Doctoral Dissertation

High-yield production of bioactive triterpenoids in heterologous hosts through protein engineering of plant-derived cytochrome P450 monooxygenase CYP716A subfamily enzyme

Romsuk Jutapat

December 2022

Department of Biotechnology

Graduate School of Engineering

Osaka University

Thesis committee

Professor Toshiya Muranaka, Ph.D.

Laboratory of Cell technology, Department of Biotechnology,

Graduate School of Engineering, Osaka University

Professor Kazuhito Fujiyama, Ph.D.

Laboratory of Applied Microbiology, International Center for Biotechnology,

Graduate School of Engineering, Osaka University

Professor Genji Kurisu, Ph.D.

Laboratory of Protein Crystallography, Institute for Protein Research,

Osaka University

List of abbreviations

aAS : α-Amyrin synthase

ATP : Adenosine triphosphate

bAS : β-Amyrin synthase

BLAST : Basic Local Alignment Search Tool

CPR : NADPH-cytochrome P450 reductase

CDS : Coding sequence

CYPs : Cytochrome P450 monooxygenases

CYP716As : Selected CYPs 716 subfamily A

DNA : Deoxyribonucleic acid

EIC : Extracted ion chromatogram

FPP : Farnesyl diphosphate

GC-MS : Gas chromatography-mass spectrometry

GW : Gateway cassette

HMGR : 3-hydroxy-3-methylglutaryl-coenzyme A reductase

LUS : Lupeol synthase

MVA : Mevalonate

NADPH : Nicotinamide adenine dinucleotide phosphate

NCBI : National Center for Biotechnology Information

OSC : Oxidosqualene cyclases

OTC : Over the counter (OTC) drugs

PCC : Pearson correlation coefficient

PCR : Polymerase chain reaction

SRSs : Substrate recognition sites

SQE : Squalene epoxidase

SQS : Squalene synthase

TIC : Total ion chromatograms

UDP : Uridine diphosphate

GT : Glucosyltransferase

WT : Wild type

Table of contents

List of tables

List of supporting tables

List of figures

List of Supporting Figures

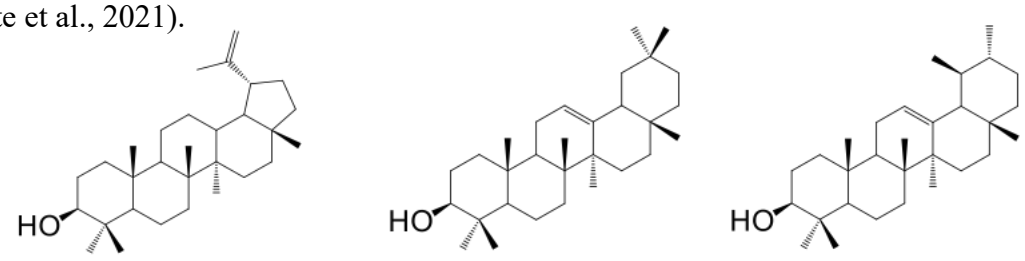
Chapter 1 : General introduction

1.1 Plant-derived triterpenoid and its important

Secondary metabolism, also known as specialized metabolism, refers to biosynthetic pathways that produce small molecules that are not essential to an organism's growth or survival (Pusztahelyi et al., 2015; Cui et al., 2018). They are a rich source of bioactive compounds with potential pharmacological properties, consisting of biosynthetic pathways of numerous chemical compounds that have been and are being used as, among other things, anti-cancer, immune-boosting, and anti-microbial drugs (de Frias et al., 2018; Bose et al., 2019). In the past decade, secondary metabolite research has been challenging, particularly regarding the relationship between biosynthetic genes and product diversity. The recent expansion of high throughput next-generation sequencing tools, metabolomics techniques, and high-performance computational resources has made it possible to conduct informational analyses (e.g., transcriptome analysis, comparative genomics) to obtain information about unannotated genes, particularly in non-model plant species. Consequently, our understanding of the coding genes involved in the biosynthesis of plant secondary metabolites has expanded (Nascimento and Fett-Neto, 2010; Seki et al., 2015; Al-Jumaili et al., 2018). Secondary metabolites perform numerous crucial roles in plant biology. They defend plants against herbivores and microbial illnesses and communicate with symbiotic bacteria and mycorrhiza. In addition, they serve as allelopathic agents in natural ecosystems and as physical and chemical barriers to abiotic stresses like as UV and evaporation. They also regulate plant growth hormones (Mizutani and Sato, 2011; Piasecka et al., 2015; Cui et al., 2018). On the basis of their biosynthetic origins, the secondary metabolites derived from plants can be divided into three major groups: (i) Flavonoids and allied phenolic and polyphenolic compounds, (ii) terpenoids (e.g. diterpenoids (C-20), sesterterpenoids (C-25), triterpenoids (C-30), and tetraterpenoids (C-40)), and (iii) nitrogen-containing alkaloids and sulfur-containing compounds (Saito, 2003; Sumner et al., 2003).

Triterpenoids are a class of specialized plant-derived metabolites that contain six isoprene units (C-30 hydrocarbon). They are produced when squalene is cyclized in numerous plant species. (Phillips et al., 2006; Sheng and Sun, 2011; Fukushima et al., 2013; Seki et al., 2015). Triterpenoids are a diverse group of compounds classified according to their skeletons as acyclic, monocyclic, bicyclic, tricyclic, tetracyclic, and pentacyclic triterpenoids. Tetracyclic triterpenoids (e.g., protostanes, cycloartanes, dammaranes, and euphanes) and pentacyclic triterpenoids (e.g., ursanes, gammaceranes, lupanes, oleananes, and hopanes) are the most extensively studied (Xu et al., 2004; Sheng and Sun, 2011; Seki et al., 2015)

The pentacyclic triterpenoid is present in medicinal plants and other types of plants. Mainly, the lupane- (lupeol, betulin, or betulinic acid aglycone), oleanane- (β-amyrin, erythodiol, or oleanolic acid aglycone), and ursane- (α-amyrin, uvaol, or ursolic acid aglycone) are the most abundant. Triterpenes display a variety of pharmacological effects devoid of toxicity. Consequently, these triterpenes have the potential to be utilized in the creation of new bioactive agents with multiple targets (**Figure 1-1**) (Jager et al., 2009). In addition to other nutraceuticals, Mediterranean spices and fruits contain pentacyclic triterpenes from the lupane, oleanane, and ursane groups. Lists of the abundant plant sources of triterpenes from the lupane, oleanane, and ursane groups as materials for triterpene extraction shown in **Table 1-1** (Jager et al., 2009). Among the many compounds found in olive tree berries and leaves are oleanolic and maslinic acids (*O. europaea*). There are small amounts of ursolic and betulinic acids in olive products. The oleanolic and maslinic acids on olive leaves are thought to help the plant's defense against water loss and pathogens by creating a physical barrier (Chen et al., 2005; Jager et al., 2009; Romero et al., 2010; Gudoityte et al., 2021).



lupane- oleanane- ursane-

Figure 1-1 Molecule structures of lupane-, oleanane- and ursane triterpenes.

Table 1-1 Triterpene distribution within various dry plant materials

(This table was modified from (Jager et al., 2009).

Plant species	Common name	Plant part	[g/100 g dry matter]								
			LUS	BE	BA	BAM	ER	OA	AM	UV	UA
Aesculus hippocastanum	horse-chestnut	leaves				< 0.3					
Aloe vera	aloe vera	leaves	0.1								
Arctostaphylos uva-ursi	bearberry	leaves	0.29		0.12	0.1	0.18	0.27	0.3	0.4	1.24
Betula alba	birch	bark	0.9–2.1	10.5–18.3	0.5–1.3		0.2–0.4	0.1–1.1			
Calendula officinalis,	pot marigold	flowers	< 0.3					2.0			
Centaurium erythraea	common centaury	herb				< 0.3		0.16	< 0.3		
Coffea arabica	coffee	leaves						< 0.3			1.8
Cornus mas	european cornel	leaves									0.15
Crataegus	hawthorn	leaves, flowers						0.1			0.52
Eucalyptus	euca-lyptus	leaves			0.84			0.31			1.17
Lavandula angustifolia	lavender	leaves			0.13			0.45			1.59
Lavandula angustifolia	lavender	flowers			0.12			0.4			1.05
Malus domestica	11 diff. apples	fruit peel			< 0.3			0.28			1.43
								0.07			0.3–0.7
Malus domestica	apple pomace	pomace						0.16			0.8
Melissa officinalis	lemon balm	*leaves*						0.16			0.67
Nerium oleander	oleander	leaves			0.11			0.37			1.27
Ocimum basilicum	sweet basil	leaves						< 0.3			< 0.3
Origanum majorana	marjoram	leaves						0.19			0.66
Origanum vulgare	oregano	leaves						< 0.3			0.28

Table 1-1 Triterpene distribution within various dry plant materials (continue)

Plant species	Common name	Plant part	[g/100 g dry matter]								
			LUS	**BE**	**BA**	**BAM**	**ER**	**OA**	**AM**	**UV**	**UA**
Olea europeae	olive	leaves					< 0.3	3.1		0.3	0.18
		bark		< 0.3	< 0.3		< 0.3	0.98			< 0.3
		fruit						0.21			
		pomace						0.18			
Pimpinella anisum	anisseed	seed				< 0.3					
Plantago major	greater plantain	*leaves*						< 0.3			0.21
Platanus acerifolia	planes	bark		< 0.3	2.44			< 0.3			
Prunus cerasus	sour cherry	unripe fruit									< 0.3
Pyrus communis	pear williams'	fruit peel									0.2
Rosmarinus officinalis	rosemary	leaves			1.53			1.23			2.95
Salvia officinalis	sage	leaves			0.02	< 0.3		0.67			1.8
Sambucus nigra	black elder	leaves						0.12			0.58
		bark		< 0.3	< 0.3			0.08			0.32
Satureja montana	winter savory	leaves			0.04			0.14			0.49
Solanum lycopersicum	tomato	fruit peel				< 0.3					
Syringa	lilac	*leaves*									< 0.3
Syzygium aromaticum	clove	flower			< 0.3			1.65			< 0.3
Thymus vulgaris	common thyme	leaves						0.37			0.94
Verbena officinalis	common vervain	herb						< 0.3			0.17
Viscum album	mistletoe	sprouts			< 0.3			0.86			
Vitis vinifera	grape vine	leaves	< 0.3								

Table Abbreviation: lupane- (lupeol (LUS), botulin (BE), or betulinic acid (BA) aglycone), oleanane- (β-amyrin (BAM), Erythodiol (ER), or Oleanolic acid (OA) aglycone), and ursane (α-amyrin (AM), Uvaol (UV), or Ursolic acid (UA) aglycone)

Triterpenoids exhibit a diverse array of unique biological properties. Extensive pharmacological and mechanistic research has been conducted to maximize the therapeutic potential of plant-derived triterpenoids (Chen et al., 2005; Dzubak et al., 2006; Sheng and Sun, 2011; Ayeleso et al., 2017) **Table 1-2**.

Numerous scientific research has shown that triterpenoids have a wide range of bioactive effects, such as anti-viral, anti-inflammatory, anti-microbial, anti-cancer, anti-parasitic, and gastroprotective. Several plant-derived triterpenoid therapies are now on the market. Clinical investigations of synthetic plant-derived triterpenoid derivatives have gained much attention due to their various pharmaceutical characteristics (Liu, 1995; Kensil et al., 1998; Chen et al., 2005; Dzubak et al., 2006; Yadav et al., 2010; Fukushima et al., 2013; Yan et al., 2014; Ayeleso et al., 2017; Jin et al., 2019).

Table 1-2 The current profile of natural plant-derived triterpenoids compounds as therapeutic agents or dietary supplements around the world

(This table was modified from (https://clinicaltrials.gov, 2000; Sheng and Sun, 2011)

Name of active compound	Category	Indications	Status	Region
Oleanolic acid	Drug	Liver diseases	Registered (OTC drug)	China
Ursolic acid	Dietary supplement Drug	Sarcopenia (Muscle loss with aging) Diabetes, obesity	Phase II Phase II	South Korea Mexico
Maslinic acid	Dietary supplement Dietary supplement	Diabetes, obesity Cardiovascular Diseases	Registered Phase III	Japan Spain
Betulinic acid	Drug	Dysplastic Nevus Syndrome	Phase III	United States
Glycyrrhizic acid	Drug	Liver diseases	Registered	China, Japan
Carbenoxolone	Drug	Gastritis and pepticulcers	Registered	Asia, Europe
Asiaticoside	Drug	Wound healing	Registered (OTC drug)	China, Europe
Corosolic acid	Dietary supplement	Diabetes, obesity	Marketed	Asia, America
Gymnemic acids	Dietary supplement	Diabetes	Marketed	Asia, America, Europe

1.1.1 Triterpenoid biosynthesis

The current knowledge reported that they are biosynthesized in three steps: cyclization of 2,3-oxidosqualene catalyzed by oxidosqualene cyclases (OSCs), site-specific oxidations catalyzed by cytochrome P450 monooxygenases (CYPs) and glycosylations catalyzed by UDP-glycosyltransferases (GTs) (Fukushima et al., 2013; Seki et al., 2015) (**Figure 1-2**).

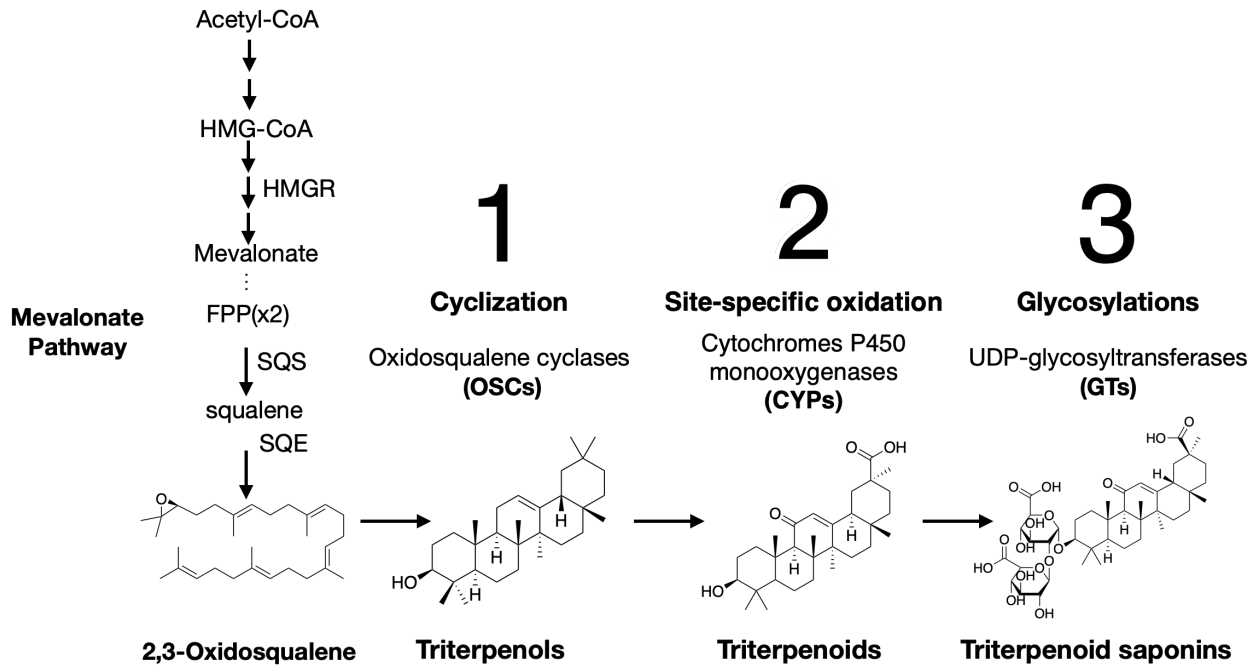


Figure 1-2 General scheme of enzymes involved in triterpenoid biosynthesis in plant cells.

However, their productivity in plant is low, their purification and chemical synthesis are difficult, and their complete biosynthetic pathway is not yet elucidated in most species. Information in the elucidation of their biosynthesis at the molecular level would help increase their production yield in both native and heterologous production (Fukushima et al., 2013; Reed et al., 2017; Dale et al., 2020).

1.2 Synthetic biology in heterologous host for triterpenoid production

Triterpenoids are a diverse class of plant-derived chemicals that are considered protective against pathogenic microorganisms and herbivores (Sawai and Saito, 2011). Due to their diverse human-beneficial qualities, triterpenoids are used in a variety of applications beyond medicine **Table1-2**. Very little is accumulated by plants, and its chemical synthesis is difficult. Consequently, the commercial potential of these advantageous compounds is minimal. (Sawai and Saito, 2011; Sheng and Sun, 2011; Yan et al., 2014).

Moreover, the extraction of triterpenoids from a plant's natural resources is unusually environmentally unfriendly and costly due to the poor yield of the compound in the plant, which

needs a considerable volume of plant biomass. Furthermore, the complexity of plant triterpenoids structure makes chemical synthesis extraordinarily difficult. Thus, there is an urgent need to develop new methods of producing plant-based products (Sawai and Saito, 2011; Moses et al., 2013; Moses et al., 2017; Cravens et al., 2019). Methods for producing triterpenoids compounds in heterologous host such as *Saccharomyces cerevisiae* and *Nicotiana benthamiana* have been studied using synthetic biology techniques (Fukushima et al., 2013; Reed et al., 2017).

1.2.1 Synthetic biology in *Saccharomyces cerevisiae*

Widespread use of microbial hosts for plant metabolite synthesis. Yeast has been found to be the most effective heterologous host to produce a wide variety of secondary metabolites from plants, including triterpenoids, when compared to other microorganisms. When comparing yeast and *E. coli*, the availability of some biosynthesis pathways in yeast provides a more suitable environment for CYPs activity in relation to plant triterpenoids biosynthesis. Yeast contains endogenous enzymes that are involved in the biosynthetic pathway of 2,3-oxidosqualene (a common C-30 precursor that yeast naturally produces) to produce triterpenoids. Like plants, yeast produces isopentenyl diphosphate and dimethylallyl diphosphate as squalene precursors from acetyl-coA through its own mevalonate pathway. Because yeast does not contain any triterpenoids biosynthetic enzymes, it is possible to express triterpenoids biosynthesis enzymes heterologously without interfering with the native metabolic pathway of yeast (Siddiqui et al., 2012; Moses et al., 2013; Moses et al., 2017; Cravens et al., 2019).

Synthetic biology techniques have been used to investigate methods for producing triterpenoids compounds in heterologous hosts such as the yeast *Saccharomyces cerevisiae*. The genome of the yeast *S. cerevisiae* has been completely sequenced and is being used as a model organism for fundamental molecular biology research. Numerous studies have been aided in recent decades by the availability of a complete yeast genomic database, which has revealed detailed models of *S. cerevisiae* metabolic pathways and processes, making it easier to control and

manipulate metabolic networks to produce a variety of plant secondary metabolites, including triterpenoids (Sheng and Sun, 2011; Fukushima et al., 2013; Xiao and Zhong, 2016; Cravens et al., 2019; Zhang and Hong, 2020; Carsanba et al., 2021).

1.2.2 Synthetic biology in *Nicotiana benthamiana*

Triterpenes (C-30) are synthesized via the mevalonate pathway, which is present in most eukaryotic organisms, including animals, fungi, and higher plants **Figure 1-2**. Triterpenes biosynthetic enzymes derived from distantly related plant species appear to be generally appropriately targeted and functional in *N. benthamiana* (Huang et al., 2017; Reed and Osbourn, 2018). Consistently, the yield of the target pathway's products can be increased by removing bottlenecks in upstream pathways or obstructing the flow of intermediates into competing pathways. A thorough understanding of the metabolic process associated with the target metabolite is essential to generate the desired molecule successfully. Understanding the regulatory enzymes required in each biosynthetic pathway and how they affect yield has consequently received considerable study. Quick screening of candidate genes and reconstruction of metabolic pathways are made possible by transient expression (Liu et al., 2011; Reed et al., 2017; Reed and Osbourn, 2018).

A 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) is a known rate-limiting enzyme for the mevalonate pathway. The N-terminus of HMGR has a crucial regulatory function but is not required for catalysis (Polakowski et al., 1998; Reed et al., 2017; Robertlee et al., 2018). This region's truncation generates a feedback-insensitive protein (tHMGR) and increases the yields of mevalonate end products like squalene (van Herpen et al., 2010; Liu et al., 2011; Cankar et al., 2015; Reed and Osbourn, 2018; Srisawat et al., 2019). This method constitutes a standard technique for boosting the production of mevalonate-derived sesquiterpenes (C-15) and triterpenes (C-30) **Table 1-3**.

Table 1-3 Examples of various mevalonate-derived terpenes produced through transient expression in *N. benthamiana*

(This table was modified from (Reed and Osbourn, 2018)

Compound	Class	Quantity	Strategy	Fold increases
Caryophyllene	Sesquiterpene	–	RNAi of *N. benthamiana* VAMP72 genes	Fivefold increase upon silencing VAMP72 genes
Costunolide	Sesquiterpene	60 ng/g FW	Targeting to mitochondria	15-fold increases from mitochondrial-targeting versus cytosol
Parthenolide	Sesquiterpene	1.4 µg/g FW	Expression of HMGR	Fourfold increases in the parthenolide precursor costunolide with HMGR
(+)-Valencene	Sesquiterpene	0.70 µg/g FW/ 24 h (unopti-mised)	Expression of tHMGR & FPS, silencing of SQS and EAS	2.9-fold increases from expression of tHMGR and FPS
Artemisinic acid	Sesquiterpene	16.6 mg/kg FW	Genes fused together with use of 2A ribosomal skipping sequences/mitoc hondrial targeting of ADS and FPS/ expression of tHMGR	Use of the fusion construct with mitochondrial FPS and tHMGR increased amorphadiene (artemisinic acid precursor) in headspace by ~ 2-fold, and internal leaf amorphadiene by ~ 7-fold
Artemisinin	Sesquiterpene	3 ng/mg DW	Expression of LTP3 and PD2 plus HMGR	Approx 50% increase in artemisinin at 13 days after infiltration
12,13-epoxy, 16-hydroxy-β-amyrin	Triterpene	1.18 mg/g DW	–	–
Various	Triterpene	0.12–3.3 mg/g DW	tHMGR, vacuum infiltration	Fourfold increases in β-amyrin upon expression of tHMGR

Table Abbreviation: FW Fresh weight, DW dry weight, VAMP vesicle-associated membrane protein, (t)HMGR HMG-CoA reductase ("t" denotes N-terminal-truncated form), FPS farnesyl diphosphate synthase, ADS amorphadiene synthase, SQS squalene synthase, EAS 5-Epi-aristolochene synthase, LTP3 lipid transfer protein 3, PD2 pleiotropic drug resistance 2

1.3 Cytochrome P450 monooxygenase

CYPs are monooxygenases that carry heme as a co-factor and play an important role in diversifying triterpenoid structures by site-specific oxidation of the cyclized backbone (Nelson and Werck-Reichhart, 2011; Fukushima et al., 2013; Seki et al., 2015). Almost all CYPs in plants are linked to the endoplasmic reticulum or chloroplast via a membrane anchor and hinge domain. Each CYPs possesses motifs that facilitate interactions with the heme cofactor, a cytochrome P450 reductase (CPR) domain, and recognition sites that interact with a wide variety of substrates (**Figure 1-3**).

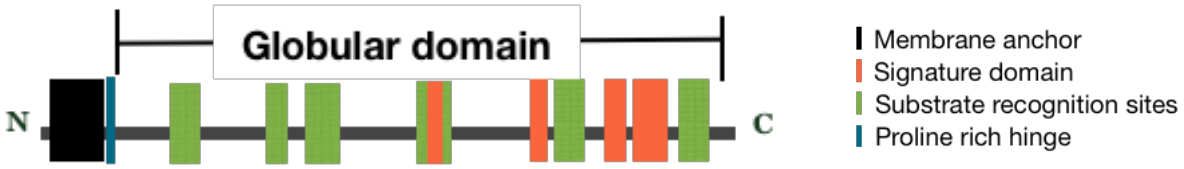


Figure 1-3 Diagram of globular domain important for functional analysis of CYPs.

The heme-binding motif is the most conserved region in the CYPs enzyme family because it contains the cysteine ligand and is strong enough to hold the heme in place during oxygenation (Fontana et al., 2005; Sirim et al., 2010; Bak et al., 2011; Sugimoto and Shiro, 2012; Poulos and Johnson, 2015; Prall et al., 2016) (**Figure 1-4**). The ExxR motif locks the heme pocket into the ligand-binding site and stabilizes the core structure (Poulos and Johnson, 2015; Prall et al., 2016) (**Figure 1-4**). The PERF motif is required for protein-protein interactions with cytochrome P450 reductase (Sirim et al., 2010; Prall et al., 2016) (**Figure 1-4**). The oxygen-activating region is positioned above the heme group in the I-helix, which extends the core region of the CYP. This domain region conserved residues that contribute to the development of a proton groove required for oxygenation via breaking the O-O bond and formation of active Fe-O hydroxylating species (Bak et al., 2011; Poulos and Johnson, 2015; Prall et al., 2016) (**Figure 1-4**). Substrate recognition sites (SRSs) are conserved in the globular domain (**Figure 1-4**). The amino acid characteristics of SRSs determine their enzymatic function in terms of substrate specificity, catalytic activity, region-

specificity, and shape of the substrate-binding pocket (Gotoh, 1992; Sugimoto and Shiro, 2012) (**Figure 1-4**).

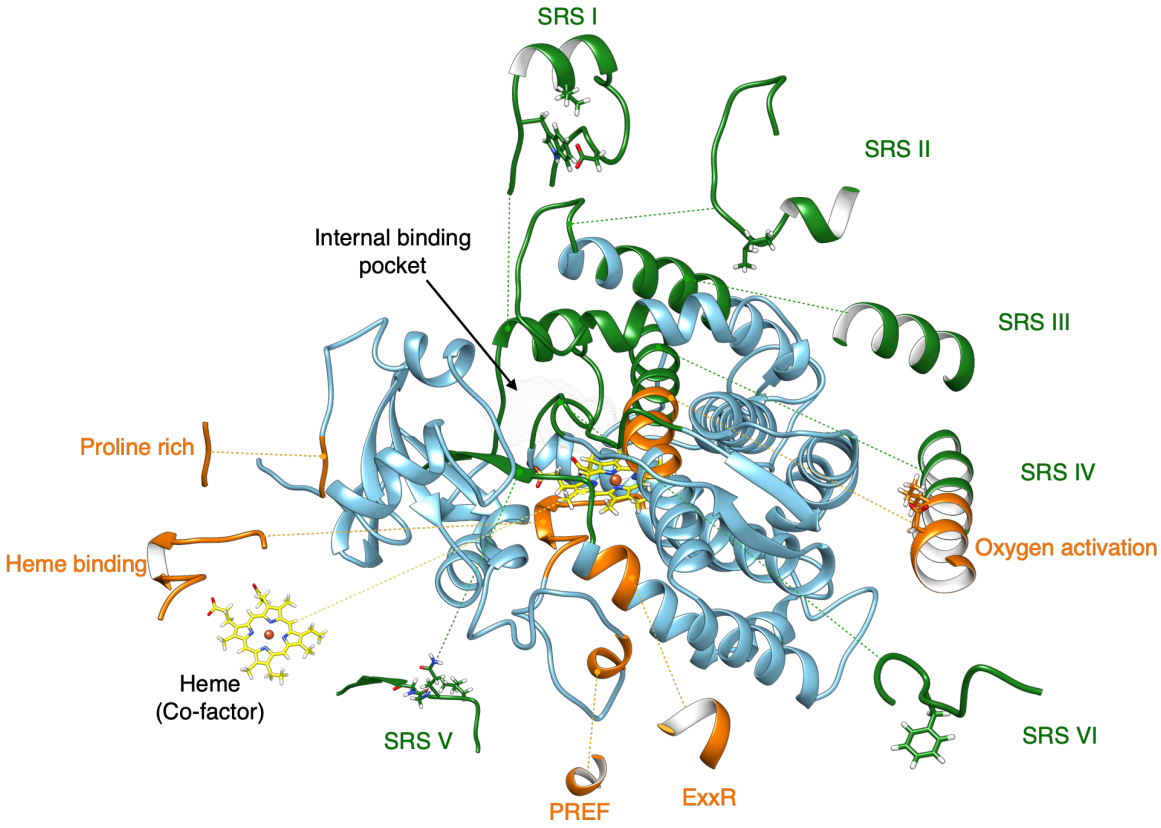


Figure 1-4 Homology modelling of CYP716A12.

Orange color highlighted signature region. Green color highlighted substrate recognition site.

The analysis of CYPs family diversity is important to understand the contributions of these enzymes to plant evolution (Nelson and Werck-Reichhart, 2011). CYPs are generally classified by their homology and named; accordingly, the first numbers following CYPs represent the family, the following letters indicate the subfamily, and the last numbers indicate the individual locus. A group of CYPs should share at least 40% of amino acid identity to belong to the same family and at least 55% to belong to the same subfamily (Nelson et al., 2004; Nelson, 2006). For Example, CYP716A is a CYPs that belongs to the subfamily A of CYP716 family.

1.3.1 CYP716 family enzyme

An enzyme of the CYP716 family belongs to the CYP85 clan of CYPs, which corresponds to non-A type P450s generally regarded as housekeeping enzymes, providing an evolutionary context for this clade's origin in the primary metabolism of triterpenoid compounds (Miettinen et al., 2017). It has been isolated from a wide range of plant families that are functionally characterized as triterpene oxidases. The CYP716 is a key enzyme family involved in the biosynthetic pathway for diversifying triterpenoid products in eudicot plants (**Figure 1-5**).

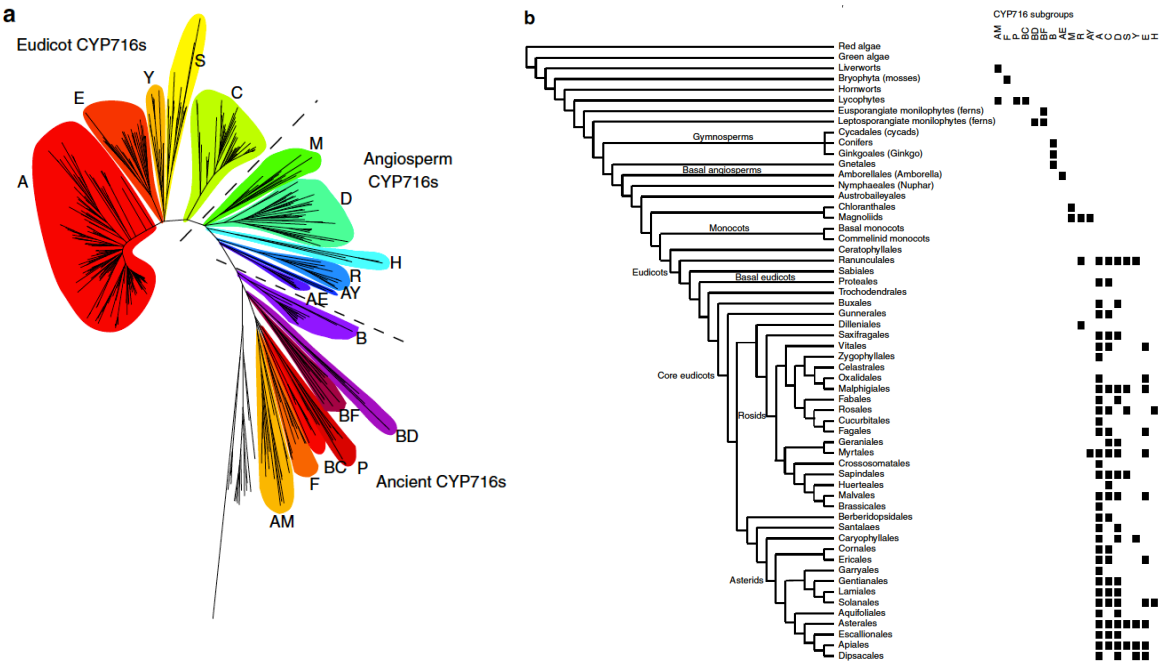


Figure 1-5 A Maximum likelihood analysis of the CYP716 family phylogeny. The picture was taken from Miettinen et al., (2017).

(a) The enzyme members of the CYP716 family form distinct subgroups that can be divided into three categories: eudicot, angiosperm, and ancient CYP716s. (b) A distribution of CYP716 subgroups by plant order, indicating the time of emergence of distinct subgroups and implying diversification of the CYP716 family in early eudicots.

To better understand the evolution of the CYP716 family enzymes, I performed a wide-ranging phylogenetic analysis using full-length CYP716 family enzyme sequences that were made publicly available. This data set contained 383 sequences from 181 plant species, all of which were identified and named in a systematic manner. The outlier CYP51G1v1, isolated from the plant

Nicotiana tabacum, was used to establish the root of the CYP716 phylogenetic tree. The CYP716 family is clearly distinguished from its closest relatives in the phylogenetic tree shown below. Even though some members of the CYP716 family share less than 40% amino acid identity, they were all classified as members of the same family because they all share the same amino acid sequence (**Figure S-1**).

1.3.2 CYP716A subfamily enzyme

The CYP716A subfamily is highly conserved in this enzyme family (Fukushima et al., 2011; Fukushima et al., 2013; Miettinen et al., 2017). The modification catalyzed by this subfamily is mainly a three-step oxidation of α-amyrin, β-amyrin, and lupeol at the C-28 position into highly valued compounds such as ursolic acid, oleanolic acid, and betulinic acid, respectively (Carelli et al., 2011; Fukushima et al., 2011; Yasumoto et al., 2016; Miettinen et al., 2017; Yasumoto et al., 2017; Sandeep et al., 2019). For example, CYP716A12 from the model legume *Medicago truncatula*, CYP716A15 from *Vitis vinifera*, CYP716A48 from *Olea europaea*, and CYP716A49 from *Beta vulgaris* are known enzymes that catalyze this three-step oxidation process (**Figure 1-6)**.

Three step site-specific oxidation at the C-28 position of α-amyrin, β-amyrin, and lupeol catalyzed by CYP716A subfamily enzyme to produce highly valued compounds such as ursolic acid, oleanolic acid, and betulinic acid, respectively. bAS, β-amyrin synthase; aAS, α-amyrin synthase; LUS, lupeol synthase.

However, there are some CYP716A subfamily enzymes that have been found to catalyze other triterpenoid oxidation reactions, including C-16α and C-22α oxidations (Yasumoto et al., 2016), C-16β of β-amyrin (Tamura et al., 2017), and C-3 oxidations of β-amyrin (Moses et al., 2015). Significant differences in catalytic activity and substrate specificity of these enzymes led to heterologous production of various oxidized triterpenoids in engineered yeast (Suzuki et al., 2018; Dale et al., 2020).

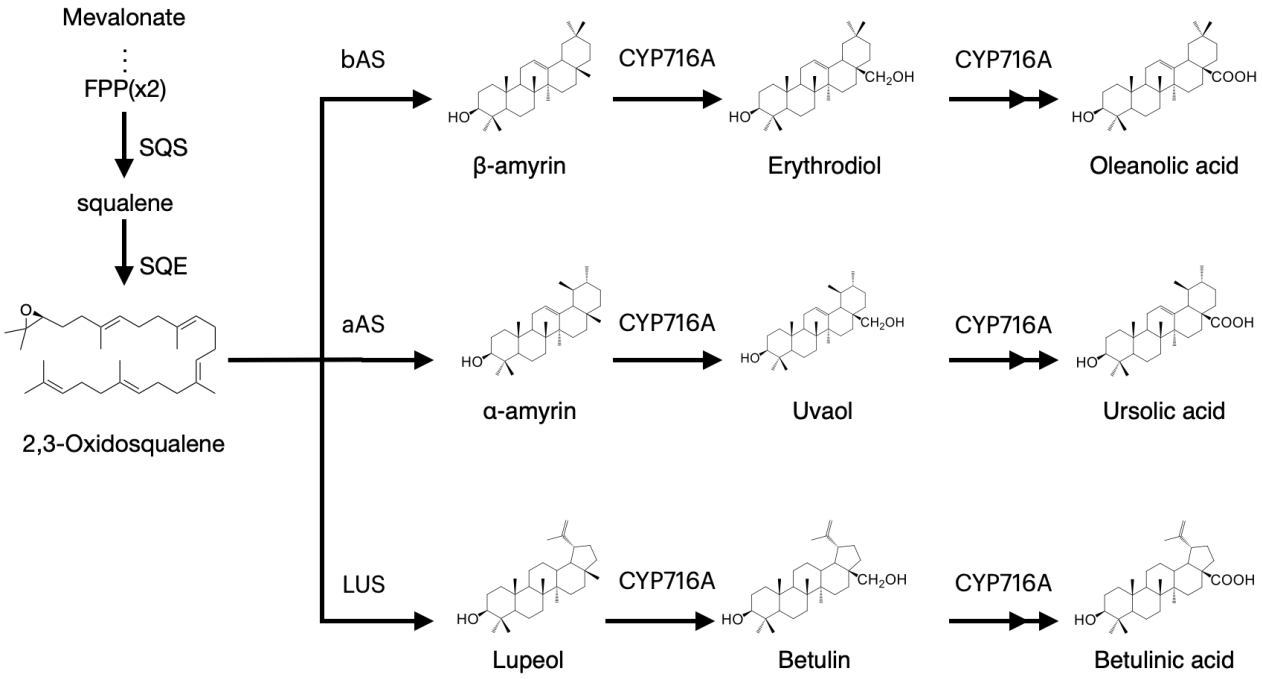


Figure 1-6 Propose of biosynthetic pathway of C-28 oxidized triterpenoids catalyzed by CYP716A subfamily enzymes. This biosynthetic pathway was modified from Suzuki et al., (2018).

