Engineering of critical enzymes and pathways for improved triterpenoid biosynthesis in yeast

Hao Guo[†], Huiyan Wang[†], Yi-xin Huo^{†,*}

[†]School of Life Science, Beijing Institute of Technology, No. 5 South Zhongguancun Street, 100081 Beijing, China

*Corresponding author: mail: huoyixin@bit.edu.cn; Tel: 86-10-68918158

Abstract

Triterpenoids represent a diverse group of phytochemicals, widely distributed in the plant kingdom with many biological activities. Recently, the heterologous production of triterpenoids in *Saccharomyces cerevisiae* has been successfully implemented by introducing various triterpenoids biosynthetic pathways. By engineering related enzymes as well as yeast metabolism, the yield of various triterpenoids is significantly improved from milligram-scale per liter to gram-scale level per liter. This achievement demonstrates that engineering of critical enzymes is considered as a potential strategy to overcome the main hurdles of translation of these potent natural products into industry. Here, we review strategies for designing enzymes to improve the yield of triterpenoids in *S. cerevisiae*, which is mainly separated into three aspects: 1. elevating the supply of the precursor—2,3-oxidosqualene, 2. optimizing triterpenoid-involved reactions, 3. lowering the competition of the native sterol pathway. And then we provide challenges and prospects on further enhancing the triterpenoid production in *S. cerevisiae*.

Key words: synthetic biology; triterpenoid; oxidosqualene cyclase; cytochrome P450; UDP-glycosyltransferase; yeast.

Introduction

Triterpenoids are a subset of terpenoids that are widely distributed in plants, microbes, and marine organisms with various structures and applications reported to date [1], e.g. artemisinin, taxol, lycopene, stevioside, perilla alcohol, ginsenosides, and azadirachtin A (Figure 1). Terpenoids are classified into 7 subgroups: hemi- (C₅), mono- (C₁₀), sesqui- (C₁₅), di- (C₂₀), tri- (C₃₀), tetra (C₄₀) and poly- (C_{>40}) terpenoids according to the number of isoprene units [2]. Triterpenoids consist of six isoprene units which are mainly synthesized from—2,3-oxidosqualene, except for a minority of triterpenoids generated from squalene and lanosterol. Most triterpenoids belong to secondary plant metabolites, which have been applied in the food, cosmetics, and pharmaceutical industry [3]. Recently, several triterpenoids have gained more attention, as they are the main active ingredients of many traditional medicines which contain a variety of pharmacological compounds, such as ginsenosides, betulinic acid, and oleanolic acid [4]. For example, ginsenosides are the main bioactive component in Asian and American ginseng that benefits the central nervous, cardiovascular, endocrine, and immune systems [5, 6]. Besides, betulinic acid is a very promising anti-tumor agent which can suppress tumor growth and induce apoptosis in human melanoma cells and other cancer cells, with relatively low cytotoxicity against healthy cells [7].

Typically, triterpenoids are produced by extraction from plant materials. However, the supply of triterpenoids is considerably limited by slow-growing plants and low extraction efficiency. Chemical synthesis is an infeasible way to substitute the plant extraction due to the complex skeletal characteristics and chiral centers of these molecules. For instance, azadirachtin A is a complex tetranor-triterpenoid from the Neem (*Azadirachta indica*), which is regarded as active ingredients of many pesticides [8]. The total chemical synthesis of azadirachtin A was completed after 22 years, which was required for 71 steps of reactions. Nonetheless, the final yield was only 0.00015% [9]. Hence, the microbial synthesis is an attractive alternative for the production of triterpenoids. Compared with current extraction from plant sources, this strategy possesses several advantages, such as short-cycle manufacture, land-saving, and controllable culture conditions.

Up to now, the production of triterpenoids has been successfully implemented in various microbes, such as *E. coli*, *S. cerevisiae*, *Y. lipolytica*, *P. pastoris*. *S. cerevisiae* is considered as an efficient homologous recombination system that can provide not only sufficient 2,3-oxidosqualene through its native sterol pathway but also the endoplasmic reticulum (ER) membrane for expression of appropriate enzymes. Hence, *S. cerevisiae* is predominately utilized in the heterologous production of triterpenoids. With the application of various metabolic engineering approaches on *S. cerevisiae*, the yield of triterpenoids, especially for dammarene-type, has been increased to the 10 gram-scale level per liter. However, the production of other types of triterpenoids is relatively lower, such as lupane-type, oleanane-type, and ursane-type triterpenoids. To sum up, there is still a large gap to realize the industrial production in place of extraction from plant sources. Here, we review the current progress in the

elucidation of strategies to increase the production of triterpenoids as well as the modification of triterpenoid-related enzymes, and then we discuss challenges and perspectives on further enhancing the triterpenoids biosynthesis in *S. cerevisiae*

Triterpenoid biosynthetic pathway

2,3-Oxidosqualene biosynthesis

The triterpenoid biosynthetic pathway can be mainly separated into two segments. 1. biosynthesis of the linear C₃₀ molecule-2,3-oxidosqualene (Figure 2). 2. the formation of cyclic triterpenes based on the 2,3-oxidosqualene and further modification by other enzymes (Figure 3). Hence, 2,3-oxidosqualene is considered as a branch-point precursor in the synthesis of the linear and cyclic compounds during the triterpenoid biosynthesis. In general, there are two routes involved in the biosynthesis of 2,3-oxidosqualene—the mevalonate (MVA) pathway and methylerythritol-phosphate (MEP) pathway. However, in *S. cerevisiae*, the MVA pathway starting from the acetyl-CoA is the unique way to synthesize 2,3-oxidosqualene. Isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are end-products of the MVA pathway that are two fundamental building blocks for 2,3-oxidosqualene biosynthesis. Two molecules of IPP and one molecule of DMAPP are condensed to generate farnesyl diphosphate (FPP) by farnesyl diphosphate synthase, and two FPP molecules are further merged by squalene synthase to produce squalene that is oxidized to 2,3-oxidosqualene by squalene epoxidase. In *S. cerevisiae*, lanosterol is the only product of the cyclization of 2,3-oxidosqualene, which is an essential precursor for the sterol biosynthesis.

85

86 2,3-Oxidosqualene cyclases

87 In plants, the cyclization of 2,3-oxidosqualene is the first committed step for triterpenoid
 88 biosynthesis which is catalyzed by a variety of oxidosqualene cyclases (OSCs). To date, numerous
 89 OSCs have been identified in plants and heterologously expressed in engineered *E. coli* and
 90 2,3-oxidosqualene producing yeast strains (Table 1). β -Amyrin, and lupeol synthases are the most
 91 abundant enzymes in the plant kingdom. In contrast, α -amyrin, dammarenediol II, thalianol, and
 92 taraxerol synthases are uncommon, especially as dammarenediol II synthase is only discovered in
 93 *Panax ginseng* and *Panax notoginseng*. These enzymes share the common conserved regions and
 94 the similar sequence. By site-specific mutagenesis of key amino acids in conserved regions, the
 95 function of these enzymes can be even switched to some extent. For example, L256W mutation
 96 of lupeol synthase from *Olea europaea* can result in the formation of β -amyrin rather than lupeol,
 97 whereas W259L mutation of β -amyrin synthase from *Panax ginseng* produces β -amyrin instead
 98 of lupeol [10](Figure 4).

99

100 The heterologous production of triterpenes has been achieved by the expression of several OSCs
 101 in *S. cerevisiae*. However, the yield of triterpenes varies depending on different OSCs of plant
 102 origin. For instance, the β -amyrin synthase from *G. glabra* yields more β -amyrin compared with
 103 other species, e.g. *A. annua*, *P. sativum*, *G. glabra*, and *P. ginseng* [11]. Also, α -amyrin synthase
 104 from *Malus × Domestica* has been characterized to exhibit high levels of specific activity which
 105 led to 5-fold increase on the yield of α -amyrin [12]. Hence, numerous isozymes provide multiple

choices for metabolic engineering which is beneficial for triterpenoid production in *S. cerevisiae*.

