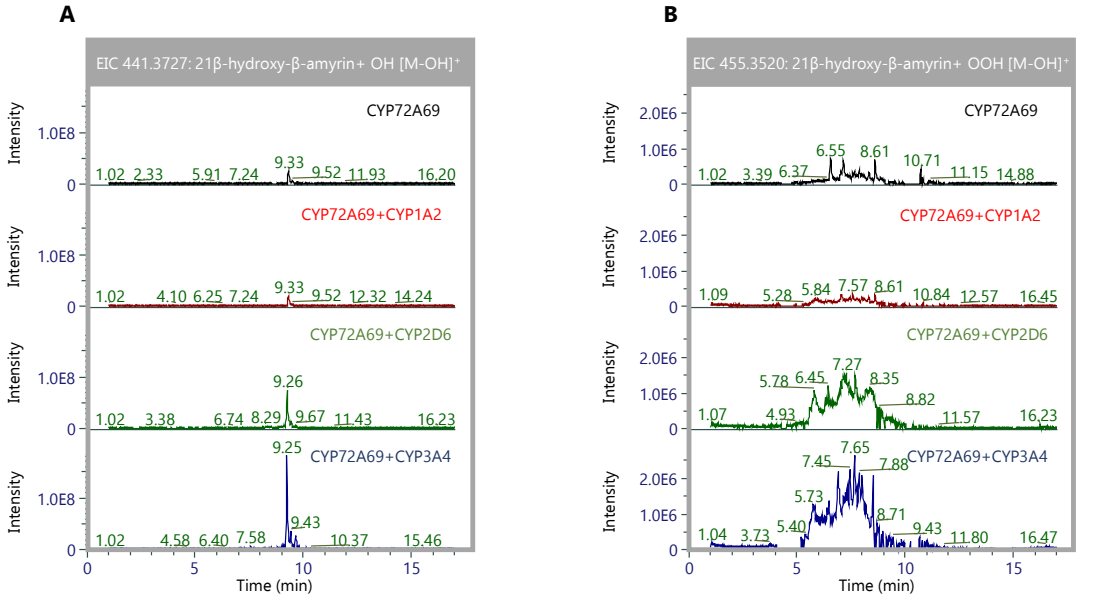


Figure A-128: CYP1A2 and CYP72A69, CYP2D6 and CYP72A69 or CYP3A4 and CYP72A69 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 21β-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 455.3520, corresponding to 21β-hydroxy-β-amyrin plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and