Additionally, cytochrome P450s catalytic activity and substrate specificity are considered as a rate-limiting factors in triterpenoid biosynthesis, affecting the production of C-28 oxidized triterpenoids in heterologous hosts (Zhou et al., 2016; Czarnotta et al., 2017; Suzuki et al., 2018; Jin et al., 2019).

1.4 Objective and strategies

Triterpenoids are a class of specialized plant metabolites with a high therapeutic value and a wide range of structural diversity. Cytochrome P450 monooxygenases (CYPs) are an enzyme family that catalyzes the site-specific oxidation of the triterpene backbone, thereby contributing to the structural diversity of triterpenoids. The CYP716 enzyme family has been isolated from various plant families as triterpenoid oxidases. However, their experimental crystal structures are unavailable, and the detailed catalytic mechanism remains elusive. Therefore, this research aims to

improve bioactive triterpenoid production in the heterologous host through protein engineering of CYP716A subfamily enzyme.

To achieve this purpose, firstly, I selected CYP716A12, isolated from a model legume plant *M. truncatula* with a plenty of genomic data available, as the first of CYP716A subfamily enzymes that have been functionally characterized. I performed homology modeling, structural alignment, *in silico* site-directed mutagenesis, and molecular docking analysis to search and screen key amino acid residues relevant to the catalytic activity and substrate specificity of the CYP716A subfamily enzyme in triterpenoid biosynthesis. An *in vivo* functional analysis using engineered yeast that endogenously produced plant-derived triterpenes was performed to elucidate the results. The product profile of CYP716A12 got modified when the amino acids in the signature region and the substrate recognition sites (SRSs) were substituted. In the results, I identified amino acid residues that controlled the substrate contraction of the enzyme (D292) and engineered the enzyme to improve its catalytic activity and substrate specificity (D122, I212, and Q358) for triterpenoid biosynthesis. In addition, I generated the mutants of Arabidopsis thaliana CYP716A1 (S356) and CYP716A2 (M206, F210) by changing the properties of key residues in substrate recognition sites (SRSs) to improve the catalytic activity at C-28 on the different triterpene backbones. Finally, I used the benefits of identified mutant CYP716A12 combined with the Tsukuba system, one of the most efficient agroinfiltration-based transient protein expression systems, to evaluate the applicability of the Tsukuba system for triterpenoid production in *N. benthamiana* through pathway reconstruction.

This research has the potential to help in the production of the desired triterpenoids in engineered yeast by increasing the catalytic activity and substrate specificity of plant CYP716A subfamily enzymes. In addition, this research provides an alternative transient expression platform for high titer of valuable triterpenoid productions through heterologous expression in *N. benthamiana* contributed to synthesizing and accessing previously inaccessible triterpenoid or the

other natural products and its analogs, as well as the possibility of revitalizing drug discovery pipelines.

1.5 Thesis outline

This thesis is divided into five chapters. Chapter 1 describes the general introduction of the importance of triterpenoids and their biosynthesis, involvement of CYPs as key important enzyme for the structural diversification in triterpenoids biosynthesis. Introduction of CYPs, CYP716 family and CYP716A subfamily enzyme for this study are also described briefly in this chapter.

In chapter 2, I used several bioinformatics approaches including, investigated through homology modeling, structural alignment, *in silico* site-directed mutagenesis, and docking profile from molecular docking analysis, to search the key amino acid residues in the ligand-binding site of CYP716A12 that required for the catalytic activity of enzyme and substrate specificity in triterpenoid biosynthesis. To elucidate the results of the bioinformatics analysis, I performed an *in vivo* functional analyses of candidate mutants. I found the different of triterpenoids product profiles in *S. cerevisiae* engineered harboring mutant CYP716A12 enzymes when key amino acids at ligand-binding sites were substituted.

In chapter 3, because I succeeded in identifications of key amino acid residues involve in catalytic activity and substrate specificity for site-specific oxidation at C-28 of different triterpene backbones in triterpenoid biosynthesis via integrating bioinformatics approaches and *in vivo* functional analysis in engineering yeast from chapter 2. In this chapter, I successfully designed CYP716A1 and CYP716A2 with improving the catalytic activity at C-28 against different triterpene backbones by following the framework from the chapter 2.

In chapter 4, I selected mutants CYP716A12 and introduced into *N. benthamiana* using agroinfiltration-based transient protein expression system via Tsukuba system. Notably, transient protein expression in *N. benthamiana* revealed that substituting amino acids on substrate recognition sites significantly increased the production of valuable triterpenoids. In this chapter, I established a general strategy for rapidly producing substantial quantities of

triterpenoids and generating valuable and difficult-to-obtain triterpenoids with potent biological activity.

Finally, in chapter 5, the conclusions of this study and the future perspectives were discussed.

Chapter 2 : Identification of key amino acid residues toward improving the catalytic activity and substrate specificity of plant-derived cytochrome P450 monooxygenases CYP716A subfamily enzyme for triterpenoid production in *Saccharomyces cerevisiae*

2.1 Introduction

Triterpenoids are a group of plant-specific metabolites that contain a variety of bioactive chemicals with pharmacological effects, including anti-cancer, anti-inflammatory, anti-viral, anti-diabetes, and anti-microbial activity (Yadav et al., 2010; Yan et al., 2014; Jin et al., 2019). However, their production in plants is extremely low, and their purification and chemical synthesis are difficult. In addition, the whole metabolic process for most species has not yet been elucidated. A synthetic biology methodology utilizing expression in heterologous hosts, such as the yeast *Saccharomyces cerevisiae*, has been examined as an alternate method of triterpenoid biosynthesis (Fukushima et al., 2011; Suzuki et al., 2018; Dale et al., 2020). Cyclization and site-specific oxidation processes, catalyzed by oxidosqualene cyclase (OSC) and cytochrome P450 monooxygenases (CYPs), convert 2,3-oxidosqualene to produce triterpenoids (**Figure 1-2**).

A CYP716 family enzyme, specifically described as a triterpene oxidase, has been isolated from a wide range of plant species. Eudicot plants use a biosynthetic route that involves the CYP716 enzyme family to produce a wide variety of triterpenoid compounds. Among these enzymes, the CYP716A subfamily is the most conserved (Fukushima et al., 2011; Nelson and Werck-Reichhart, 2011; Fukushima et al., 2013; Miettinen et al., 2017). To produce useful chemicals including ursolic acid, oleanolic acid, and betulinic acid, this subfamily primarily catalyzes the three-step oxidation of α-amyrin, β-amyrin, and lupeol at the C-28 position respectively (**Figure 1-6**) (Carelli et al., 2011; Fukushima et al., 2011; Yasumoto et al., 2016; Miettinen et al., 2017; Yasumoto et al., 2017; Sandeep et al., 2019). Examples of enzymes that catalyze this three-step C-28 oxidation process

include CYP716A12 from the model legume *Medicago truncatula*, CYP716A15 from *Vitis vinifera*, CYP716A48 from *Olea europaea*, and CYP716A49 from *Beta vulgaris*. Some CYP716A subfamily enzymes have been found to catalyze other triterpenoid oxidation reactions, including C-16 and C-22 oxidations (Yasumoto et al., 2016), C-16β of β-amyrin (Tamura et al., 2017), and C-3 oxidations of β-amyrin (Moses et al., 2015). Nonetheless, engineered yeast can produce a wide variety of oxidized triterpenoids due to the divergent of enzyme catalytic activities and substrate specificities (Suzuki et al., 2018; Dale et al., 2020). The production of C-28 oxidized triterpenoids in heterologous hosts is also affected by the catalytic activity and substrate specificity of cytochrome P450s, which is a rate-limiting factor in triterpenoid biosynthesis (Zhou et al., 2016; Czarnotta et al., 2017; Suzuki et al., 2018; Jin et al., 2019). It has been discovered that numerous C-28 oxidized triterpenoids serve as precursors to various bioactive triterpenoids of commercial interest, such as platycodin D, which has been reported to have anti-cancer and anti-viral properties (Liu, 1995; Jeon et al., 2019; Kim et al., 2021). Onjisaponin F is utilized to boost nasal anti-influenza virus IgA antibody titers (Li et al., 2016; Peng et al., 2020), and QS-21 is used as a vaccine adjuvant in clinical trials against HIV and malaria (Kensil et al., 1998; Liu et al., 2002; Zhu and Tuo, 2016).

Despite their widespread distribution and functional characterization of numerous CYP716A subfamily enzymes, their crystal structure is not yet available. Little is known about the molecular mechanisms that create the significant differences in catalytic activity and substrate specificity between these subfamilies (Yasumoto et al., 2016; Miettinen et al., 2017; Suzuki et al., 2018). However, bioinformatics approaches combined with knowledge from other CYP families may mitigate these issues (Fukushima et al., 2013; Seki et al., 2015; Miettinen et al., 2017). This computational analysis is more efficient than conventional methods like random mutagenesis at identifying candidate amino acid residues likely to be involved in catalytic activity of CYP716 family enzyme.

The goal of this chapter is to correctly identify key amino acid residues for enhancing the catalytic activity and substrate specificity of CYP716As in *S. cerevisiae* triterpenoid production. To accomplish this and serve as a model for this investigation, I selected CYP716A12, which was isolated from the model legume *Medicago truncatula*. To date, CYP716A12 is the first member of the CYP716A subfamily to have performed functional characterization. The enzyme catalyzes the regioselective oxidation of α-amyrin, β-amyrin, and lupeol at C-28 (Fukushima et al., 2011; Suzuki et al., 2018). Homology modeling, structural alignment, *in silico* site-directed mutagenesis, and a docking profile from molecular docking analysis were used to determine the important amino acid residues in the ligand-binding site of CYP716A12 required for catalytic activity of the enzyme and substrate specificity. *In vivo* functional analyses of candidate mutants were carried out to gain insight on the findings of the bioinformatics investigation. *S. cerevisiae* engineered to harbor mutant CYP716A12 enzymes showed altered product profiles after key amino acids at ligand-binding sites were altered.

This investigation has the potential to improve protein engineering techniques for producing desirable triterpenoids.

2.2 Materials and methods

2.2.1 Chemical authentic standards

All the triterpenoid standards (α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid, lupeol, betulin, and betulinic acid) were purchased from Extrasynthase (France).

2.2.2 Homology modeling

I chose the CYP716A subfamily enzymes CYP716A1, CYP716A2, CYP716A12, CYP716A15, CYP716A48, and CYP716A49, which were functionally characterized as C-28 multifunctional oxidases. The protein sequences of the selected CYP716As are presented in Supporting information Section 2. A Protein BLAST (BlastP) search with a BLOSUM62 matrix

was conducted against the experimental structure in the Protein Data Bank (PDB) database to discover homologous sequences (Madden et al., 1996). The structure model with the highest amino acid sequence homology was chosen as a template. The target protein sequences were aligned with the model structure, and construction, refinement, and validation of the homology models were performed using Modeller 9.21 software (Webb and Sali, 2017). The model's non-water HETATM residues were incorporated into the homology model. The structure was optimized using a multi-domain assembler (MDA). Chimera 1.14 was used to visualize all homology models (Pettersen et al., 2004). VERIFY3D was used to evaluate the quality of homology modeling, with at least 80% of the amino acids scoring >= 0.2 in the 3D/1D profile. (Eisenberg et al., 1997).

2.2.3 Molecular docking analysis

Molecular docking analysis was used to determine the possible interactions between CYP716As and their putative ligands (α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid, lupeol, betulin, and betulinic acid). The chemical structures of the putative ligands were obtained from the Cambridge Structural Database (CSD) and the PubChem database (https://pubchem.ncbi.nih.gov). After molecular docking analysis, the scoring function (docking score) was used to predict the binding affinity between the two molecules. A I-TASSER modeling pipeline was utilized to predict the ligand-binding site area for molecular docking analysis (Yang et al., 2015). An analysis of molecular docking was performed using Auto Dock Vina 1.1.2 (Trott and Olson, 2010) via Chimera 1.14 (Pettersen et al., 2004) with the box center: 50 × -5 × 15 and box dimensions of 30 × 30 × 30 along X, Y, and Z, respectively, in the receptor coordinate system. As criteria for the selection of binding poses, the lowest docking score, the shortest distance between the oxidation target sites (C-28 of triterpene backbones), and the heme reaction center to the ferrous iron fragment (Fe^{2+}) moiety were chosen (Yuki et al., 2012; Geisler et al., 2013). The catalytic site environment of the enzyme, including catalytic site capacity, amino acid side chain

position and conformation, ligand orientation, and ligand position, was also chosen as a criterion for docking. (Docking profiles).

2.2.4 Structural alignment

Homology modeling of CYP716A12 was used as the reference structure. CYP716A1, CYP716A2, CYP716A15, CYP716A48, and CYP716A49 were matched and aligned with the ClustalX algorithm for homology modeling comparison (Thompson et al., 2002). The residue-residue distance cut-off was set at 5 Å. The residue aligned in a column if it was within the cut-off criteria of at least one other. A Chimera 1.14 was used to create visualizations of the analysis results (Pettersen et al., 2004).

2.2.5 *In silico* site-directed mutagenesis

Candidate residues for site-directed mutagenesis were decided by using the PyMol 2.0.5 software program (Schrodinger, 2015). The native residue of the target position in the CYP716A homology model was replaced with a representative amino acid based on its properties; non-polar alanine (A) and tyrosine (Y), non-charged polar glutamine (Q), basic amino acid lysine (K), acidic amino acid aspartic acid (D), and residues in the loop have been replaced with proline (P). The substitution of proline in the loop enhances the protein's stability and kinetics (Matthews et al., 1987; Yuan et al., 2017). To keep the hydrogen bond in the homology model, an additional hydrogen molecule was introduced and retrained. The PDB file was generated from the exported structure (.pdb). A Chimera 1.14 was used to optimize and display the protein's structures (Pettersen et al., 2004). At last, VERIFY3D was used to evaluate the quality of the structures (Eisenberg et al., 1997).

2.2.6 Plasmid construction

Bioinformatic predictions were further used experimentally to verify the function of CYP716A12 mutations. The entry clones of CYP716As, pENTR-CYP716A12, pENTR-

CYP716A15, pENTR-CYP716A48, and pENTR-CYP716A49 were obtained from previous research (Seki et al., 2008; Fukushima et al., 2011; Yasumoto et al., 2016). The mutations were introduced into CYP716A12 through site-directed mutagenesis using the PrimeSTAR® Mutagenesis Basal Kit (TaKaRa Bio, Shiga, Japan). The specific primers used for site-directed mutagenesis were detailed in (**Table S-1)**. Using the Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific), yeast expression clones were generated by transferring the coding sequence (CDS) of CYP716A12 into pELC-GW (Seki et al., 2008), pYES-DEST52 (Thermo Fisher Scientific), and a gateway-compatible version of pESC-HIS (Agilent Technologies) as destination vectors (Seki et al., unpublished). Using the same method, yeast expression clones of the other CYP716As were constructed, including pELC-CYP716As, pYES-DEST52-CYP716As, and pESC-HIS-CYP716As.

2.2.7 *In vivo* functional analysis in engineering yeast

The expression clones for OSC (pYES3-ADH-*OSC1*, pYES3-ADH-*aAS,* or pYES3-ADH-*LUS*) (Fukushima et al., 2011) were transformed into *Saccharomyces cerevisiae* INV*Sc*1 (*MATa his3D1 leu2 trp1-289 ura3-52 MATα his3D1 leu2 trp1-289 ura3-52*; Invitrogen) using the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA, USA). Transformants were chosen on SD selection medium with 2% glucose and incubated for three days at 30 °C. Each transformant that had been chosen was grown in 2 ml of the SD selection medium. It was incubated overnight at 30 °C, shaking at 200 rpm. The overnight culture (500 l) was transferred into 5 ml of the same medium broth and then incubated at 30 °C for two days at 200 rpm. To induce CYP716As expression, yeast cells were harvested and resuspended in 5 ml of an appropriate SD medium containing 2% galactose and grown at 30 °C for two days at 200 rpm. The yeast cultures were then frozen at -80 °C until extraction could be performed. All assays were performed in triplicates (three independent assays were obtained from different colonies to confirm the results). *S. cerevisiae*

INV*Sc*1 harbors the OSC expression vector (pYES3-ADH-*OSC1*, pYES3-ADH-*aAS,* or pYES3-ADH-LUS), with the empty vectors pELC-GW, pYES-DEST52, and pESC-HIS used as controls.

2.2.8 Analysis of triterpenoid production in engineered yeast

Before yeast culture extraction, 50 µl of an internal standard (uvaol or betulin, 100 ppm in methanol) was added to quantify triterpenoid production. A 5 ml of yeast cultures (cells and spent media) were extracted with 5 ml of ethyl acetate (Wako, Osaka, Japan). The mixture was then vortexed for 30 minutes and sonicated. Using a Pasteur pipette, the organic phase was transferred to a new tube following centrifugation at 9000 g for 5 minutes. Using a centrifugal evaporator, the extracted samples were evaporated for 45 minutes or until completely dry. This procedure was repeated three times. The remaining powder was resuspended in methanol (1 ml). A new Pasteur pipette was used to transfer the obtained samples into vials. Until the next analysis, the samples were stored at 4 °C.

To quantify triterpenoid production in engineered yeast, 50 µl of the sample solution (β-amyrin-producing yeast extracts) or 100 µl of the sample solution (α-amyrin or lupeol-producing yeast extracts) was transferred to vial inserts and then to a centrifugal evaporator for 60 min or until completely dry. Finally, the evaporated pellet was derivatized with 50 µl of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (Sigma-Aldrich) for 30 min at 80 °C before Gas Chromatography-Mass Spectrometry (GC-MS) analysis. Like the method described above, Fifty microliters of authentic standard solutions (α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid, lupeol, betulin, and betulinic acid, 10 ppm. in methanol) were used.

The metabolite analysis of sample extracts from lupeol-producing yeast was performed using GC-MS analysis with a 5977A MSD (Agilent Technologies) coupled with a 7890B GC system (Agilent Technologies) and a DB-1ms (length 30 m, 0.25 mm internal diameter, 0.25 µm film thickness: Agilent Technologies). However, an HP-5ms (length 30 m, 0.25 mm internal diameter, 0.25 µm film thickness; Agilent Technologies) capillary column was used for metabolite analysis

of extracting the sample from α-amyrin or β-amyrin-producing yeast extracts. The injection component and the MSD transfer line were set to 250 °C. The oven temperature was programmed as follows: 80 °C for 1 min, followed by a rise to 300 °C at a rate of 20 °C min^{-1}, and held at 300 °C for 14 min. The carrier gas was helium (He), and the flow rate was 1 ml min^{-1}. Mass spectra were recorded by scanning the *m/z* range of 50–750. Retention times and mass spectral patterns were compared to reference standards to locate the peaks (Fukushima et al., 2011; Yasumoto et al., 2016). The relative concentrations of the oxidized triterpenoids in the extracted samples were determined by comparing the analyte's peak area to the internal standard's peak area. To quantify the concentration of triterpenoids in the engineered yeast, a standard curve was constructed using α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid, lupeol, betulin, and betulinic acid. The absolute quantity of triterpenoids in yeast was determined by comparing the standard curve to the experimental data (**Supporting information Section 4, S5, Figure S-2).**

2.2.9 Statistical analysis

One-way analysis of variance (ANOVA) was used to determine the difference in oxidized triterpenoid levels between the engineered yeast strains, and Tukey's test was used to determine the significance of the differences between the means. *p*-values less than 0.05 ($p \leq 0.05$) were considered significant in this study. All statistical analyses were conducted using JASP 0.16 for macOS (JASP Team, University of Amsterdam, Amsterdam, the Netherlands).

2.3 Results

2.3.1 Comparative protein homology modeling of *M. truncatula* CYP716A12

CYP716A12 was selected as a model because it is the first CYP716A subfamily enzyme to be functionally characterized; however, there is no experimental crystal structure for a CYP716A subfamily enzyme. Enzymatic catalytic activity and substrate specificity cannot be determined purely based on the primary protein structure (amino acid sequence). To visualize the location and

orientation of each amino acid in protein folding structures, particularly at the ligand-binding site, it is important to model and analyze the tertiary structure. (Prall et al., 2016). As a result, a comparative homology modeling strategy was utilized to construct a homology model of CYP716A12. CYP716A12 homology modeling was designed and modeled based on its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (template), CYP120A1 (2VE3) chain A (Kuhnel et al., 2008), 38.1% identical (**Figure 2-1A)**. During homology modeling, unmatched template residues were removed.

Finally, 446 amino acid residues (30-475) were added to the optimized homology model. The functionally relevant conserved CYP domain was highlighted in the CYP716A12 homology model. Substrate recognition sites (SRSI–VI) are highlighted in green. Alternatively, the signature regions are highlighted in orange (**Figure 2-1B)**. The quality of this homology modeling surpassed VERIFY3D's quality standards (Eisenberg et al., 1997) evaluation criteria of 89.24% of residues with an average 3D-1D score greater than 0.2. The same strategy was applied to generate homology models for CYP716A15 **(Figure S-3A)**, CYP716A48 **(Figure S-3B)**, CYP716A49 **(Figure S-3C)**.

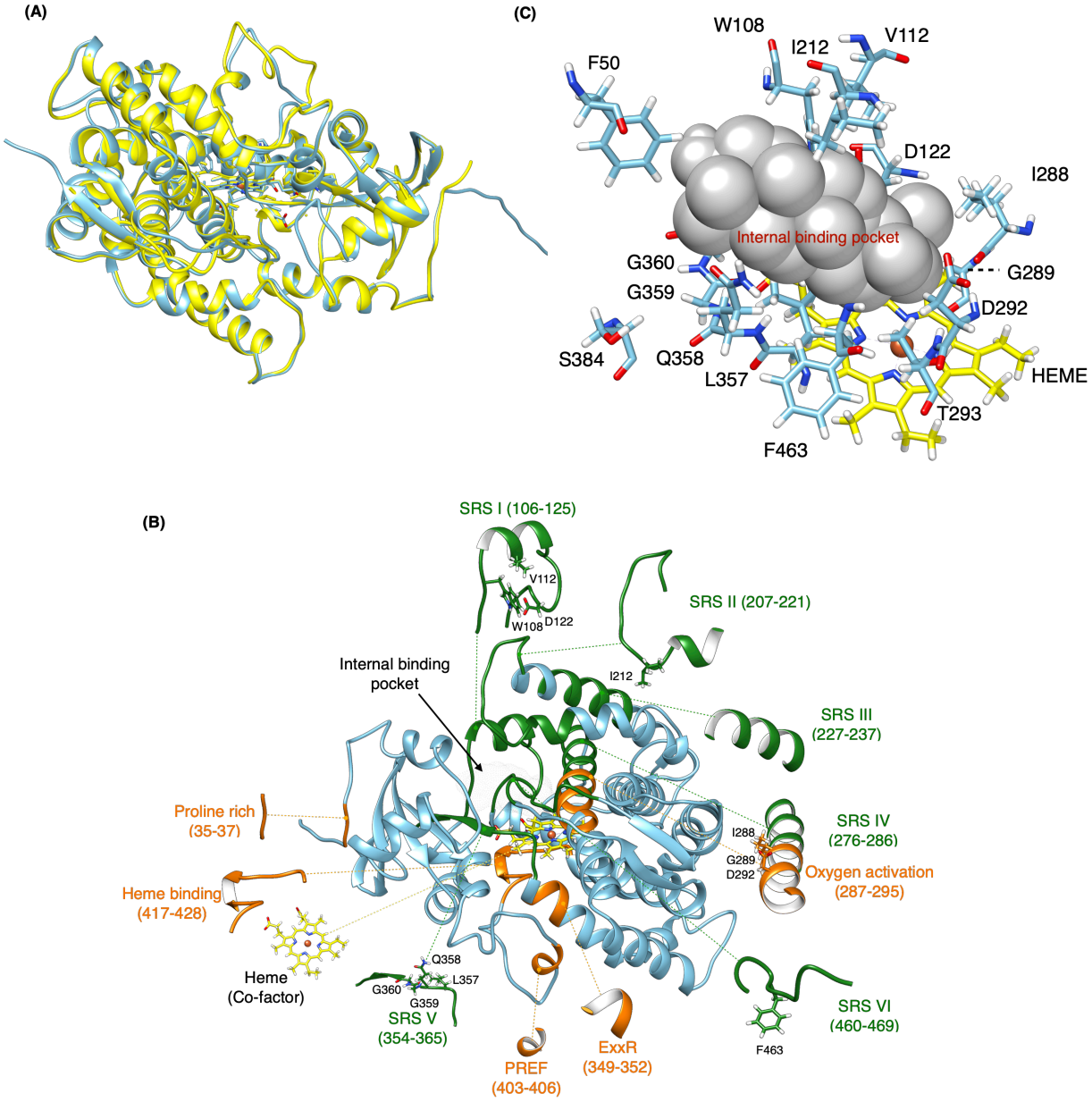


Figure 2-1 Comparative homology modeling of CYP716A12.

(A) The structure of retinoic acid bound cyanobacterial CYP120A1 chain A (2VE3_A) shows the highest similarity in amino acid sequence among of the experimental structures. It was selected as a template model (Yellow) and homology model of CYP716A12 was in blue. (B) Homology modelled CYP716A12 with a predicted ligand-binding site from molecular docking analysis. Substrate recognition sites (SRS) was highlighted green and Signature region was highlighted in orange. This homology model exceeded the quality criteria verified by VERIFY3D by 89.24%. (C) The spatial positions of 15 (15 of 479) candidate amino acid residues were putatively involved in the catalytic activity of the CYP716A12 enzyme in the ligand-binding site via a molecular docking analysis and the I-TASSER pipeline.

2.3.2 Identification of candidate amino acid residues putatively involved in the catalytic activity of the CYP716A12 enzyme

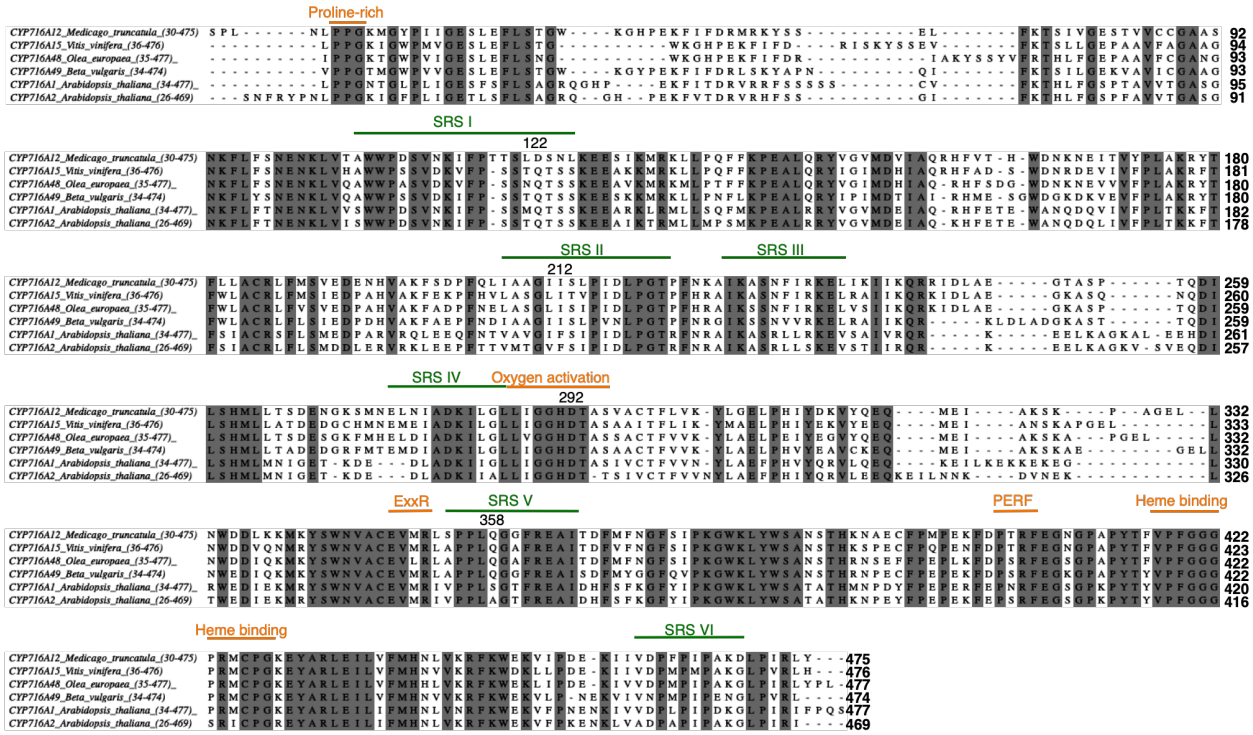
Using the I-TASSER pipeline, the small spheres of the homology model were placed in the ligand-binding pocket to highlight the substrate-binding region above the heme cofactor (Yang et al., 2015). This area was used for molecular docking analysis (**Figure 2-1B, C**). Using the I-TASSER pipeline, the spatial positions of 15 candidate amino acid residues (15 of 479) were predicted as ligand binding site residues. Most of these residues were conserved at the catalytic site. Thirteen of the 15 residues in the conserved CYP motifs were in SRSI (W108, V112, D122), SRSII (I212), oxygen activation (I288, G289, D292, T293), SRSV (L357, Q358, G359, G360), and SRSVI (L357, Q358, G359, G360) (F463). Two residues were discovered simultaneously at the entrance of the binding pocket (F50 and S384) (**Figure 2-1B, C**). The catalytic site was not predictive of the amino acids discovered on SRSIII and SRSIV. In addition, the vast majority of amino acid residues in these regions are located opposite or away from the ligand-binding site. To investigate the evolutionary and functional relationships that may not be discernible by comparing CYP716As, structural alignment was performed utilizing the side-chain orientation and structural conformation of the homology model. The structure-based amino acid sequence alignment based on structural alignment is shown in **Figure 2-2A**. Most of the amino acid sequences in the signature motif were conserved, as determined by structural alignment. Aspartic acid 292 is a conserved residue in oxygen activation motif of the CYP716A subfamily. Its side chain is pointing to the reaction center of the catalytic site (**Figure 2-2B–2-2G**). In contrast, the side chains of the remaining conserved residues (I288, G289, T293) pointed away from the ligand-binding pocket.

Interestingly, the amino acid residues in SRS were variable. D122 in the SRSI of CYP716A12; nevertheless, glutamine was a conserved in several other CYP716As. I212 in SRSII of CYP716A12 was also a conserved SRSII residue in CYP716As. Phenylalanine was a conserved in this region, except in CYP716A1 and CYP716A2. On CYP716As SRS V, Q358 was a conserved residue. The conserved residues in these aligned CYP716A1 and CYP716A2 are S356 and A352,

respectively. F463, located in SRSVI, was not a structurally conserved residue; however, methionine (M) was a conserved in CYP716A15 (M464), CYP716A48 (M463), and CYP716A49 (M463).

The most well-characterized CYP716A catalyzes a three-step oxidation reaction at the C-28 position of β-amyrin, α-amyrin, and lupeol, yielding hydroxyl, aldehyde, and carboxylic moieties. However, there is a distinction between the enzyme's catalytic activity and substrate specificity. whereas certain CYP716A enzymes oxidize triterpene backbones at other carbon positions (Fukushima et al., 2011; Miettinen et al., 2017; Yasumoto et al., 2017). Thus, the predicted residues may have a direct impact on catalytic mechanisms and enzymatic activity. Thus, all predicted ligand-binding site residues were considered potential candidates for amino acid mutations.

(A)



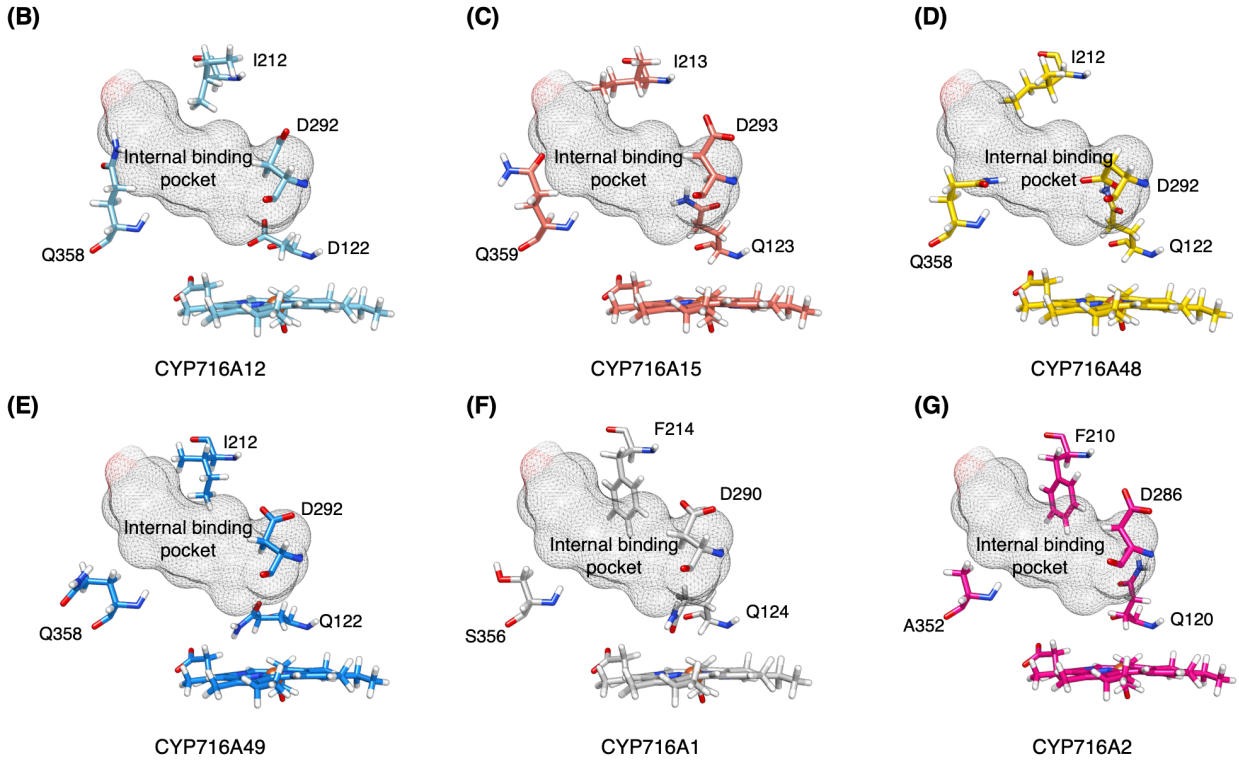


Figure 2-2 Amino acid residues predicted to be in the substrate recognition sites and signature region via structural alignment of CYP716A12 and selected CYP716As.

Structural alignment of selected CYP716As. (A) Structure-based Amino acid sequence alignment of CYP716A12 against selected CYP716As with conserved CYP domains relevant to functional analysis. Predicted substrate recognition sites (SRS), (green); SRSI (106-125); SRSII (207-221); SRSIII (227-237); SRSIV (276-286); SRSV (354-365); SRSVI (460-469), and Signature region (orange); Proline-rich (35-37); Oxygen activation (287-295); ExxR (349-352); PREF (403-406); Heme-Binding (417-428) are labeled. Comparison of focusing residues on catalytic activity site of (B) CYP716A12 against (C) CYP716A15, (D) CYP716A48, (E) CYP716A49, (F) CYP716A1, and (G) CYP716A2.

2.3.3 Screening candidate of amino acid residues for site-directed mutagenesis

To elucidate the substrate contacts of CYP716A12 that may affect the catalytic activity of the enzyme, a molecular docking analysis was conducted on the wild-type and mutant versions of the candidate residues identified above. Using *in silico* site-directed mutagenesis, the enzyme's catalytic activity and substrate specificity-related ligand-binding sites were screened for important

amino acid residues. Then, a docking analysis was performed on molecules. When only four of the fifteen native amino acid residues at the ligand-binding site were replaced, the docking profiles were modified.

When D292 was substituted in the oxygen activation motif, the docking profile closely resembled that of the wild-type CYP716A12. When D292 was replaced with alanine, however, the hydrophobicity of the amino acid residue and the environment of the ligand-binding pocket changed (**Figure 2-3B)**. Due to its location in the oxygen activation motif, this amino acid residue may impact the oxygen activation and electron transfer processes of the enzyme, thereby influencing its catalytic activity and substrate specificity.

Similarly, when the amino acid residues at SRSI, II, and V were changed, the docking profiles changed. When D122 was substituted with glutamine, the direction and distance from the target site (C-28) were straighter than when they were centered on the heme-Fe^{2+} reaction center. This substitution replaces the terms distal to the polarity and distance from the ligand-binding pocket (**Figure 2-3C**). When I212 in SRSII was replaced with proline, the binding orientation and substrate distance changed. Additionally, the position of C-28 in the reaction center was altered (**Figure 2-3D**). C-28 was oriented more directly toward the reaction center when proline 358 was substituted for Q358 at SRS V. The ligand-binding pocket's capacity increased in part due to a change in the orientation of the side chain (**Figure 2-3E)**.