Also, some OSCs are identified as multifunctional enzymes with mixed end-products [13]. In particular, baccharis oxide synthase from *A. thaliana* can make baruol as the primary product with up to 22 additional by-products through the cyclization of 2,3-oxidosqualene [14]. However, the mechanism of generation of mixed-products is unclear, it is probably due to the deprotonation of numerous sites during the cyclization process [15]. The metabolic diversity may be caused by the evolution, which allow plants to produce various triterpenoids against disease or insect attack. Nevertheless, it seriously influences the concentration of targeted products and causes a complicated purification process.

Table 1 Verified 2,3-oxidosqualene cyclases for triterpene synthesis

OSC	Source	Products	Ring conformation	Application in host	Ref
AaBAS	<i>Artemisia annua</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[16]
AsbAS1	<i>Avena sativa</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[17]
AtLUP4	<i>Arabidopsis thaliana</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[18]
BgbAS	<i>Bruguiera gymnorrhiza</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[19]
BPY	<i>Betula platyphylla</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[20]
CqbAS1	<i>Chenopodium quinoa</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[21]
EtAS	<i>Euphorbia tirucalli</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i> /E.coli	[22, 23]
GgbAS	<i>Glycyrrhiza glabra</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[11]
LjAMY1	<i>Lotus japonicus</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[24]
MtAMY1	<i>Medicago truncatula</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[25]
NsbAS1	<i>Nigella sativa</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[26]
PNY1/PNY2	<i>Panax ginseng</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[27]
PtBS	<i>Polygala tenuifolia</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[28]
PSY	<i>Pisum sativum</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[29]
TTS1	<i>Solanum lycopersicum</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[30]
BS	<i>Vaccaria hispanica</i>	β -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[31]
AtLUP1	<i>Arabidopsis thaliana</i>	lupeol	6/6/6/6/5	<i>S. cerevisiae</i>	[32]
BgLUS	<i>Bruguiera gymnorrhiza</i>	lupeol	6/6/6/6/5	<i>S. cerevisiae</i>	[19]
BPW	<i>Betula platyphylla</i>	lupeol	6/6/6/6/5	<i>S. cerevisiae</i>	[20]
OEW	<i>Olea europaea</i>	lupeol	6/6/6/6/5	<i>S. cerevisiae</i>	[32]
OSC3	<i>Lotus japonicus</i>	lupeol	6/6/6/6/5	<i>S. cerevisiae</i>	[24]
GgLUS1	<i>Glycyrrhiza glabra</i>	lupeol	6/6/6/6/5	<i>S. cerevisiae</i>	[33]

TRW	<i>Taraxacum officinale</i>	lupeol	6/6/6/6/5	<i>S. cerevisiae</i>	[34]
PgDDS	<i>Panax ginseng</i>	dammarene diol II	6/6/6/5	<i>S. cerevisiae</i>	[35]
PqDS	<i>Panax quinquefolius</i>	dammarene diol II	6/6/6/5	<i>S. cerevisiae</i> / <i>E. coli</i>	[36]
CPQ	<i>Cucurbita pepo</i>	cucurbitadienol	6/6/6/5	<i>S. cerevisiae</i>	[37]
THAS1	<i>Arabidopsis thaliana</i>	thalianol	6/6/5	<i>S. cerevisiae</i>	[38]
AiOSC1	<i>Azadirachta indica</i>	tirucalla-7,24-dien-3 β -ol	6/6/6/5	<i>S. cerevisiae</i>	[39]
OsIAS	<i>Oryza sativa</i>	isoarborinol	6/6/6/6/5	<i>S. cerevisiae</i>	[40]
CrAS	<i>Catharanthus roseus</i>	β -amyrin/ α -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[41]
IaAS1	<i>Ilex asprella</i>	β -amyrin/ α -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[30]
MdOSC1	<i>Malus domestica</i>	β -amyrin/ α -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[42]
OEA	<i>Olea europaea</i>	β -amyrin/ α -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[43]
LjAMY2	<i>Lotus japonicus</i>	β -amyrin/ α -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[44]
AtLUP2	<i>Arabidopsis thaliana</i>	β -amyrin/ α -amyrin	6/6/6/6/6	<i>S. cerevisiae</i>	[45]
KcMS	<i>Kandelia candel</i>	lupeol/ β -amyrin/ α -amyrin	6/6/6/6/6 and 6/6/6/6/5	<i>S. cerevisiae</i>	[46]
BARS1	<i>Arabidopsis thaliana</i>	baccharis oxide with 22 mixed products		<i>S. cerevisiae</i>	[14]

116

117 P450 monooxygenases

118 P450 monooxygenases decorate basic triterpene skeletons by the introduction of hydroxyl,
119 ketone, aldehyde, carboxyl, or epoxy groups. To date, a large number of P450s have been
120 characterized as triterpene-oxidizing enzymes in various plant species (Table 2). Based on
121 homology and phylogenetic criteria, the P450 family has been classified into 10 clans: CYP71,
122 CYP72, CYP85, CYP86, CYP51, CYP74, CYP97, CYP710, CYP711, and CYP727 [47]. Many P450s
123 exhibit substrate diversity compared with OSCs. Typically, pentacyclic triterpenoids, α -amyrin,
124 β -amyrin, and lupeol, can be oxidized at the position of C12, C13, C-24, C-28, C-30, to generate
125 mixed corresponding products by various P450s, such as CYP51H10, CYP93E1, CYP72A63,
126 CYP716A12, and CYP716A15. A subgroup of the CYP85 clade, the CYP716, dominates the
127 oxidation of triterpene skeletons in plants [48]. In contrast, a tetracyclic
128 triterpenoid—dammarene diol II can be merely oxidized at the C-12 and C-6 position in sequence,
129 to generate protopanaxadiol and protopanaxatriol respectively catalyzed by CYP716A47 and

CYP716A53v2 [49].

P450 typically catalyze the following reaction: $\text{RH} + \text{O}_2 + 2\text{e}^- + 2\text{H}^+ \rightarrow \text{ROH} + \text{H}_2\text{O}$. This reaction relies on the P450 reductase (CPR) for electron transfer as well as NAD(P)H for the electron donor. In some cases, cytochrome b_5 is required as an alternative electron donor for P450-catalyzed reactions [50]. The formation of mixed end-products is possibly due to P450-CPR pairs, the lack of appropriate cytochrome b_5 , and the insufficiency of NADPH and oxygen supply.

As the numerous P450s and CPRs have been characterized in plants, the varying P450-CPR pairs were examined for improving the yield of triterpenoids. Dai et al. introduced different combinations of CYP450s and CPRs in yeast, such as MtCYP716A12 from *M. truncatula*, CrCYP716AL1 from *C. roseus*, VvCYP716A15 from *V. vinifera*, CPR from *A. thaliana*, *X. dendrorhous*, *L. japonicas*, *P. ginseng*, and *V. vinifera*. Finally, the yeast strain bearing the CYP716A15 and CPR from *V. vinifera* exhibited the highest yield of lupane-type triterpenoids [51].

In addition to screen different combinations of P450-CPR pairs, the protein engineering strategies have been utilized to modify the P450 reaction. For instance, a fusion protein strategy was applied on the protopanaxadiol synthase which was fused to its cofactor-truncated CPR1 from *A. thaliana* (ATR1, 46 amino acids were deleted at the N-terminus). The fusion enzyme (PPDS-ATR1) leads to a 71% increase in protopanaxadiol production compared with the separate expression of PPDS and ATR1 genes [52]. Besides, an additional cytochrome b_5 might benefit or inhibit the reaction depending on different P450s [53]. For example, CYP71AV1 is an amorpha-4,11-diene oxidase which catalyzes three oxidation reactions from amorpha-4,11-diene to artemisinic acid, and expression of a cytochrome b_5 from can benefit this reaction, which can increase the yield of artemisinic acid [54].