CYP72A69 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

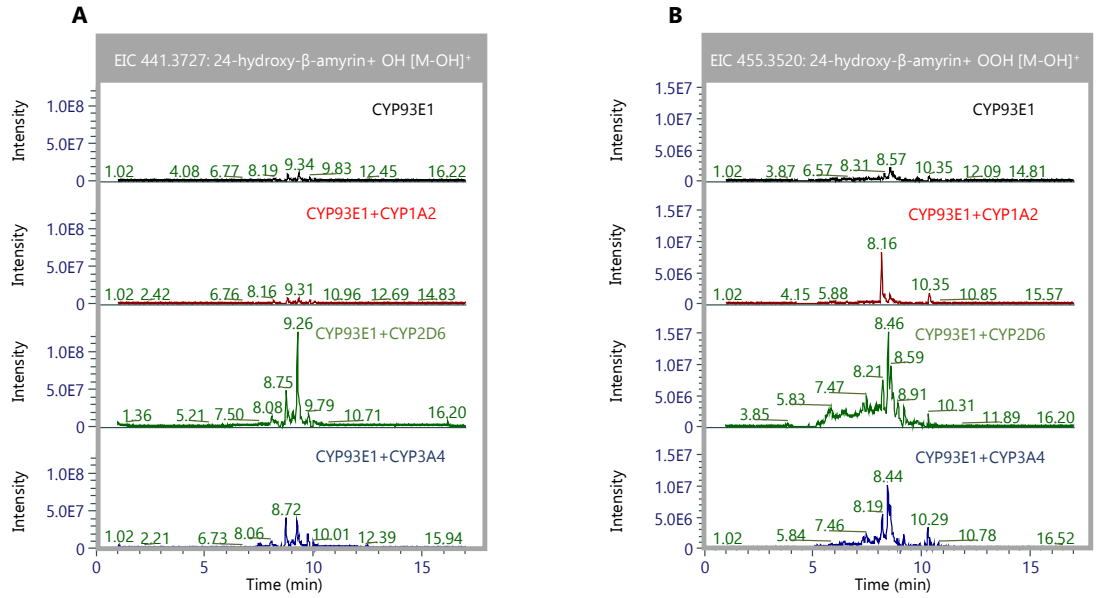


Figure A-129: CYP1A2 and CYP93E1, CYP2D6 and CYP93E1, or CYP3A4 and CYP93E1 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 24-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 455.3520, corresponding to 24-hydroxy-β-amyrin plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and CYP93E1 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

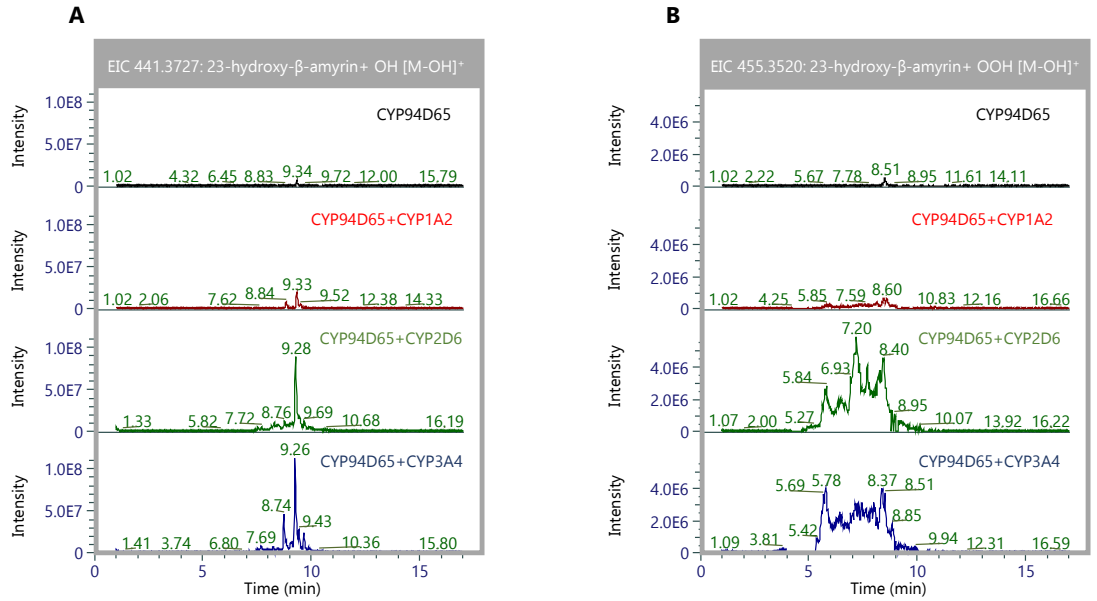


Figure A-130: CYP1A2 and CYP94D65, CYP2D6 and CYP94D65, or CYP3A4 and CYP94D65 produce new products in co-expression. Extracted ion chromatograms (EIC) for A) *m/z* = 441.3727, corresponding to hydroxylated 23-hydroxy-β-amyrin $[M\text{-}OH]^+$, or B) *m/z* = 455.3520, corresponding to 23-hydroxy-β-amyrin plus a carboxylic acid group $[M\text{-}OH]^+$. Traces are of *N. benthamiana* extracts expressing tHMGR, β-amyrin synthase (SAD1), and

CYP94D65 as a control (black), and with addition of either CYP1A2 (red), CYP2D6 (green), or CYP3A4 (blue).

A.4 List of CYPs in the Triterpene Toolkit

Table A-15: List of the CYPs in the triterpene toolkit that show activity towards β-amyrin or its derivatives. Where no reference is given, CYP was obtained through unpublished work in the Osbourn lab.