A Q358 (SRSV) was identified at an opposite position from D122 (SRS I) and isoleucine 212. (SRSII). As previously stated, docking profiles were altered when these amino acid residues were replaced. Therefore, it is reasonable to hypothesize that the double replacement of residues on either side of the ligand-binding site is key for modifying the enzyme's catalytic activity and substrate specificity. Consequently, I designed a replacement variant of CYP716A12_Q358P containing either Q122 or P212, resulting in the formation of CYP716A12 D122Q_Q358P and CYP716A12_I212P_Q358P. The substrate orientation profile in the binding pocket of the CYP716A12_D122Q _Q358P variant did not differ significantly from the profile of the substituted

residue, as stated previously. However, the properties, environment, and binding capacity of the ligand-binding site were altered. (**Figure 2-3F)**. CYP716A12_I212P_Q358P has a β-amyrin-binding orientation like that of a single isoleucine 212 substitution; however, the orientation of erythrodiol binding was significantly altered in this variant (**Figure 2-3G)**.

Comparing the four substituted residues to the homology model of CYP716A12 showed a docking profile that was different from the other chosen CYP716As (**Figure 2-2B–2-2G)**. The selected CYP716A molecular docking analysis of the β-amyrin backbone was shown in **Figure S3**. The ligand-target sites (β-amyrin and erythrodiol) were within the range and orientation for site-specific oxidation in all homology models using the docking profile. A homology model revealed a similar docking profile for ligands in wild-type of CYP716A12 (**Figure S-4A)** and CYP716A15 (**Figure S-4B)** whereas, it was different from CYP716A48 (**Figure S-4C)** and CYP716A49 (**Figure S-4D)**, respectively.

The same analyses were performed on molecular docking analysis to reveal the docking profile against the other putative substrates, α-amyrin **(Figure S-5, S-6)**, and lupeol backbones **(Figure S-7, S-8)**. This analysis revealed differences in the docking profile of the engineered variant of CYP716A12, which further altered the catalytic activity of enzyme and substrate specificity.

The preliminary results obtained from homology modeling, structural alignment, *in silico* site-directed mutagenesis, and the docking profile from molecular docking analysis suggested a strategy for understanding the amino acid differences at each CYP716A ligand-binding site. The ligand-binding site's properties were changed when candidate amino acid residues were substituted. This would also influence the enzyme's ability to generate vast variations in its catalytic activity and substrate specificity. Altering the properties of key residues and the environment surrounding the ligand-binding site may have a significant effect on substrate recognition, and binding may affect the catalytic activity (Narayan et al., 2015; Blaha-Nelson et al., 2017; Li et al., 2017; Baek et al., 2020; Sun et al., 2020). I hypothesized that altering the docking profile would affect substrate

recognition, thereby altering the interaction between the binding pocket and the triterpene backbone and the enzyme's catalytic activity.

To test the hypothesis, I conducted an *in vivo* functional analysis to determine the catalytic activity of enzyme and substrate specificity in order to provide a more detailed explanation of enzymatic activity and quantify the desired product.

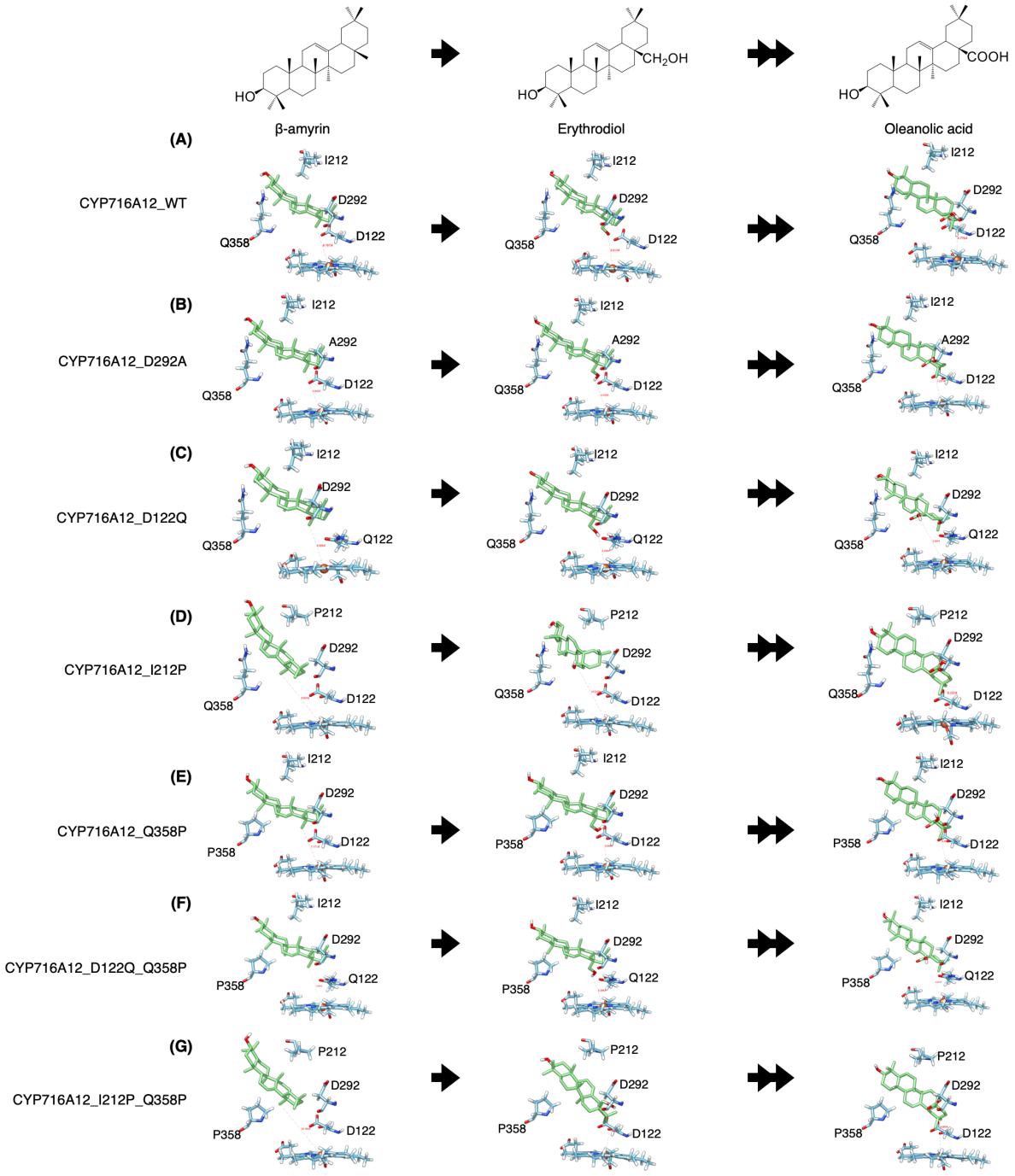


Figure 2-3 Representative interactions of selected CYP716A12 and its variants against the well-known substrate β-amyrin backbone and its derivatives (Erythrodiol and oleanolic acids).

The structure homology modeling showed β-amyrin backbone and its derivatives in wild type of (A) CYP716A12, and CYP716A12 variant (B) CYP716A12_D292A, (C) CYP716A12 _D122Q, (D) CYP716A12_I212P, (E) CYP716A12_Q358P, (F) CYP716A12_D122Q_Q358P, and (G) CYP716A12_I212P_Q358P.

2.3.4 *In vivo* functional analysis in engineered yeast harboring the CYP716A sub-family enzymes

When candidate key amino acid residues at the ligand-binding site were replaced, bioinformatics-based integration results altered the docking profile pattern. To demonstrate the bioinformatics prediction and investigate their enzymatic activities, an *in vivo* functional analysis of the yeast enzyme was conducted. Yeast strains expressing one β-amyrin synthase (bAS) expression vector along with each of the three CYP expression vectors of CYP716A12 wild-type and their variants (D122Q, I212P, D292A, Q358P, D122Q_Q358P, or I212P_Q358P), selected CYP716As (CYP716A15 wild-type, CYP716A48 wild-type, and CYP716A49 wild-type), or the control vector (empty expression vector) were generated. To observe the triterpenoid profile of β-amyrin-producing yeast extract, product analysis was carried out using GC-MS **(Figure 2-4A)**. The extremely similar strategy was applied to engineered yeast, which generates the remaining triterpene backbones, α-amyrin **(Figure S-9A)** and lupeol **(Figure S-11A),** harboring three target CYPs expression vectors. GC-MS analysis results showed that the metabolite profile of CYP716A12 variants against all triterpene backbones was altered in engineered yeast when native amino acids at SRSs were substituted **(Figure 2-4A, S-9A,** and **S-11A)**.

To determine the *in vivo* catalytic activity and substrate specificity of the engineered yeast by comparing the production of triterpenoids, the concentration of triterpenes in a yeast culture harboring CYP716As against β-amyrin backbones was quantified. Consequently, the concentration

of the C-28 oxidized triterpenoid product (oleanolic acid) was significantly higher in yeast cultures containing the wild-type CYP716A12 (5.22 ± 0.08 μg/ml) than in those culture harboring CYP716A15 (4.70 ± 0.25 μg/ml). Nevertheless, yeast cultures harboring CYP716A48 (9.14 ± 0.15 μg/ml) and CYP716A49 (7.15 ± 0.31 μg/ml) produced more oleanolic acid than those harboring the wild type of CYP716A12 (**Figure 2-4B)**. Compared to the wild type, the triterpene concentrations of CYP716A12 variants were significantly different. When CYP716A12 variants D122Q (1.85 times), Q358P (1.06 times), or D122Q_Q358P (2.65 times) were co-expressed in engineered yeast, oleanolic acid production was increased relative to wild-type yeast (**Figure 2-4B)**. The CYP716A12 variant D122Q_Q358P contained the highest amounts of oleanolic acid, 13.82 ± 0.13 μg/ml. In yeast, co-expression of the CYP716A12_I212P variants (4.11 times) or I212P_Q358P (4.85 times) significantly increased erythrodiol production (**Figure 2-4B)**.

When CYP716A12 variants were expressed in α-amyrin-producing yeast, the concentration of the C-28 oxidized triterpenoid product (ursolic acid) was lower in the yeast culture harboring the wild-type CYP716A12 (2.58 ± 0.17 μg/ml) compared to other selected CYP716As. The highest concentration of ursolic acid (8.57 ± 0.23 μg/ml) was observed in yeast containing the wild-type CYP716A48 **(Figure S-9B)**. Ursolic acid concentration increased in yeast strains carrying variants of CYP716A12_D122Q, and D122Q_Q358P **(Figure S-9B)**. Although α-amyrin is the major product of this yeast system, β-amyrin has also been detected as a minor product (Suzuki et al., 2018), thus determining the substrate specificity of CYP716As in yeast. Since there is no α-amyrin synthase that produces only α-amyrin, this occurs. In yeast co-expressing α-amyrin synthase and CYP716As, I was able to detect both β-amyrin (Oleanane-Type) and α-amyrin (Ursane-Type) derivatives, with α-amyrin being the more abundant of the two. Due to this discovery, yeast strains containing the wild types CYP716A15, CYP716A48, and CYP716A49 were preferred for producing ursolic acid (Ursane-type) over oleanolic acid (Oleanane-type). Surprisingly, yeast strains harboring CYP716A12, or its variants were found to preferentially produce oleanolic acid (Oleanane-type) over ursolic acid (Ursane-type). The substrate specificity

of CYP716A12 remained unchanged despite the substitution of key amino acid residues; however, the enzyme's catalytic activity was enhanced **(Figure S-9A, S-9B, S-10)**.

Although the engineered yeast produced lupeol, the concentration of the C-28 oxidized triterpenoid product (betulinic acid) was significantly lower (1.39 ± 0.10 µg/ml) in the yeast culture harboring wild-type CYP716A12 than in the cultures harboring the other selected CYP716As. However, the concentration of betulinic acid was enhanced in yeast harboring the CYP716A12 variants D122Q, Q358P, and D122Q_Q358P. The CYP716A12 variant with the highest concentration of betulinic acid (8.75 ± 0.17 µg/ml) was D122Q_Q358P **(Figure S-11A, B).**

Surprisingly, when the amino acid in the residual oxygen activation region 292 (D292A) was substituted, no oxidized triterpenoid products were detected in any yeast strain producing triterpene backbones (β-amyrin, α-amyrin, and lupeol) that possessed this variant **(Figure 2-5**, **S-9, S-10, S-11).**

These data support the hypothesis that the observed changes in product profiles were the result of a change in the docking profile following in silico site-directed mutagenesis and molecular docking analysis performed.

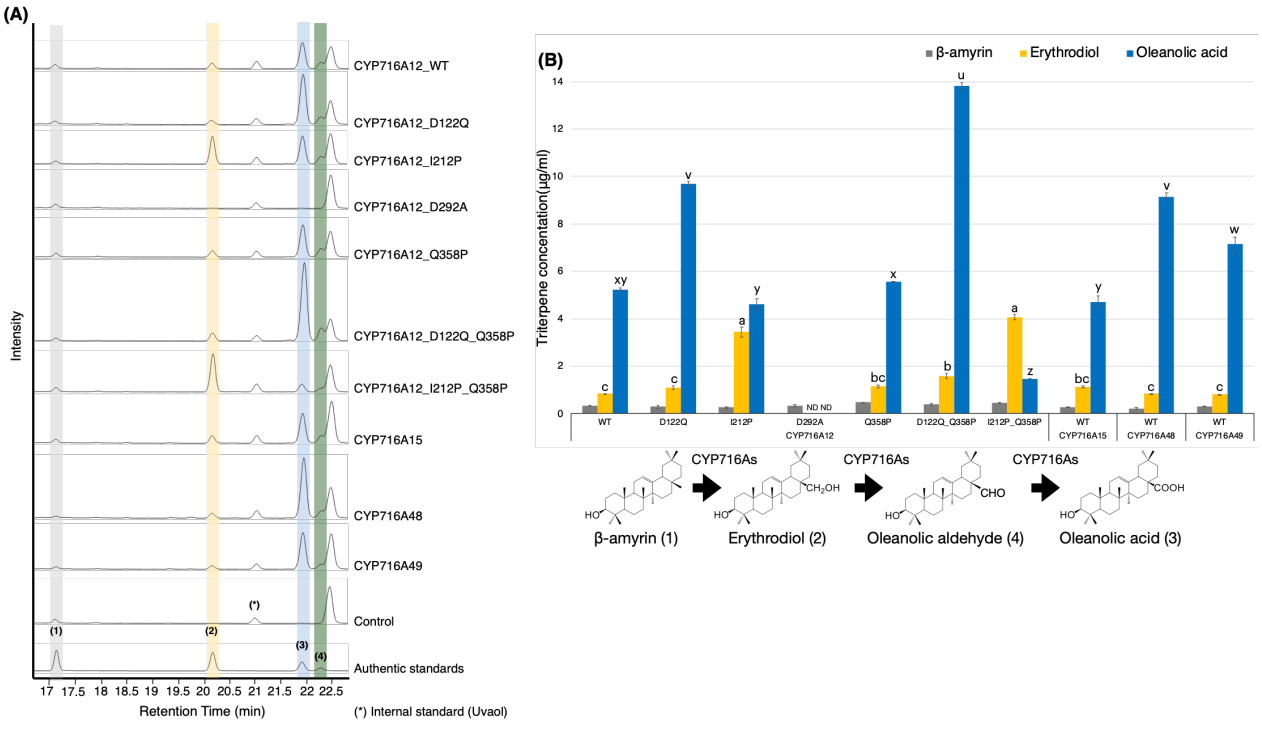


Figure 2-4 *In vivo* activities of selected CYP716As and CYP716A12 and its variants against triterpene skeletons.

(A) Total ion chromatograms (TICs) of extracts from yeast harboring one *bAS* expression vector for producing β-amyrin and three CYPs expression vectors (pELC-CYP716As, pYES-DEST52-CYP716As, and pESC-HIS-CYP716As) were used. (B) Triterpene concentration quantification in engineered yeast harboring CYP716As relative to β-amyrin backbone. Quantification and error bars correspond, respectively, to the mean and standard deviation. The information is representative of at least three biological replications (n=3). Letters indicate statistical differences between the different oxidized triterpenoids levels (a-c: erythrodiol levels, u-z: oleanolic acid levels) for each sample (one-way ANOVA; Tukey's post-hoc test, $p \leq 0.05$). ND, not detected.

2.4 Discussion

The CYP716 enzyme family has been discovered in numerous plant species. These family members catalyze a variety of triterpene backbone modifications. The most prevalent modification of CYP716A in triterpenoids is a three-step site-specific oxidation at the C-28 position of the triterpene backbone (Carelli et al., 2011; Fukushima et al., 2011; Yasumoto et al., 2016; Miettinen et al., 2017; Yasumoto et al., 2017; Sandeep et al., 2019). This analysis demonstrates that even though the structural alignment of the signature motif is highly conserved among CYP716As, their SRS is highly variable. Variations in SRS determine the variety of enzyme functions in terms of substrate specificity, catalytic activity, regio-specificity, and substrate-binding pocket properties (Gotoh, 1992; Sugimoto and Shiro, 2012).

In this chapter, I revealed the amino acid responsible for the previously unknown significant difference between catalytic activity and substrate specificity in C-28 oxidases. I was able to identify the amino acids responsible for the differences between the catalytic activity and substrate specificity of C-28 oxidases by combining bioinformatics techniques with *in vivo* functional analysis in yeast. I observed that altering the properties of a key amino acid residue in the ligand-

binding site could alter the product profile of CYP716A12, specifically the key amino acids present on SRSs.

Aspartic acid 122, located in SRSI of CYP716A12, is a key residue for enhancing the enzyme's catalytic activity. This residue is modifiable to control the direction of the target site toward the reaction center, specifically in the β-amyrin and lupeol backbones, thereby increasing the accumulation of oxidized products. Notably, engineered yeast increased the production of oleanolic acid and betulinic acid significantly.

C-28 oxidation efficiency of triterpene backbones was significantly influenced by isoleucine 212 and neighboring residues in SRSII of CYP716A12. The amino acid side chain of a residue in this region may alter the orientation of triterpene backbone binding in the internal binding pocket, thereby altering the yield of triterpene-oxidized products in yeast.

While modifying glutamine 358 in SRSV has no effect on the orientation of the substrate, it can slightly alter the catalytic activity of an enzyme. Specifically, the CYP716A12_D122Q_Q358P variant improved the catalytic activity of the enzyme and increased its ligand-binding pocket capacity. The observed increase in oxidized triterpenoid yield was the result of the double substitution of two essential residues. Using my current data, I can hypothesize that the Q358P variant of CYP716A48 will be more active than its wild type. However, experimental work is necessary to confirm this hypothesis. The Q358P and D122Q_Q358P variants of CYP716A12 accumulate more of the alcohol intermediate (Uvaol) than the acid product (Ursolic acid) in α-amyrin-producing yeast. The molecular docking research reveals that the Q358P substitution altered the binding capability and environment of the ligand-binding site. It would enhance the binding effectiveness of substrate α-amyrin at the ligand-binding site. Therefore, a high level of uvaol (oxidized product) was identified in co-expression with two variations. To design an enzyme with a function that can be altered to stop at the alcohol (Uvaol) (C-28 hydroxy function), molecular docking analysis suggests that substituting I212 with proline may increase alcohol synthesis by

interfering with substrate binding. Therefore, triple substitution (D122Q_I212P_Q358P) may also aid in stopping at alcohol production (Uvaol).

The CYP716A12 I212P Q358P variant significantly increased the ligand-binding site capacity. However, the amino acid side chain at position 212 (proline 212) reduced the substrate's binding effectiveness. Therefore, this enzyme's catalytic activity may be limited.

The side chain of proline is connected to its nitrogen backbone, forming a five-membered pyrrolidine ring. Compared to other amino acids, this pyrrolidine ring decreases the backbone conformational entropy of the unfolded protein. Proline substitution promotes protein stability by decreasing the entropic difference between the unfolded and folded forms and increasing the free energy difference (Matthews et al., 1987; Choi and Mayo, 2006; Yuan et al., 2017). Although, the results from *in vivo* functional analysis demonstrate that the accumulations of triterpenoids production were increased in engineered yeast when key amino acid residues were substituted with proline. However, the molecular docking analysis showed that the capacity of the ligand-binding site was modified when substituted with proline. To in-depth understand the interaction between triterpene backbones (substrates) and the CYP716A12-variants that key residues were substituted with proline, the analysis using another bioinformatics approach, such as molecular dynamics simulation or related tools, is required in future.

Within the oxygen activation motif, the aspartic acid residue at position 292 is an extremely important component. When alanine was substituted for this residue, the orientation of the substrate, as determined by the docking profile, revealed that this possibility remained unaltered. This was demonstrated by the fact that this possibility remained unchanged. However, the amino acid characteristics of the ligand-binding pocket were altered, which would prevent the enzyme from contracting with the substrate if it were to occur normally. In addition, the substrate's ability to bind to the enzyme-binding pocket was thwarted. During the *in vivo* functional examination performed on the engineered yeast that produced triterpene backbones and harbored this variation, there was no sign of any oxidation product. As a direct consequence of this finding, this residue

was identified as being essential for the proper functioning of the enzyme. In a prior study, it was determined that this location in relation to aspartic acid residue 301, which is positioned in the oxygen activation region, was significant in determining substrate specificity and CYP2D6 activity (Ellis et al., 1995).

In this study, the variants of CYP716A12 were designed by substituting key amino acid residues with only a representative amino acid based on its properties and proline at the loop. Although, the integrated results of this research suggested that this strategy helped to increase knowledge about key amino acids of CYP716A12 enzyme that were important in triterpenoids biosynthesis toward improving the accumulation of triterpenoids production in engineered yeast. However, the substitutions of key amino acid residues of CYP716A12 with all amino acids in all properties are required in future experiments to provide complete coverage data. In addition, the specific catalytic mechanism of the CYP716A12-variants is currently unknown. Therefore, in-depth analysis by bioinformatics approaches and experimental studies are required to study the function of enzymes following the substitution of key residues.

2.5 Conclusion

In a summary, I was able to determine the key amino acid residues in the CYP716A12 that contribute to the enzyme's catalytic activity as well as its substrate specificity. Through enzyme engineering, this study provides a framework for improving the heterologous production of oxidized C-28 triterpenoids in engineered yeast. The modification of the findings obtained has a wide range of applications in protein engineering, particularly in the directed production of triterpenoids, which makes it appropriate for a variety of biotechnological and medicinal uses.

Chapter 3 : Improving the C-28 catalytic activity of *Arabidopsis thaliana* CYP716A1 and CYP716A2

3.1 Introduction

The three-step site-specific oxidation that occurs at the C-28 position of the triterpene backbone is the most common kind of CYP716A modification that can be detected in the process of biosynthesis of triterpenoids (Carelli et al., 2011; Fukushima et al., 2011; Yasumoto et al., 2016; Miettinen et al., 2017; Yasumoto et al., 2017; Sandeep et al., 2019). Catalytic activity at C-28 of the triterpene backbone varied significantly between *Arabidopsis thaliana* CYP716A1 and CYP716A2 isoforms. CYP716A1 catalyzes the site-specific oxidation of α-amyrin and β-amyrin at C-28, resulting in the formation of ursolic acid and oleanolic acid, respectively. Although several types of alcohol were found, ursolic acid and oleanolic acid were not detected in yeast containing the CYP716A2 enzyme. Lupeol-producing yeasts harboring CYP716A1 and CYP716A2 do not produce betulinic acid (Yasumoto et al., 2016) **Figure 3-1**.

Using bioinformatics and *in vivo* functional analysis in engineered yeast, I was able to identify critical amino acids involved in catalytic activity and substrate specificity for site-specific oxidation at C-28 of triterpene backbones in triterpenoid biosynthesis in chapter 2.

In this chapter, I aim to design mutants of *A. thaliana* CYP716A1 and CYP716A2 with improved enzymatic activity at C-28 against diverse triterpene backbones. Through homology modeling, structural alignment, *in silico* site-directed mutagenesis, and docking profile from molecular docking analysis, the key amino acid residues in the ligand-binding site of CYP716A1 and CYP716A2 required for the catalytic activity and substrate specificity of enzyme were identified. To explore the result of the bioinformatics approach, in vivo functional analyses of potential mutants were conducted. When key amino acids at ligand-binding sites were substituted, distinct product profiles were detected in *S. cerevisiae* engineered to contain mutant CYP716A1 or CYP716A2 enzymes.

Finally, I successfully engineered CYP716A1 and CYP716A2 to improve the catalytic activity at C-28 against different triterpene backbones by using the analysis framework from chapter 2.

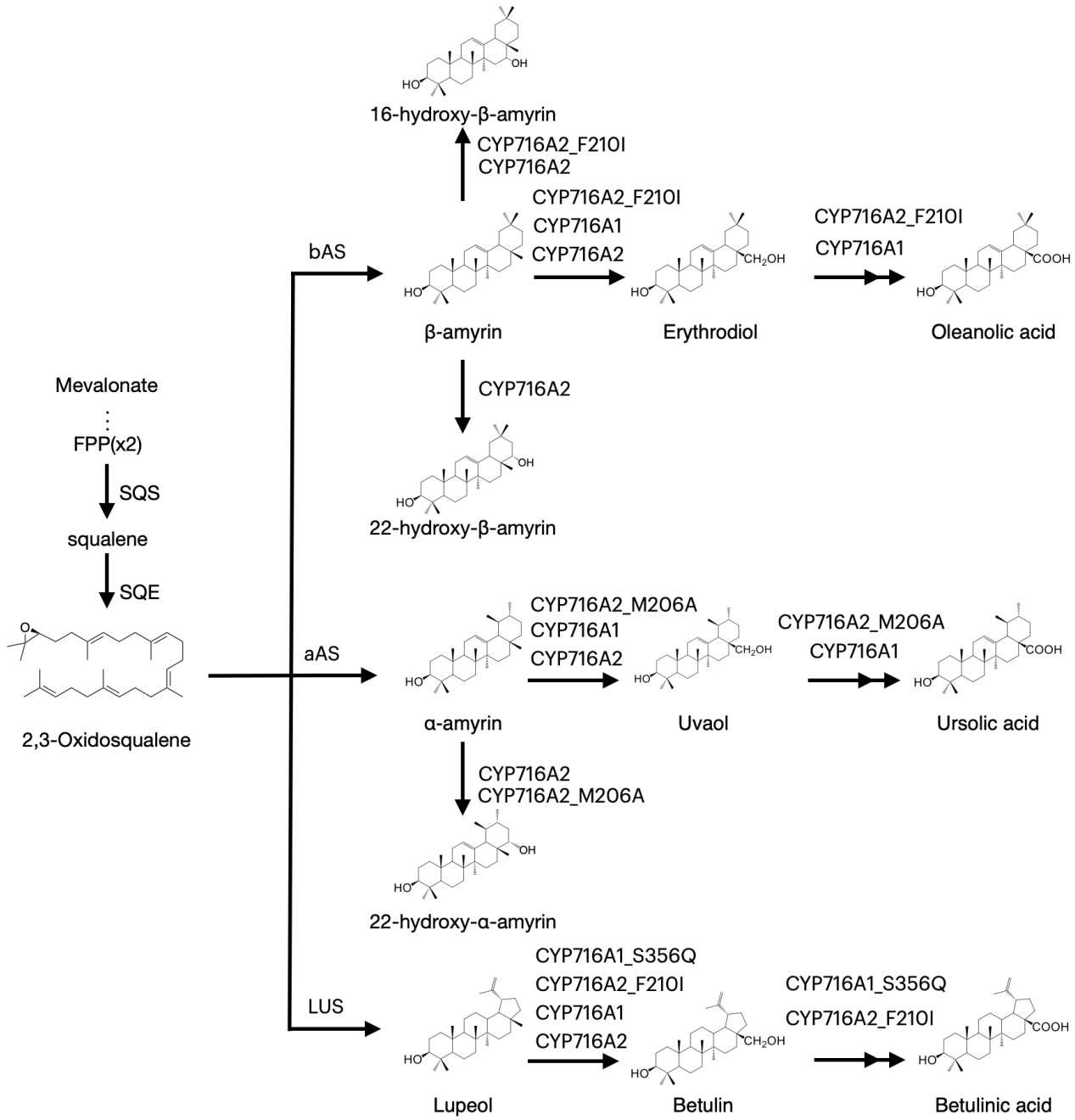


Figure 3-1 Proposed biosynthetic pathway of the triterpenoids detected in chapter 3.

3.2 Materials and methods

3.2.1 Chemical authentic standards

All triterpenoid standards (α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid, lupeol, betulin, and betulinic acid) were purchased from Extrasynthase (France).

3.2.2 Homology modeling

I selected six enzymes belonging to CYP716A subfamily functionally characterized as C-28 multifunctional oxidases. *A. thaliana* CYP716A1 and CYP716A2. The protein sequences of the selected CYP716As are shown in **Supporting information Section 2**. To explore homologous structure in the protein structure database, a Protein BLAST (BlastP) search with a BLOSUM62 matrix was performed against the sequence available in the Protein Data Bank (PDB) database (Madden et al., 1996). The structure model with the highest homology with experimental data was chosen as a template. The target protein sequences were aligned with the model structure. The construction, refinement, and validation of the homology models were performed using Modeller 9.21 software (Webb and Sali, 2017). The non-water HETATM residues of the model were included in the homology model. Structure optimization was carried out using a multi-domain assembler (MDA). All homology models were visualized using Chimera 1.14 (Pettersen et al., 2004). The quality of homology modeling was evaluated using VERIFY3D, with at least 80% of the amino acids having scored >= 0.2 in a 3D/1D profile (Eisenberg et al., 1997).

3.2.3 Molecular docking analysis

Molecular docking analysis revealed interactions between CYP716A1, CYP716A2 and their putative ligands (α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid, lupeol, betulin, and betulinic acid) for attempting to determine the best match interaction between the two molecules. The chemical structures of the putative ligands were obtained from the Cambridge Structural Database (CSD) (https://www.ccdc.cam.ac.uk/theccdcprofile/) and the PubChem

database (https://pubchem.ncbi.nlm.nih.gov). After performing a docking analysis, the scoring function known as the docking score was utilized to provide predictions regarding the level of binding affinity between two molecules. The I-TASSER modeling pipeline was utilized to make a prediction regarding the ligand-binding site region in preparation for the molecular docking experiment (Yang et al., 2015). Molecular docking analysis was performed using AutoDock Vina 1.1.2 (Trott and Olson, 2010) via Chimera 1.14 (Pettersen et al., 2004) with the box center: 50 × -5 × 15 and box dimensions of 30 × 30 × 30 along X, Y, and Z, respectively, in the receptor coordinate system. As criteria for the selection of binding poses, the lowest docking score, the shortest distance between the oxidation target sites (C-28 of triterpene backbones), and the heme reaction center to the ferrous iron fragment (Fe^{2+}) moiety were chosen (Yuki et al., 2012; Geisler et al., 2013). The catalytic site environment of the enzyme, including catalytic site capability, amino acid side chain position and conformation, ligand orientation, and ligand position, was also chosen as an evaluation criterion for docking (docking profiles).

3.2.4 Structural alignment

CYP716A12 homology model was used as the reference structure. For the homology modeling, CYP716A1, CYP716A2, CYP716A15, CYP716A48, and CYP716A49 were matched and aligned with the ClustalX algorithm (Thompson et al., 2002). The residue-residue distance cut-off was set at 5 Å. The residue was correctly placed in the appropriate column if it met the cut-off criteria of at least one other. A Chimera 1.14 was used to create a visualization of the findings of the analysis (Pettersen et al., 2004).

3.2.5 *In silico* site-directed mutagenesis

Candidate residues for site-directed mutagenesis were decided by using the PyMol 2.0.5 software program (Schrodinger, 2015). The target residue in the CYP716A1 or CYP716A2 homology model was replaced with a representative amino acid based on its properties following the analysis framework from the chapter 2 (**Materials and methods section 2.2.5**). The substitution

of proline in the loop increases the stability and kinetics of a protein (Matthews et al., 1987). Hydrogen atoms were added and retrained to maintain hydrogen bond in homology model. The structure was exported in PDB format (.pdb). A Chimera 1.14 was used to perform structural analysis, optimization, and visualization on the designed protein (Pettersen et al., 2004). VERIFY3D was then utilized to perform an evaluation of the overall quality of homology modelling (Eisenberg et al., 1997).

3.2.6 Plasmid construction

Bioinformatic predictions were used to experimentally verify the CYP716A1 or CYP716A2 mutations. The entry clones of CYP716As, pENTR-CYP716A1, pENTR-CYP716A2 (Yasumoto et al., 2016). The mutations were introduced to CYP716A1, and CYP716A2 by site-directed mutagenesis using the PrimeSTAR® Mutagenesis Basal Kit (TaKaRa Bio, Shiga, Japan). The specific primers used for site-directed mutagenesis are shown in (**Table S-1)**. Yeast expression clones were constructed using the Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific) by transferring the coding sequence (CDS) of CYP716A1 or CYP716A2 into pELC-GW (Seki et al., 2008), pYES-DEST52 (Thermo Fisher Scientific), and a gateway-compatible version of pESC-HIS (Agilent Technologies) as destination vectors (Seki et al., unpublished). Yeast expression clones of the other CYP716As were constructed using the same approach, including pELC-CYP716As, pYES-DEST52-CYP716As, and pESC-HIS-CYP716As.

3.2.7 *In vivo* functional analysis in engineered yeast

Saccharomyces cerevisiae INV*Sc*1 (*MATa his3D1 leu2 trp1-289 ura3-52 MATα his3D1 leu2 trp1-289 ura3-52*; Invitrogen) harboring the expression clones for OSC (pYES3-ADH-*OSC1*, pYES3-ADH-*aAS,* or pYES3-ADH-*LUS*) (Fukushima et al., 2011) were further transformed with the constructed expression clone(s) (pELC-CYP716As, pDEST52-CYP716As, and pESC-HIS-CYP716As) using the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA, USA). Transformants were selected on SD selection medium containing 2% glucose, incubated at 30 °C

for three days. Each selected transformant was cultured in 2 ml of the SD selection medium. It was cultured at 30 °C overnight at 200 rpm. The overnight culture (500 µl) was transferred into 5 ml of the same medium broth and then cultured at 30 °C for two days, shaking at 200 rpm. Yeast cells were harvested and resuspended in 5 ml of an appropriate SD medium containing 2% galactose to induce CYP716As expression and cultured at 30 °C for two days at 200 rpm. Yeast cultures were then stored at -80 °C until extraction was performed. All assays were performed in triplicates (three independent assays were obtained from different colonies to confirm the results). *S. cerevisiae* INV*Sc*1 harbors the OSC expression vector (pYES3-ADH-*OSC1*, pYES3-ADH-*aAS,* or pYES3-ADH-LUS), with the empty vectors pELC-GW, pYES-DEST52, and pESC-HIS used as controls.

3.2.8 Analysis of triterpenoid production in engineering yeast

To quantify triterpenoid production in the yeast culture, 50 µl of internal standard (uvaol or betulin, 100 ppm in methanol) was added to the yeast culture before extraction. Yeast cultures (cells and spent media, 5 ml) were extracted with 5 ml of ethyl acetate (Wako, Osaka, Japan). The mixture was then vortexed and sonicated for 30 min. After centrifugation at 9000 g for 5 min, the organic phase was transferred to a new tube using a Pasteur pipette. The extracted samples were evaporated using a centrifugal evaporator for 45 min or until they were dry. This procedure was repeated three times. The remaining powder was resuspended in methanol (1 ml). The obtained samples were transferred into vials using a new Pasteur pipette. The samples were stored at 4 °C until the subsequent analysis.

To quantify triterpenoid production in engineered yeast, 50 µl of the sample solution (β-amyrin-producing yeast extracts) or 100 µl of the sample solution (α-amyrin or lupeol-producing yeast extracts) was transferred to vial inserts and then to a centrifugal evaporator for 60 min, or until dry. Finally, the evaporated pellet was derivatized with 50 µl of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (Sigma-Aldrich) for 30 min at 80 °C before GC-MS analysis. Fifty microliters of authentic standard solutions (α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid,

lupeol, betulin, and betulinic acid, 10 ppm. in methanol) were applied with the same method as stated previously.

For quantification of triterpenoid production in engineered yeast, GC-MS analysis using a 5977A MSD (Agilent Technologies) coupled with a 7890B GC system (Agilent Technologies) and a DB-1ms (length 30 m, 0.25 mm internal diameter, 0.25 μm film thickness; Agilent Technologies) capillary column was used for metabolite analysis of sample extracts from lupeol-producing yeast extracts. Nevertheless, an HP-5ms (length 30 m, 0.25 mm internal diameter, 0.25 μm film thickness; Agilent Technologies) capillary column was utilized to sample extracts from α-amyrin or β-amyrin-producing yeast. The injection component and the MSD transfer line were set to 250 °C. The oven temperature was programmed as follows: 80 °C for 1 min, followed by a rise to 300 °C at a rate of 20 °C min^{-1}, and held at 300 °C for 14 min. The carrier gas was helium (He), and the flow rate was 1 ml min^{-1}. Mass spectra were recorded by scanning the *m/z* range of 50–750. Peaks were identified by comparing retention times and mass spectra patterns to those of authentic standards (Fukushima et al., 2011; Yasumoto et al., 2016) **Supporting information Section 4.**

3.3 Result

3.3.1 Screening candidate amino acid residues for site-directed mutagenesis of CYP716A1 and CYP716A2

A previous study indicated that CYP716A1 and CYP716A2 catalyzed the C-28 oxidant against a similar triterpene backbone in many enzymes belonging to the CYP716A subfamily (Yasumoto et al., 2016). In engineered yeast harboring CYP716A1, oleanolic acid, and ursolic acid are detected when co-expressed with β-amyrin synthase and α-amyrin synthase, respectively. However, these products were undetectable in the engineered yeast expressing CYP716A2. Surprisingly, a variety of alcohols, including erythrodiol, 16-hydroxy-β-amyrin, 22α-hydroxy-β-amyrin, and 22α-hydroxy-α-amyrin, are detected in engineered yeast harboring CYP716A2

(Yasumoto et al., 2016). Despite this, betulinic acid cannot be detected in engineered yeast that produce lupeol and contain either CYP716A1 or CYP716A2 enzymes (Yasumoto et al., 2016).