Table 2 Plant triterpenoid-related cytochrome P450s in yeast

Clade	P450	Source	Host	Substrates	Products	Ref
CYP51	CYP51H10	<i>Avena strigosa</i>	<i>S. cerevisiae</i>	β -amyirin	12,13-epoxide-16-OH- β - amyirin	[55]
	CYP716S5	<i>Platycodon grandiflorus</i>	<i>S. cerevisiae</i>	β -amyirin	12,13 α -epoxy oleanolic acid	[48]
CYP71	CYP93E1	<i>Glycine max</i>	<i>S. cerevisiae</i>	β -amyirin, sophoradiol	24-OH- β -amyirin, soyasapogenol B	[56]
	CYP93E2	<i>Medicago truncatula</i>	<i>S. cerevisiae</i>	β -amyirin	24-OH- β -amyirin, β -amyirin-24-oic acid	[57]
	CYP93E3	<i>Glycyrrhiza uralensis</i>	<i>S. cerevisiae</i>	β -amyirin	24-OH- β -amyirin	[58]
CYP72	CYP72A63	<i>Medicago truncatula</i>	<i>S. cerevisiae</i>	β -amyirin	30-hydroxy- β -amyirin	[59]
	CYP72A154	<i>Glycyrrhiza uralensis</i>	<i>S. cerevisiae</i>	β -amyirin, 11-oxo- β -amyirin	glycyrrhetinic acid	[59]
	CYP72A68v2	<i>Medicago truncatula</i>	<i>S. cerevisiae</i>	oleanolic acid	gypsogenic acid	[60]
CYP85	CYP716A12	<i>Medicago truncatula</i>	<i>S. cerevisiae</i>	β -amyirin, α -amyirin, lupeol	oleanolic acid, ursolic acid, betulinic acid,	[61]
	CYP716A15 CYP716A17	<i>Vitis vinifera</i>	<i>S. cerevisiae</i>	β -amyirin, α -amyirin, lupeol	oleanolic acid, ursolic acid, betulinic acid,	[57]
	CYP716AL1	<i>Catharanthus roseus</i>	<i>S. cerevisiae</i>	β -amyirin, α -amyirin,	oleanolic acid, ursolic acid, betulinic acid,	[62]
	CYP716Y1	<i>Bupleurum falcatum</i>	<i>S. cerevisiae</i>	β -amyirin, α -amyirin	16a-OH- β -amyirin, 16a-OH- α -amyirin	[63]
	CYP716A19	<i>Glycyrrhiza uralensis</i>	<i>S. cerevisiae</i>	lupeol, β -amyirin	betulinic acid, oleanolic acid	[64]
	CYP716A41	<i>Bupleurum chinense</i>	<i>S. cerevisiae</i>	lupeol	betulinic acid	[65]
	CYP708A2	<i>Arabidopsis thaliana</i>	<i>S. cerevisiae</i>	thalianol	7 β -hydroxythalianol	[66]
	CYP716A155	<i>Rosmarinus officinalis</i>	<i>S. cerevisiae</i>	β -amyirin, lupeol	betulinic acid, ursolic acid	[67]
	CaCYP716C49	<i>Crataegus pinnatifida</i>	<i>S. cerevisiae</i>	oleanolic acid, ursolic acid, betulinic acid	maslinic acid, corosolic, acid and alphetolic acid	[51]
	CYP716A47	<i>Panax ginseng</i>	<i>S. cerevisiae</i>	dammarenediol II	protopanaxadiol	[49]
	CYP716A53v2	<i>Panax ginseng</i>	<i>S. cerevisiae</i>	protopanaxadiol	protopanaxatriol	[49]

UDP-glycosyltransferases

Triterpenoid saponins are generated by the glycosylation of triterpenoids, which is realized by transferring a sugar from UDP-sugar catalyzed by UGTs [68]. This modification step confers such compounds more water solubility, stability, and pharmacological properties. The group UGTs 71, 73, 74, 85, 91, and 94, are thought to be involved in the glycosylation of triterpenoid scaffolds (Table 3)[69]. Similar to P450s, UGTs also shows regioselectivity which catalyzes the glycosylation of triterpenoids at the position of C-3, C-28, C-6, C-20, and C-22 [70].

To date, numerous dammarene-type related UGTs have been identified in *P. ginseng*, which

considerably enrich the structural diversity of ginsenosides. For example, PgUGT74AE2 can decorate the C3 hydroxyl groups of the protopanaxadiol and ginsenoside compound K to form Rh2 and F2 [71]. Moreover, the unnatural ginsenoside—3 β ,12 β -Di-O-glc-protopanaxadiol has been synthesized from protopanaxadiol by UGT109A1 from *B. subtilis* [72]. In contrast, the pentacyclic triterpenoids-oleanolic acid and betulinic acid are mainly glycosylated by the UGT73 family (such as UGT73C10, UGT73C13) at their position of C-3 and C-28, to respectively generate 3-glc-oleanolic acid, 28-glc-betulinic acid [73].

Although the expression of UGTs has been successfully implemented in engineered yeast. the yield of these compounds is considerably low, only reaching to the milligram per liter. However, with the application of random mutations in the UGTPg45 gene, 2.25 g/L ginsenoside Rh2 in 10 L fed-batch fermentation was obtained based on the high-yield of its precursor in yeast [74]. Furthermore, the UGT73C5 coupled with sucrose synthase (from *A. thaliana*) was applied to generate the ginsenoside Rh2 from protopanaxadiol *in vitro*. This is due to that sucrose synthase can convert sucrose and UDP into UDPG which is indispensable for the glycosylation. Finally, the highest production of ginsenoside Rh2 reaches to 3.2 g/L by gradually adding protopanaxadiol in the reaction [75].

Table 3 Plant UGTs verified in yeast

Clade	P450	Source	Host	Substrates	Products	Ref
UGT71	UGT71G1	<i>Medicago truncatula</i>	<i>S. cerevisiae</i>	medicagenic acid	medicagenic acid-C28-O- glc	[76]
UGT73	UGT73C10, UGT73C11, UGT73C12, UGT73C13	<i>Barbarea vulgaris</i>	<i>S. cerevisiae</i>	oleanolic acid and hederagenin	3-o- β -glc-oleanolic acid, 3-o- β -glc-hederagenin	[73]
	UGT73K1	<i>Medicago truncatula</i>	<i>S. cerevisiae</i>	hederagenin, soyasapogenol B/E	hederagenin-glc, soyasapogenol B/E-glc	[76]
	UGT73F2 UGT73F4	Glycine max	<i>S. cerevisiae</i>	saponin A0-ag	saponin deacetyl Ab	[77]
	UGT73F3	<i>Medicago truncatula</i>	<i>S. cerevisiae</i>	hederagenin	hederagenin-C28-O-glc	[78]

	UGT73P2	<i>Glycine max</i>	<i>S. cerevisiae</i>	soyasapogenol B 3-o-glucuronide	soyasaponinβg	[79]
UGT74	UGT74M1	<i>Saponaria vaccaria</i>	<i>S. cerevisiae</i>	gypsogenin, gypsogenic acid	28-glc-gypsogenin, 28-glc-gypsogenic acid	[31]
	UGT74AE2	<i>Panax ginseng</i> ,	<i>S. cerevisiae</i>	ginsenoside compound K	ginsenoside F2	[71]
UGT91	UGT91H4	<i>Glycine max</i>	<i>S. cerevisiae</i>	soyasaponin III	soyasaponin βg	[31]
	UGT94Q2	<i>Panax. ginseng</i> ,	<i>S. cerevisiae</i>	ginsenoside F2	ginsenoside Rd	[71]
	UGTPg100	<i>Panax. ginseng</i> ,	<i>S. cerevisiae</i>	protopanaxatriol	ginsenoside Rh1	[80]
	UGTPg1	<i>Panax. ginseng</i> ,	<i>S. cerevisiae</i>	dammarenediol II protopanaxadiol	ginsenoside Rh2	[74]
	UGTPg45	<i>Panax. ginseng</i> ,	<i>S. cerevisiae</i>	protopanaxadiol	ginsenoside Rh2	[74]
	UGTPn50	<i>Panax notoginseng</i>	<i>S. cerevisiae</i>	protopanaxadiol	ginsenoside Rh2	