Name	Species	Substrate	Major Product	Ref
CYP714	*Quillaja saponaria*	oleanolic acid/echinocystic acid	gypsogenin/quillaic acid	
CYP716A12	*Medicago truncatula*	α/β-amyrin and lupeol	ursolic/oleanolic/betulinic acid	[114]
CYP716A14v2	*Artemesia annua*	α/β/δ-amyrin	α/β/δ-amyrone	[115]
CYP716A141	*Platycodon grandiflorus*	β-amyrin/ oleanolic acid	16β-hydroxy-β-amyrin	[109]
CYP716C11	*Centella asiatica*	oleanolic acid	2α-hydroxy-oleanolic acid (maslinic acid)	[117]
CYP716E26	*Solanum lycopersicum*	α/β-amyrin	6β-hydroxy-β-amyrin	[116]
CYP716S5	*Platycodon grandiflorus*	β-amyrin	12,13-epoxy-β-amyrin	[117]
CYP72A65	*Medicago truncatula*	β-amyrin	30-hydroxy-β-amyrin	[118]
CYP716Y1	*Bupleurum falcatum*	β-amyrin/oleanolic acid	16-hydroxy-β-amyrin /echinocystic acid	[8]
CYP716A224	*Quillaja saponaria*	β-amyrin	oleanolic acid	
CYP716-2012090	*Quillaja saponaria*	β-amyrin/oleanolic acid	16-hydroxy-β-amyrin /echinocystic acid	
CYP72A475	*Avena strigosa*	12,13-epoxy-16β-hydroxy-β-amyrin	12,13-epoxy-16β,21β-dihydroxy-β-amyrin	[197]
CYP72A476	*Avena strigosa*	12,13-epoxy-16β-hydroxy-β-amyrin	12,13-epoxy-16β,30-dihydroxy-β-amyrin	[123]
CYP72A61v2	*Medicago truncatula*	24-hydroxy-β-amyrin	22,24-dihydroxy-β-amyrin	[198]
CYP72A63	*Medicago truncatula*	β-amyrin	30-hydroxy-β-amyrin	[198]
CYP72A69	*Glycine max*	β-amyrin	21β-hydroxy-β-amyrin	[119]
CYP87D16	*Maesa lanceolata*	β-amyrin/oleanolic acid	16-hydroxy-β-amyrin/echinocystic acid	[199]
CYP51H10	*Avena strigosa*	β-amyrin	12,13β-epoxy-16β-hydroxy-β-amyrin	[120]
CYP51H10 A354L	*Avena strigosa*	β-amyrin	16β-hydroxy-β-amyrin	

Table A-15 (continued): List of the CYPs in the triterpene toolkit that show activity towards β-amyrin or its derivatives. Where no reference is given, CYP was obtained through unpublished work in the Osbourn lab.

Name	Species	Substrate	Major Product	Ref
CYP51H10 I471M	*Avena strigosa*	β-amyrin	12,13β-epoxy-β-amyrin	
CYP88D6	*Glycyrrhiza uralensis*	β-amyrin	11-oxo-β-amyrin	[121]
CYP93E1	*Glycine max*	β-amyrin	24-hydroxy-β-amyrin	[122]
CYP94D65	*Avena strigosa*	β-amyrin	23-hydroxy-β-amyrin	[123]
CYP106A1	*Bacillus megaterium*	β-amyrin (and derivatives)	7β-(Primarily), also some 15α-hydroxylation	[141]
CYP106A2	*Bacillus megaterium*	β-amyrin (and derivatives)	7β- and 15α-hydroxylation	[142]
CYP1A2	*Homo sapiens*	oxidised β-amyrin	unknown	[149]
CYP2D6	*Homo sapiens*	oxidised β-amyrin	unknown	[200]
CYP3A4	*Homo sapiens*	oxidised β-amyrin	C-21 and C-28 oxidation	[150]