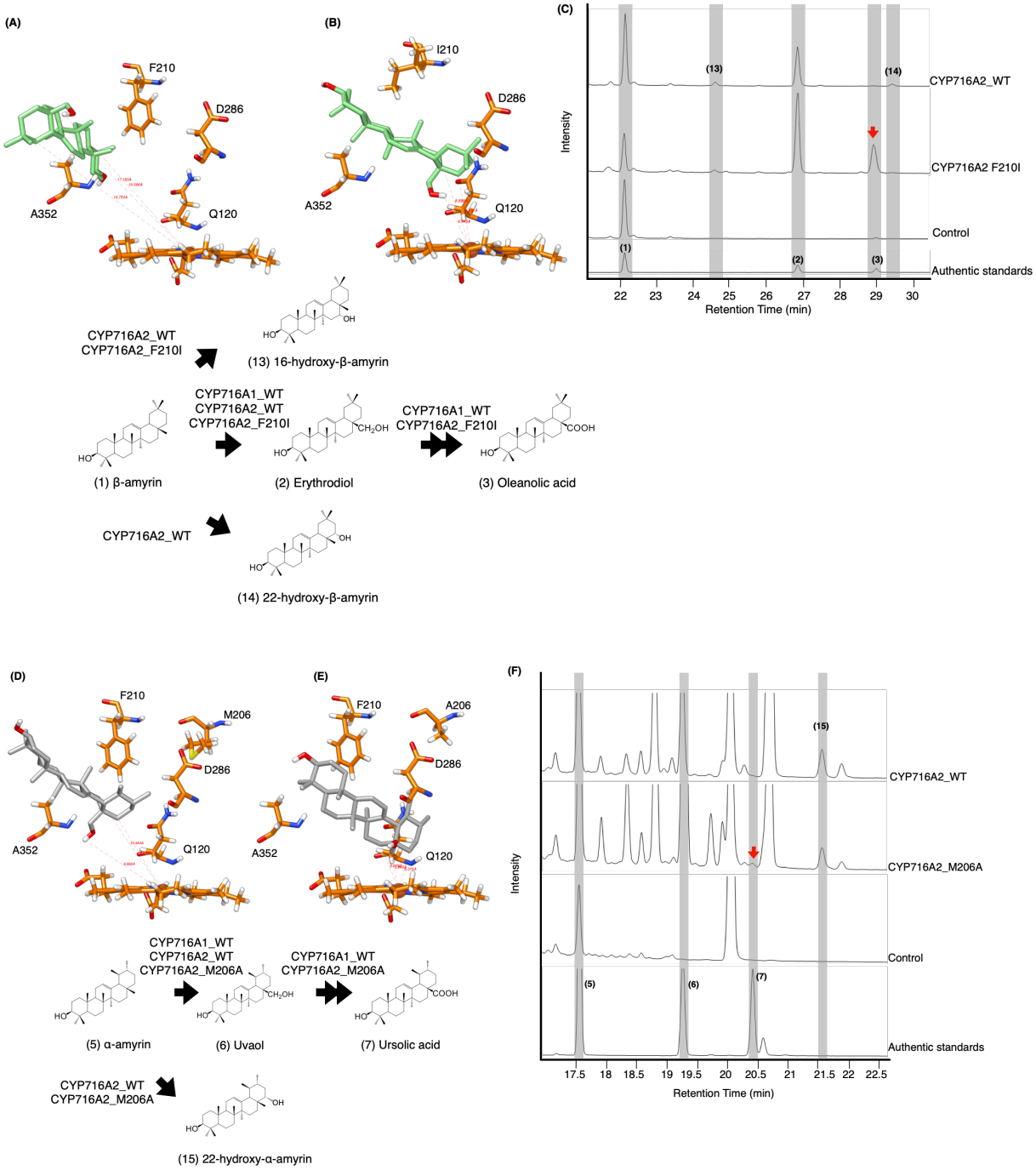
When the amino acid in the SRS of CYP716A12 was replaced, the product profile of C-28 oxidases was altered depending on the triterpene backbones, according to the findings of chapter 2. Specifically, amino acid variations in SRSII and SRSV were replaced. Isoleucine 212 was a conserved residue in SRSII of CYP716A12 and the other CYP716As, as described in chapter 2. However, phenylalanine (F) was a conserved in SRSII of CYP716A1 (F214) and CYP716A2 (F210). On SRSV of the other CYP716As, glutamine 358 was a conserved residue. Nevertheless, the conserved residues in these aligned with CYP716A1 and CYP716A2 were a serine 356 and an alanine 352, respectively. Therefore, the amino acid sequences on SRSII and SRSV of CYP716A1 and CYP716A2 were structurally conserved with CYP716A12 to increase the catalytic activity at C-28 of triterpene backbones (**Figure 2-2A).**

3.3.2 Improving the C-28 catalytic activity of *Arabidopsis thaliana* CYP716A1 and CYP716A2 inspired by CYP716A12 variants

When isoleucine was substituted for F210 in SRS II of CYP716A2, the docking profile of the variant against erythrodiol was altered. The C-28 direction and distance were almost perfectly centered at the reaction center **(Figure 3-2A, B)**. An *in vivo* functional analysis of β-amyrin-producing yeast harboring this variant revealed the production of oleanolic acid (red arrow) and 16-hydroxy-β-amyrin. In comparison, 22-hydroxy-β-amyrin was not detected **(Figure 3-2C)**.

Additionally, when alanine (A) was substituted for M206 on SRS II of CYP716A2 and the docking profile was compared to that of the wild-type, the docking profile of the variant was altered **(Figure 3-2D, E)**. Ursolic acid (red arrow) and 22α-hydroxy-α-amyrin production were detected in α-amyrin-producing yeast harboring this variant **(Figure 3-2F)**. Likewise, when glutamine was substituted for S356 on SRSV of CYP716A1, the docking profile against betulin was altered **(Figure 3-2G, H)**. Similarly, the docking profile indicated that the orientation of betulin in the

ligand-binding pocket of CYP716A2 was different in the variant in which isoleucine was substituted with F210 on SRS II. The direction and distance of C-28 were determined using points centered on the Fe^{2+} moiety heme reaction center **(Figure 3-2I, J)**. However, betulinic acid production (red arrow) was observed in the engineered lupeol-producing yeast harboring these variants (CYP716A1_S356Q and CYP716A2_F210I) **(Figure 3-2K)**.



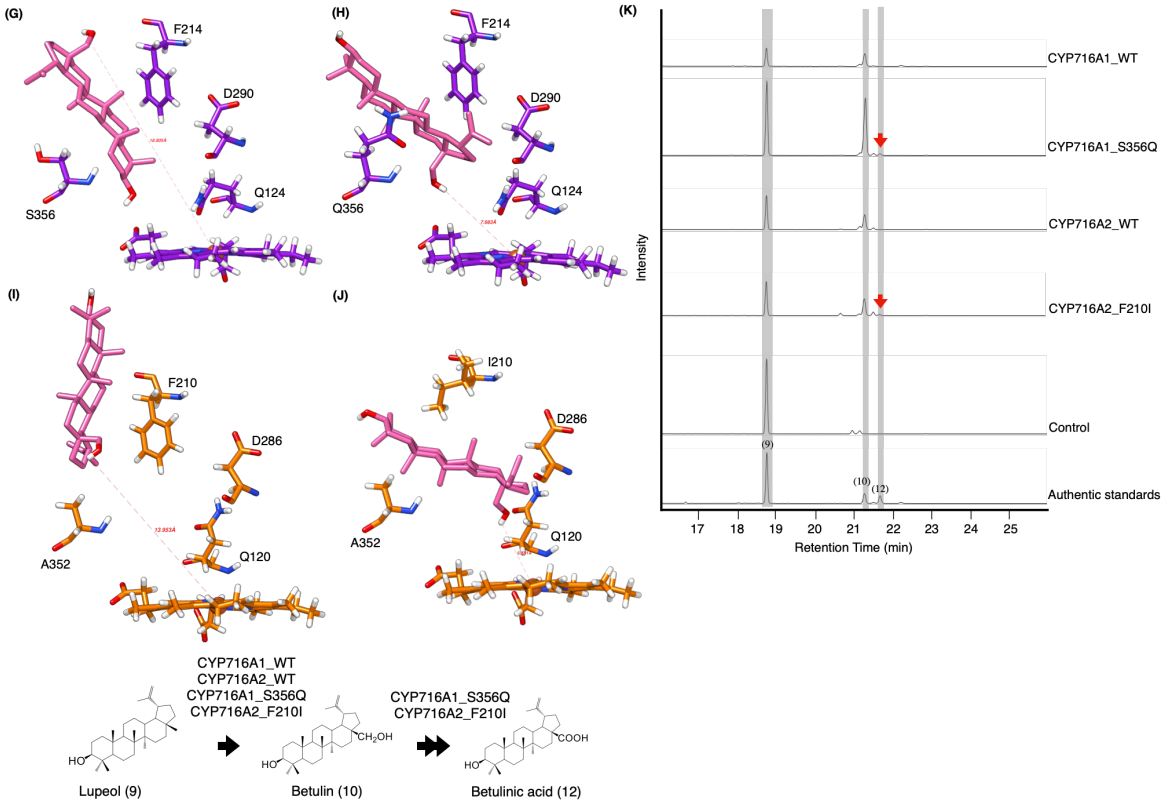


Figure 3-2 *In-silico* study and *In vivo* activity of CYP716A1 and CYP716A2 toward triterpene skeletons in co-expressing yeast strain.

Representative interactions of (A) CYP716A2 wild-type, and (B) its variant (CYP716A2_F210I) against erythrodiol backbone. (C) TICs of extracts from yeast harboring one *bAS* expression vector for producing β-amyrin and three CYPs expression vectors of CYP716A2 or its variant (CYP716A2_F210I). Representative interactions of (D) CYP716A2 wild-type, and (E) its variant (CYP716A2_M206A) against uvaol backbone. (F) TICs of extracts from yeast harboring *aAS* expression vector for producing α-amyrin as a major substrate and three CYPs expression vectors CYP716A2 or its variant (CYP716A2_M206A). Representative interactions of (G) CYP716A2 wild-type, and (H) its variant (CYP716A2_M206A) (I) CYP716A1 wild-type, and (J) its variant (CYP716A1_S356Q) against betulin backbone. (K) EIC of extracts from yeast harboring one *LUS* expression vector for producing lupeol and three CYPs expression vectors of CYP716A2 or CYP716A2_M206A or CYP716A1 or CYP716A1_S356Q. Data are representative of at least three biological replicates (n=3).

3.4 Discussion

Previously undiscovered amino acid residues in C-28 oxidases are responsible for the considerable difference in catalytic activity and substrate specificity identified in chapter 2. A key amino acid residue in the ligand-binding site of CYP716A12 can change the product's profile by modifying its characteristics. Especially, the key amino acids were found on SRSs.

I demonstrated that changing the characteristics of SRS II residues (M206A, F210I) can enhance the catalytic activity of CYP716A2 for heterologous synthesis of oxidized C-28 triterpenoids in engineered yeast. Nonetheless, it impaired the CYP716A2 variant site-specific oxidation activity at C-22 of the β-amyrin backbone. Similarly, engineering of SRSV serine 356 changed the catalytic activity of CYP716A1 on the lupeol backbone. I propose engineering by double mutation at SRSII and SRSV to improve CYP716A1 and CYP716A2 more functionally efficient at C-28 oxidases.

The results suggest a strategy for increasing knowledge about critical amino acid residues involved in the catalytic activity and substrate specificity of CYP716A1 and CYP716A2 in triterpenoid biosynthesis. However, the specific mechanism of the enzymatic catalysis following substitution of critical residues is currently unknown.

3.5 Conclusion

In summary, I identified critical amino acid residues in the CYP716A1 and CYP716A2 that contribute to its catalytic activity and substrate specificity. In chapter 2 provides a framework for enhancing heterologous production of oxidized C-28 triterpenoids in engineered yeast through enzyme engineering of CYP716As. I demonstrated the enormous utility of this approach by engineering *A. thaliana* CYP716A1 and CYP716A2 to modify the catalytic activity at C-28 on triterpene backbones.

Chapter 4 : High-yield bioactive triterpenoid production by heterologous expression in *Nicotiana benthamiana* using the Tsukuba system

4.1 Introduction

Oleanolic acid is a pentacyclic triterpenoid that occurs naturally in over 200 plant species. Numerous studies indicate that it is a precursor to a quantity of bioactive triterpenoids of commercial interest, such as bardoxolone methyl, a powerful activator of the Nrf2 pathway and inhibitor of the NF-κB pathway that is undergoing a phase 2 clinical study for the treatment of chronic disorders (Liby and Sporn, 2012). Additionally, oleanolic acid is a precursor of onjisaponin F, which boosts nasal anti-influenza virus IgA antibody titers (Li et al., 2016; Peng et al., 2020); platycodin D, which has anti-cancer and anti-viral properties (Jeon et al., 2019; Kim et al., 2021); and maslinic acid, which is now being researched as a potential novel medicinal drug, possesses anti-inflammatory, hepatoprotective, analgesic, anti-bacterial, anti-mycotic, virostatic, and immunomodulatory properties, as well as a few others (Fukumitsu et al., 2016; Reyes-Zurita et al., 2016). On the other hand, oleanolic acid only accumulates to very low levels in plant cells, and its chemical synthesis is challenging due to the complexity of its structure (Dale et al., 2020).

Oleanolic acid is biosynthesized from 2,3-oxidosqualene via cyclization catalyzed by oxidosqualene cyclase (OSC), i.e., β-amyrin synthase (bAS), to produce β-amyrin, followed by three-step oxidation at C-28 of the β-amyrin backbone through erythrodiol (28-hydroxy-β-amyrin) and oleanolic aldehyde as reaction intermediates **Figure 4-1**. This oxidation process is mostly catalyzed by enzymes belonging to the CYP716A subfamily (Fukushima et al., 2011; Suzuki et al., 2018). The CYP716A subfamily is a member of the CYP716 family of enzymes, which has highly conserved functionality in the process of triterpenoid biosynthesis. The three-step site-specific

oxidation of the C-28 position of α-amyrin, β-amyrin, and lupeol is the most common modification of the triterpene backbone that is catalyzed by CYP716A (Fukushima et al., 2011; Fukushima et al., 2013; Miettinen et al., 2017). In the CYP716A subfamily of enzymes, *Medicago truncatula* CYP716A12 was the first enzyme to be functionally described as a multifunctional triterpene oxidase able to generate these C-28-oxidized products (Carelli et al., 2011; Fukushima et al., 2011).

In chapter 2, I identified the key amino acid residues responsible for determining the catalytic activity and substrate specificity of CYP716A12 using *in vivo* functional analysis utilizing *Saccharomyces cerevisiae* engineered to produce the β-amyrin backbone. Changing the product profiles of CYP716A12 by replacing key amino acid residues in substrate recognition sites (SRS). Three mutant CYP716A12s (D122Q, Q358P, and D122Q Q358P) enhanced oleanolic acid product yield, whilst two others (I212P and I212P Q358P) produced erythrodiol predominately.

As an alternate approach of triterpenoid production, a synthetic biology strategy that involves pathway reconstruction in heterologous hosts, such as yeast and *Nicotiana benthamiana*, has been researched and examined (Fukushima et al., 2011; Buyel et al., 2017; Reed et al., 2017; Suzuki et al., 2018; Dale et al., 2020). Transient protein expression in *N. benthamiana* leaves using agroinfiltration is a simpler, more flexible, and cost-effective methodology for researching cell biology and physiology and producing excessive quantities of recombinant proteins than stable transgenic procedures or conventional cell culture systems (Buyel et al., 2017; Suzaki et al., 2019).

Several natural product pathways have been reconstructed because of transient expression experiments in *N. benthamiana*. *N. benthamiana* has emerged as a good host to produce terpenes, particularly triterpenoids (Reed et al., 2017; De La Pena and Sattely, 2021). A high yield of the desired compound is required to be generated by a transient expression method that is both high level and efficient. Utilizing the high-expression binary vector pBYR2HS, the Tsukuba system is one of the most effective transient protein expression systems for various plant cells (Yamamoto et al., 2018). It has been demonstrated that this vector, which contains geminiviral replication machinery and a double terminator, considerably improves the transient protein expression

efficiency in lettuce, *N. benthamiana,* tomatoes, eggplant, hot pepper, melon, and orchid. When compared to a single terminator, the double terminator can eliminate transcriptional interference, which leads to an increase in the expression level. Additionally, the geminiviral replication machinery leads to a high yield of expression of foreign proteins. In addition to this, the pBYR2HS vector has an expression cassette for the gene-silencing suppressor p19 protein from tomato bushy stunt virus. This protein can also contribute to an increase in the expression of target proteins (Yamamoto et al., 2018; Suzaki et al., 2019). However, it has not yet been used to reconstruct plant triterpenoid biosynthesis pathways.

Through pathway reconstruction, I examined the applicability of the Tsukuba system for triterpenoid biosynthesis in *N. benthamiana* in this study. I evaluated the efficacy of pBYR2HS in enhancing oleanolic acid production by comparing the pBYR2HS-based expression construct to a conventional binary vector without geminiviral replication machinery and a double terminator. I also selected the host based on the oleanolic acid-producing qualities of two *Nicotiana* species. To prevent leaf necrosis and enhance oleanolic acid production, the appropriate incubation period and application of ascorbic acid following agroinfiltration were also evaluated. In addition, I proved the need for an additional NADPH-cytochrome P450 reductase (CPR) to improve the microenvironment of CYPs and enhance the production of oleanolic acid. I chose five mutant CYP716A12s with altered product profiles based on chapter 2, along with several wild-type CYP716A enzymes, such as *Vitis vinifera* CYP716A15, *Olea europaea* CYP716A48, and *Beta vulgaris* CYP716A49, to compare the potential for high-titer oleanolic acid production using the Tsukuba system. The enzyme mutation CYP716A12_D122Q has the greatest potential for heterologous oleanolic acid synthesis. I also optimized the biosynthetic pathway to improve oleanolic acid biosynthesis through the co-introduction of mutant *Arabidopsis thaliana* HMGR1 catalytic domain (AtHMGR1cd-S577A) to enhance the mevalonate pathway. I finally established the applicability of this expression system for increasing production in the bioactive triterpenoid pathway by using maslinic acid as a case study **Figure 4-1**.

This study developed a general technique for efficiently synthesizing substantial quantities of triterpenoids and producing important and difficult-to-obtain triterpenoids with potent biological activity.

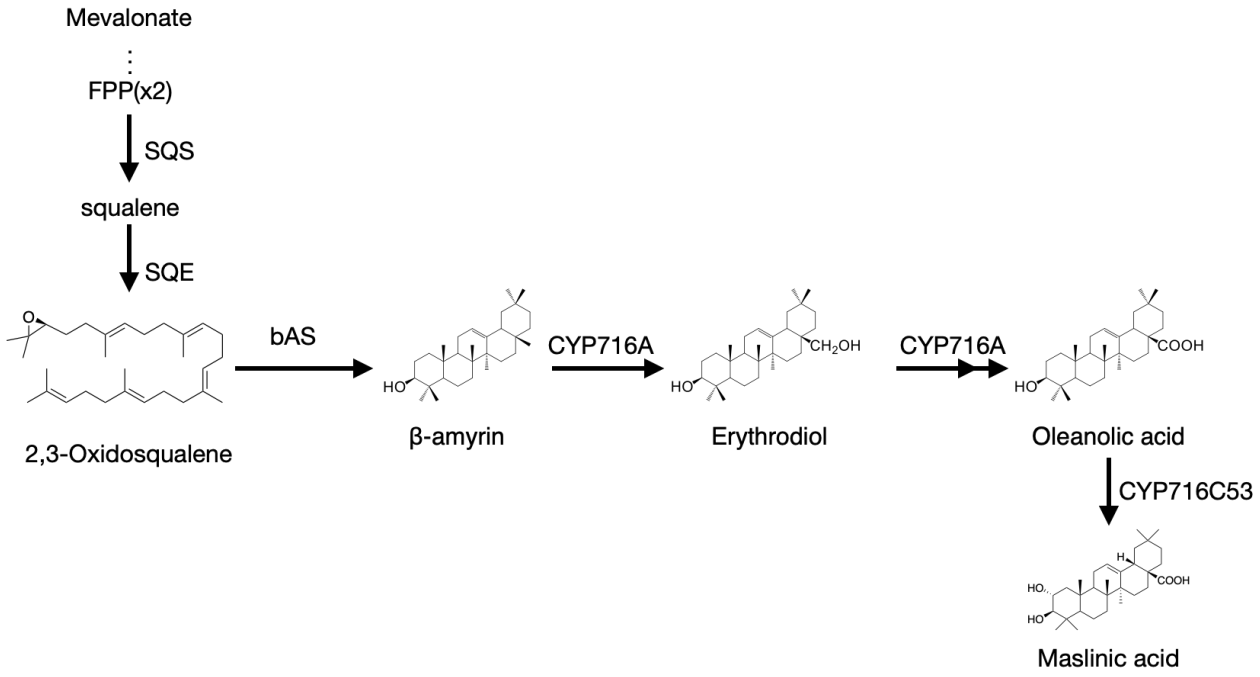


Figure 4-1 Proposed biosynthetic pathway of the triterpenoids detected in chapter 4.

4.2 Material and method

4.2.1 Chemical Authentic Standards

The following triterpenoid standards were purchased from Extrasynthese: uvaol, β-amyrin, erythrodiol, oleanolic acid, and maslinic acid (Genay, France).

4.2.2 Plasmid Construction

pRI 201-AN (TaKaRa Bio, Shiga, Japan), a standard binary vector developed for high-level foreign gene expression in dicotyledonous plants, was utilized in this investigation. To develop pYS 015, a gateway cassette was introduced into pRI 201-AN digested with NdeI and SacI to construct a gateway-compatible version (Srisawat et al., 2019). pRI201-AN was also applied as a template vector for cassette amplification during the construction of pBYR2HS (Yamamoto et al., 2018). The structures of the T-DNA regions of the expression constructs pBYR2HS and pRI201-

AN are displayed in **Figure S-12**. To construct a gateway-compatible version of pBYR2HS, pBYR2HS-SalI (Yamamoto et al., 2018) was digested with SalI, blunted with KOD polymerase, and ligated with GATEWAY conversion cassette frame A (Invitrogen, Carlsbad, CA, USA).

The entry clones of pENTR-CYP716A12, pENTR-selected mutant CYP716A12s, pENTR-CYP716A15, pENTR-CYP716A48, pENTR-CYP716A49, mutant *A. thaliana* HMGR1 catalytic domain (AtHMGR1cd-S577A), and *Lotus japonicus* NADPH-cytochrome P450 reductase class II (LjCPR2) were obtained in previous studies (Seki et al., 2008; Fukushima et al., 2011; Robertlee et al., 2018; Suzuki et al., 2018; Istiandari et al., 2021). The amino acid sequences of the enzymes in this chapter are shown in **Supporting information Section 2**. The mutant CYP716A12s were generated by site-directed mutagenesis using a PrimeSTAR® Mutagenesis Basal Kit (TaKaRa Bio) in chapter 2. Complete codon-optimized L. japonicus β-amyrin synthase (LjOSC1) was synthesized for expression in *N. benthamiana* (**Supporting information Section 10**). Using Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific, Waltham, MA, USA), the coding sequences (CDSs) of the target genes were transferred to gateway-compatible versions of pBYR2HS and pYS 015 to generate plasmids for transient expression in plant cells.

4.2.3 Transient expression in *Nicotiana benthamiana*

Electroporation was used to transfer expression constructs for target genes into *Agrobacterium tumefaciens* strain GV3101 (pMP90) (MicroPulser™; Bio-Rad, Hercules, CA, USA). Three days of incubation at 28°C on LB selection medium with 50 mg/l kanamycin, 25 mg/l gentamycin, and 100 mg/l rifampicin was used to select transformants. Each selected transformant was cultured overnight in 2 ml LB selection medium at 28°C with 200 rpm shaking. Transferred 75 µl of the overnight cultures into 5 ml of the same medium and then cultured overnight at 28°C with 200 rpm shaking. Freshly grown *A. tumefaciens* cells were resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES, pH 5.6, 100 µM acetosyringone) to adjust OD_{600} to ~1.00 as described previously (Yamamoto et al., 2018). A 1 ml needleless syringe was used to infiltrate

equal volumes of A. tumefaciens harboring target constructions into the abaxial air spaces of the leaves of 5-week-old *N. benthamiana* plants.

To prevent necrosis, I sprayed each leaf with 0.3 ml of 200 mM ascorbic acid or 200 mM ascorbic acid sodium salt every two days (i.e., on days 1, 3, and 5) after infiltration. Seven days after infiltration, the leaves of three plants were harvested for triterpenoid analysis.

4.2.4 Metabolite extraction of *Nicotiana benthamiana* leaves

To quantify triterpenoid production in the *N. benthamiana* leaves, 20 µl internal standard (uvaol, 100 ppm in methanol) was added to the powdered freeze-dried leaf samples before extraction. The samples (10 mg) were extracted with 1 ml methanol (Wako, Osaka, Japan). Then the mixtures were vortexed and sonicated for 60 min (45% intensity: Sharp, Osaka, Japan). Using a Pasteur pipette, the organic phase was transferred to a fresh tube after centrifugation at 12,000 rpm for 5 min. Using a centrifugal evaporator, the extracted samples were evaporated for 45 min or until completely dry. One milliliter of methanol was used to resuspend the remaining powder. To remove macromolecules such as sugar and chlorophyll from the extracted sample, 1 ml methanol and 1 ml 4 M HCl were used to perform saponification. The mixtures were briefly vortexed, heated at 80°C for 1 h, and left to stand at room temperature for 10 min. In the last extraction step, 2 ml of ethyl acetate and hexane (1:1 v/v) were added. Using a Pasteur pipette, the organic phase was transferred to a fresh tube following three minutes of centrifugation at 1,500 rpm. The extracted samples were evaporated using a centrifugal evaporator for 60 min or until completely dry. A 500 µl of methanol were used to resuspend the remaining powder. The samples were transferred into vials using a new Pasteur pipette and kept at 4°C until analysis.

4.2.5 GC-MS analysis of leaf extracts and standards

To quantify triterpenoid production in *N. benthamiana*, 50 µl of sample solution was transferred to vial inserts and subsequently evaporated in a centrifugal evaporator for 60 min, or until dry. Then the evaporated pellet was derivatized with 50 µl *N*-methyl-*N*-(trimethylsilyl)

trifluoroacetamide (Sigma-Aldrich, St. Louis, MO, USA) for 30 min at 80°C before GC-MS analysis. Authentic standard solutions (uvaol, β-amyrin, erythrodiol, oleanolic acid, maslinic acid, 10 ppm in methanol) were applied in aliquots of 50 µl using the same method as described previously. GC-MS analysis was conducted utilizing a 5977A MSD coupled with a 7890B GC system and HP-5MS capillary column (length 30 m, 0.25 mm internal diameter, 0.25 µm film thickness) (Agilent Technologies, Santa Clara, CA, USA). The injection component and the MSD transfer line were set to 250°C. The oven temperature was programmed as follows: 80°C for 1 min, followed by an increase to 300°C at a rate of 20°C min^{-1}, and held at 300°C for 18 min. The carrier gas was helium (He), and the flow rate was 1 ml min^{-1}.

Mass spectra were acquired by scanning the 50–750 *m/z* range. The retention times and mass spectra of the peaks were compared to those of authentic standards. The relative concentrations of triterpenoids in extracted samples were determined by comparing the peak areas of the analyte and internal standard. The concentrations of triterpenoids in the *N. benthamiana* leaves were determined by comparison with authentic standard curves constructed using β-amyrin, erythrodiol, oleanolic acid, and maslinic acid (**Supporting information Section 11 and Section13, Figure S-13**).

4.2.6 Statistical analysis

The significance of the differences in oxidized triterpenoid concentrations between samples extracted from each combination or condition was evaluated using Tukey's test and one-way analysis of variance (ANOVA). Using the unpaired Student's t test, the differences in oxidized triterpenoid levels between two groups were compared. Statistical analyses were conducted on macOS using JASP 0.16 (JASP Team, University of Amsterdam, Amsterdam, the Netherlands). In all analyses, a *p*-value ($p < 0.05$) was considered as statistically significant.

4.3 Results

4.3.1 Establishing a method for improving oleanolic acid production

To compare the production of oleanolic acid in transiently transfected *N. benthamiana* leaves containing pBYR2HS-based expression construct to the conventional pYS_015-based expression construct, leaves were infiltrated with a mixture of *A. tumefaciens* harboring pBYR2HS expressing LjOSC1 (β-amyrin synthase), LjCPR2, and CYP716A12_D122Q or a combination of pYS_015 vector expressing LjOSC1, LjCPR2, CYP716A12_D122Q, and p19 by infiltration. Given that the pBYR2HS vector contains a p19 expression cassette, I introduced a separate p19 expression construct into the pYS_015 vector-based infiltration combination.

As a background control, a combination of pBYR2HS expressing; LjOSC1, LjCPR2, and empty vector or a combination of pYS_015 vector expressing; LjOSC1, LjCPR2, empty vector, and p19 was used. I detected a single major product peak in the GC-MS chromatogram of the extracts of transiently expressed CYP716A12_D122Q in *N. benthamiana* leaves, but not in the background control (**Figure 4-2A**). This peak was determined to be oleanolic acid (3) by comparison to the authentic standard (**Figure 4-2A**). I analyzed the triterpenoid content of N. *benthamiana* leaf extracts to compare oleanolic acid production (**Figure 4-2B**). The accumulation of oleanolic acid was detected using the conventional binary vector pYS_015 as 2.1 ± 0.7 mg/g dry weight (dw). When the pBYR2HS (Tsukuba system) vector was used, the accumulation of oleanolic acid increased by 13.1-fold (27.3 ± 2.6 mg/g dw) (**Figure 4-2B**).

The transformation of *Nicotiana* species, particularly *N. tabacum* and *N. benthamiana*, using *Agrobacterium* utilizing leaf disks as the target explant is an incredibly important technique for the speedy evaluation of transgenes in higher plants (Clemente, 2006). To choose the superior *Nicotiana* species for generating oleanolic acid via agroinfiltration-based transient protein expression, *A. tumefaciens* containing expression vectors expressing LjOSC1, LjCPR2, and CYP716A12_D122Q was infiltrated in equal amounts into *N. benthamiana* and *N. tabacum.*

The quantities of oleanolic acid in the infiltrated leaves of various *Nicotiana* species were determined by GC-MS (**Figure S-14**). The yield of oleanolic acid in the leaves of *N. benthamiana* was 37.9 ± 0.9 mg/g dw, which was 3.3 times that of *N. tabacum* (11.4 ±1.7 mg/g dw) (**Figure 4-2C**). This finding implies that both *N. benthamiana* and *N. tabacum* could serve as heterologous hosts for oleanolic acid production using the Tsukuba system. In terms of product yield per leaf, however, this discovery suggests that, of the two species, *N. benthamiana* provides the more promising production platform.

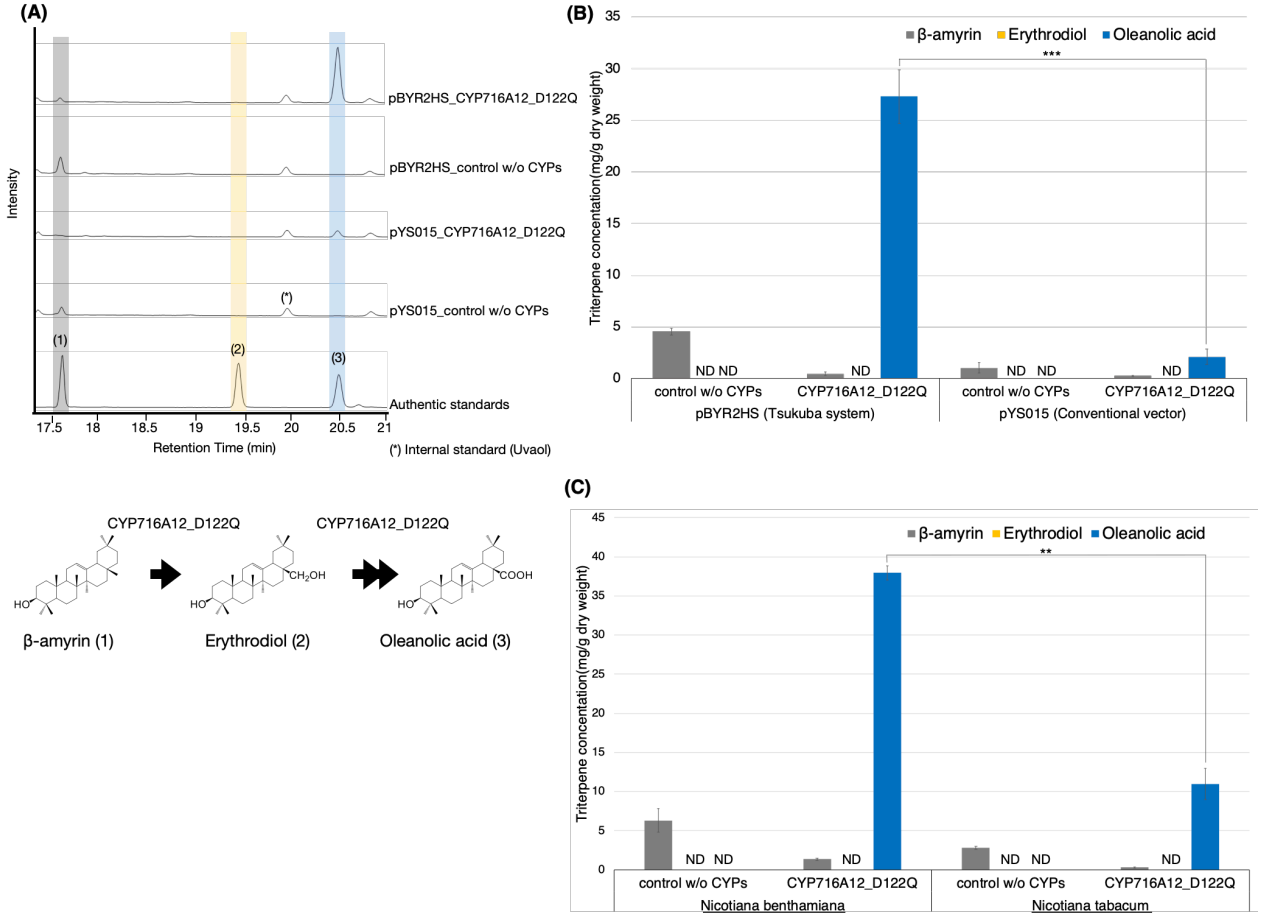


Figure 4-2 Comparison of the accumulation of oleanolic acid in leaves from different expression systems and *Nicotiana* species.

(A) TICs of extracts from *N. benthamiana* leaves infiltrated with a mixture of *A. tumefaciens* harboring pBYR2HS expressing LjOSC1, LjCPR2, and CYP716A12_D122Q or a combination of the pYS_015 vector expressing LjOSC, LjCPR2, CYP716A12_D122Q, and p19. As a background control, a combination of pBYR2HS expressing LjOSC, LjCPR2, and empty vector or pYS015 binary vector expressing LjOSC1, LjCPR2, empty vector, and p19 was used. β-Amyrin (**1**),

erythrodiol (**2**), and oleanolic acid (**3**) were used as authentic standards. Uvaol (*) was used as an internal standard. (B) Comparison of triterpene concentrations in *N. benthamiana* leaves infiltrated with different expression vectors. (C) Comparison of the triterpene concentrations in infiltrated leaves of *N. benthamiana* and *N. tabacum*. Data are representative of at least three biological replicates ($n = 3$) and are expressed as the mean ± SEM. Differences in oleanolic acid production were examined using Student's *t* test for each sample. $^{**}p < 0.01$, $^{***}p < 0.001$, ND, not detected.

4.3.2 Requirement of additional CPR for optimization of the CYP microenvironment to enhance oleanolic acid production

For site-specific oxidation, CYPs need electrons transported via NADPH-CPR. In a heterologous host, native CPRs are probably insufficient for heterologous CYPs to reach optimum efficiency (Zhu et al., 2018; Liu et al., 2019; Theron et al., 2019; Istiandari et al., 2021; Liu et al., 2021). To determine the efficacy of co-expression of additional CPR in this system, equal volumes of *A. tumefaciens* expressing LjOSC1, CYP716A12_D122Q with or without additional LjCPR2 were mixed and infiltrated into *N. benthamiana*. As a background control, a combination of *A. tumefaciens* with pBYR2HS expressing LjOSC1, LjCPR2, and an empty vector was used. The GC-MS analysis and quantification of oleanolic acid in leaf extracts were performed (**Figure 4-3A**). The leaves of *N. benthamiana* that transiently expressed LjOSC1 and CYP716A12_D122Q with LjCPR2 generated 39.3 ± 3.4 mg/g dw oleanolic acid, which was 2.7-fold greater than the amount produced without LjCPR2 co-expression (14.7 ± 0.3 mg/g dw) (**Figure 4-3B**).