179

180 Engineering yeast metabolism for improving the triterpenoids production

181 Enhancing the precursor supply

182 A metabolic engineering approach is a powerful tool that can improve the yield and productivity
183 of target products, reduce by-products, and increase cell viability in various microbes. Up to now,
184 the most successful example is engineering *S. cerevisiae* for the heterologous production of
185 artemisinic acid, which is the precursor of artemisinin—an effective drug for people suffering from
186 malaria [81]. To achieve this, *S. cerevisiae* was engineered by up-regulating the expression of
187 genes (overexpression of *tHMG1* gene) in the MVA pathway, repressing the competing reaction
188 (transcriptional restriction of *ERG9* expression), improving the catalytic efficiency of P450
189 reactions (additional expression of *ADH1* and *ALDH1* genes from *A. annua*). Finally, 25 g/L
190 artemisinic acid was achieved by an ethanol pulse feed process which has been applied in
191 industrial production [54](Figure 5). This case sheds the light on the production of triterpenoids.
192 Currently, the strategy for improving the production of triterpenoids is mainly implemented in
193 three steps: 1. increasing the supply of the precursor—2,3-oxidosqualene. 2. optimizing the
194 triterpenoid biosynthetic pathway. 3. lowering the competition of the native sterol pathway
195 (Table 4).

196

197 Overexpression of the *tHMG1* gene results in the accumulation of squalene, which is a common
198 and efficient strategy for enhancing triterpenoid production [82, 83]. It is due to that the
199 reduction of hydroxymethylglutaryl-CoA (HMG-CoA) to mevalonate catalyzed by HMG-CoA
200 reductase is generally considered as a rate-limiting step in the MVA pathway. As squalene is an
201 indirect precursor of triterpenoids biosynthesis. squalene epoxidase (encoded by *ERG1* gene) is
202 also deemed to catalyze another rate-limiting step in the biosynthesis of triterpenoids. Dai et al.
203 demonstrated that overexpression of the *ERG1* gene (encoding squalene synthase) could result in
204 a 10-fold increase in the yield of protopanaxadiol [35].

205 Besides the rate-limiting step, elevating the expression level of all enzymes involved in the MVA
206 pathway is also beneficial for increasing the carbon flux in yeast. Up to 18 g/L dammarene-type
207 triterpenoids (dammarenediol II and protopanaxadiol) was achieved based on multi-copy
208 integration of the whole 10 genes in the mevalonate pathway in the chromosome of *S. cerevisiae*
209 using delta DNA site [74]. However, a similar strategy seems to have limited effects on the
210 production of oleanane-type, ursane-type, and lupane-type triterpenoids. With the high
211 expression level of all genes of the MVA pathway as well as CYP716A155 and CPR, approximately
212 1 g/L betulin and 1.5 g/L betulinic acid were achieved though the fed-batch fermentation [67].
213 Also, 600 mg/L oleanolic acid and 130 mg/L ursolic acid are respectively the highest yield
214 achieved through the fed-batch fermentation reported to date [12], which is much lower than
215 dammarene-type triterpenoids. It is probably caused by different OSCs, in which dammarenediol
216 synthase exhibited higher enzyme activities than other cyclases in *S. cerevisiae*. Despite this,
217 higher-level expression of rate-limiting enzymes, as well as other enzymes, is considered as an

essential and efficient strategy for increasing the production of triterpenoids.

The reactions catalyzed by P450s and CPR are also vital for achieving the high-yield of triterpenoids, which requires extra carbon, NADPH and oxygen. The poor coupling of P450-CPR may result in the excess generation of the reactive oxygen species (ROS) which can damage DNA, proteins, lipids, even cause apoptosis in the cell [24]. Hence, regulation of yeast metabolism by engineering related enzymes is an alternative to benefit the yield of triterpenoids. Zhao et al. showed that expression of *SSD1* and *YBP1* genes (involved in the cellular response to oxidative stress), effectively decreased the ROS level and increased cell viability, which further improved the efficiency of protopanaxadiol biosynthesis [23]. Enhancing the supply of acetyl-CoA by overexpression of both the *ACS1* gene (acetyl-CoA synthetase) from *S. enterica* and the native *ALD6* gene could also increase the amorphaadiene (a sesquiterpene) production [84]. Moreover, the deletion of *GDH1* gene (encoding glutamate dehydrogenase) enhanced the available NADPH in the cytosol which also resulted in an approximately 85% increase in the production of cubebol (a sesquiterpene) [85].

Restraining the native sterol pathway

S. cerevisiae is not only used for triterpenoid but also for other terpenoid production based on intermediates of the MVA pathway. Hence, the competition between the native pathway and the heterologous pathway is always regarded as a hurdle for achieving the high yield of targeted products. The strategy for down-regulation of protein activity has been already applied to other metabolic nodes of the MVA pathway, such as Erg9p and Erg20p. Repressing the *ERG9* expression by replacing its native promoter with the copper-regulated promoter can efficiently redirect the

carbon flow to the production of artemisinic acid [54]. Also, the ER-associated protein degradation system was applied to shorten the half-life of Erg9p which could improve the production of trans-nerolidol (a sesquiterpene) by 86% [86]. For the downregulation of *ERG20* gene, the expression of Erg20p was weakened by N-degron mediated protein degradation and sterol-responsive transcriptional regulation, which could significantly increase the yield of monoterpene (limonene or geraniol) production in engineered yeast strains [87].

Similarly, many protein down-regulation strategies have been already applied to Erg7p. Broker et al. inhibited the expression of the *ERG7* gene by the copper-regulated promoter, which resulted in the accumulation of high specific 2,3-oxidosqualene titer (about 200 mg/g DCW) in the presence of 150 μ M CuSO₄. However, the bulk of the accumulated 2,3-oxidosqualene seemed to contribute less to improve the production of triterpenoids [88]. Similarly, Kirby et al presented a similar strategy that repression of *ERG7* gene was achieved by the methionine-repressible promoter. However, they only observed a 50% increase in the yield of β -amyrin [16]. Moreover, the TetR-TetO based gene regulation system was used to manipulate the transcriptional efficiency of *ERG7* gene, which only led to an about 10% increase in the yield of protopanaxadiol [89]. It appears that the down-regulation of the *ERG7* gene has a limited effect on improving triterpenoid production.