A.5 Table of Cytotoxicity Data

Table A-16: Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50										Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa2	M202	MCF-7	BV2	HeLa	
Oleanolic Acid	47.61	32.885	61.95	#N/A	#N/A	50.58	43.37	44.95	#N/A	43.44	
Avenacin A-1	1.51	0.222	0.756	#N/A	#N/A	#N/A	#N/A	0.785	#N/A	#N/A	
Avenacin A-2	#N/A	0.306	0.896	#N/A	#N/A	#N/A	#N/A	0.804	#N/A	#N/A	
Avenacin B-1	#N/A	0.232	0.589	#N/A	#N/A	#N/A	#N/A	0.954	#N/A	#N/A	
Avenacin B-2	#N/A	0.377	0.434	#N/A	#N/A	#N/A	#N/A	0.792	#N/A	#N/A	
13-epi-Avenestergenin A2	#N/A	35.38	30.81	#N/A	#N/A	#N/A	#N/A	42.78	#N/A	#N/A	
Avenestergenin A2	#N/A	15.88	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
13-epi-Avenastergenin B2	#N/A	7.48	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
11-a-hydroxy-12-ene-Avenestergenin A1	#N/A	6.78	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
13-epi-avenetergenin A1	#N/A	13.03	12.82	#N/A	#N/A	#N/A	#N/A	14.16	#N/A	#N/A	
Avenestergenin A1	#N/A	3.01	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
13-epi-30-hydroxy-Avenestergenin A1	#N/A	10.09	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Betulinic Acid	20.445	49.2	16.71	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Ursolic Acid	10.185	48.96	33.34	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	33.12	
Alpha Hederin	4.66	9.94	3.7	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
18α-glycyrrhetinic Acid	100	100	#N/A	100	100	#N/A	#N/A	100	#N/A	100	
α-Hederagenin	38.2	72	#N/A	100	46.4	#N/A	#N/A	#N/A	#N/A	#N/A	

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50										Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa2	M202	MCF-7	BV2	HeLa	
Madecassic Acid	100	98.7	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	#N/A	
Digoxigenin	3.1	1.4	#N/A	2.8	100	#N/A	#N/A	#N/A	#N/A	#N/A	
Tomatidine	100	100	#N/A	100	56.7	#N/A	#N/A	#N/A	#N/A	#N/A	
Erythrodiol	#N/A	14.32	#N/A	#N/A	#N/A	100	20.72	15.9	#N/A	#N/A	
Protopanaxadiol	#N/A	27.9	#N/A	#N/A	#N/A	43.7	41.08	45.72	#N/A	#N/A	
11-oxo-β-amyrin	60.68	19.66	#N/A	#N/A	#N/A	64.77	57.38	62.72	100	#N/A	
11-oxo-epi-β-amyrin	#N/A	30.44	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
11-oxo-β-amyrone	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
11α-hydroxy-β-amyrin	#N/A	100	#N/A	#N/A	#N/A	100	100	100	#N/A	#N/A	
30-hydroxy-β-amyrin	100	100	#N/A	#N/A	#N/A	100	100	100	100	#N/A	
24-hydroxy-β-amyrin	100	100	#N/A	#N/A	#N/A	100	100	100	100	#N/A	
12,13-β-epoxy-16β-hydroxy-β-amyrin	#N/A	100	#N/A	#N/A	#N/A	100	100	100	#N/A	#N/A	
12,13-β-epoxy-16β-hydroxy-β-amyrin-3-Ara	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	
12,13-β-epoxy-16β,21β-dihydroxy-β-amyrin-3-Ara	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50										Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa2	M202	MCF-7	BV2	HeLa	
epi-β-amyrin	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
β-amyrin	100	100	#N/A	#N/A	#N/A	100	100	100	100	#N/A	
11-oxo-oleanolic Acid	#N/A	30.6	#N/A	#N/A	#N/A	100	79.38	100	#N/A	#N/A	
18β-glycyrrhetinic Acid	100	25.7	#N/A	#N/A	#N/A	76.59	59.41	88.78	100	#N/A	
16β-hydroxy-β-amyrin	25.64	100	#N/A	#N/A	#N/A	100	100	100	100	#N/A	
α-amyrin	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Lupeol	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
12,13-β-epoxy-16β,24-dihydroxy-β-amyrin	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	29.62	#N/A	#N/A	
12β-hydroxyoleanane-13β:28-olide	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	
Oleanolic Acid-3-glc,-28-glc	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Oleanolic Acid-3-glc,-28-2xglc	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	
Oleanolic Acid-3-glc,-28-3xglc	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	
Oleanolic Acid-28-2xglc	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	83.84	#N/A	#N/A	
Oleanolic Acid-28-3xglc	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50										Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa 2	M202	MCF-7	BV2	HeLa	
Oleanolic Acid-28-4xglc	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	
12,13-β-epoxy-11β,16β-dihydroxy-β-amyrin	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	61.77	#N/A	#N/A	
11-oxo-12,13-β-epoxy-16β-hydroxy-β-amyrin	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	82.16	#N/A	#N/A	
11-oxo-24-hydroxy-β-amyrin	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	
24-hydroxy-11-oxo-epi-β-amyrin	#N/A	18.82	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
11-oxo-16β-hydroxy-β-amyrin	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	28.13	#N/A	#N/A	
21β-hydroxy-β-amyrin	100	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
21β-hydroxy-11-oxo-β-amyrin	#N/A	66.04	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Ellarinacin	33.17	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
6β-hydroxy-β-amyrin	46.89	17.56	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
7β-hydroxy-β-amyrin	#N/A	28.13	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Asiatic Acid	34.4	36.2	#N/A	39.4	30.8	#N/A	#N/A	68.5	#N/A	52.47	