These findings suggest that additional CPR co-expression was helpful in enhancing the microenvironment and optimizing the performance of heterologous CYPs in *N. benthamiana*, resulting in an increase in oleanolic acid production.

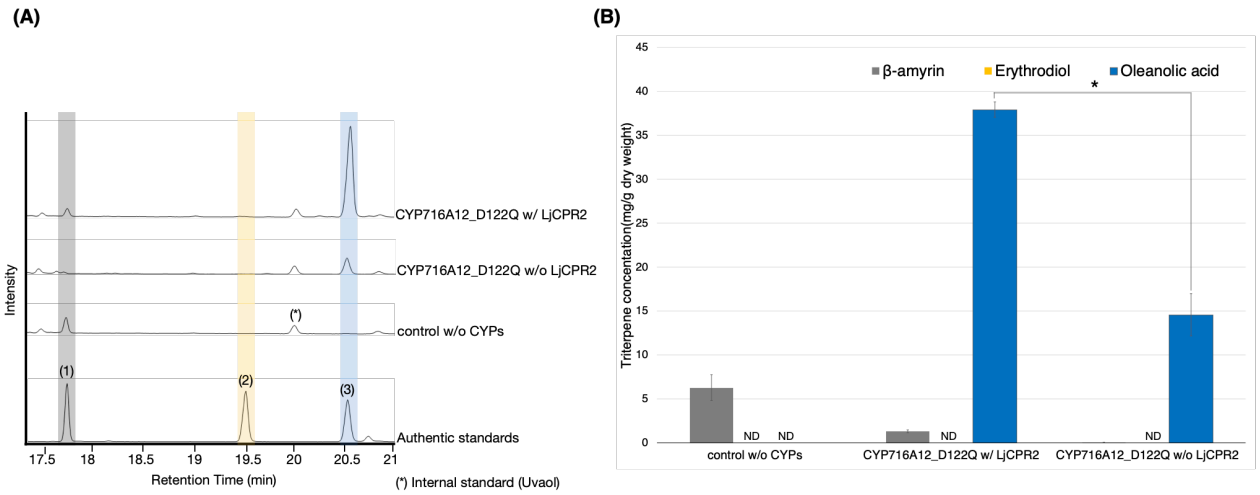


Figure 4-3 Requirement of additional CPR to optimize the CYP microenvironment for enhancement of oleanolic acid production in *N. benthamiana.*

(A) TICs of extracts from *N. benthamiana* leaves transiently expressing a combination of LjOSC, CYP716A12_D122Q with or without LjCPR2 using pBYR2HS. A combination of LjOSC, LjCPR2, and empty vector was used as a background control. β-Amyrin (**1**), erythrodiol (**2**), and oleanolic acid (**3**) were used as authentic standards. Uvaol (*) was used as an internal standard. (B) Quantification of triterpene concentration in *N. benthamiana* leaf extracts. Data are representative of at least three biological replicates ($n = 3$) and are expressed as the mean ± SEM. Differences in oleanolic acid production were examined using Student's *t* test for each sample. *$p < 0.05$, ND, not detected.

4.3.3 Application of ascorbic acids for higher rates of oleanolic acid production

Agrobacterium-induced leaf necrosis, or leaf death, can occur in *N. benthamiana* leaves following infiltration (Nosaki et al., 2021). To prevent necrosis following agroinfiltration, ascorbic acid was sprayed on the leaves, and optimal incubation durations for enhancing the generation of oleanolic acid were determined. The leaves were infiltrated with a mixture of equal volumes of *A. tumefaciens* harboring pBYR2HS expressing LjOSC1, LjCPR2, and CYP716A12_D122Q. Equal volumes of *A. tumefaciens* harboring pBYR2HS expressing LjOSC1, LjCPR2, and empty vector was used as the background control. Leaves were sprayed with a 0.3 ml of 200 mM ascorbic acid or 200 mM ascorbic acid sodium salt (sodium L-ascorbate) every 2 days (i.e., days 1, 3, 5, and 7)

post-infiltration. Negative controls consisted of leaves that were not treated with ascorbic acid. The leaves of three plants were harvested for metabolites analysis at 3-, 5-, 7-, and 9-days post-infiltration. GC-MS analysis and triterpenoid quantitation were conducted on leaf extracts (**Figure 4-4, Figure S-15**). Oleanolic acid production was greatly increased by spraying leaves with 200 mM ascorbic acid (30.3 ± 3.3 mg/g dw) or 200 mM sodium L-ascorbate (24.7 ± 2.8 mg/g dw) after 7 days of agroinfiltration, compared to production without ascorbic acid spraying (15.3 ± 1.7 mg/g dw). There was no significant variation in yield between 7- and 9-days following infiltration (**Figure 4-4**).

As a finding, foliar spraying with 200 mM ascorbic acid and incubation for 7 days following agroinfiltration resulted in the highest oleanolic acid production titer; hence, these conditions were utilized in further experiments.

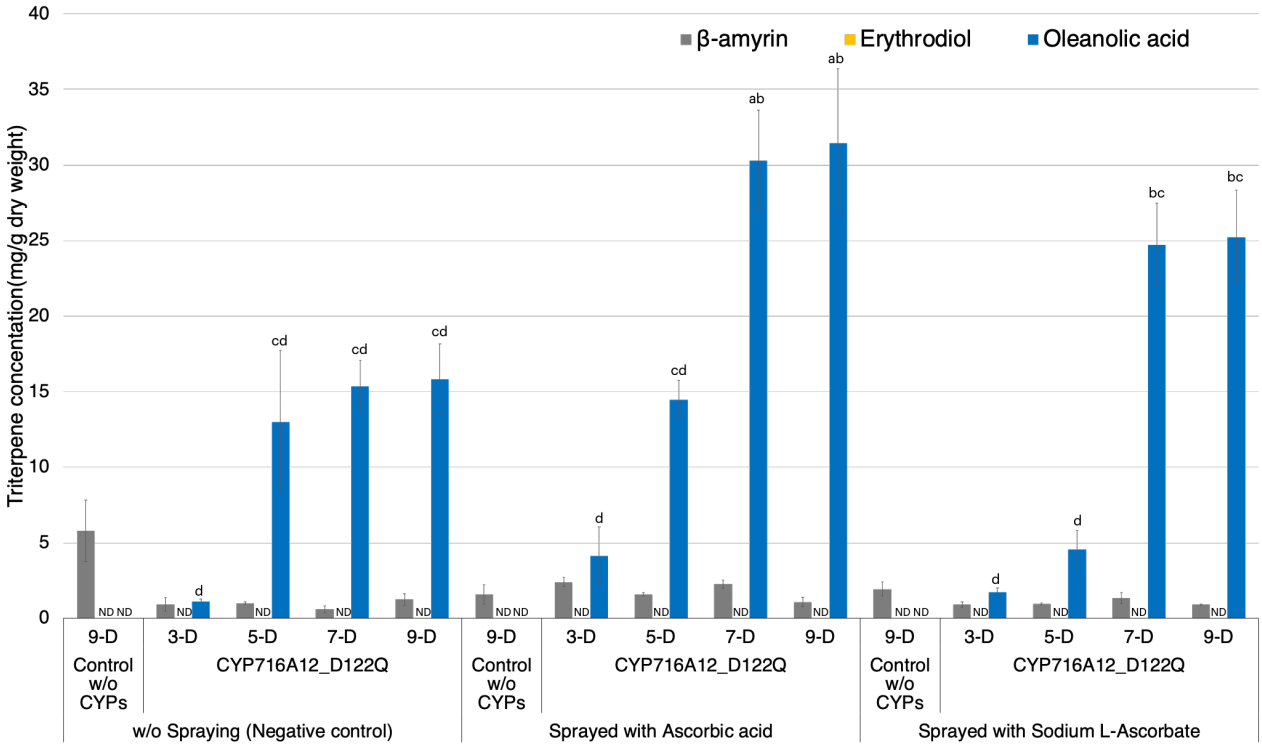


Figure 4-4 Improving production of oleanolic acid in *N. benthamiana* leaves by application of ascorbic acid.

Quantification of triterpene concentration in *N. benthamiana* leaf extracts transiently expressing a combination of LjOSC, LjCPR2, and CYP716A12_D122Q or pBYR2HS empty vector (control)

with or without foliar application of 200 mM ascorbic acid or ascorbic acid sodium salt (sodium L-ascorbate). Metabolites were extracted from leaves at 3-, 5-, 7-, or 9-days after agroinfiltration. As a background control, a combination of pBYR2HS expressing LjOSC, LjCPR2, and an empty vector was used (9 days). The mean and standard deviation are indicated by the values and error bars, and the data are representative of at least three biological replicates (n = 3). Letters indicate significant differences in the levels of oleanolic acid (a–d) among samples (one-way ANOVA with Tukey's *post hoc* test, $p < 0.05$). ND, not detected.

4.3.4 Identification of CYP716A12 mutants with the highest potential for oleanolic acid production

In chapter 2, I determined the key amino acid residues for enhancing the catalytic activity and substrate specificity of CYP716A for triterpenoid synthesis in *S. cerevisiae*. I selected five mutant CYP716A12s with changed product profiles when key amino acid residues in the substrate recognition site were substituted, namely D122Q, I212P, Q358P, D122Q_Q358P, and I212Q_Q358P. In addition, I compared the catalytic activity of wild-type CYP716A12, CYP716A15, CYP716A48, and CYP716A49 in the production of oleanolic acid to those of chosen mutant CYP716A12s. Using the Tsukuba system, all candidate genes were cloned into the pBYR2HS vector to select the mutation with the greatest potential for oleanolic acid production in *N. benthamiana.* A *N. benthamiana* was infiltrated with equal amounts of *A. tumefaciens* harboring a pBYR2HS expressing LjOSC1, LjCPR2, and candidate CYP716As. As a control for the background, leaves were infiltrated with *A. tumefaciens* with pBYR2HS expressing LjOSC1, LjCPR2, and an empty vector. The triterpenoid profile of *N. benthamiana* leaf extract was determined by GC-MS analysis. A single major peak of oxidized triterpenoid was observed in the metabolite profiles of *N. benthamiana* leaves transiently expressing all candidates, but not in the profiles of the background control. An oleanolic acid (3) was identified as the major peak based on comparison to the authentic standard (**Figure 4-5A**). Erythrodiol (2) was identified as a minor oxidized triterpenoid peak only when CYP716A12_I212P or CYP716A12_I212P_Q358P was

transiently expressed (**Figure 4-5A**). The *N. benthamiana* leaf extracts transiently expressing each of the five mutant CYP716A12s or wild-type CYP716A12, CYP716A15, CYP716A48, or CYP716A49 were analyzed for their quantities of triterpenoids production. The highest amount of oleanolic acid was found in transiently expressed CYP716A12 D122Q in *N. benthamiana* leaves (36.9 ± 1.6 mg/g dw) (**Figure 4-5B**).

In previous research, co-expression of the *A. thaliana* HMGR1 catalytic domain with the S577A mutation (AtHMGR1cd-S577A, lacking an inactivating phosphorylation site and demonstrating more in vitro activity than AtHMGR1cd) enhanced triterpenoid production in agroinfiltrated *N. benthamiana* leaves (Robertlee et al., 2018; Srisawat et al., 2019). To increase oleanolic acid production utilizing the Tsukuba system, I applied this strategy. A combination of *A. tumefaciens* containing pBYR2HS and expressing LjOSC1, LjCPR2, and CYP716A12 D122Q, with or without AtHMGR1cd-S577A, was used to infiltrate leaves. As controls, I used leaves infiltrated with a combination of *A. tumefaciens* lacking CYPs.

Seven days after agroinfiltration, the leaves were harvested and processed for metabolite extraction, GC-MS analysis, and quantification of oleanolic acid (**Figure 4-5C, Figure S-16**). The accumulation of β-amyrin was increased almost 21-fold because of co-expression of AtHMGR1cd-S577A (**Figure 4-5C**). The accumulation of oleanolic acid tended to increase when AtHMGR1cd-S577A was co-expressed (30.8 ± 6.0 mg/g dw) (**Figure 4-5C**).

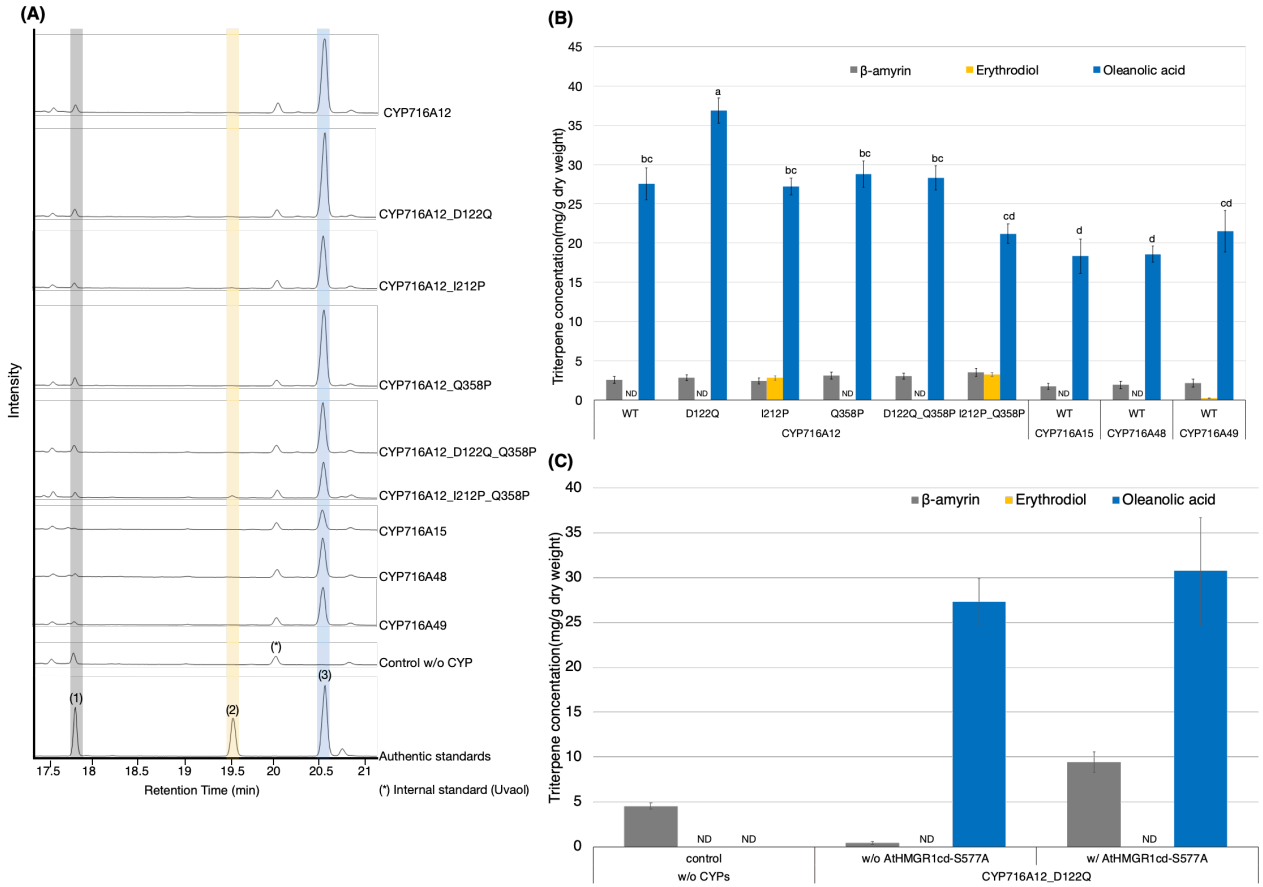


Figure 4-5 Identification of CYP716A with the highest potential for high-level oleanolic acid production in *N. benthamiana* using the Tsukuba system.

(A) TICs of extracts from *N. benthamiana* leaves transiently expressing a combination of LjOSC, LjCPR2, and mutant CYP716A12s (D122Q, I212P, Q358P, D122Q Q358P, and I212P Q358P), wild-type CYP716As (CYP716A12, CYP716A15, CYP716A48, and CYP716A49), or empty vector as a control. β-Amyrin (**1**), erythrodiol (**2**), and oleanolic acid (**3**) were used as authentic standards. Uvaol (*) was used as an internal standard. (B) Quantification of triterpene concentration in *N. benthamiana* leaf extracts analyzed in (A). Letters indicate significant differences in the levels of oleanolic acid (a–d) among samples (one-way ANOVA with Tukey's *post hoc* test, $p < 0.05$). (C) The accumulation of triterpenes in *N. benthamiana* leaves transiently expressing a combination of LjOSC1, LjCPR2, and CYP716A12_D122Q with or without AtHMGR1cd-S577A. Values and error bars represent the mean and standard deviation, and the data are representative of at least three biological replicates with similar results from three independent experiments ($n = 9$). ND, not detected.

4.3.5 Combinatorial biosynthesis of maslinic acid in *Nicotiana benthamiana*

I had a hypothesis that this transient expression platform utilizing the Tsukuba system would be appropriate for high-titer production of oleanolic acid as a precursor of other bioactive triterpenoids. The pentacyclic triterpene maslinic acid, which has a history of usage in alternative medical practices, can be derived from a wide variety of naturally occurring sources (Lozano-Mena et al., 2014; Reyes-Zurita et al., 2016). In engineered yeast, our previously found CYP716C53 from gray mangrove (*Avicennia marina*) demonstrated that it can catalyze the C-2α hydroxylation of oleanolic acid to produce maslinic acid (Nakamura et al., 2018). To produce maslinic acid in *N. benthamiana* leaves using the Tsukuba system, leaves were infiltrated with *A. tumefaciens* harboring pBYR2HS expressing LjOSC1, LjCPR2, CYP716A12_D122Q, and CYP716C53 with or without AtHMGR1cd-S577A. Seven days following agroinfiltration, the leaves were harvested and processed for metabolite extraction. I analyzed and quantified maslinic acid production in leaf extracts using GC-MS analysis (**Figure 4-6A**). A maslinic acid (16) was found as a major peak of oxidized triterpenoid in the leaves of *N. benthamiana* only when CYP716C53 was transiently expressed, based on comparison with the authentic standard. Three novel minor peaks (peaks 17, 18, and 19) were discovered when CYP716C53 was expressed; nevertheless, their chemical structures are unknown. According to the mass spectrometry results, peaks 17 and 19 are most likely isomers of maslinic acid that differ in the position of the hydroxyl group added by CYP716C53 from C-2. Peak 18 was presumed to be a mono-hydroxylated product of β-amyrin (**Figure 4-6A, Figure S-17**).

Maslinic acid production was measured to compare co-expression conditions with and without AtHMGR1cd-S577A (**Figure 4-6B**). *N. benthamiana* leaves transiently expressing LjOSC1, LjCPR2, CYP716A12_D122Q, and CYP716C53 without AtHMGR1cd-S577A showed maslinic acid production of 16.6 ± 1.0 mg/g dw, which was increased 1.6-fold (27.2 ± 3.0 mg/g dw) after co-expressing AtHMGR1cd-S577A (**Figure 4-6B**).

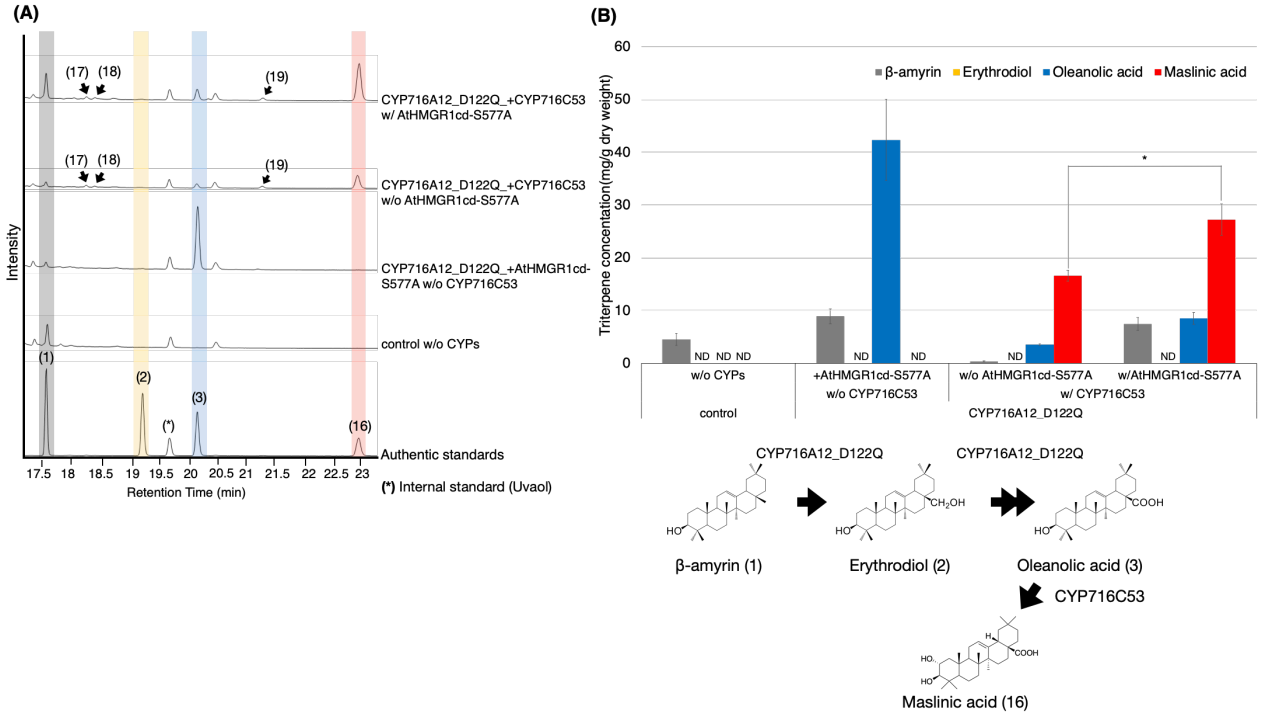


Figure 4-6 Production of maslinic acid in *N. benthamiana* leaves using the Tsukuba system.

TICs of extracts from *N. benthamiana* leaves transiently expressing a combination of LjOSC, LjCPR2, CYP716A12_D122Q, and CYP716C53 with or without AtHMGR1cd-S577A. A combination of pBYR2HS expressing LjOSC, LjCPR2, and empty vector was used as a background control. β-Amyrin (**1**), erythrodiol (**2**), oleanolic acid (**3**), and maslinic acid (**16**) were used as authentic standards. Uvaol (*) was used as an internal standard. (B) Quantification of the concentration of maslinic acid in *N. benthamiana* leaf extract. Data are representative of at least three biological replicates ($n = 3$) and are expressed as the mean ± SEM. Differences in maslinic acid concentration were compared using Student's *t* test for each sample. $*p < 0.05$, ND, not detected.

4.4 Discussion

The bioactive pentacyclic triterpenoids are abundantly found in plants. For instance, olive tree (*Olea europaea*) fruits and leaves are particularly rich in oleanolic and maslinic acids. Concentrated oleanolic and maslinic acids produce a physical barrier on the surface of olive leaves that protects the plant from water loss and infections (Chen et al., 2005; Jager et al., 2009; Romero

et al., 2010; Gudoityte et al., 2021). Recent research has shown that oleanolic acid possesses a wide array of biological actions, including anti-cancer, anti-diabetic, anti-oxidant, cardioprotective, neuroprotective, anti-parasitic, and growth-stimulating characteristics. Additionally, oleanolic acid is a precursor of several industrially useful bioactive triterpenoids (Liu, 1995; Jeon et al., 2019; Kim et al., 2021). CYP716A12 catalyzes the three-step oxidation of C-28 of β-amyrin to produce oleanolic acid via the intermediates erythrodiol (28-hydroxy-β-amyrin) and oleanolic aldehyde (Fukushima et al., 2011; Suzuki et al., 2018). In chapter 2, I found that the product profile of CYP716A12 was modified by substituting key amino acid residues on substrate recognition sites in yeast engineered to produce β-amyrin; three mutant CYP716A12s (D122Q, Q358P, D122Q_Q358P) had the potential for high-level oleanolic acid production, and two mutant CYP716A12s (I212P, I212P_Q358) had the potential for high-level erythrodiol production.

Significant progress in triterpenoid production has recently been reported using microorganisms, particularly *S. cerevisiae*, as alternative hosts (Yao et al., 2021). To achieve high-yield production with such microbial systems, considerable metabolic engineering to increase upstream carbon flux flow from acetyl-CoA to 2,3-oxidosqualene and codon tuning of enzyme activity are typically necessary. In comparison to microbial expression systems, *Nicotiana* species provides the optimal environment for plant-derived biosynthetic enzymes. The chassis of a plant is a rich source of precursor molecules, intermediates from primary metabolism, and enzyme cofactors required for the proper functioning of biosynthetic pathways. Furthermore, several biosynthetic enzymes can be expressed by co-infiltration of various *A. tumefaciens* strains, each harboring a unique design, without the requirement to develop multi-gene constructs (Christ et al., 2019; De La Pena and Sattely, 2021).

Using *N. benthamiana* as a heterologous expression host for rapid, cost - effective production and reconstruction of plant triterpenoids pathways is a translational synthetic biology platform due to its sensitivity to the scalable and adaptable *Agrobacterium*-based transient expression system (Geisler et al., 2013; Reed et al., 2017; Reed and Osbourn, 2018). The Tsukuba system is one of

the most effective transient protein expression systems for plant cells, employing agroinfiltration to transport desired genes into plant cells for efficient recombinant protein production using geminiviral replication machinery (Yamamoto et al., 2018; Suzaki et al., 2019). Heterologous expression in *N. benthamiana* utilizing the Tsukuba system showed that mutant CYP716A12 had potential for high-level oleanolic acid production. The mutant CYP716A12 in which aspartic acid 122 in SRSI was substituted with glutamine exhibited the highest quantity of oleanolic acid production (D122Q). Using this expression system, I show that oleanolic acid production was enhanced 13.1-fold (27.3 ± 2.6 mg/g dw) compared to the conventional binary vector (2.1 ± 0.7 mg/g dw) (**Figure 4-2B**).

As indicated previously, CYP716A12_D122Q showed the highest production in *N. benthamiana*, although CYP716A12_D122Q_Q358P showed the highest production in engineered yeast producing β-amyrin, followed by CYP716A12_D122Q. This might be due to differences in the expression systems, which may affect the catalytic activity of enzyme (Czarnotta et al., 2017), resulting in different oxidized product yields. An erythrodiol was detected as a minor product only when mutant CYP716A12s, in which the key amino acid residue isoleucine 212 (SRSII) was substituted with proline (I212P or I212P_Q358P). In the chapter 2, isoleucine 212 and its neighbors in SRSII of CYP716A12 were identified as residues that significantly influenced the efficacy of C-28 triterpene backbone oxidation. Therefore, it is potential to change the biosynthesis of triterpene-oxidized compounds in *N. benthamiana* by substituting these residues in the ligand-binding pocket.

I compared *N. benthamiana* and *N. tabacum* as hosts for oleanolic acid production in this work. Although there were no statistically significant differences in β-amyrin production between the leaves of *N. tabacum* and *N. benthamiana* (**Figure 4-2C**), the levels of oleanolic acid production in the leaves of *N. benthamiana* were 3,3-fold more than those of *N. tabacum*. In terms of product yield per leaf, this observation suggests that *N. benthamiana* may be a more suitable host for producing oleanolic acid.

Infection by *Agrobacterium* induces necrosis and/or dehydration, which reduces the

transformation and expression efficiency of recombinant proteins (Herlache et al., 2001; Gils et al., 2005; Magdum, 2013). Spraying leaves with sodium L-ascorbate following agroinfiltration reduced leaf necrosis and increased the concentration of human recombinant proteins transiently produced in *N. benthamiana* (Nosaki et al., 2021). This finding reveals that spraying 7–9 days after agroinfiltration with 200 mM ascorbic acid caused a nearly two - fold increase in oleanolic acid accumulation (**Figure 4-4)**.

CYPs are involved for the structural diversity of triterpenoids by modifying the triterpene scaffold. To catalyze site-specific oxidation, these enzymes require receive electrons from NADPH-CPR (Istiandari et al., 2021). Triterpenoid production can be increased when CYPs are established in a microenvironment conducive to their optimal performance (Zhu et al., 2018; Liu et al., 2019; Theron et al., 2019; Istiandari et al., 2021; Liu et al., 2021). I observed a 2.70-fold increase in oleanolic acid production following co-expression with LjCPR2 in this investigation (**Figure 4-3B**). Additionally, our previous research demonstrated that the co-expression of CPRs derived from plant species may or may not be the appropriate partners for CYPs in yeast (Istiandari et al., 2021). For high-yield triterpenoid production in *N. benthamiana*, the appropriate combination of CYPs and CPRs is essential. In addition, for optimal CYP activity, future research must establish the appropriate CPR:CYP ratio.

In a previous report, the triterpenoid 12,13-epoxy,16-hydroxy-β-amyrin was obtained at a yield of 1.18 mg/g dw from leaves following transient expression in *N. benthamiana* (Geisler et al., 2013; Reed et al., 2017). Moreover, co-expression of the HMGR catalytic domain improved the triterpenoid production; the levels of simple (β-amyrin) and oxidized triterpenoid (i.e., 12,13-epoxy,16-hydroxy-β-amyrin) production reached 3.3 and 3.9 mg/g dw, respectively (Reed et al., 2017; Reed and Osbourn, 2018). In our previous studies using a conventional binary vector, a α-amyrin production was increased in *N. benthamiana* after transient co-expression of AtHMGR1cd with the S577A mutation (AtHMGR1cd-S577A) lacking the inactivating phosphorylation site (Robertlee et al., 2018; Srisawat et al., 2019). Using the Tsukuba system, I demonstrated that

transient expression of AtHMGR1cd-S577A increased triterpenoid synthesis in *N. benthamiana*. After transient co-expression of AtHMGR1cd-S577A, the concentration of both simple and oxidized triterpenoids increased, resulting in the production of β-amyrin (9.4 mg/g dw, **Figure 4-5C**), oleanolic acid (30.8 mg/g dw, **Figure 4-5C**), and maslinic acid (27.2 mg/g dw, **Figure 4-6B**). Therefore, this transient expression system in *N. benthamiana* resulted in more triterpenoid synthesis than the previously reported yield (Reed et al., 2017; Reed and Osbourn, 2018).

Maslinic acid, a pentacyclic triterpene used for centuries in traditional medicine, possesses anticancer, anti-diabetic, antioxidant, cardioprotective, anti-parasitic, and growth-stimulating activities (Jager et al., 2009; Romero et al., 2010; Reyes-Zurita et al., 2016). Maslinic acid was obtained as a minor product in our previous in vivo functional analysis of CYP716C53 in engineered yeast when co-expressed with bAS, CPR, and CYP716A259 (β-amyrin C-28 oxidase of *A. marina*) (Nakamura et al., 2018). When CYP716C53 was expressed in *N. benthamiana* leaves using the Tsukuba system, maslinic acid was identified as the major oxidized product **(Figure 4-6)**. Maslinic acid is found in a variety of natural sources, including olives, which are a particularly rich source (Romero et al., 2010; Lozano-Mena et al., 2014). Age of the plant may also influence the accumulation of triterpenoids in raw plant materials (Jager et al., 2009; Romero et al., 2010; Reyes-Zurita et al., 2016). Maslinic acid was generated in the leaves of *N. benthamiana* in this investigation, yielding 27.2 ± 3.0 mg/g dw 7 days after agroinfiltration **(Figure 4-6B).** This quantity was 20.7 times that of olive fruit (cv. Kalamata), which has been reported to contain a high concentration of maslinic acid (1.3 mg/g dw) (Romero et al., 2010; Lozano-Mena et al., 2014).

4.5 Conclusion

In conclusion, our investigation developed a framework for enhancing oleanolic acid production as a precursor of bioactive triterpenoids via *Agrobacterium*-mediated transient expression in *N. benthamiana* using the Tsukuba system. I validated the efficacy of this method by synthesizing considerable quantities of maslinic acid, an example of a triterpenoid produced from

a plant with medicinal activity. This research will lead to the development of a more efficient platform for synthesizing and gaining access to previously unavailable natural products and analogues, as well as the potential revitalization of drug discovery pipelines.

Chapter 5 : General conclusion

Plants produce triterpenoids, which are structurally and pharmacologically active metabolites. Despite this, their purification and chemical synthesis are difficult, and their productivity in plants is limited. The site-specific oxidation of the triterpenoid backbone is catalyzed by cytochrome P450 monooxygenases (CYPs), an enzyme family important in triterpenoid biosynthesis. Many plant species have been found to contain enzymes belonging to the CYP716 family. The triterpene backbone is most commonly oxidized at the C-28 position by CYP716A in a three-step process (Carelli et al., 2011; Fukushima et al., 2011; Yasumoto et al., 2016; Miettinen et al., 2017; Yasumoto et al., 2017; Sandeep et al., 2019). However, their crystal structures are not publicly available, and the specific catalytic mechanism that they implement is an unknown.

Protein engineering of the CYP716A subfamily enzyme is being used in this study with the intention of achieving the general goal of increasing the production of bioactive triterpenoids in heterologous hosts. To achieve this objective, I decided to use the *M. truncatula* CYP716A12 enzyme as the first CYP716A subfamily enzyme to be functionally characterized.

Due to the findings of this study, I am now able to determine the key amino acid residues in C-28 oxidases that are important for the catalytic activity and substrate specificity of CYP716A12 (D122, I212, and Q358). These key amino acid residues were previously unknown. A CYP716A12 was then designed to increase its catalytic activity of enzymes in triterpenoids biosynthesis. The substrate specificity of CYP716A12 remained unchanged despite substituting key amino acid residues; nevertheless, the enzyme's catalytic activity was increased. In addition, I discovered the amino acid residues that control the substrate contraction of the enzyme (D292). I also generated mutants by modifying the characteristics of key residues in SRSs to enhance the catalytic activity of CYP716A1 and CYP716A2 at C-28 on diverse triterpene backbones. Finally, all C-28 oxidase

products (oleanolic acid, ursolic acid, and betulinic acid) were detected in yeast engineered to harbor CYP716A1 or CYP716A2 mutants.

This study led to the development of a plant-platform for high-yield production of valuable triterpenoids via heterologous expression in *Nicotiana benthamiana* through the identification and engineering of plant-derived cytochrome P450 monooxygenases. The advantages of identified mutant CYP716A12 were combined with the Tsukuba system, one of the most efficient agroinfiltration-based transient protein expression systems, to evaluate the applicability of the Tsukuba system for triterpenoid production in *N. benthamiana* via pathway reconstruction. Using this plant platform, I was able to detect the high yield of oleanolic and maslinic acids, which are being studied as potential new therapeutic treatments.

This research has the potential to increase the production yield of desired triterpenoids in engineered yeast and *N. benthamiana* by increasing the catalytic activity and substrate specificity of CYP716A subfamily enzymes. This research will also contribute to the synthesis and accessibility of previously unavailable triterpenoids and analogs, as well as the revitalization of drug discovery pipelines.

5.1 Future perspectives

In this research suggests a strategy for increasing the understanding of critical amino acid residues involved in the catalytic activity and substrate specificity of CYP716A subfamily enzymes to improve triterpenoid production in heterologous host. However, the low similarity of the homology model to the experimental template might have been a constraint on accurately identifying of orientation of amino acid residues, as well as the interaction and contraction between substrate and enzyme. Therefore, the enzyme's specific catalytic mechanism following substitution of critical residues are required to study in future experiment. The future update of the experimental crystal structure of plant derived CYPs involved in triterpenoid biosynthesis might provide more

accurate information for constructing the homology model of CYP716A subfamily enzymes and suggest more knowledge about their roles in triterpenoid biosynthesis.

The quantifying of CYPs enzymes via *in vitro* experiment remains a difficult task. To achieve and facilitate optimal CYPs activity in the heterologous host, the optimal pair match and sustainable molecular ratio of CPR as an electron transfer partner of CYPs must be considered at the molecular level.

Reference

Al-Jumaili, A., Kumar, A., Bazaka, K., and Jacob, M.V. (2018). Plant Secondary Metabolite-Derived Polymers: A Potential Approach to Develop Antimicrobial Films. *Polymers (Basel)* 10(5). doi: 10.3390/polym10050515.

Ayeleso, T.B., Matumba, M.G., and Mukwevho, E. (2017). Oleanolic Acid and Its Derivatives: Biological Activities and Therapeutic Potential in Chronic Diseases. *Molecules* 22(11). doi: 10.3390/molecules22111915.

Baek, I., Choi, H., Yoon, S., and Na, S. (2020). Effects of the Hydrophobicity of Key Residues on the Characteristics and Stability of Glucose Oxidase on a Graphene Surface. *ACS Biomater Sci Eng* 6(4)**,** 1899-1908. doi: 10.1021/acsbiomaterials.9b01763.

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Hofer, R., Paquette, S., et al. (2011). Cytochromes p450. *Arabidopsis Book* 9**,** e0144. doi: 10.1199/tab.0144.