Table 4 Metabolic engineering for production of triterpenoids in various microbes

Host strains	Metabolic engineering strategies	Yield	Ref
DS1	<i>E. coli</i> BL21 (DE3), harboring pOSQ8 (<i>P_{T7}:: idi-ispA- tScERG9-McSE-tAtCPR</i>) and pDS (<i>P_{T7}::DS</i>)	8.63 mg /L dammarenediol II in batch culture	[90]

<i>E. coli</i> BL21 (DE3)	<i>E. coli</i> BL21 (DE3) harboring pAC-HpIDI/AtSQS/SQE, pRSFD-camAB, pRSFD-NsRED/FER and pETD-EtAS	β-amyrin about 175 μg/g FCW (fresh cell weight) in batch culture	[22]
YY16	INVSc1, pYYG-ERG20-ERG9-ERG1-MdOSC1, 2μ	11.97±0.61 mg/L α-amyrin, in batch culture	[91]
ScLCZ08	<i>S. cerevisiae</i> W303-1a harboring pRS403-tHMG1 and pRS405-ERG20; δDNA::P _{TEF1} -CgERG1, P _{GAL1} -CrMAS	175.15 mg/L α-amyrin, 44.92 mg/L β-amyrin in fed-batch fermentation	[12]
ScLCZ11	ScLCZ08; δDNA :: P _{PGK1} -AtCPR1, P _{TDH3} -MtOAS	123.27 mg/L ursolic acid and 155.58 mg/L oleanolic acid in fed-batch fermentation	[12]
BY-OA	<i>S. cerevisiae</i> BY4742, δDNA :: P _{PGK1} -tHMG1 rDNA:: P _{TDH3} -ERG1 P _{TEF1} -ERG9 P _{PGK1} -GgbAS P _{TDH3} -AtCPR1 P _{TEF1} -MtOAS	88.6 mg/L β-amyrin 71.0 mg/L oleanolic acid in batch culture	[11]
BA4	<i>S. cerevisiae</i> CEN.PK, multi-copy integration of AtLUP1 and CYP716A11, single-copy integration of tHMG1, ERG1 and AtATR2	182 mg/L betulinic acid, 854 mg/L total lupane-type triterpenoids in fed-batch fermentation	[32]
BA-2	<i>S. cerevisiae</i> CEN.PK2-1D his3 ::TRP1-P _{PGK1} -AtSQS2, P _{TDH3} -AtSQE2, P _{TEF1} -SmFPS, Gal7::URA3- P _{PGK1} -tHMG1, P _{PGK1} -ERG12, P _{ENO2} -IDI1, P _{PGK1} -ERG19, P _{FBA1} -ERG13, P _{TDH3} -ERG8, P _{TEF1} -ERG10, NDT80::HIS3-P _{PGK1} -RoCPR01, P _{TDH3} -optiRoCYP01, P _{TEF1} -optiAtLUP, GAL80::LEU2-P _{PGK1} -RoCPR01, P _{TDH3} -optiRoCYP01, P _{TEF1} -optiAtLUP	193.5 mg/L betulinic acid yield, 59.5 mg/L betulin and 23.6 mg/L lupeol in batch culture. 0.1 g/L lupeol, 1.0 g/L betulin, and 1.5 g/L betulinic acid in fed-batch fermentation,	[67]
YLCC56	<i>Y. lipolytica</i> , ATCC 201249, P _{EXP} -ERG1, P _{TEF1} -ERG9, P _{TEF1} -HMG1, IntF::P _{TEF1} -MTR, Δku70::P _{TEF1} -RCLUS, hisG	51.87±2.77 mg/L betulinic acid, 204.89 ± 11.56 mg/L total triterpenoids with 65.44% betulin, 23.71% betulinic acid and 10.85% betulinic aldehyde in fed-batch fermentation	[92]
ZD-PPD-018	<i>S. cerevisiae</i> BY4742, δDNA :: P _{PGK1} -tHMG1, P _{TDH3} -ERG1, P _{TEF1} -ERG9, P _{PGK1} -ERG20, P _{PGK1} -AtCPR1, rDNA:: P _{PGK1} -PgDDS, P _{TDH3} -AtCPR1, P _{TEF1} -PgPPDS	1548 mg/L dammarenediol II, 1189 mg/L protopanaxadiol in fed-batch fermentation	[35]
W3a	<i>S. cerevisiae</i> W303-1a, rDNA :: P _{TDH3} -tHMG1, P _{PGK1} -ERG20, P _{TEF1} -ERG9, P _{TEF1} -ERG1, P _{PGK1} -DS, P _{PGK1} -PPDS-linker1- 46tATR1	1436.6 mg/L protopanaxadiol in fed-batch fermentation	[52]
W3a-ssPy	W3a, ura3 :: P _{PGK1} -YBP1 his3:: P _{PGK1} -SSD1	4.25±0.18 g/L protopanaxadiol in fed-batch fermentation	[86]
WLT-MVA5	<i>S. cerevisiae</i> W303-1a, ura3, his3 :: P _{ALD6} -DS, P _{TDH3} -PPDS-ATR1, P _{TEF1} -ACSse ^{L641P} P _{TEF1} -YBP1, P _{PGK1} -SSD1, P _{PGK1} -NCP1, P _{ACS1} -TetR, P _{ERG7C} -ERG7 with overexpression of every gene in the MVA pathway	8.09 g/L protopanaxadiol in fed-batch fermentation	[89]
ZWDRH2-10	<i>S. cerevisiae</i> BY4742, overexpression of every gene in the MVA and the triterpenoid biosynthetic pathway δDNA:: P _{UAS-TDH3} -UGTPn50-HV	8088.8 mg/L dammarenediol II, 9054.5 mg/L protopanaxadiol and 2252.3 mg/L ginsenoside Rh2 in fed-batch fermentation	[74]
GSKED-PTH11	<i>P. pastoris</i> strain GS115, pPIC3.5K-ERG1-PgDDS, GSKED, with P _{ERG7} replaced by P _{TH11}	1.073 mg/gDCW dammarenediol II in batch culture	[93]
YL-MVA-CK	<i>Y. lipolytica</i> ATCC 201249 zeta :: P _{FBA1} -tHMG1 P _{EXP} 1-ERG9, P _{GPD1} -ERG20 rDNA:: P _{TEF1} -OpDS, P _{EXP1} -PPDS-linker2-ATR1, P _{GPD1} -UGT1	30.4 mg/L ginsenoside CK in batch culture, 161.8 mg/L ginsenoside CK in fed-batch fermentation	[74]
SGibS	<i>S. cerevisiae</i> INVSc1, rDNA :: P _{FBA1} -bAS, P _{TYL1} -ERG1, P _{GPM1} -ERG9, P _{ALAI} -ERG20, P _{ADH1} -IDI	138.8 mg/L β-amyrin in fed-batch fermentation	[94]
OA07	<i>S. cerevisiae</i> JDY52, overexpression of GgbAS, MtCPR1 and CYP716A12, ΔGAL1 and ΔGAL80, P _{TEF1} -tHMG1, MtCPR1, P _{GAL1} -ERG1, P _{GAL10} -ERG9.	186.1 mg/L oleanolic acid in shake flask, 606.9 ± 9.1 mg/L oleanolic acid with 16.0 ± 0.8 mg/g DCW in the fed-batch fermentation.	[12]

258

259 Challenges for triterpenoids production in yeast

260 To date, most of the metabolic engineering approaches focus on elevating the carbon flux of the

261 MVA pathway. However, there are fewer successful cases on the manipulation of Erg7, OSCs,

262 P450s, and UGTs, which are still challenges that need to be overcome. These challenges also

263 represent the potential for continued improvement for the heterologous production of

264 triterpenoids in *S. cerevisiae*.

Although microbial synthesis is relatively easier than the chemical synthesis, there are still many uncharacterized synthetic pathways of triterpenoids. Due to the loss of intermediates, it is challenging to speculate an unknown synthetic pathway based on the limited intermediates and the end-products. Even though several triterpenoid biosynthetic pathways have been identified and characterized, we still lack appropriate enzymes expressed in yeast. For example, a monofunctional α -amyrin synthase has not yet been identified in plants, which causes the production of α -amyrin frequently with the accumulation of β -amyrin, lupeol, and other triterpenes. In our study, CYP716A15 catalyzed the conversion of lupeol to betulinic acid, with the concomitant of betulin and betulinic aldehyde due to incomplete oxidation. This phenomenon seriously influences the yield of targeted products and causes a high-cost purification process. Besides, there is currently a lack of a chassis capable of producing high-level production of 2,3-oxidosqualene, which is simultaneously beneficial for triterpenoids production. In general, squalene accumulation is a representative indicator for the carbon flux of the MVA pathway. However, no 2,3-oxidosqualene producing yeast strain has been achieved through metabolic engineering, probably due to that there is not an effective strategy for down-regulation of *ERG7* gene.