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50										Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa 2	M202	MCF-7	BV2	HeLa	
23-hydroxy-β-amyrin	#N/A	19.14	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
α-boswellic acid	100	30.16	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
3-O-Acetyl-α-boswellic acid	#N/A	24.22	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
β-boswellic acid	100	53.99	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
11-oxo-β-boswellic acid	100	55.22	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	20	#N/A	
3-O-acetyl-β-boswellic acid	100	36.32	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
3-O-acetyl-11-oxo-β-boswellic acid	20	15.62	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	20	31	
Caulophyllogenin	100	78.1	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
Augustic Acid	#N/A	16.52	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Maslinic acid	20	24.12	4.5	#N/A	#N/A	#N/A	#N/A	#N/A	20	#N/A	
Medicagenic acid	100	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
Gypsogenin	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
β-amyrone	100	100	#N/A	#N/A	#N/A	100	100	100	100	#N/A	
Quillaic Acid	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	
Echinocystic Acid	20	16.14	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	20	#N/A	

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50									Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa2	M202	MCF-7	BV2	HeLa	
Quillaic Acid-TriR	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Quillaic Acid-TriX	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Quillacid Acid-TriR-FR	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
Frondoside A	3	2.5	2.5	#N/A	#N/A	0.5	#N/A	2	#N/A	#N/A	[72]
Lepiotaprocerin A	#N/A	40	40	#N/A	#N/A	#N/A	#N/A	40	#N/A	#N/A	[3]
Lepiotaprocerin B	#N/A	40	40	#N/A	#N/A	#N/A	#N/A	40	#N/A	#N/A	[3]
Lepiotaprocerin C	#N/A	40	40	#N/A	#N/A	#N/A	#N/A	40	#N/A	#N/A	[3]
Lepiotaprocerin D	#N/A	40	40	#N/A	#N/A	#N/A	#N/A	40	#N/A	#N/A	[3]
Lepiotaprocerin E	#N/A	40	40	#N/A	#N/A	#N/A	#N/A	40	#N/A	#N/A	[3]
Lepiotaprocerin F	#N/A	40	40	#N/A	#N/A	#N/A	#N/A	40	#N/A	#N/A	[3]
Lepiotaprocerin G	#N/A	3.5	7	#N/A	#N/A	#N/A	#N/A	13	#N/A	#N/A	[3]
Lepiotaprocerin H	#N/A	2.9	5	#N/A	#N/A	#N/A	#N/A	8	#N/A	#N/A	[3]
Lepiotaprocerin I	#N/A	14.6	18.9	#N/A	#N/A	#N/A	#N/A	25.9	#N/A	#N/A	[3]
Lepiotaprocerin J	#N/A	12.7	17.7	#N/A	#N/A	#N/A	#N/A	15.5	#N/A	#N/A	[3]
Lepiotaprocerin K	#N/A	14.1	21	#N/A	#N/A	#N/A	#N/A	23.2	#N/A	#N/A	[3]
Lepiotaprocerin L	#N/A	3.1	5.7	#N/A	#N/A	#N/A	#N/A	12.5	#N/A	#N/A	[3]
Palatanic acid	#N/A	#N/A	47.3	#N/A	#N/A	#N/A	#N/A	34.2	#N/A	#N/A	[201]
Rhabdastrellin G	#N/A	56.1	#N/A	#N/A	#N/A	#N/A	#N/A	77.9	#N/A	#N/A	[202]
Rhabdastrellin H	#N/A	14.8	#N/A	#N/A	#N/A	#N/A	#N/A	16	#N/A	#N/A	[202]
Rhabdastrellin I	#N/A	46.3	#N/A	#N/A	#N/A	#N/A	#N/A	63.8	#N/A	#N/A	[202]
Rhabdastrellin J	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[202]