Blaha-Nelson, D., Kruger, D.M., Szeler, K., Ben-David, M., and Kamerlin, S.C. (2017). Active Site Hydrophobicity and the Convergent Evolution of Paraoxonase Activity in Structurally Divergent Enzymes: The Case of Serum Paraoxonase 1. *J Am Chem Soc* 139(3)**,** 1155-1167. doi: 10.1021/jacs.6b10801.

Bose, B., Tripathy, D., Chatterjee, A., Tandon, P., and Kumaria, S. (2019). Secondary metabolite profiling, cytotoxicity, anti-inflammatory potential and in vitro inhibitory activities of Nardostachys jatamansi on key enzymes linked to hyperglycemia, hypertension and cognitive disorders. *Phytomedicine* 55**,** 58-69. doi: 10.1016/j.phymed.2018.08.010.

Buyel, J.F., Twyman, R.M., and Fischer, R. (2017). Very-large-scale production of antibodies in plants: The biologization of manufacturing. *Biotechnol Adv* 35(4)**,** 458-465. doi: 10.1016/j.biotechadv.2017.03.011.

Cankar, K., Jongedijk, E., Klompmaker, M., Majdic, T., Mumm, R., Bouwmeester, H., et al. (2015). (+)-Valencene production in Nicotiana benthamiana is increased by down-regulation of competing pathways. *Biotechnol J* 10(1)**,** 180-189. doi: 10.1002/biot.201400288.

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., et al. (2011). Medicago truncatula CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell* 23(8)**,** 3070-3081. doi: 10.1105/tpc.111.087312.

Carsanba, E., Pintado, M., and Oliveira, C. (2021). Fermentation Strategies for Production of Pharmaceutical Terpenoids in Engineered Yeast. *Pharmaceuticals (Basel)* 14(4). doi: 10.3390/ph14040295.

Chen, Y., Liu, J., Yang, X., Zhao, X., and Xu, H. (2005). Oleanolic acid nanosuspensions: preparation, in-vitro characterization and enhanced hepatoprotective effect. *J Pharm Pharmacol* 57(2)**,** 259-264. doi: 10.1211/0022357055407.

Choi, E.J., and Mayo, S.L. (2006). Generation and analysis of proline mutants in protein G. *Protein Eng Des Sel* 19(6)**,** 285-289. doi: 10.1093/protein/gzl007.

Christ, B., Xu, C., Xu, M., Li, F.S., Wada, N., Mitchell, A.J., et al. (2019). Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat Commun* 10(1)**,** 3206. doi: 10.1038/s41467-019-11286-7.

Clemente, T. (2006). Nicotiana (Nicotiana tobaccum, Nicotiana benthamiana). *Methods Mol Biol* 343**,** 143-154. doi: 10.1385/1-59745-130-4:143.

Cravens, A., Payne, J., and Smolke, C.D. (2019). Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat Commun* 10(1)**,** 2142. doi: 10.1038/s41467-019-09848-w.

Cui, J.L., Zhang, Y.Y., Vijayakumar, V., Zhang, G., Wang, M.L., and Wang, J.H. (2018). Secondary Metabolite Accumulation Associates with Ecological Succession of Endophytic Fungi in Cynomorium songaricum Rupr. *J Agric Food Chem* 66(22)**,** 5499-5509. doi: 10.1021/acs.jafc.8b01737.

Czarnotta, E., Dianat, M., Korf, M., Granica, F., Merz, J., Maury, J., et al. (2017). Fermentation and purification strategies for the production of betulinic acid and its lupane-type precursors

in Saccharomyces cerevisiae. *Biotechnol Bioeng* 114(11)**,** 2528-2538. doi: 10.1002/bit.26377.

Dale, M.P., Moses, T., Johnston, E.J., and Rosser, S.J. (2020). A systematic comparison of triterpenoid biosynthetic enzymes for the production of oleanolic acid in Saccharomyces cerevisiae. *PLoS One* 15(5)**,** e0231980. doi: 10.1371/journal.pone.0231980.

de Frias, U.A., Pereira, G.K.B., Guazzaroni, M.E., and Silva-Rocha, R. (2018). Boosting Secondary Metabolite Production and Discovery through the Engineering of Novel Microbial Biosensors. *Biomed Res Int* 2018**,** 7021826. doi: 10.1155/2018/7021826.

De La Pena, R., and Sattely, E.S. (2021). Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nat Chem Biol* 17(2)**,** 205-212. doi: 10.1038/s41589-020-00669-3.

Dzubak, P., Hajduch, M., Vydra, D., Hustova, A., Kvasnica, M., Biedermann, D., et al. (2006). Pharmacological activities of natural triterpenoids and their therapeutic implications. *Nat Prod Rep* 23(3)**,** 394-411. doi: 10.1039/b515312n.

Eisenberg, D., Luthy, R., and Bowie, J.U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277**,** 396-404. doi: 10.1016/s0076-6879(97)77022-8.

Ellis, S.W., Hayhurst, G.P., Smith, G., Lightfoot, T., Wong, M.M., Simula, A.P., et al. (1995). Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6. *J Biol Chem* 270(49)**,** 29055-29058. doi: 10.1074/jbc.270.49.29055.

Fontana, E., Dansette, P.M., and Poli, S.M. (2005). Cytochrome p450 enzymes mechanism based inhibitors: common sub-structures and reactivity. *Curr Drug Metab* 6(5)**,** 413-454. doi: 10.2174/138920005774330639.

Fukumitsu, S., Villareal, M.O., Aida, K., Hino, A., Hori, N., Isoda, H., et al. (2016). Maslinic acid in olive fruit alleviates mild knee joint pain and improves quality of life by promoting weight loss in the elderly. *J Clin Biochem Nutr* 59(3)**,** 220-225. doi: 10.3164/jcbn.16-40.

Fukushima, E.O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., et al. (2011). CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol* 52(12), 2050-2061. doi: 10.1093/pcp/pcr146.

Fukushima, E.O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., et al. (2013). Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol* 54(5), 740-749. doi: 10.1093/pcp/pct015.

Geisler, K., Hughes, R.K., Sainsbury, F., Lomonossoff, G.P., Rejzek, M., Fairhurst, S., et al. (2013). Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc Natl Acad Sci U S A* 110(35), E3360-3367. doi: 10.1073/pnas.1309157110.

Gils, M., Kandzia, R., Marillonnet, S., Klimyuk, V., and Gleba, Y. (2005). High-yield production of authentic human growth hormone using a plant virus-based expression system. *Plant Biotechnol J* 3(6), 613-620. doi: 10.1111/j.1467-7652.2005.00154.x.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* 267(1), 83-90.

Gudoityte, E., Arandarcikaite, O., Mazeikiene, I., Bendokas, V., and Liobikas, J. (2021). Ursolic and Oleanolic Acids: Plant Metabolites with Neuroprotective Potential. *Int J Mol Sci* 22(9). doi: 10.3390/ijms22094599.

Herlache, T.C., Zhang, H.S., Ried, C.L., Carle, S.A., Zheng, D., Basaran, P., et al. (2001). Mutations that Affect Agrobacterium vitis-Induced Grape Necrosis also Alter Its Ability to Cause a Hypersensitive Response on Tobacco. *Phytopathology* 91(10), 966-972. doi: 10.1094/PHYTO.2001.91.10.966.

https://clinicaltrials.gov (2000). Available: https://clinicaltrials.gov [Accessed April 28 2022].

Huang, A.C., Kautsar, S.A., Hong, Y.J., Medema, M.H., Bond, A.D., Tantillo, D.J., et al. (2017). Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome

mining reveals convergent evolution. *Proc Natl Acad Sci U S A* 114(29)**,** E6005-E6014. doi: 10.1073/pnas.1705567114.

Istiandari, P., Yasumoto, S., Srisawat, P., Tamura, K., Chikugo, A., Suzuki, H., et al. (2021). Comparative Analysis of NADPH-Cytochrome P450 Reductases From Legumes for Heterologous Production of Triterpenoids in Transgenic Saccharomyces cerevisiae. *Front Plant Sci* 12**,** 762546. doi: 10.3389/fpls.2021.762546.

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(6)**,** 2016-2031. doi: 10.3390/molecules14062016.

Jeon, D., Kim, S.W., and Kim, H.S. (2019). Platycodin D, a bioactive component of Platycodon grandiflorum, induces cancer cell death associated with extreme vacuolation. *Anim Cells Syst (Seoul)* 23(2)**,** 118-127. doi: 10.1080/19768354.2019.1588163.

Jin, C.C., Zhang, J.L., Song, H., and Cao, Y.X. (2019). Boosting the biosynthesis of betulinic acid and related triterpenoids in Yarrowia lipolytica via multimodular metabolic engineering. *Microb Cell Fact* 18(1)**,** 77. doi: 10.1186/s12934-019-1127-8.

Kensil, C.R., Wu, J.Y., Anderson, C.A., Wheeler, D.A., and Amsden, J. (1998). QS-21 and QS-7: purified saponin adjuvants. *Dev Biol Stand* 92**,** 41-47.

Kim, T.Y., Jeon, S., Jang, Y., Gotina, L., Won, J., Ju, Y.H., et al. (2021). Platycodin D, a natural component of Platycodon grandiflorum, prevents both lysosome- and TMPRSS2-driven SARS-CoV-2 infection by hindering membrane fusion. *Exp Mol Med* 53(5)**,** 956-972. doi: 10.1038/s12276-021-00624-9.

Kuhnel, K., Ke, N., Cryle, M.J., Sligar, S.G., Schuler, M.A., and Schlichting, I. (2008). Crystal structures of substrate-free and retinoic acid-bound cyanobacterial cytochrome P450 CYP120A1. *Biochemistry* 47(25)**,** 6552-6559. doi: 10.1021/bi800328s.

Li, G., Yao, P., Gong, R., Li, J., Liu, P., Lonsdale, R., et al. (2017). Simultaneous engineering of an enzyme's entrance tunnel and active site: the case of monoamine oxidase MAO-N. *Chem Sci* 8(5)**,** 4093-4099. doi: 10.1039/c6sc05381e.

Li, X., Cui, J., Yu, Y., Li, W., Hou, Y., Wang, X., et al. (2016). Traditional Chinese Nootropic Medicine Radix Polygalae and Its Active Constituent Onjisaponin B Reduce beta-Amyloid Production and Improve Cognitive Impairments. *PLoS One* 11(3)**,** e0151147. doi: 10.1371/journal.pone.0151147.

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(4)**,** 972-1003. doi: 10.1124/pr.111.004846.

Liu, G., Anderson, C., Scaltreto, H., Barbon, J., and Kensil, C.R. (2002). QS-21 structure/function studies: effect of acylation on adjuvant activity. *Vaccine* 20(21-22)**,** 2808-2815. doi: 10.1016/s0264-410x(02)00209-8.

Liu, J. (1995). Pharmacology of oleanolic acid and ursolic acid. *J Ethnopharmacol* 49(2)**,** 57-68. doi: 10.1016/0378-8741(95)90032-2.

Liu, J., Zhang, C., and Lu, W. (2019). Biosynthesis of Long-Chain omega-Hydroxy Fatty Acids by Engineered Saccharomyces cerevisiae. *J Agric Food Chem* 67(16)**,** 4545-4552. doi: 10.1021/acs.jafc.9b00109.

Liu, Q., Majdi, M., Cankar, K., Goedbloed, M., Charnikhova, T., Verstappen, F.W., et al. (2011). Reconstitution of the costunolide biosynthetic pathway in yeast and Nicotiana benthamiana. *PLoS One* 6(8)**,** e23255. doi: 10.1371/journal.pone.0023255.

Liu, T., Huang, Y., Jiang, L., Dong, C., Gou, Y., and Lian, J. (2021). Efficient production of vindoline from tabersonine by metabolically engineered Saccharomyces cerevisiae. *Commun Biol* 4(1)**,** 1089. doi: 10.1038/s42003-021-02617-w.

Lozano-Mena, G., Sanchez-Gonzalez, M., Juan, M.E., and Planas, J.M. (2014). Maslinic acid, a natural phytoalexin-type triterpene from olives--a promising nutraceutical? *Molecules* 19(8)**,** 11538-11559. doi: 10.3390/molecules190811538.

Madden, T.L., Tatusov, R.L., and Zhang, J. (1996). Applications of network BLAST server. *Methods Enzymol* 266**,** 131-141. doi: 10.1016/s0076-6879(96)66011-x.

Magdum, S.S. (2013). Effect of Agrobacterium Induced Necrosis, Antibiotic Induced Phytotoxicity and Other Factors in Successful Plant. *Journal of Stress Physiology & Biochemistry* 9(3)**,** 98-112.

Matthews, B.W., Nicholson, H., and Becktel, W.J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci U S A* 84(19)**,** 6663-6667. doi: 10.1073/pnas.84.19.6663.

Miettinen, K., Pollier, J., Buyst, D., Arendt, P., Csuk, R., Sommerwerk, S., et al. (2017). The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nat Commun* 8**,** 14153. doi: 10.1038/ncomms14153.

Mizutani, M., and Sato, F. (2011). Unusual P450 reactions in plant secondary metabolism. *Arch Biochem Biophys* 507(1)**,** 194-203. doi: 10.1016/j.abb.2010.09.026.

Moses, T., Mehrshahi, P., Smith, A.G., and Goossens, A. (2017). Synthetic biology approaches for the production of plant metabolites in unicellular organisms. *J Exp Bot* 68(15)**,** 4057-4074. doi: 10.1093/jxb/erx119.

Moses, T., Pollier, J., Shen, Q., Soetaert, S., Reed, J., Erffelinck, M.L., et al. (2015). OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of Artemisia annua. *Plant Cell* 27(1)**,** 286-301. doi: 10.1105/tpc.114.134486.

Moses, T., Pollier, J., Thevelein, J.M., and Goossens, A. (2013). Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. *New Phytol* 200(1)**,** 27-43. doi: 10.1111/nph.12325.

Nakamura, M., Linh, T.M., Lien, L.Q., Suzuki, H., Mai, N.C., Giang, V.H., et al. (2018). Transcriptome sequencing and identification of cytochrome P450 monooxygenases involved in the biosynthesis of maslinic acid and corosolic acid in Avicennia marina. *Plant Biotechnol (Tokyo)* 35(4)**,** 341-348. doi: 10.5511/plantbiotechnology.18.0810a.

Narayan, A.R., Jimenez-Oses, G., Liu, P., Negretti, S., Zhao, W., Gilbert, M.M., et al. (2015). Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations. *Nat Chem* 7(8)**,** 653-660. doi: 10.1038/nchem.2285.

Nascimento, N.C., and Fett-Neto, A.G. (2010). Plant secondary metabolism and challenges in modifying its operation: an overview. *Methods Mol Biol* 643**,** 1-13. doi: 10.1007/978-1-60761-723-5_1.

Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J* 66(1)**,** 194-211. doi: 10.1111/j.1365-313X.2011.04529.x.

Nelson, D.R. (2006). Cytochrome P450 nomenclature, 2004. *Methods Mol Biol* 320**,** 1-10. doi: 10.1385/1-59259-998-2:1.

Nelson, D.R., Zeldin, D.C., Hoffman, S.M., Maltais, L.J., Wain, H.M., and Nebert, D.W. (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* 14(1)**,** 1-18.

Nosaki, S., Kaneko, M.K., Tsuruta, F., Yoshida, H., Kato, Y., and Miura, K. (2021). Prevention of necrosis caused by transient expression in Nicotiana benthamiana by application of ascorbic acid. *Plant Physiol* 186(2)**,** 832-835. doi: 10.1093/plphys/kiab102.

Peng, F., Lu, L., Wei, F., Wu, D., Wang, K., and Tang, J. (2020). The onjisaponin B metabolite tenuifolin ameliorates dopaminergic neurodegeneration in a mouse model of Parkinson's disease. *Neuroreport* 31(6)**,** 456-465. doi: 10.1097/WNR.0000000000001428.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13)**,** 1605-1612. doi: 10.1002/jcc.20084.

Phillips, D.R., Rasbery, J.M., Bartel, B., and Matsuda, S.P. (2006). Biosynthetic diversity in plant triterpene cyclization. *Curr Opin Plant Biol* 9(3)**,** 305-314. doi: 10.1016/j.pbi.2006.03.004.

Piasecka, A., Jedrzejczak-Rey, N., and Bednarek, P. (2015). Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytol* 206(3)**,** 948-964. doi: 10.1111/nph.13325.

Polakowski, T., Stahl, U., and Lang, C. (1998). Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. *Appl Microbiol Biotechnol* 49(1)**,** 66-71. doi: 10.1007/s002530051138.

Poulos, T.L., and Johnson, E.F. (2015). "Structures of Cytochrome P450 Enzymes," in *Cytochrome P450: Structure, Mechanism, and Biochemistry,* ed. P.R. Ortiz de Montellano. (Cham: Springer International Publishing), 3-32.

Prall, W., Hendy, O., and Thornton, L.E. (2016). Utility of a Phylogenetic Perspective in Structural Analysis of CYP72A Enzymes from Flowering Plants. *PLoS One* 11(9)**,** e0163024. doi: 10.1371/journal.pone.0163024.

Pusztahelyi, T., Holb, I.J., and Pocsi, I. (2015). Secondary metabolites in fungus-plant interactions. *Front Plant Sci* 6**,** 573. doi: 10.3389/fpls.2015.00573.

Reed, J., and Osbourn, A. (2018). Engineering terpenoid production through transient expression in Nicotiana benthamiana. *Plant Cell Rep* 37(10)**,** 1431-1441. doi: 10.1007/s00299-018-2296-3.

Reed, J., Stephenson, M.J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., et al. (2017). A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab Eng* 42**,** 185-193. doi: 10.1016/j.ymben.2017.06.012.

Reyes-Zurita, F.J., Rufino-Palomares, E.E., Garcia-Salguero, L., Peragon, J., Medina, P.P., Parra, A., et al. (2016). Maslinic Acid, a Natural Triterpene, Induces a Death Receptor-Mediated Apoptotic Mechanism in Caco-2 p53-Deficient Colon Adenocarcinoma Cells. *PLoS One* 11(1), e0146178. doi: 10.1371/journal.pone.0146178.

Robertlee, J., Kobayashi, K., Tang, J., Suzuki, M., and Muranaka, T. (2018). Evidence that the Arabidopsis thaliana 3-hydroxy-3-methylglutaryl-CoA reductase 1 is phosphorylated at Ser577 in planta. *Plant Biotechnol (Tokyo)* 35(1), 1-7. doi: 10.5511/plantbiotechnology.17.1208a.

Romero, C., Garcia, A., Medina, E., Ruiz-Mendez, M.V., de Castro, A., and Brenes, M. (2010). Triterpenic acids in table olives. *Food Chemistry* 118(3), 670-674. doi: 10.1016/j.foodchem.2009.05.037.

Saito, K. (2003). [Plant metabolomics]. *Tanpakushitsu Kakusan Koso* 48(15 Suppl), 2199-2204.

Sandeep, Misra, R.C., Chanotiya, C.S., Mukhopadhyay, P., and Ghosh, S. (2019). Oxidosqualene cyclase and CYP716 enzymes contribute to triterpene structural diversity in the medicinal tree banaba. *New Phytol* 222(1), 408-424. doi: 10.1111/nph.15606.

Sawai, S., and Saito, K. (2011). Triterpenoid biosynthesis and engineering in plants. *Front Plant Sci* 2, 25. doi: 10.3389/fpls.2011.00025.

Schrodinger, LLC (2015). "The PyMOL Molecular Graphics System, Version 1.8".).

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2008). Licorice beta-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proc Natl Acad Sci U S A* 105(37), 14204-14209. doi: 10.1073/pnas.0803876105.

Seki, H., Tamura, K., and Muranaka, T. (2015). P450s and UGTs: Key Players in the Structural Diversity of Triterpenoid Saponins. *Plant Cell Physiol* 56(8), 1463-1471. doi: 10.1093/pcp/pcv062.

Sheng, H., and Sun, H. (2011). Synthesis, biology and clinical significance of pentacyclic triterpenes: a multi-target approach to prevention and treatment of metabolic and vascular diseases. *Nat Prod Rep* 28(3)**,** 543-593. doi: 10.1039/c0np00059k.

Siddiqui, M.S., Thodey, K., Trenchard, I., and Smolke, C.D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res* 12(2)**,** 144-170. doi: 10.1111/j.1567-1364.2011.00774.x.

Sirim, D., Widmann, M., Wagner, F., and Pleiss, J. (2010). Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Struct Biol* 10**,** 34. doi: 10.1186/1472-6807-10-34.

Srisawat, P., Fukushima, E.O., Yasumoto, S., Robertlee, J., Suzuki, H., Seki, H., et al. (2019). Identification of oxidosqualene cyclases from the medicinal legume tree Bauhinia forficata: a step toward discovering preponderant alpha-amyrin-producing activity. *New Phytol* 224(1)**,** 352-366. doi: 10.1111/nph.16013.

Sugimoto, H., and Shiro, Y. (2012). Diversity and substrate specificity in the structures of steroidogenic cytochrome P450 enzymes. *Biol Pharm Bull* 35(6)**,** 818-823. doi: 10.1248/bpb.35.818.

Sumner, L.W., Mendes, P., and Dixon, R.A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62(6)**,** 817-836.

Sun, W., Xue, H., Liu, H., Lv, B., Yu, Y., Wang, Y., et al. (2020). Controlling Chemo- and Regioselectivity of a Plant P450 in Yeast Cell toward Rare Licorice Triterpenoid Biosynthesis. *ACS Catalysis* 10(7)**,** 4253-4260. doi: 10.1021/acscatal.0c00128.

Suzaki, T., Tsuda, M., Ezura, H., Day, B., and Miura, K. (2019). Agroinfiltration-based efficient transient protein expression in leguminous plants. *Plant Biotechnol (Tokyo)* 36(2)**,** 119-123. doi: 10.5511/plantbiotechnology.19.0220b.

Suzuki, H., Fukushima, E.O., Umemoto, N., Ohyama, K., Seki, H., and Muranaka, T. (2018). Comparative analysis of CYP716A subfamily enzymes for the heterologous production of

C-28 oxidized triterpenoids in transgenic yeast. *Plant Biotechnol (Tokyo)* 35(2)**,** 131-139. doi: 10.5511/plantbiotechnology.18.0416a.

Tamura, K., Teranishi, Y., Ueda, S., Suzuki, H., Kawano, N., Yoshimatsu, K., et al. (2017). Cytochrome P450 Monooxygenase CYP716A141 is a Unique beta-Amyrin C-16beta Oxidase Involved in Triterpenoid Saponin Biosynthesis in Platycodon grandiflorus. *Plant Cell Physiol* 58(5)**,** 874-884. doi: 10.1093/pcp/pcx043.

Theron, C.W., Labuschagne, M., Albertyn, J., and Smit, M.S. (2019). Heterologous coexpression of the benzoate-para-hydroxylase CYP53B1 with different cytochrome P450 reductases in various yeasts. *Microb Biotechnol* 12(6)**,** 1126-1138. doi: 10.1111/1751-7915.13321.

Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2**,** Unit 2 3. doi: 10.1002/0471250953.bi0203s00.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2)**,** 455-461. doi: 10.1002/jcc.21334.

van Herpen, T.W., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H.J., and Beekwilder, J. (2010). Nicotiana benthamiana as a production platform for artemisinin precursors. *PLoS One* 5(12)**,** e14222. doi: 10.1371/journal.pone.0014222.

Webb, B., and Sali, A. (2017). Protein Structure Modeling with MODELLER. *Methods Mol Biol* 1654**,** 39-54. doi: 10.1007/978-1-4939-7231-9_4.

Xiao, H., and Zhong, J.J. (2016). Production of Useful Terpenoids by Higher-Fungus Cell Factory and Synthetic Biology Approaches. *Trends Biotechnol* 34(3)**,** 242-255. doi: 10.1016/j.tibtech.2015.12.007.

Xu, R., Fazio, G.C., and Matsuda, S.P. (2004). On the origins of triterpenoid skeletal diversity. *Phytochemistry* 65(3)**,** 261-291. doi: 10.1016/j.phytochem.2003.11.014.

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(10)**,** 2428-2466. doi: 10.3390/toxins2102428.

Yamamoto, T., Hoshikawa, K., Ezura, K., Okazawa, R., Fujita, S., Takaoka, M., et al. (2018). Improvement of the transient expression system for production of recombinant proteins in plants. *Sci Rep* 8(1)**,** 4755. doi: 10.1038/s41598-018-23024-y.

Yan, X.J., Gong, L.H., Zheng, F.Y., Cheng, K.J., Chen, Z.S., and Shi, Z. (2014). Triterpenoids as reversal agents for anticancer drug resistance treatment. *Drug Discov Today* 19(4)**,** 482-488. doi: 10.1016/j.drudis.2013.07.018.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12(1)**,** 7-8. doi: 10.1038/nmeth.3213.

Yao, L., Wang, J., He, J., Huang, L., and Gao, W. (2021). Endophytes, biotransforming microorganisms, and engineering microbial factories for triterpenoid saponins production. *Crit Rev Biotechnol* 41(2)**,** 249-272. doi: 10.1080/07388551.2020.1869691.

Yasumoto, S., Fukushima, E.O., Seki, H., and Muranaka, T. (2016). Novel triterpene oxidizing activity of Arabidopsis thaliana CYP716A subfamily enzymes. *FEBS Lett* 590(4)**,** 533-540. doi: 10.1002/1873-3468.12074.

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. *Front Plant Sci* 8**,** 21. doi: 10.3389/fpls.2017.00021.

Yuan, H., Wu, J., Wang, X., Chen, J., Zhong, Y., Huang, Q., et al. (2017). Computational Identification of Amino-Acid Mutations that Further Improve the Activity of a Chalcone-Flavonone Isomerase from Glycine max. *Front Plant Sci* 8**,** 248. doi: 10.3389/fpls.2017.00248.

Yuki, H., Honma, T., Hata, M., and Hoshino, T. (2012). Prediction of sites of metabolism in a substrate molecule, instanced by carbamazepine oxidation by CYP3A4. *Bioorg Med Chem* 20(2)**,** 775-783. doi: 10.1016/j.bmc.2011.12.004.

Zhang, C., and Hong, K. (2020). Production of Terpenoids by Synthetic Biology Approaches. *Front Bioeng Biotechnol* 8**,** 347. doi: 10.3389/fbioe.2020.00347.

Zhou, C., Li, J., Li, C., and Zhang, Y. (2016). Improvement of betulinic acid biosynthesis in yeast employing multiple strategies. *BMC Biotechnol* 16(1)**,** 59. doi: 10.1186/s12896-016-0290-9.

Zhu, D., and Tuo, W. (2016). QS-21: A Potent Vaccine Adjuvant. *Nat Prod Chem Res* 3(4). doi: 10.4172/2329-6836.1000e113.

Zhu, M., Wang, C., Sun, W., Zhou, A., Wang, Y., Zhang, G., et al. (2018). Boosting 11-oxo-beta-amyrin and glycyrrhetinic acid synthesis in Saccharomyces cerevisiae via pairing novel oxidation and reduction system from legume plants. *Metab Eng* 45**,** 43-50. doi: 10.1016/j.ymben.2017.11.009.

Publication

Romsuk, J., Yasumoto, S., Fukushima, E.O., Miura, K., Muranaka, T., and Seki, H. (2022). High-yield bioactive triterpenoid production by heterologous expression in Nicotiana benthamiana using the Tsukuba system. *Frontiers in Plant Science* 13. doi: 10.3389/fpls.2022.991909.

Romsuk, J., Yasumoto, S., Seki, H., Fukushima, E.O., and Muranaka, T. (2022). Identification of key amino acid residues toward improving the catalytic activity and substrate specificity of plant-derived cytochrome P450 monooxygenases CYP716A subfamily enzyme for triterpenoid production in Saccharomyces cerevisiae. *Frontiers in Bioengineering and Biotechnology* 10. doi: 10.3389/fbioe.2022.955650.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Toshiya Muranaka, Assoc. Prof. Dr. Hikaru Seki, Assoc. Prof. Dr. Ery Odette Fukushima and Asst. Prof. Dr. Shuhei Yasumoto, for patience, motivation, enthusiasm, and immense knowledge on every major point of my work. Their valuable suggestions, technical support, and discussion have been the most rewarding and informative to my thesis. Besides my advisor, I would like to thank the rest of my thesis committee: Professor Kazuhito Fujiyama, and Professor Genji Kurisu, for their encouragement, incisive comments, and insightful questions.

I would like to thank the Biotechnology Global Human Resource Development Program, Department of Biotechnology, Graduate School of Engineering, Osaka University, and full scholarship supported by MEXT (Monbukagakusho Japanese Government), for allowing me to carry out my master courses and doctoral courses. I am very grateful to all the faculty staffs, all guests for giving valuable knowledge during my study.

I would like to express my gratitude to my seniors and alumni of Muranaka Laboratory, including Pisanee Srisawat, Paskorn Muangphrom, Much Zaenal Fanani, Soo Yeon Chung, Suzuki Hayato, and Jekson Robertlee, for their kind co-operation, encouragement, moral support, and confidence in me, which enabled me to persevere through the difficulties of my studies, as well as for their continuous advice on how to improve the quality of my research and publication.

To my great friend Pramesti Istiandari (Isti), my cohort partner in the same batch and laboratory, who has fought alongside me in this journey since the beginning and has always had my back and believed in me both technically and psychologically. I am grateful for her kindness and positive attitude at work as well as her optimism and perseverance, which constantly inspire me to do better. Besides my advisor, I would like to thank Callan, Yen, and others Muranaka lab member, as for the kindness to support my research, valuable comments, and encouragement throughout my thesis work.

Some special words of gratitude go to my Thai student friends who have always been a major source of support when things would get a bit discouraging. Thanks, guys, for always being there for me.

Finally, a very special word of thanks goes for my family: Mr. Prapas Romsuk, Mrs. Somchit Romsuk, Mrs. Sribamphen Phiphatkommalakhet, and Mr. Pratompong Romsuk, for their love, support, and encouragement throughout my life.

Supporting information

Section 1. phylogenetic analysis of CYP716 Family

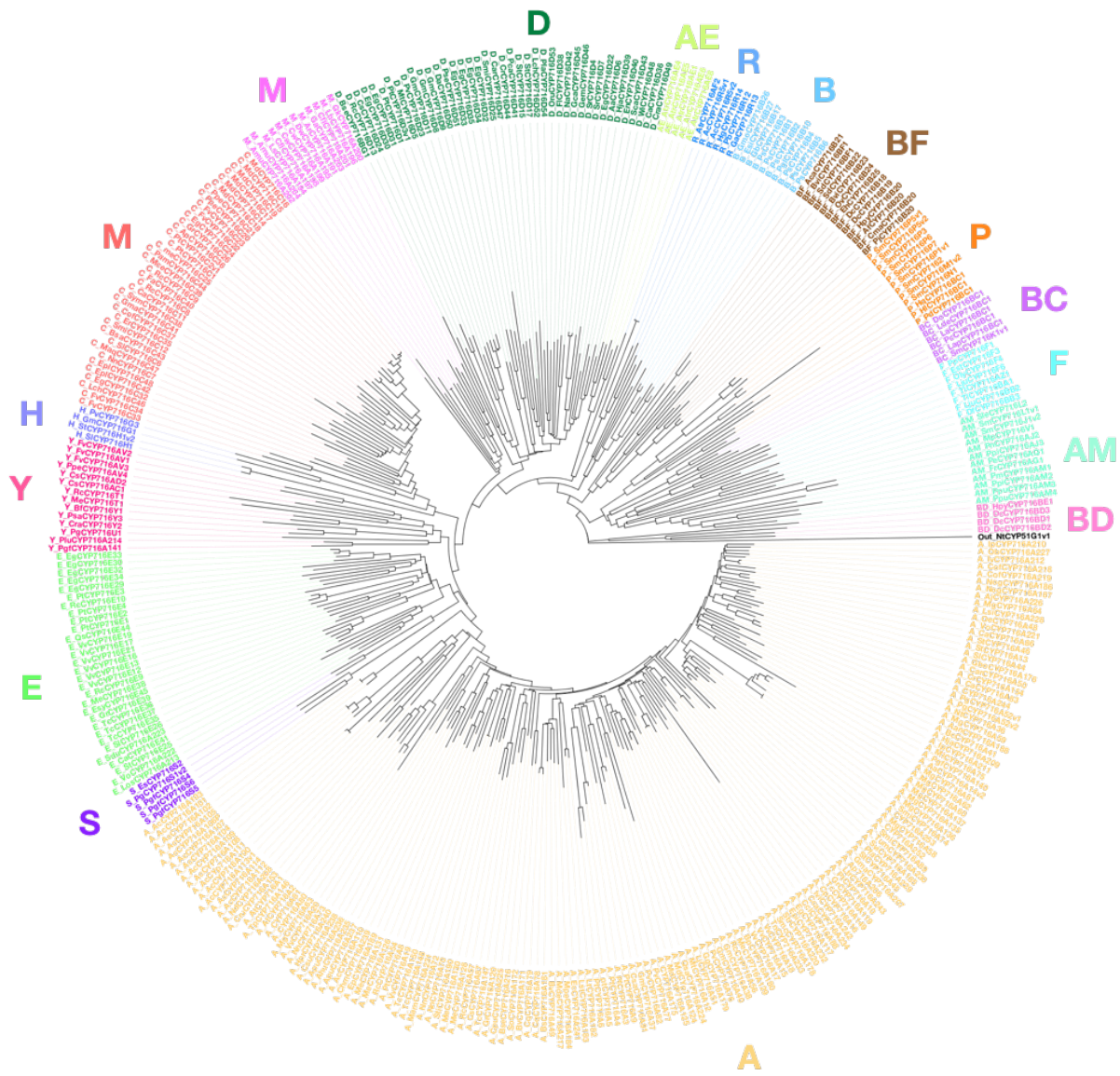


Figure S-1 Neighbor-joining phylogenetic analysis of CYP716 family.