Although ergosterol is essential for yeast cell viability, a decrease of ergosterol content concurrent with 2,3-oxidosqualene accumulation can be realized by the addition of U14266A (3 β -(2-dimethylaminoethoxy)-androst-5-en-17-one, an inhibitor of Erg7p) [95]. This result indicated that there is some space to repress the Erg7p activity in the condition of healthy cells. Nonetheless, Erg7p is not a single enzyme that has complicated interactions with other enzymes in the sterol pathway, such as Erg6p, Erg11p, Erg27p [96]. For example, deletion of 3-keto sterol

reductase (Erg27p) leads to a concomitant loss of lanosterol synthase (Erg7p) [97], it is probably a hurdle to manipulate the Erg7 for redirection the carbon flow to the triterpenoid biosynthesis rather than the native sterol pathway.

Perspective

Identification of novel genes is a vital tool for understanding the triterpenoid synthetic pathway, especially as often more than one triterpenoid is synthesized and the enzyme classes (OSCs, P450s, UGTs) contribute to biochemical diversity. Besides, a transporter may exist in the plant cell, which can transport betulin and even betulinic acid to the bark of white birch trees. A functional transporter probably reduces the metabolic burden, improves the production of triterpenoids and benefits for the extraction process in yeast. With the development of synthetic biology, high-throughput sequencing is a potential method to dig novel genes for efficient production of triterpenoids.

Apart from the discovery of the novel gene, application of protein engineering on the existing Erg7, OSCs, P450s, and UGTs is also a promising strategy to facilitate the triterpenoid biosynthesis. With the development of protein engineering, many mature strategies have been applied to control protein activity. For example, several conditional degrons have been designed to control the abundance of targeted protein by adapting experimental conditions, which is a promising strategy to repress the Erg7 expression [98, 99]. Active-site mutagenesis of triterpenoid-related P450s might increase the catalytic efficiency and decrease the generation of by-products, which has been utilized to improve the selectivity and catalytic activity of the bacterial P450s [100].

Considering that OSCs and P450s are respectively localized in lipid droplets and ER membranes [101], the co-localization of these enzymes is probably beneficial for the yield of triterpenoids. For instance, relocation of the MVA pathway to the mitochondria could enhance the carbon flux which further improved the yield of sesquiterpene [102]. Moreover, compartmentalization of isobutanol biosynthetic pathway into the mitochondria was also able to improve the isobutanol yields by 260% in *S. cerevisiae* [103].

In conclusion, *S. cerevisiae* is an excellent alternative for biotechnological production of valuable triterpenoids and identification of putative triterpenoid-related enzymes. Metabolic engineering and protein engineering allow further modification of yeast to efficiently produce targeted triterpenoids, which paves the way for the industrial production in place of extraction from plant sources.

Conflicts of interest

The authors declare no conflicts of interests

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant No. 21676026), and the National Key R&D Program of China (grant No. 2017YFD0201400).

References

- 327 1. Bohlmann, J. and C.I. Keeling, Terpenoid biomaterials. The Plant Journal, 2008. **54**(4): p.
- 328 656-669.
- 329 2. Springob, K. and T.M. Kutchan, Introduction to the different classes of natural products, in
- 330 *Plant-Derived Natural Products*. 2009, Springer. p. 3-50.
- 331 3. Mizutani, K., et al., Sweetness of glycyrrhetic acid 3-O- β -D-monoglucuronide and the related
- 332 glycosides. Bioscience, biotechnology, and biochemistry, 1994. **58**(3): p. 554-555.
- 333 4. Ghosh, S., Biosynthesis of structurally diverse triterpenes in plants: the role of oxidosqualene
- 334 cyclases. Proceedings of the Indian National Science Academy, 2016. **82**(4): p. 1189-1210.
- 335 5. Christensen, L.P., Ginsenosides: chemistry, biosynthesis, analysis, and potential health effects.
- 336 *Advances in food and nutrition research*, 2008. **55**: p. 1-99.
- 337 6. Shibata, S., Chemistry and cancer preventing activities of ginseng saponins and some related
- 338 triterpenoid compounds. Journal of Korean medical science, 2001. **16**(Suppl): p. S28.
- 339 7. Zhao, H., et al., Efficient synthesis and anti-fungal activity of oleanolic acid oxime esters.
- 340 *Molecules*, 2013. **18**(3): p. 3615-3629.
- 341 8. Nisbet, A.J., Azadirachtin from the neem tree *Azadirachta indica*: its action against insects.
- 342 *Anais da Sociedade Entomológica do Brasil*, 2000. **29**(4): p. 615-632.
- 343 9. Jauch, J., Total synthesis of azadirachtin—finally completed after 22 years. *Angewandte*
- 344 *Chemie International Edition*, 2008. **47**(1): p. 34-37.
- 345 10. Kushiro, T., et al., Mutational studies on triterpene synthases: engineering lupeol synthase
- 346 into β -amyrin synthase. Journal of the American Chemical Society, 2000. **122**(29): p.
- 347 6816-6824.

- 348 11. Dai, Z., et al., Producing aglycons of ginsenosides in bakers' yeast. Scientific reports, 2014. **4**:
349 p. 3698.
- 350 12. Lu, C., et al., Biosynthesis of ursolic acid and oleanolic acid in *Saccharomyces cerevisiae*.
351 *AIChE Journal*, 2018. **64**(11): p. 3794-3802.
- 352 13. Haralampidis, K., M. Trojanowska, and A.E. Osbourn, Biosynthesis of triterpenoid saponins in
353 plants, in *History and Trends in Bioprocessing and Biotransformation*. 2002, Springer. p.
354 31-49.
- 355 14. Lodeiro, S., et al., An oxidosqualene cyclase makes numerous products by diverse
356 mechanisms: a challenge to prevailing concepts of triterpene biosynthesis. *Journal of the*
357 *American Chemical Society*, 2007. **129**(36): p. 11213-11222.
- 358 15. Thimmappa, R., et al., Triterpene biosynthesis in plants. *Annual review of plant biology*, 2014.
359 **65**: p. 225-257.
- 360 16. Kirby, J., et al., Engineering triterpene production in *Saccharomyces cerevisiae*– β -amyrin
361 synthase from *Artemisia annua*. *The FEBS journal*, 2008. **275**(8): p. 1852-1859.
- 362 17. Pecchia, P., et al., Quinic Acids from *Aster caucasicus* and from Transgenic Callus Expressing a
363 β -Amyrin Synthase $\S$. *Natural product communications*, 2011. **6**(11): p.
364 1934578X1100601125.
- 365 18. Shibuya, M., et al., Identification of a product specific β -amyrin synthase from *Arabidopsis*
366 *thaliana*. *Plant Physiology and Biochemistry*, 2009. **47**(1): p. 26-30.
- 367 19. Basyuni, M., et al., Triterpene synthases from the Okinawan mangrove tribe, Rhizophoraceae.
368 *The FEBS journal*, 2007. **274**(19): p. 5028-5042.