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50									Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa2	M202	MCF-7	BV2	HeLa	
Rhabdastrellin K	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[202]
Jaspolide C	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[202]
Globostelletin C	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[202]
Globostelletin D	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[202]
2a-O-benzoyl-19a-hydroxy-ursolic acid	#N/A	#N/A	7.2	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
2a-O-benzoyl-19a-hydroxy-3-keto-ursolic acid	#N/A	#N/A	5.1	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Schmidic Acid	#N/A	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Ceropiacic Acid-3-methyl Ester	#N/A	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
2-oxo-Pomolic Acid b-D-glucopyranosyl Ester	#N/A	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Ursolazuroside 1	#N/A	#N/A	100	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Corosolic Acid	#N/A	24.4	23.5	#N/A	#N/A	#N/A	#N/A	40.3	#N/A	#N/A	[154]
23-hydroxy-11-keto-erythrodiol	#N/A	#N/A	21.5	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50										Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa 2	M202	MCF-7	BV2	HeLa	
11-keto-erythrodiol	#N/A	#N/A	15.4	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Morolic Acid-3-O-Caffeate	#N/A	#N/A	1.8	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Ambradiolic Acid	#N/A	#N/A	2.1	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Betulin	#N/A	#N/A	14.1	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	[154]
Alatavolide D	#N/A	#N/A	78.6	#N/A	#N/A	#N/A	#N/A	63.1	#N/A	#N/A	[203]
Hebecarposide A	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide B	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide C	#N/A	16.9	18.6	#N/A	#N/A	#N/A	#N/A	22.1	#N/A	#N/A	[204]
Hebecarposide D	#N/A	16.6	27.6	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide E	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide F	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide G	#N/A	13.5	100	#N/A	#N/A	#N/A	#N/A	27.7	#N/A	#N/A	[204]
Hebecarposide H	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide I	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide J	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[204]
Hebecarposide K	#N/A	25.6	28.3	#N/A	#N/A	#N/A	#N/A	35	#N/A	#N/A	[204]
Lupenone	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	19	#N/A	100	[205]
Pardinol A	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[206]
Pardinol B	#N/A	8.3	14.4	#N/A	#N/A	#N/A	#N/A	12.7	#N/A	#N/A	[206]
Pardinol C	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[206]

Table A-16 (continued): Cell viability data used for triterpene SAR. Where no reference is given, data was obtained in the O'Connell lab (unpublished)

Compound	IC50										Ref
	THP-1	HL-60	A549	A375	SJSA	Mia PaCa 2	M202	MCF-7	BV2	HeLa	
Pardinol D	#N/A	100	100	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[206]
Pardinol E	#N/A	9.8	9.8	#N/A	#N/A	#N/A	#N/A	11.9	#N/A	#N/A	[206]
Pardinol F	#N/A	11.2	12.6	#N/A	#N/A	#N/A	#N/A	10.5	#N/A	#N/A	[206]
Pardinol G	#N/A	18.4	25.6	#N/A	#N/A	#N/A	#N/A	32.8	#N/A	#N/A	[206]
Pardinol H	#N/A	15.5	11.8	#N/A	#N/A	#N/A	#N/A	11.8	#N/A	#N/A	[206]
Saponaceol B	#N/A	100	28.2	#N/A	#N/A	#N/A	#N/A	100	#N/A	#N/A	[206]
Xuedanencin A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	49.2	[207]
Xuedanencin B	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	[207]
Xuedanencin C	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	[207]
Xuedanencin D	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	[207]
Xuedanencin E	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	100	[207]
Xuedanencin F	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	35.8	[207]
Xuedanencin G	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	1.8	[207]
Xuedanencin H	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	2.5	[207]
Xuedanencin I	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	15.8	[207]
Xuedanencin J	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	28.3	[207]
Xuedanencin K	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	35.7	[207]
epi-maslinic acid	#N/A	58.9	#N/A	#N/A	#N/A	#N/A	#N/A	67.8	#N/A	#N/A	[70]