Section 2. The protein sequence of the CYP716A subfamily used in this study

> A_AtCYP716A1_*Arabidopsis_thaliana* (accession no. NM_123002.2)

MYMAIMIILFLSSILLSLLLLLRKHLSHFSYPNLPPGNTGLPLIGESFSFLSAGRQGHPEKF
ITDRVRRFSSSSSCVFKTHLFGSPTAVVTGASGNKFLFTNENKLVVSWWPDSVNKIFPSSMQ
TSSKEEARKLRMLLSQFMKPEALRRYVGVMDEIAQRHFETEWANQDQVIVFPLTKKFTFSIA
CRSFLSMEDPARVRQLEEQFNTVAVGIFSIPIDLPGTRFNRAIKASRLLRKEVSAIVRQRKE
ELKAGKALEEHDILSHMLMNIGETKDEDLADKIIGLLIGGHDTASIVCTFVVNYLAEFPHVY
QRVLQEQKEILKEKKEKEGLRWEDIEKMRYSWNVACEVMRIVPPLSGTFREAIDHFSFKGFY
IPKGWKLYWSATATHMNPDYFPEPERFEPNRFEGSGPKPYTYVPFGGGPRMCPGKEYARLEI
LIFMHNLVNRFKWEKVFPNENKIVVDPLPIPDKGLPIRIFPQSX

> A_AtCYP716A2_*Arabidopsis_thaliana* (accession no. LC106013.1)

MYLTIIFLFISSIIFPLLFFLGKHLSNFRYPNLPPGKIGFPLIGETLSFLSAGRQGHPEKFV
TDRVRHFSSGIFKTHLFGSPFAVVTGASGNKFLFTNENKLVISWWPDSVNKIFPSSTQTSSK
EEAIKTRMLLMPSMKPEALRRYVGVMDEIAQKHFETEWANQDQLIVFPLTKKFTFSIACRLF
LSMDDLERVRKLEEPFTTVMTGVFSIPIDLPGTRFNRAIKASRLLSKEVSTIIRQRKEELKA
GKVSVEQDILSHMLMNIGETKDEDLADKIIALLIGGHDTTSIVCTFVVNYLAEFPHIYQRVL
EEQKEILNNKDVNEKLTWEDIEKMRYSWNVACEVMRIVPPLAGTFREAIDHFSFKGFYIPKG
WKLYWSATATHKNPEYFPEPEKFEPSRFEGSGPKPYTYVPFGGGSRICPGREYARLEILIFM
HNLVKRFKWEKVFPKENKLVADPAPIPAKGLPIRIFPQS

> A_MtCYP716A12_*Medicago_truncatula* (accession no. DQ335781.1)

MEPNFYLSLLLLFVSFISLSLFFIFYKQKSPLNLPPGKMGYPIIGESLEFLSTGWKGHPEKF
IFDRMRKYSSELFKTSIVGESTVVCCGAASNKFLFSNENKLVTAWWPDSVNKIFPTTSLDSN
LKEESIKMRKLLPQFFKPEALQRYVGVMDVIAQRHFVTHWDNKNEITVYPLAKRYTFLLACR
LFMSVEDENHVAKFSDPFQLIAAGIISLPIDLPGTPFNKAIKASNFIRKELIKIIKQRRIDL
AEGTASPTQDILSHMLLTSDENGKSMNELNIADKILGLLIGGHDTASVACTFLVKYLGELPH
IYDKVYQEQMEIAKSKPAGELLNWDDLKKMKYSWNVACEVMRLSPPLQGGFREAITDFMFNG
FSIPKGWKLYWSANSTHKNAECFPMPEKFDPTRFEGNGPAPYTFVPFGGGPRMCPGKEYARL
EILVFMHNLVKRFKWEKVIPDEKIIVDPFPIPAKDLPIRLYPHKA

> A_VvCYP716A15_*Vitis_vinifera* (accession no. AB619802.1)

MEVFFLSLLLIFVLSVSIGLHLLFYKHRSHFTGPNLPPGKIGWPMVGESLEFLSTGWKGHPE
KFIFDRISKYSSEVFKTSLLGEPAAVFAGAAGNKFLFSNENKLVHAWWPSSVDKVFPSSTQT
SSKEEAKKMRKLLPQFFKPEALQRYIGIMDHIAQRHFADSWDNRDEVIVFPLAKRFTFWLAC
RLFMSIEDPAHVAKFEKPFHVLASGLITVPIDLPGTPFHRAIKASNFIRKELRAIIKQRKID
LAEGKASQNQDILSHMLLATDEDGCHMNEMEIADKILGLLIGGHDTASAAITFLIKYMAELP
HIYEKVYEEQMEIANSKAPGELLNWDDVQNMRYSWNVACEVMRLAPPLQGAFREAITDFVFN
GFSIPKGWKLYWSANSTHKSPECFPQPENFDPTRFEGNGPAPYTFVPFGGGPRMCPGKEYAR
LEILVFMHNVVKRFKWDKLLPDEKIIVDPMPMPAKGLPVRLHPHKP

>A_OeCYP716A48_*Olea_europaea* (accession no. G0244188.1)
MEFFYVSLLCLFVFLISLSLHFLFYKNKSSFSGQIPPGKTGWPVIGESLEFLSNGWKGHPEK
FIFDRIAKYSSYVFRTHLFGEPAAVFCGANGNKFLFSNENKLVQAWWPASVDKVFPSSNQTS
SKEEAVKMRKMLPTFFKPEALQRYVGIMDHIAQRHFSDGWDNKNEVVVFPLAKRYTFWLACR
LFVSVEDPAHVAKFADPFNELASGLISIPIDLPGTPFHRAIKSSNFIRKELVSIIKQRKIDL
AEGKASPTQDILSHMLLTSDESGKFMHELDIADKILGLLVGGHDTASSACTFVVKYLAELPE
IYEGVYQEQMEIAKSKAPGELLNWDDIQKMKYSWNVACEVLRLAPPLQGAFREAITDFMFNG
FSIPKGWKLYWSANSTHRNSEFFPEPLKFDPSRFEGSGPAPYTFVPFGGGPRMCPGKEYARL
EILVFMHHLVKRFKWEKLIPDEKIVVDPMPIPAKGLPIRLYPLNA

>A_BvCYP716A49_*Beta_vulgaris* (accession no. FG44256.1)
MELFFLCGLILFLSLSLASLYLLYNHNSTKGYRVPPGTMGWPVVGESLEFLSTGWKGYPEKF
IFDRLSKYAPNQIFKTSILGEKVAVICGAAGNKFLYSNENKLVQAWWPSSVDKIFPSSTQTS
SKEESKKMRKLLPNFLKPEALQRYIPIMDTIAIRHMESGWDGKDKVEVFPLAKRYTFWLACR
LFLSIEDPDHVAKFAEPFNDIAAGIISLPVNLPGTPFNRGIKSSNVVRKELRAIIKQRKLDL
ADGKASTTQDILSHMLLTADEDGRFMTEMDIADKILGLLIGGHDTASAACTFVVKYLAELPH
VYEAVCKEQMEIAKSKAEGELLNWEDIQKMKYSWNVACEVMRLAPPLQGGFREAISDFMYGG
FQVPKGWKLYWSANSTHRNPECFPEPEKFDPSRFEGKGPAPYTYVPFGGGPRMCPGKEYARL
EILVFMHNVVKRFKWEKVLPNEKVIVNPMPIPENGLPVRLFPHPQIVAA

Section 3. The specific primers for site-directed mutagenesis

Table S-1 Specific primer for site-directed mutagenesis

Primer name	Forward (5'->3')	Reverse (5'->3')
CYP716A12_D122Q	TCTCTTCAATCTAACTTGAAGGAAGAA	GTTAGATTGAAGAGAAGTAGTAGGGAA
CYP716A12_I212P	GGAATCCCATCTCTACCAATTGATTTG	TAGAGATGGGATTCCGGCTGCAATTAA
CYP716A12_D292A	GGACATGCTACTGCTAGCGTCGCATGC	AGCAGTAGCATGTCCTCCGATCAAAAG
CYP716A12_Q358P	CCACTCCCAGGAGGTTTCAGGGAAGCC	ACCTCCTGGGAGTGGAGGGGAAAGTCT
CYP716A1_S356Q	CCTCTTCAAGGCACTTTTCGTGAGGCC	AGTGCCTTGAAGAGGAGGAACAATTCT
CYP716A2_M206A	ACGGTTGCTACGGGTGTCTTCTCAATC	ACCCGTAGCAACCGTAGTGAATGGCTC
CYP716A2_F210I	GGTGTCATTTCAATCCCAATAGATTTA	GATTGAAATGACACCCGTCATAACCGT

Section 4. Experimental Procedures

Section 4-1 GC–MS analysis for the functional analysis of CYP716A2 in yeasts producing β-amyrin

Gas chromatography-mass spectrometry (GC-MS) analysis was performed using an HP-5MS capillary column. The injection component and the MSD transfer line were set to 250 °C. The oven temperature was programmed as follows: 150°C for 1 min, followed by an increase to 260°C at a rate of 30°C min–1, followed by an increase to 300°C at a rate of 1°C min–[1].

Section 4-2 Standard curve for calculating the quantifications of triterpene concentrations in engineered yeast

The relationship between the quantities of the two compounds was determined using standard curves. They are used to calculate the value of an unknown quantity (triterpene concentration) of metabolites extracted from engineered yeast relative to one that can be measured more conveniently (triterpene standard), whose concentration is known. To quantify triterpene concentrations extracted from engineered yeast, an authentic standard curve of triterpene compounds was constructed, and the relative concentration was determined by comparing the peak area of the authentic standard to that of an internal standard. The relative peak area of the internal standard was plotted against a known but variable triterpene concentration using an authentic standard curve.

Using the standard curve equation, the actual triterpene concentration in yeast was calculated.

Section 5. The authentic standard curve of the triterpenoid

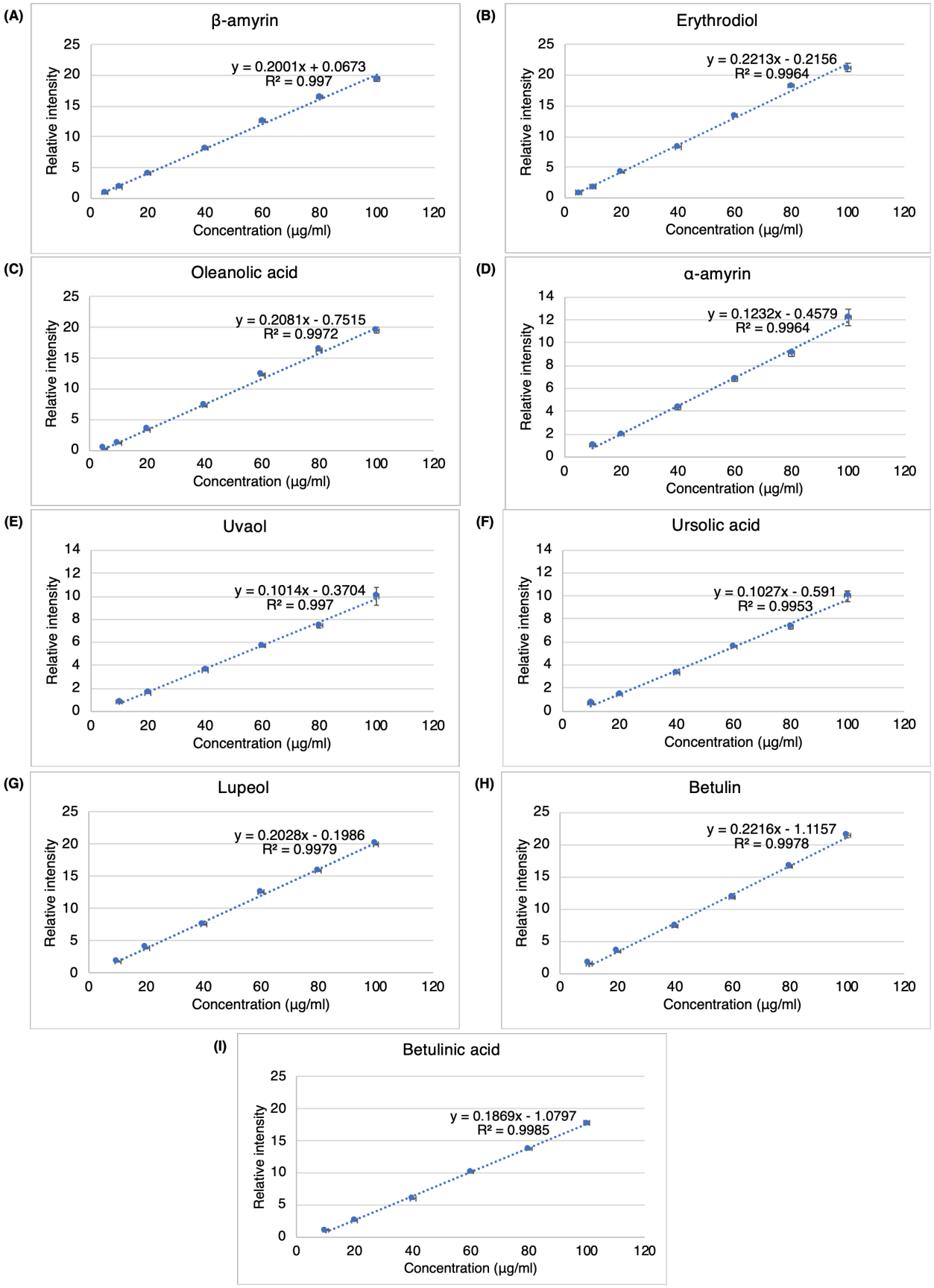


Figure S-2 The authentic standard curve of the triterpenoid used in Chapter 2 and Chapter 3.

(A) β-amyrin, (B) erythrodiol, (C) oleanolic acid, (D) α-amyrin, (E) uvaol, (F) ursolic acid, (G) lupeol, (H) betulin, and (I) betulinic acid. Data are representative of at least three biological replicates (n=3). Error bars correspond to the mean and standard deviation.

Section 6. List of variants

Table S-2 CYP716As wild-type, Variants, and mutations in *in vivo* functional analysis

Sample	Origin	Mutations
CYP716A12_WT	CYP716A12	Wild type
CYP716A12_D122Q		D122Q
CYP716A12_I212P		I212P
CYP716A12_D292A		D292A
CYP716A12_Q358P		Q358P
CYP716A12_D122Q_Q358P		D122Q_Q358P
CYP716A12_I212P_Q358P		I212P_Q358P
CYP716A15_WT	CYP716A15	Wild type
CYP716A48_WT	CYP716A48	Wild type
CYP716A49_WT	CYP716A49	Wild type
CYP716A1_WT	CYP716A1	Wild type
CYP716A1_S356Q		S356Q
CYP716A2_WT	CYP716A2	Wild type
CYP716A2_M206A		M206A
CYP716A2_F210I		F210I

Section 7. Comparative homology modeling of CYP716As

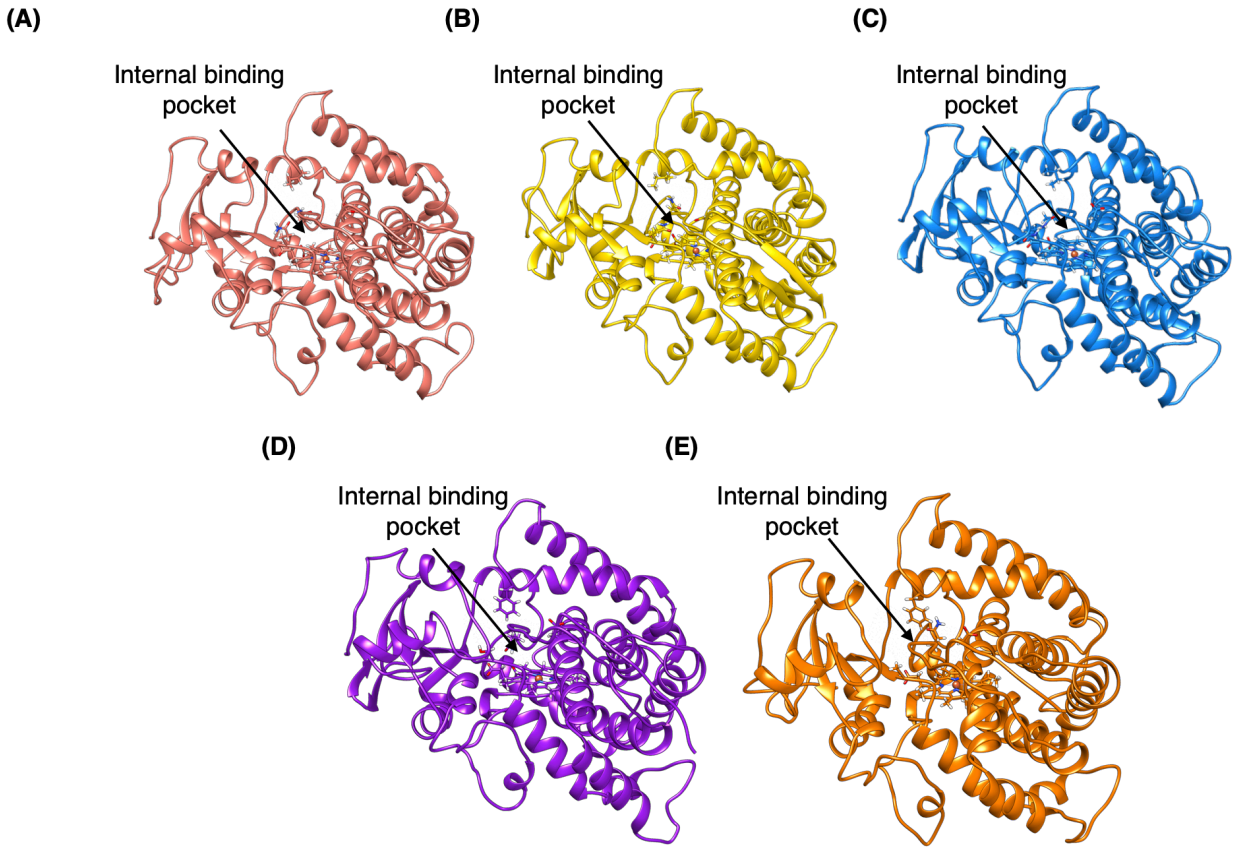


Figure S-3 Comparative homology modeling of CYP716As with identified internal binding pocket.

The crystal structure of retinoic acid bound cyanobacterial CYP120A1(2VE3) chain A was selected as a template by homology sequence searching via Protein BLAST. (A) Homology modeling of CYP716A15 was 37.3% identical against the template. (B) Homology modeling of CYP716A48 was 37.8% identical against the template. (C) Homology modeling of CYP716A49 was 36.9% identical against the template. (D) Homology modeling of CYP716A1 was 36.2% identical against the template. (E) Homology modeling of CYP716A2 was 36.6% identical against the template. All homology modeling was passed to pass quality criteria verified by VERIFY3D, with at least 80% of the amino acids having scored >= 0.2 in the 3D/1D profile.

Section 8. Representative interactions of selected CYP716As against triterpene skeleton

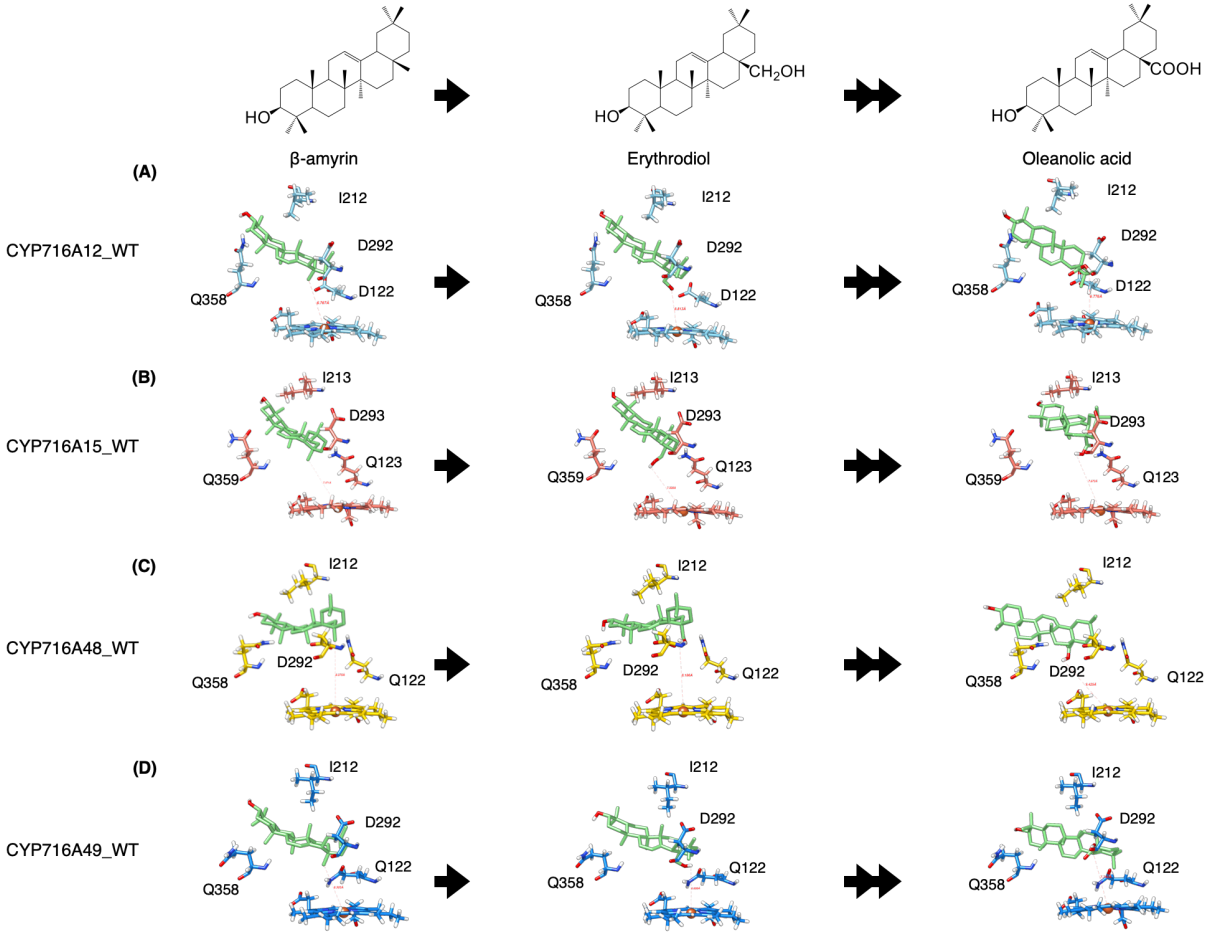


Figure S-4 Representative interactions of selected CYP716As against β-amyrin backbone and its derivatives (erythrodiol and oleanolic acids).

The structure homology modeling showed the β-amyrin backbone and its derivatives in wild type of (A) CYP716A12, (B) CYP716A15, (C) CYP716A48, and (D) CYP716A49.

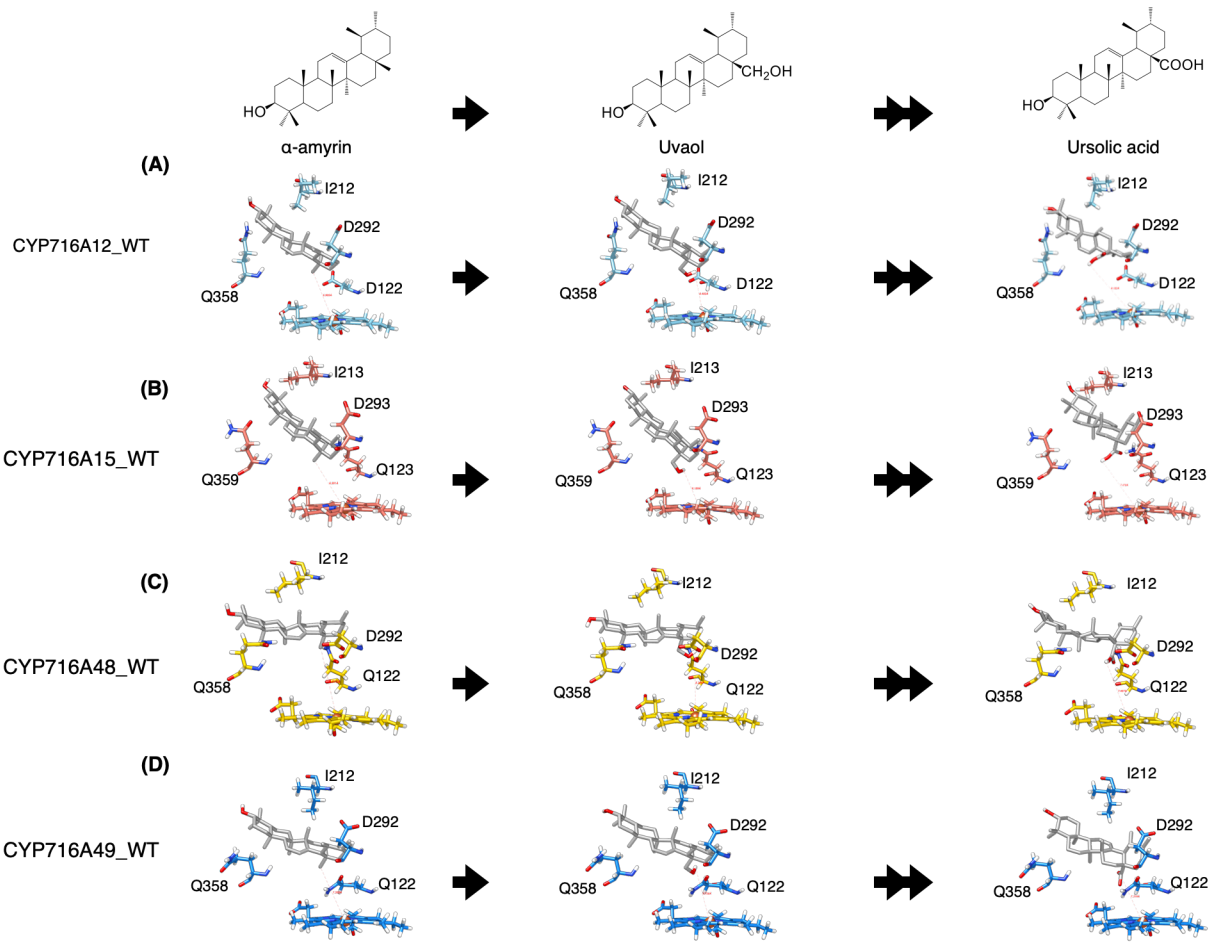


Figure S-5 Representative interactions of CYP716As against α-amyrin backbone and its derivatives (uvaol, ursolic acid).

The structure homology modeling showed α-amyrin backbone and its derivatives in wild type of (A) CYP716A12, (B) CYP716A15, (C) CYP716A48, and (D) CYP716A49**.**

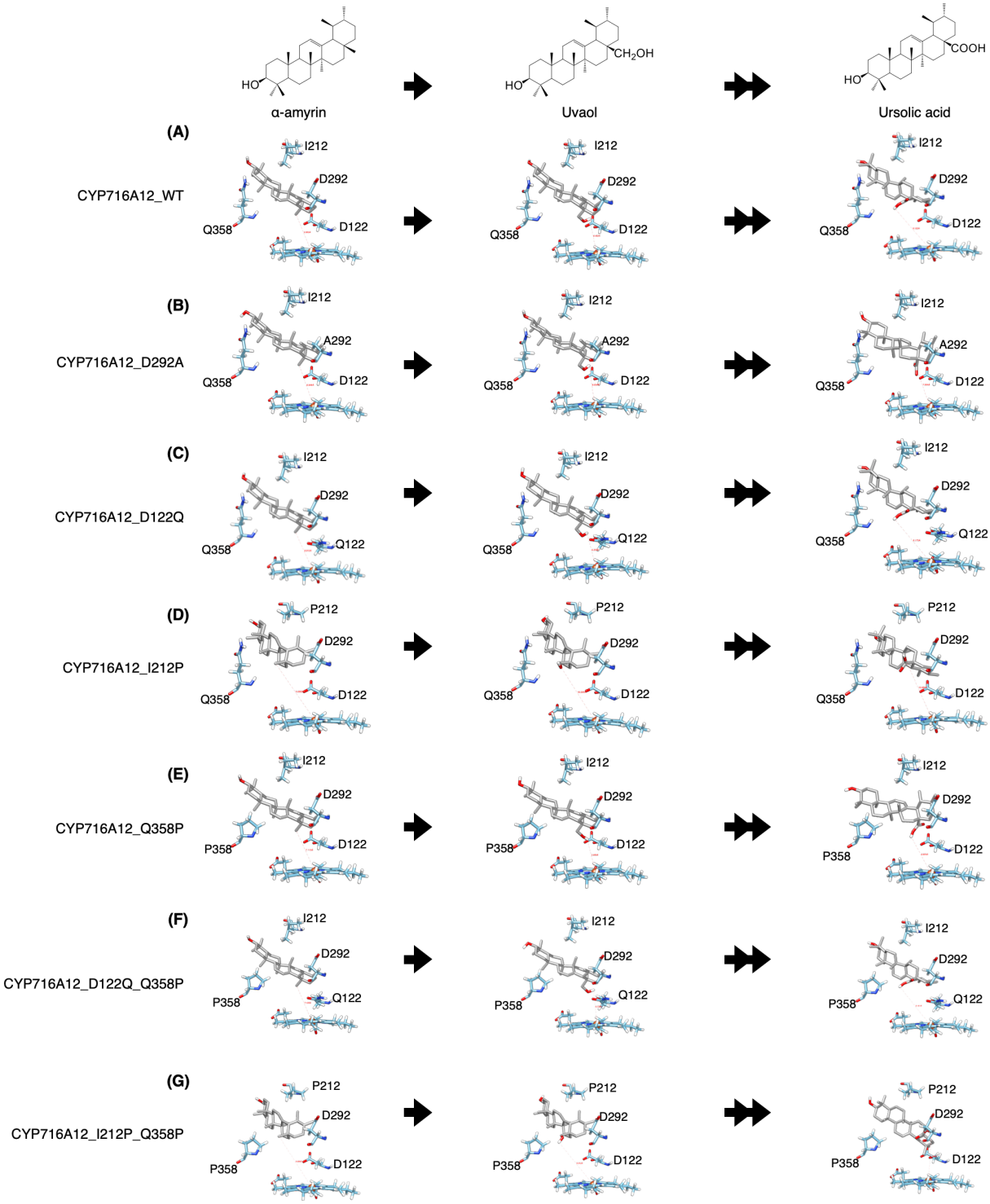


Figure S-6 Representative interactions of CYP716A12 and its variants against α-amyrin backbone and its derivatives (uvaol, ursolic acid).

The structure homology modeling showed α-amyrin backbone and its derivatives in wild-type of (A) CYP716A12, and CYP716A12 variant (B) CYP716A12_D292A, (C) CYP716A12_D122Q, (D) CYP716A12_I212P (E) CYP716A12_Q358P, (F) CYP716A12_D122Q_Q358P, and (G) CYP716A12_I212P_Q358P

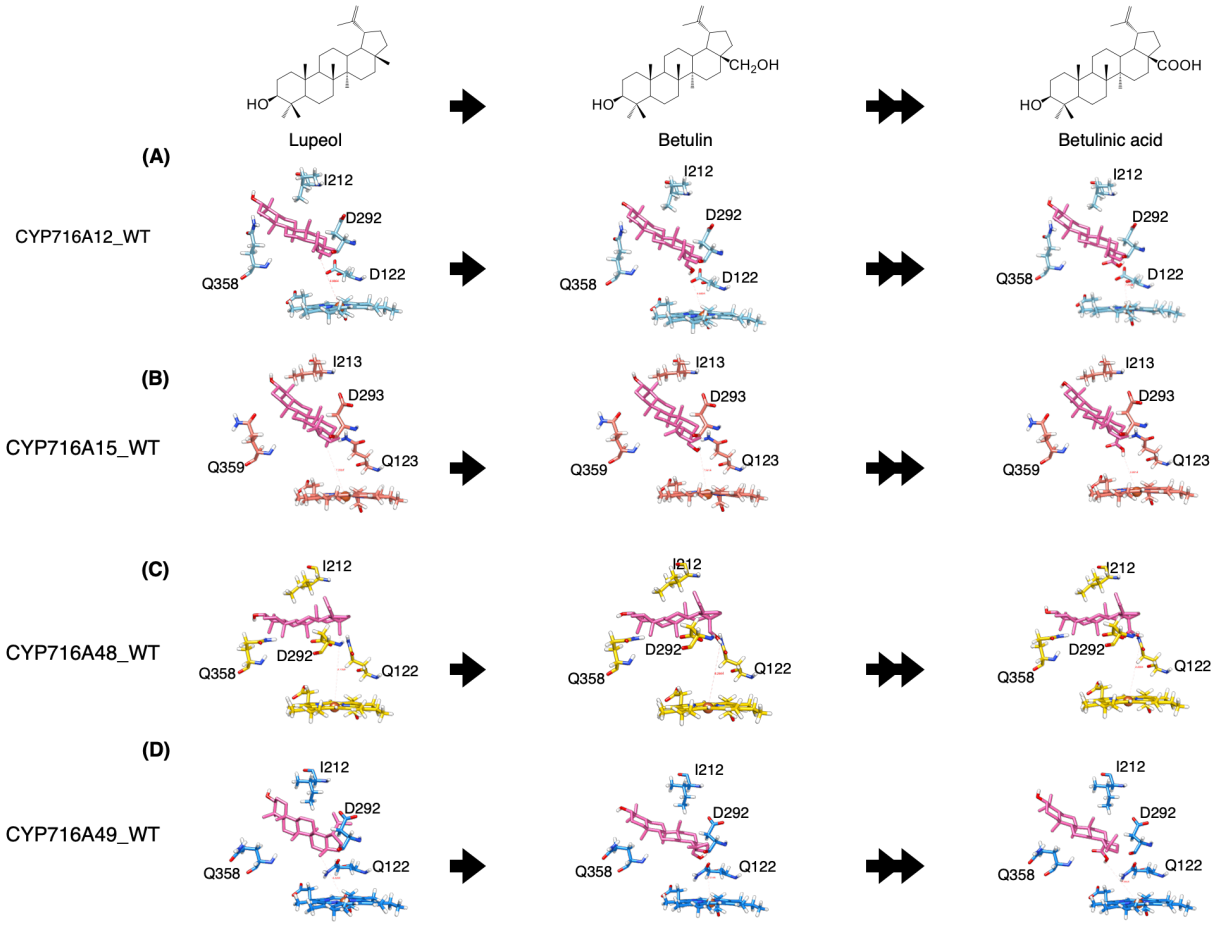


Figure S-7 Representative interactions of selected CYP716As against lupeol backbone and its derivatives (betulin, betulinic acid).

The structure homology modeling showed the lupeol backbone and its derivatives in wild type of (A) CYP716A12, (B) CYP716A15, (C) CYP716A48, and (D) CYP716A49.

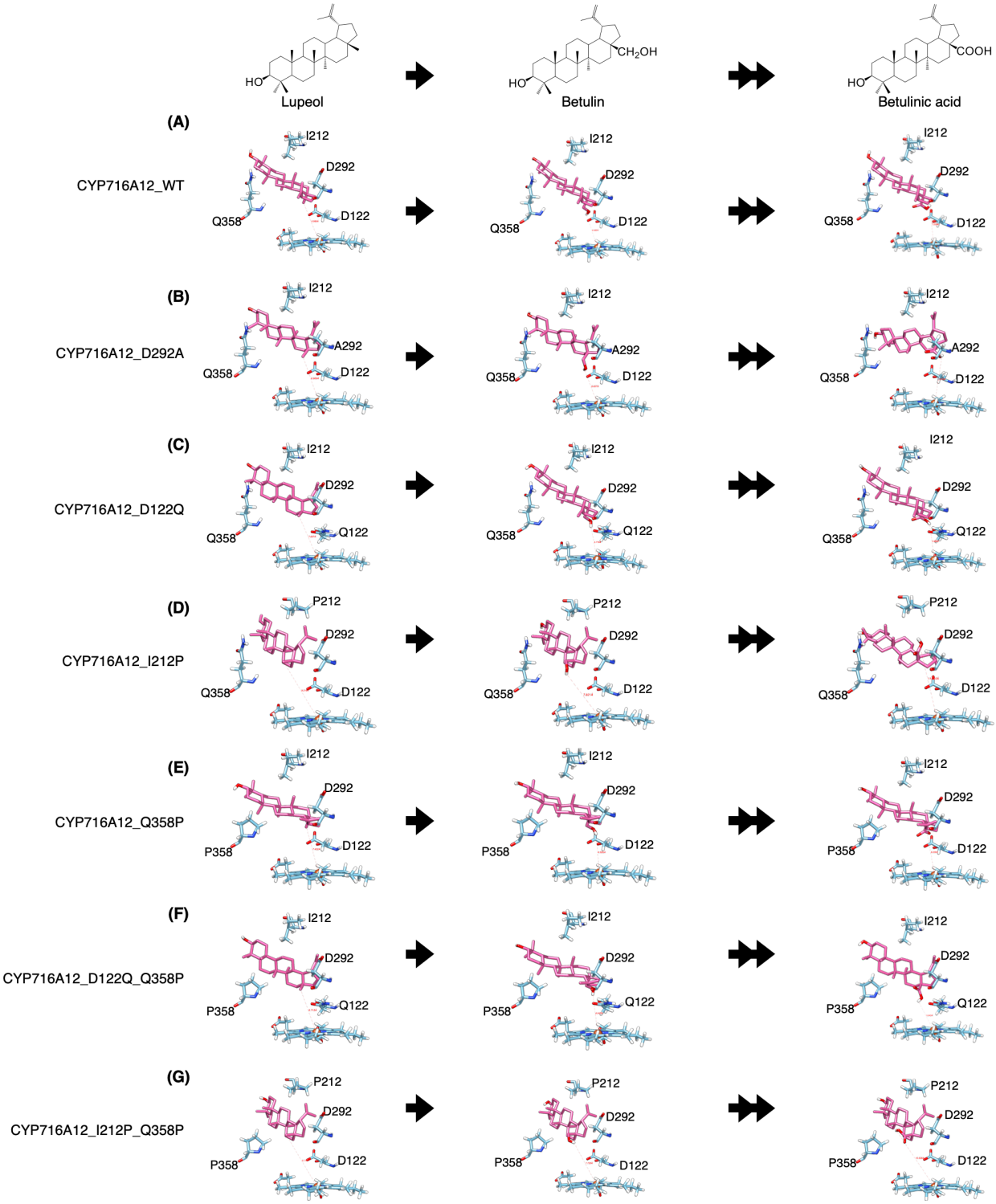


Figure S-8 Representative interactions of CYP716A12 and its variants against lupeol backbone and its derivatives (betulin, betulinic acid).

The structure homology modeling showed lupeol backbone and its derivatives in wild type of (A) CYP716A12, and CYP716A12 variant (B) CYP716A12_D292A, (C) CYP716A12_D122Q, (D) CYP716A12_I212P, (E) CYP716A12_Q358P, (F) CYP716A12_D122Q_Q358P, and (G) CYP716A12_I212P_Q358P

Section 9. *In vivo* functional analysis in yeast results

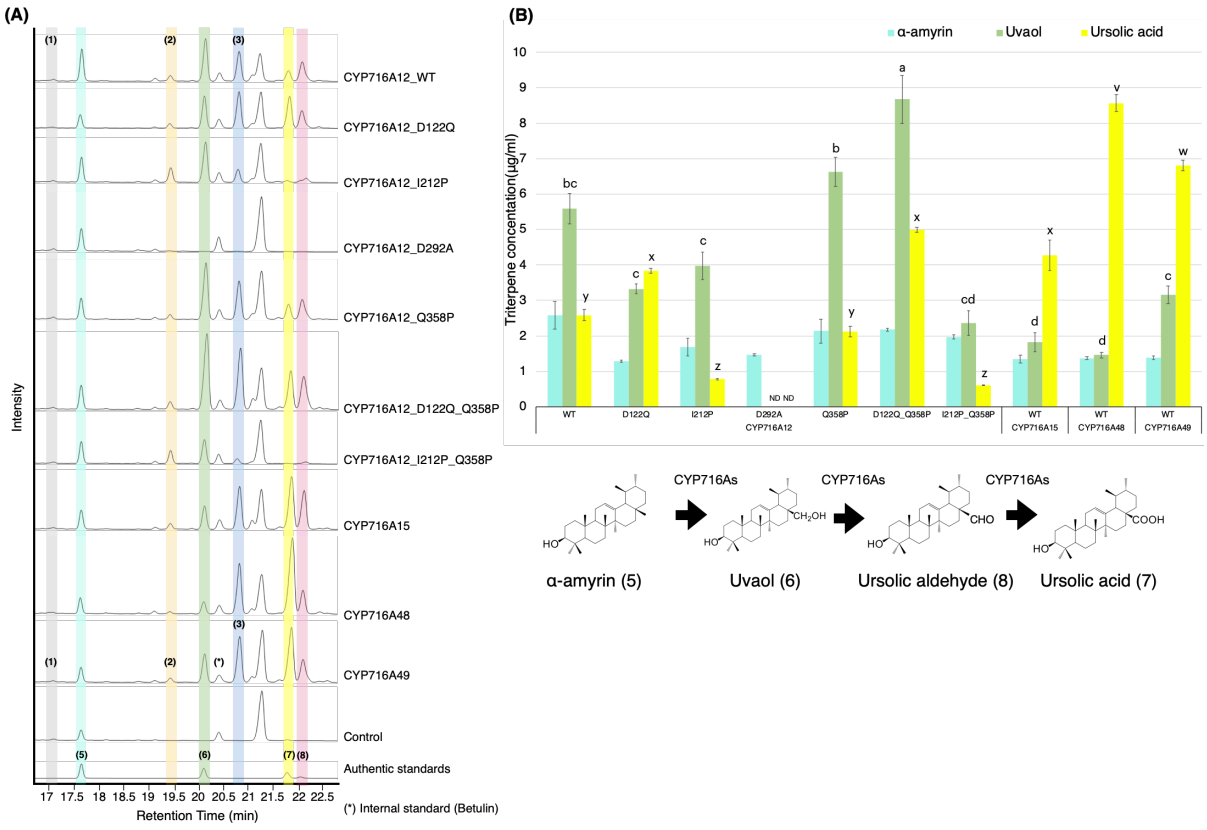


Figure S-9 *In vivo* activities of selected CYP716As and CYP716A12, and its variants against α-amyrin skeletons.