- 369 20. Zhang, H., et al., Oxidosqualene cyclases from cell suspension cultures of *Betula platyphylla*
370 var. *japonica*: molecular evolution of oxidosqualene cyclases in higher plants. *Biological and*
371 *Pharmaceutical Bulletin*, 2003. **26**(5): p. 642-650.
- 372 21. Fiallos-Jurado, J., et al., Saponin determination, expression analysis and functional
373 characterization of saponin biosynthetic genes in *Chenopodium quinoa* leaves. *Plant Science*,
374 2016. **250**: p. 188-197.
- 375 22. Takemura, M., R. Tanaka, and N. Misawa, Pathway engineering for the production of β -amyrin
376 and cycloartenol in *Escherichia coli*—a method to biosynthesize plant-derived triterpene
377 skeletons in *E. coli*. *Applied microbiology and biotechnology*, 2017. **101**(17): p. 6615-6625.
- 378 23. Ito, R., Y. Masukawa, and T. Hoshino, Purification, kinetics, inhibitors and CD for recombinant
379 β -amyrin synthase from *Euphorbia tirucalli* L and functional analysis of the DCTA motif, which
380 is highly conserved among oxidosqualene cyclases. *The FEBS journal*, 2013. **280**(5): p.
381 1267-1280.
- 382 24. Sawai, S., et al., Functional and structural analysis of genes encoding oxidosqualene cyclases
383 of *Lotus japonicus*. *Plant Science*, 2006. **170**(2): p. 247-257.
- 384 25. Iturbe-Ormaetxe, I., et al., Molecular cloning and characterization of triterpene synthases
385 from *Medicago truncatula* and *Lotus japonicus*. *Plant molecular biology*, 2003. **51**(5): p.
386 731-743.
- 387 26. Scholz, M., et al., Methyl jasmonate induced accumulation of kalopanaxsaponin I in *Nigella*
388 *sativa*. *Phytochemistry*, 2009. **70**(4): p. 517-522.

- 389 27. Kushiro, T., M. Shibuya, and Y. Ebizuka, β -Amyrin synthase: cloning of oxidosqualene cyclase
390 that catalyzes the formation of the most popular triterpene among higher plants. European
391 Journal of Biochemistry, 1998. **256**(1): p. 238-244.
- 392 28. Jin, M.L., et al., Isolation and characterization of an oxidosqualene cyclase gene encoding a
393 β -amyrin synthase involved in *Polygala tenuifolia* Willd. saponin biosynthesis. Plant cell
394 reports, 2014. **33**(3): p. 511-519.
- 395 29. Morita, M., et al., Molecular cloning of pea cDNA encoding cycloartenol synthase and its
396 functional expression in yeast. Biological and Pharmaceutical Bulletin, 1997. **20**(7): p.
397 770-775.
- 398 30. Wang, Z., et al., Two oxidosqualene cyclases responsible for biosynthesis of tomato fruit
399 cuticular triterpenoids. Plant Physiology, 2011. **155**(1): p. 540-552.
- 400 31. Meesapyodsuk, D., et al., Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding
401 β -amyrin synthase and a triterpene carboxylic acid glucosyltransferase. Plant physiology, 2007.
402 **143**(2): p. 959-969.
- 403 32. Czarnotta, E., et al., Fermentation and purification strategies for the production of betulinic
404 acid and its lupane-type precursors in *Saccharomyces cerevisiae*. Biotechnology and
405 bioengineering, 2017. **114**(11): p. 2528-2538.
- 406 33. Hayashi, H., et al., Differential expression of three oxidosqualene cyclase mRNAs in
407 *Glycyrrhiza glabra*. Biological and Pharmaceutical Bulletin, 2004. **27**(7): p. 1086-1092.
- 408 34. Shibuya, M., et al., Two branches of the lupeol synthase gene in the molecular evolution of
409 plant oxidosqualene cyclases. European Journal of Biochemistry, 1999. **266**(1): p. 302-307.

- 410 35. Dai, Z., et al., Metabolic engineering of *Saccharomyces cerevisiae* for production of
411 ginsenosides. *Metabolic engineering*, 2013. **20**: p. 146-156.
- 412 36. Wang, L., et al., The isolation and characterization of dammarenediol synthase gene from
413 *Panax quinquefolius* and its heterologous co-expression with cytochrome P450 gene PqD12H
414 in yeast. *Functional & integrative genomics*, 2014. **14**(3): p. 545-557.
- 415 37. Shibuya, M., S. Adachi, and Y. Ebizuka, Cucurbitadienol synthase, the first committed enzyme
416 for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol
417 biosynthesis. *Tetrahedron*, 2004. **60**(33): p. 6995-7003.
- 418 38. Fazio, G.C., R. Xu, and S.P. Matsuda, Genome mining to identify new plant triterpenoids.
419 *Journal of the American Chemical Society*, 2004. **126**(18): p. 5678-5679.
- 420 39. Hodgson, H., et al., Identification of key enzymes responsible for protolimonoid biosynthesis
421 in plants: Opening the door to azadirachtin production. *Proceedings of the National Academy*
422 *of Sciences*, 2019. **116**(34): p. 17096-17104.
- 423 40. Xue, Z., et al., Divergent evolution of oxidosqualene cyclases in plants. *New Phytologist*, 2012.
424 **193**(4): p. 1022-1038.
- 425 41. Basyuni, M., et al., Molecular cloning and functional expression of a multifunctional
426 triterpene synthase cDNA from a mangrove species *Kandelia candel* (L.) Druce.
427 *Phytochemistry*, 2006. **67**(23): p. 2517-2524.
- 428 42. Zheng, X., et al., Characterisation of two oxidosqualene cyclases responsible for triterpenoid
429 biosynthesis in *Ilex asprella*. *International journal of molecular sciences*, 2015. **16**(2): p.
430 3564-3578.

- 431 43. Hayashi, H., et al., Cloning and characterization of a cDNA encoding β -amyrin synthase
432 involved in glycyrrhizin and soyasaponin biosyntheses in licorice. Biological and
433 Pharmaceutical Bulletin, 2001. **24**(8): p. 912-916.
- 434 44. Yu, F., et al., Functional characterization of amyrin synthase involved in ursolic acid
435 biosynthesis in Catharanthus roseus leaf epidermis. Phytochemistry, 2013. **91**: p. 122-127.
- 436 45. Brendolise, C., et al., An unusual plant triterpene synthase with predominant
437 α -amyrin-producing activity identified by characterizing oxidosqualene cyclases from Malus
438 domestica. The FEBS journal, 2011. **278**(14): p. 2485-2499.
- 439 46. Morita, M., et al., Molecular cloning and functional expression of triterpene synthases from
440 pea (*Pisum sativum*) New α -amyrin-producing enzyme is a multifunctional triterpene
441 synthase. European Journal of Biochemistry, 2000. **267**(12): p. 3453-3460.
- 442 47. Nelson, D. and D. Werck-Reichhart, A P450-centric view of plant evolution. The Plant Journal,
443 2011. **66**(1): p. 194-211.
- 444 48. Miettinen, K., et al., The ancient CYP716 family is a major contributor to the diversification of
445 eudicot triterpenoid biosynthesis. Nature communications, 2017. **8**(1): p. 1-13.
- 446 49. Han, J.-Y., et al., Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol
447 from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. Plant and cell
448 physiology, 2012. **53**(9): p. 1535-1545.
- 449 50. Renault, H., et al., Cytochrome P450-mediated metabolic engineering: current progress and
450 future challenges. Current opinion in plant biology, 2014. **19**: p. 27-34.

- 451 51. Dai, Z., et al., Identification of a novel cytochrome P450 enzyme that catalyzes the C-2 α
452 hydroxylation of pentacyclic triterpenoids and its application in yeast cell factories. *Metabolic*
453 *engineering*, 2019. **51**: p. 70-78.
- 454 52. Zhao, F., et al., Optimization of a cytochrome P450 oxidation system for enhancing
455 protopanaxadiol production in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*,
456 2016. **113**(8): p. 1787-1795.
- 457 53. Brugliera, F., et al., Violet/blue chrysanthemums—metabolic engineering of the anthocyanin
458 biosynthetic pathway results in novel petal colors. *Plant and Cell Physiology*, 2013. **54**(10): p.
459 1696-1710.
- 460 54. Paddon, C.J., et al., High-level semi-synthetic production of the potent antimalarial
461 artemisinin. *Nature*, 2013. **496**(7446): p. 528-532.
- 462 55. Geisler, K., et al., Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme
463 required for synthesis of antimicrobial triterpenes in plants. *Proceedings of the National*
464 *Academy of Sciences*, 2013. **110**(35): p. E3360-E3367.
- 465 56. Shibuya, M., et al., Identification of β -amyrin and sophoradiol 24-hydroxylase by expressed
466 sequence tag mining and functional expression assay. *The FEBS journal*, 2006. **273**(5): p.
467 948-959.
- 468 57. Fukushima, E.O., et al., CYP716A subfamily members are multifunctional oxidases in
469 triterpenoid biosynthesis. *Plant and Cell Physiology*, 2011. **52**(12): p. 2050-2061.
- 470 58. Seki, H., et al., Licorice β -amyrin 11-oxidase, a cytochrome P450 with a key role in the
471 biosynthesis of the triterpene sweetener glycyrrhizin. *Proceedings of the National Academy*
472 *of Sciences*, 2008. **105**(37): p. 14204-14209.