(A) TICs of extracts from yeast harboring one *aAS* expression vector for producing α-amyrin as a major substrate and three CYPs expression vectors (pELC-CYP716As, pYES-DEST52-CYP716As, and pESC-HIS-CYP716As) were used. (B) Quantification of triterpene concentration in yeast harboring CYP716As against α-amyrin skeletons. Quantitation and error bars correspond to the mean and standard deviation, respectively. Data are representative of at least three biological replicates (n=3). Letters indicate statistical differences between the different oxidized triterpenoids levels (a-d: uvaol levels, v-z: ursolic acid levels) for each sample (one-way ANOVA; Tukey's post-hoc test, $p \leq 0.05$). ND, not detected.

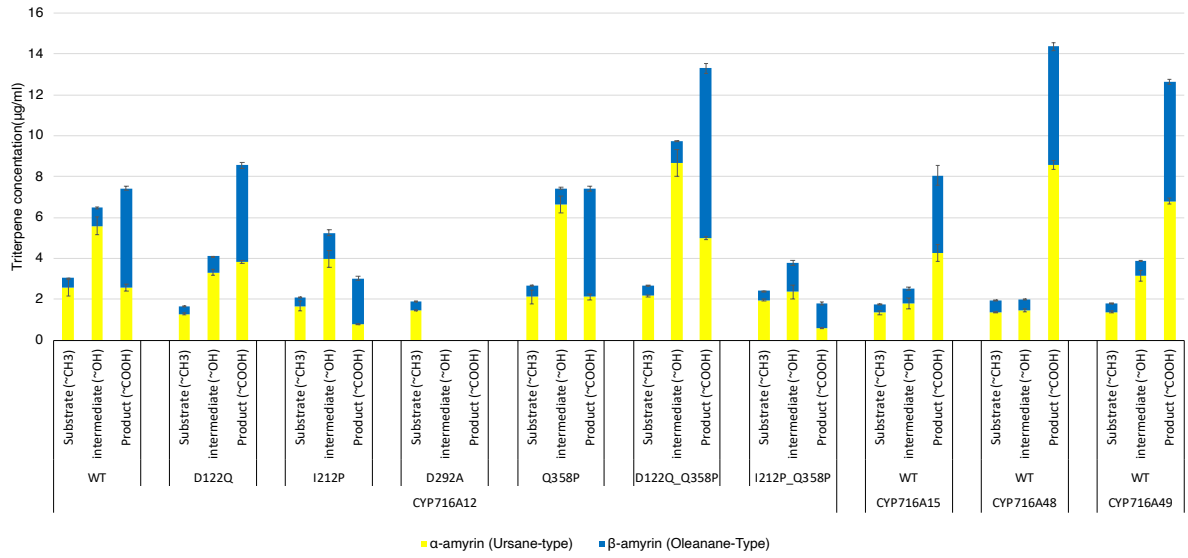


Figure S-10 Comparison of triterpene concentration from extracts of yeast producing α-amyrin as a major substrate and β-amyrin as a minor substrate.

Yeast harboring one *aAS* expression vector for making α-amyrin as a major substrate and three CYPs expression vectors (pELC-CYP716A, pYES-DEST52-CYP716As, and pESC-HIS-CYP716As) were used. Quantitation and error bars correspond to the mean and standard deviation, respectively. Data are representative of at least three biological replicates (n=3). ND, not detected.

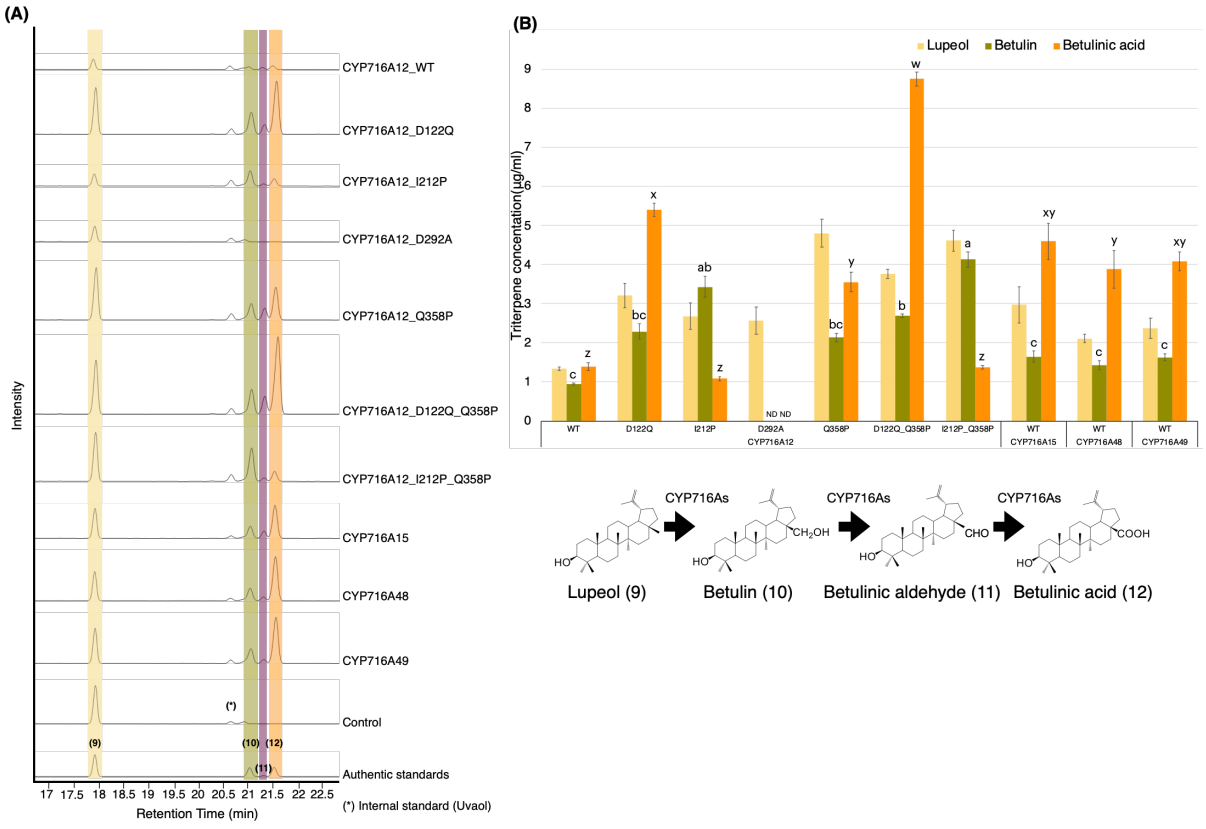


Figure S-11 *In vivo* activities of selected CYP716As and CYP716A12, and its variants against lupeol skeletons.

(A) Extracted ion chromatogram (EIC) by selecting *m*/*z* of extracts from yeast harboring one *LUS* expression vector for producing lupeol, and three CYPs harboring expression vectors (pELC-CYP716As, pYES-DEST52-CYP716A, and pESC-HIS-CYP716A) were used. (B) Quantification of triterpene concentration in yeast harboring CYP716As against lupeol skeletons. Quantitation and error bars correspond to the mean and standard deviation, respectively. Data are representative of at least three biological replicates (n=3). Letters indicate statistical differences between the different oxidized triterpenoids levels (a-c: betulin levels, w-z: betulinic levels) for each sample (one-way ANOVA; Tukey's post-hoc test, $p \leq 0.05$). ND, not detected.

Section 10. The nucleotide sequence of the codon-optimized CDS of LjOSC1 used in chapter 4

> *Lotus japonicus*_ β-Amyrin synthase I (LjOSC1)

ATGTGGAAACTTAAAGTTGCTGATGGCGGCAAGGACCCGTACATCTTCTCTACTAACAACTT
CGTTGGCCGGCAGACCTGGGAGTATGATCCTGATGCTGGTACTCCTGAAGAGAGAGCACAGG
TTGAAGAAGCTAGGCAGGACTTCTACAACAACCGGTACAAGGTTAAGCCTTGCGGCGATCTT
CTTTGGAGGTTCCAGGTTCTGCGTGAGAACAACTTCAAGCAGACCATTCCGAGCGTGAAGAT
CGAGGATGGTGAAGAGATTACCTACGAGAAGGCTACCACCACCTTGAAGAGAGCTGCTCATC
ATCTTGCTGCTCTGCAGACTTCTGATGGTCATTGGCCTGCTCAAATTGCTGGCCCTTTGTTT
TTCCAGCCTCCTCTGGTTTTCTGCATGTACATTACCGGCCACCTGAACTCTGTGTTCCCTGA
AGAGTACCGGAAAGAGATCCTGCGGTACATCTACGTGCACCAGAATGAAGATGGTGGTTGGG
GTCTGCATATTGAGGGACACTCTACTATGTTCTGCACCGCTCTGAACTACATCTGCATGAGG
ATGCTTGGTGAGGGTCCTGATGGTGGTCAGGATAATGCTTGTGCTAGAGCCCGGAAGTGGAT
TCTTGATCATGGTGGTGTGACCCACATTCCGTCTTGGGGTAAGACCTGGCTTAGCATTCTTG
GCATCTTCGACTGGAAGGGCTCTAACCCTATGCCTCCTGAGTTCTGGATTCTGCCTTCTTTC
CTTCCTATGCACCCTGCCAAGATGTGGTGCTATTGCAGGCTTGTGTACATGCCGATGAGCTA
CCTGTACGGCAAGAGATTCGTGGGTCCTATCACTCCTCTTATCCTGCAGCTTCGGGAAGAGT
TGTTCACTCAGCCTTATGAGAAGGTGAACTGGAAGAAAGCAAGGCACCAGTGCGCTAAAGAG
GACATCTATTACCCTCATCCGCTGATCCAGGATCTGATGTGGGATTCTCTGTACCTGTTCAC
CGAGCCTTTGCTTACTAGGTGGCCTTTCAACAAGCTGGTGAGAGAGAAGGCTCTCGAGGTTA
CCATGAAGCACATTCACTACGAGGACGAGAACAGCCGGTACATCACTATTGGCTGCGTTGAG
AAGGTTCTGTGCATGCTTGCTTGCTGGGTCGAAGATCCTAACGGGGACGCTTTTAAGAAGCA
CCTTGCTAGGATCCCTGACTACCTTTGGGTTAGCGAAGATGGCATGTGCATGCAGAGCTTCG
GTTCTCAAGAATGGGATGCTGGTTTCGCTGTGCAGGCTCTTTTGGCTACTAACCTTGTGGAT
GAGCTGGGTCCTACTCTTGCTAAGGGTCACGACTTCATCAAGAAGTCCCAGGTTAGGGATAA
CCCGAGCGGCGATTTCAAGAATATGCACCGGCACATCTCCAAAGGCAGCTGGACTTTTTCTG
ATCAGGACCACGGTTGGCAGGTTTCAGATTGCACTGCTGAGGGTCTTAAGTGCTGCCTTCTG
CTTTCTATGCTCCCTCCTGATATCGTGGGCGAGAAGATGGAACCTGAGTGCCTTTTCGATAG
CGTGAACCTTCTGCTGAGCCTGCAGTCTAAGAAAGGTGGACTTGCTGCTTGGGAACCTGCTG
GTGCTCAAGAGTGGCTTGAGCTTCTTAATCCGACCGAGTTCTTCGCTGACATCGTGGTTGAG
CATGAGTACGTTGAGTGCACCGGTTCTGCTATTGGTGCTCTTGTGCTGTTCAAGAAGCTGTA

CCCTGGCCACCGGAAGAAAGAGATTGAGAACTTCATCAGCGAGGCCGTGAGATTCCTTGAGG
ATACTCAAACCGCTGACGGTAGCTGGTATGGTAATTGGGGAGTGTGCTTCACCTACGGCTCT
TGGTTTGCTCTTGGTGGTTTGGCTGCTGCTGGTAAGACTTACGCTAACTGCGCTGCTATCCG
GAAGGCTGTGAAGTTCCTTTTGACTACCCAAAGAGGTGACGGCGGTTGGGGAGAGTCTTATC
TGTCAAGCCCGAAGAAGATCTACGTTCCGTTCGAGGGCAACCGTTCTAATGTGGTTCATACC
GCTTGGGCTCTGATGGGTCTTATTCATTCAGGTCAGGCTGAGAGGGATCCAACACCTCTTCA
TAGAGCTGCTAAGCTGCTGATCAACAGCCAGCTTGAAGAAGGCGATTGGCCTCAGCAAGAGA
TCACTGGTGTGTTCATGAAGAACTGCATGCTGCACTACCCGATGTACAGGGACATCTACCCT
ATGTGGGCTCTCGCTGAGTATAGAAGAAGGGTGCCATTGCCTTCTACCGCCGTTTAG

Section 11. Standard curve for calculating the quantifications of triterpene concentrations in *Nicotiana benthamiana* leaves extracts

The relationship between the quantities of the two compounds was determined using standard curves. They are used to calculate the value of an unknown quantity (triterpene concentration) of metabolite extracted from *N. benthamiana* and *N. tabacum* leaf extracts relative to the one that is more directly measured (triterpene standard), whose concentration is known. To quantify the concentrations of triterpenes extracted from leaf extracts, an authentic standard curve of triterpene compounds was constructed, and the relative concentrations were determined by comparing the peak area of an authentic standard to that of an internal standard. The relative peak area of the internal standard was obtained by plotting against a known but variable triterpene concentration using an authentic standard curve.

Using the standard curve equation, the actual triterpene concentration in leaf extracts was calculated.

Section12. Schematic diagram of the T-DNA region of the plasmids

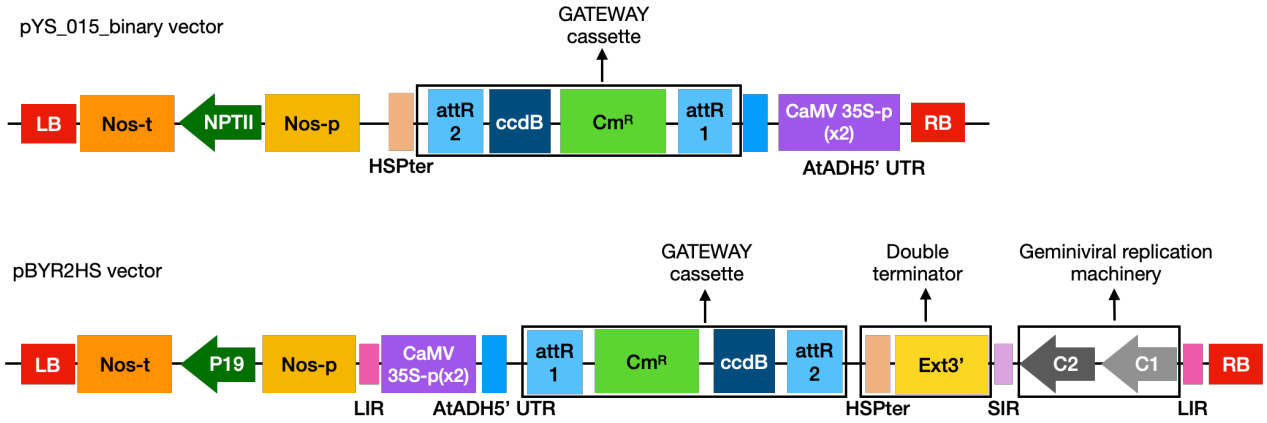


Figure S-12 Schematic diagram of the T-DNA region of the conventional binary vector (pYS_015) or gateway-compatible version of pBYR2HS.

CaMV 35S-p: Cauliflower mosaic virus (CaMV) 35S promoter with double-enhanced element, AtADH5': 5'-untranslated region (UTR) of *Arabidopsis thaliana* alcohol dehydrogenase gene, HSPter: terminator of heat shock protein gene, Ext3′: tobacco extensin gene 3′ element, Nos-t: NOS terminator, LIR: long intergenic region of bean yellow dwarf virus (BeYDV) genome, SIR: short intergenic region of BeYDV genome, C1/C2: BeYDV ORFs C1 and C2 encoding for replication initiation protein (Rep) and RepA, LB and RB: the left and right borders of the T-DNA region, respectively; Nos-p and Nos-t: NOS promoter and terminator, respectively; p19: a gene-silencing suppressor gene from tomato bushy stunt virus; and NPTII: neomycin phosphotransferase II enzyme of *Escherichia coli* Tn5 transposon; attR1 and attR2: attachment sites that allow recombination cloning of the gene of interest from an entry clone; Cm^R: chloramphenicol resistance gene; ccdB gene: a lethal gene that targets DNA gyrase that allows negative selection of plasmid.

Section 13. The authentic standard curve of the triterpenoid used in chapter 4

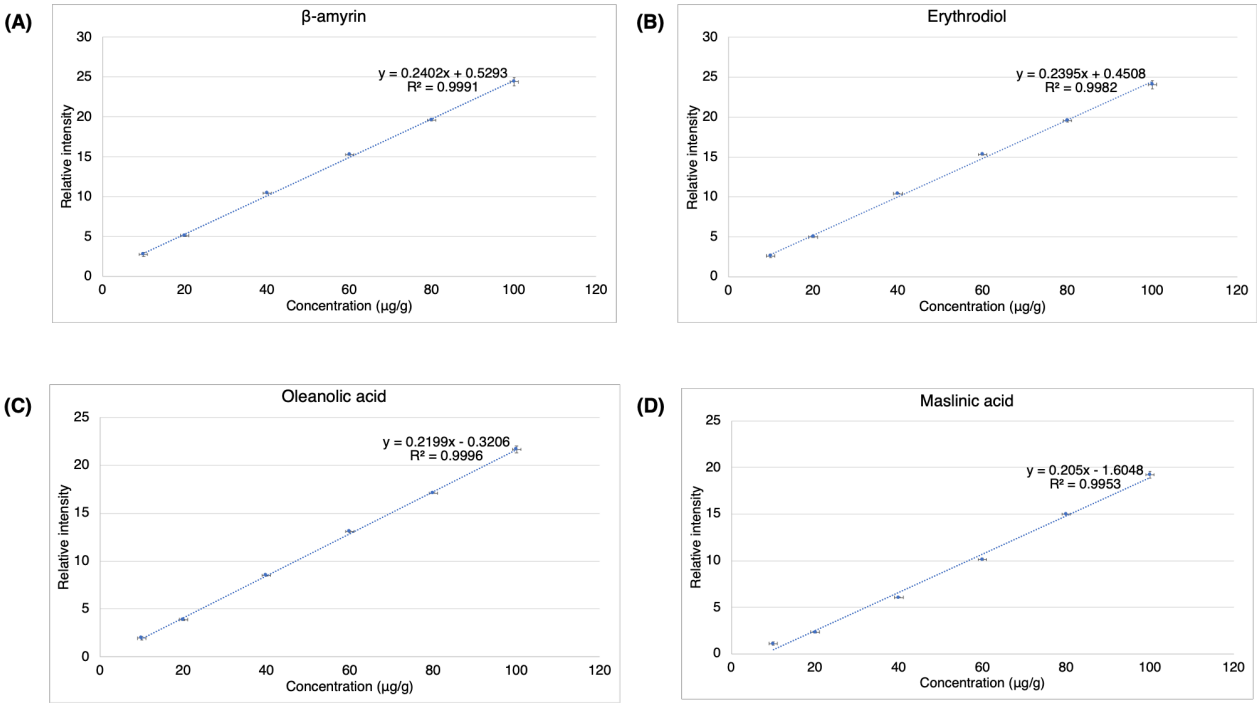


Figure S-13 Standard curves of authentic used in chapter 4.

(A) β-amyrin (1), (B) erythrodiol (2), (C) oleanolic acid (3) and (D) Maslinic acid (16). Mean values and standard deviation (error bars) for three replicates are shown.

Section 14. GC-MS analysis and mass fragmentation patterns of triterpenoids detected in leaves extract via transiently expression

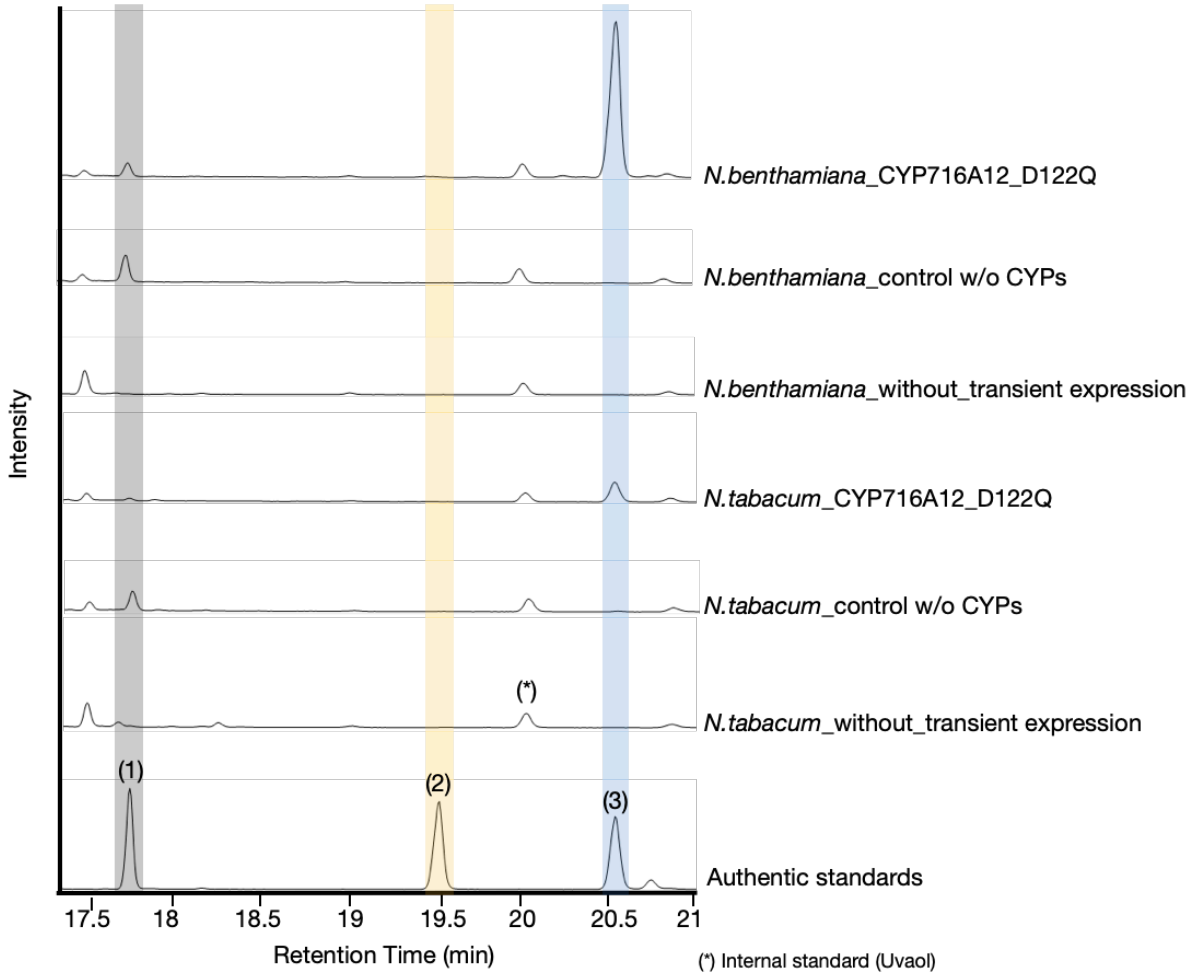


Figure S-14 Total ion chromatograms of the extracts from different Nicotiana species.

Total ion chromatograms (TICs) of extracts from *N. benthamiana* and *N. tabacum* leaves transiently expressing a combination of LjOSC1, LjCPR2, and with or without CYP716A12_D122Q. As a background control, leaves without transient expression were used. The quantification results are shown in **Figure 4-2C**. β-Amyrin (**1**), erythrodiol (**2**), and oleanolic acid (**3**) were used as authentic standards. Uvaol (*) was used as an internal standard.

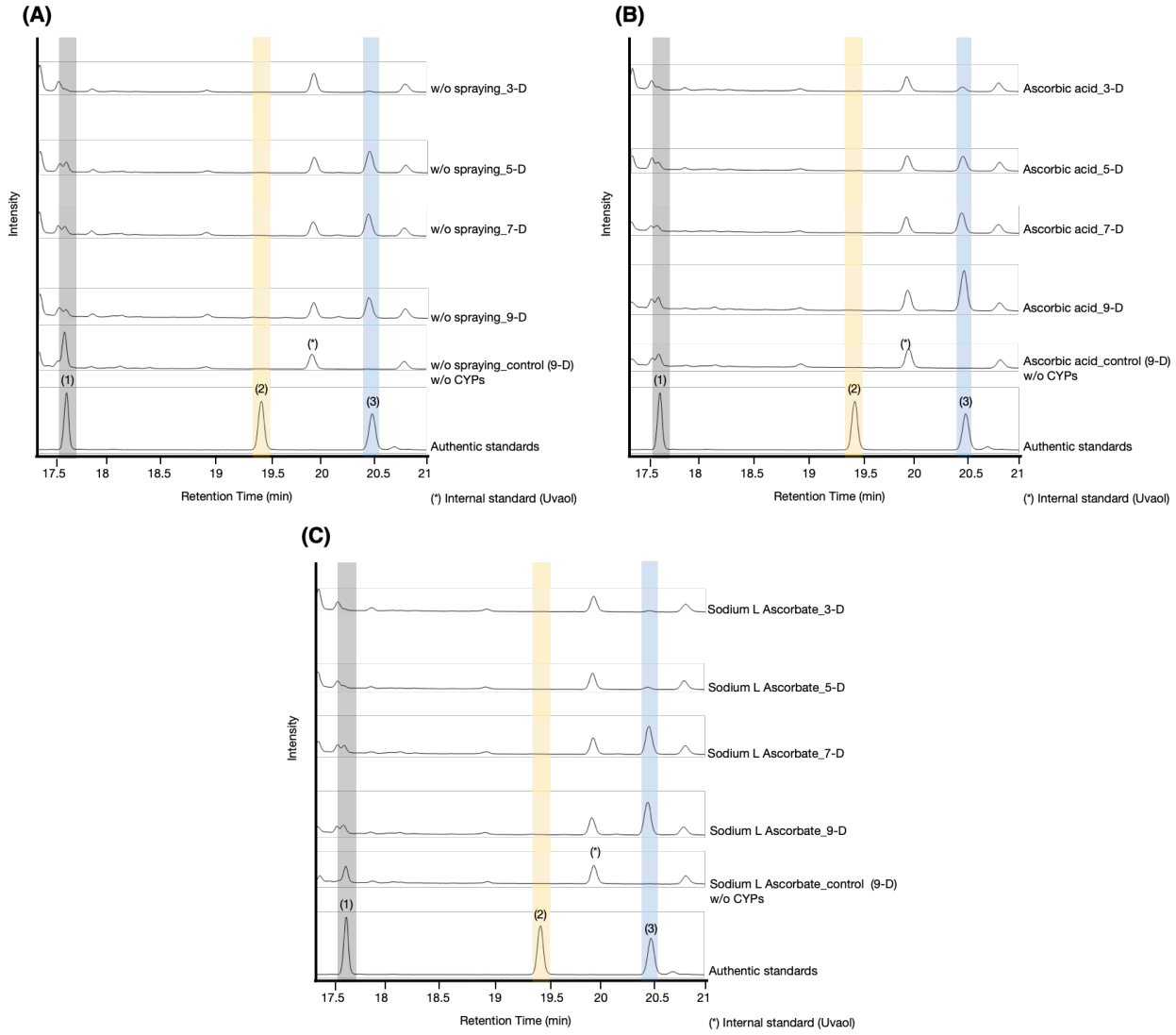


Figure S-15 TIC of the extracts from *N. benthamiana* leaves with or without application of ascorbic acid after agroinfiltration.

TICs of extracts from *N. benthamiana* leaves transiently expressing a combination of LjOSC, LjCPR2, and CYP716A12_D122Q or pBYR2HS empty vector (w/o CYP control). (A) Leaves without sprayed with ascorbic acid as negative control. (B) Leaves with foliar application of 200 mM ascorbic acid or (C) ascorbic acid sodium salt (sodium L-ascorbate). Metabolites were extracted from leaves at 3, 5, 7, or 9 days after agroinfiltration. The quantification results are shown in **Figure 4-4**. β-Amyrin (**1**), erythrodiol (**2**), and oleanolic acid (**3**) were used as authentic standards. Uvaol (*) was used as an internal standard.

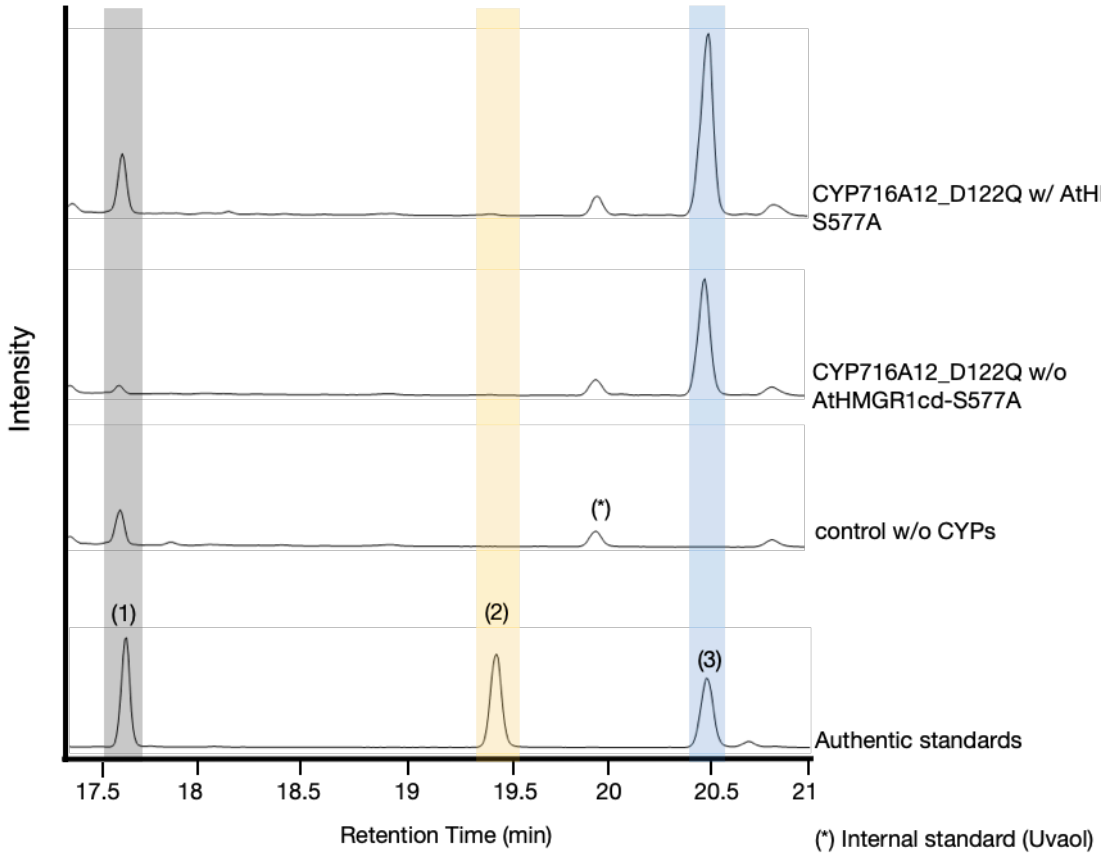


Figure S-16 TICs of the extracts from *N. benthamiana* leaves transiently expressing a combination of LjOSC1, LjCPR2, CYP716A12_D122Q, and with or without *AtHMGR1*cd-S577A.

The quantification results are shown in **Figure 4-5C**. β-Amyrin (**1**), erythrodiol (**2**), and oleanolic acid (**3**) were used as authentic standards. Uvaol (*) was used as an internal standard.

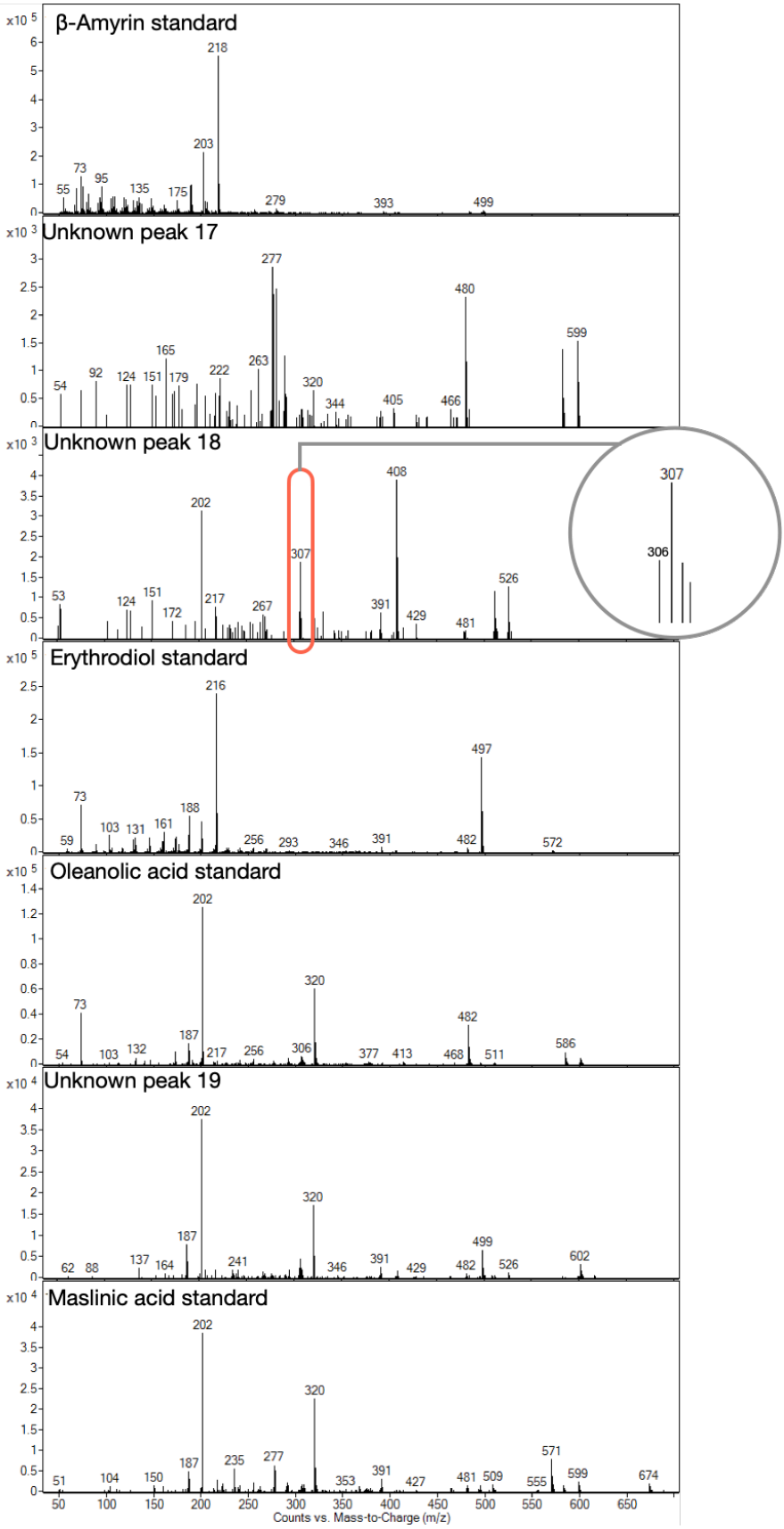


Figure S-17 Mass fragmentation pattern of three minor peaks from the GC profile shown in Figure 4-6A.

The mass spectra of peaks 17 and 19 showed fragment ion at *m/z* 320 (a carboxyl group was added to the C-28 of the β-amyrin backbone), suggesting that both products were derived from oleanolic acid. The mass spectra of peak 18 showed fragment ion at *m/z* 306, suggesting that one hydroxyl group was introduced into the β-amyrin backbone, probably at C-2.