- 473 59. Seki, H., et al., Triterpene functional genomics in licorice for identification of CYP72A154
474 involved in the biosynthesis of glycyrrhizin. *The Plant Cell*, 2011. **23**(11): p. 4112-4123.
- 475 60. Fukushima, E.O., et al., Combinatorial biosynthesis of legume natural and rare triterpenoids
476 in engineered yeast. *Plant and cell physiology*, 2013. **54**(5): p. 740-749.
- 477 61. Carelli, M., et al., *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the
478 biosynthesis of hemolytic saponins. *The plant cell*, 2011. **23**(8): p. 3070-3081.
- 479 62. Huang, L., et al., Molecular characterization of the pentacyclic triterpenoid biosynthetic
480 pathway in *Catharanthus roseus*. *Planta*, 2012. **236**(5): p. 1571-1581.
- 481 63. Moses, T., et al., Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces*
482 *cerevisiae* using a C-16 α hydroxylase from *Bupleurum falcatum*. *Proceedings of the National*
483 *Academy of Sciences*, 2014. **111**(4): p. 1634-1639.
- 484 64. Tamura, K., et al., CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured
485 stolons of *Glycyrrhiza uralensis*. *Plant cell reports*, 2017. **36**(3): p. 437-445.
- 486 65. Sui, C., et al., Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in
487 the biosynthesis of saikosaponins. *BMC genomics*, 2011. **12**(1): p. 539.
- 488 66. Field, B. and A.E. Osbourn, Metabolic diversification—-independent assembly of operon-like
489 gene clusters in different plants. *Science*, 2008. **320**(5875): p. 543-547.
- 490 67. Huang, J., et al., Identification of RoCYP01 (CYP716A155) enables construction of engineered
491 yeast for high-yield production of betulinic acid. *Applied microbiology and biotechnology*,
492 2019. **103**(17): p. 7029-7039.

- 493 68. Gachon, C.M., M. Langlois-Meurinne, and P. Saindrenan, Plant secondary metabolism
494 glycosyltransferases: the emerging functional analysis. Trends in plant science, 2005. **10**(11):
495 p. 542-549.
- 496 69. Brazier-Hicks, M., et al., Substrate specificity and safener inducibility of the plant
497 UDP-glucose-dependent family 1 glycosyltransferase super-family. Plant biotechnology
498 journal, 2018. **16**(1): p. 337-348.
- 499 70. Osmani, S.A., S. Bak, and B.L. Møller, Substrate specificity of plant UDP-dependent
500 glycosyltransferases predicted from crystal structures and homology modeling.
501 Phytochemistry, 2009. **70**(3): p. 325-347.
- 502 71. Jung, S.-C., et al., Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd.
503 Plant and Cell Physiology, 2014. **55**(12): p. 2177-2188.
- 504 72. Liang, H., et al., Production of a bioactive unnatural ginsenoside by metabolically engineered
505 yeasts based on a new UDP-glycosyltransferase from *Bacillus subtilis*. Metabolic engineering,
506 2017. **44**: p. 60-69.
- 507 73. Augustin, J.M., et al., UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea*
508 *vulgaris* catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. Plant
509 physiology, 2012. **160**(4): p. 1881-1895.
- 510 74. Wang, P., et al., Synthesizing ginsenoside Rh2 in *Saccharomyces cerevisiae* cell factory at
511 high-efficiency. Cell discovery, 2019. **5**(1): p. 1-14.
- 512 75. Hu, Y., et al., Biocatalytic synthesis of ginsenoside Rh2 using *Arabidopsis thaliana*
513 glucosyltransferase-catalyzed coupled reactions. Journal of Biotechnology, 2020.

514 76. Achnine, L., et al., Genomics-based selection and functional characterization of triterpene
515 glycosyltransferases from the model legume *Medicago truncatula*. *The Plant Journal*, 2005.
516 **41**(6): p. 875-887.

517 77. Sayama, T., et al., The Sg-1 glycosyltransferase locus regulates structural diversity of
518 triterpenoid saponins of soybean. *The Plant Cell*, 2012. **24**(5): p. 2123-2138.

519 78. Naoumkina, M.A., et al., Genomic and coexpression analyses predict multiple genes involved
520 in triterpene saponin biosynthesis in *Medicago truncatula*. *The Plant Cell*, 2010. **22**(3): p.
521 850-866.

522 79. Shibuya, M., et al., Identification and characterization of glycosyltransferases involved in the
523 biosynthesis of soyasaponin I in *Glycine max*. *FEBS letters*, 2010. **584**(11): p. 2258-2264.

524 80. Wei, W., et al., Characterization of *Panax ginseng* UDP-glycosyltransferases catalyzing
525 protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically
526 engineered yeasts. *Molecular plant*, 2015. **8**(9): p. 1412-1424.

527 81. Gu, S. and J. Pei, Innovating chinese herbal medicine: From traditional health practice to
528 scientific drug discovery. *Frontiers in pharmacology*, 2017. **8**: p. 381.

529 82. Donald, K., R.Y. Hampton, and I.B. Fritz, Effects of overproduction of the catalytic domain of
530 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces*
531 *cerevisiae*. *Appl. Environ. Microbiol.*, 1997. **63**(9): p. 3341-3344.

532 83. Polakowski, T., U. Stahl, and C. Lang, Overexpression of a cytosolic hydroxymethylglutaryl-CoA
533 reductase leads to squalene accumulation in yeast. *Applied microbiology and biotechnology*,
534 1998. **49**(1): p. 66-71.

- 535 84. Chen, Y., et al., Establishing a platform cell factory through engineering of yeast acetyl-CoA
536 metabolism. *Metabolic engineering*, 2013. **15**: p. 48-54.
- 537 85. Asadollahi, M.A., et al., Enhancing sesquiterpene production in *Saccharomyces cerevisiae*
538 through in silico driven metabolic engineering. *Metabolic engineering*, 2009. **11**(6): p.
539 328-334.
- 540 86. Zhao, F., et al., Enhancing *Saccharomyces cerevisiae* reactive oxygen species and ethanol
541 stress tolerance for high-level production of protopanaxadiol. *Bioresource technology*, 2017.
542 **227**: p. 308-316.
- 543 87. Peng, B., et al., Engineered protein degradation of farnesyl pyrophosphate synthase is an
544 effective regulatory mechanism to increase monoterpene production in *Saccharomyces*
545 *cerevisiae*. *Metabolic engineering*, 2018. **47**: p. 83-93.
- 546 88. Bröker, J.N., et al., Upregulating the mevalonate pathway and repressing sterol synthesis in
547 *Saccharomyces cerevisiae* enhances the production of triterpenes. *Applied microbiology and*
548 *biotechnology*, 2018. **102**(16): p. 6923-6934.
- 549 89. Zhao, F., et al., A modular engineering strategy for high-level production of protopanaxadiol
550 from ethanol by *Saccharomyces cerevisiae*. *AIChE Journal*, 2019. **65**(3): p. 866-874.
- 551 90. Li, D., et al., Heterologous biosynthesis of triterpenoid dammarenediol-II in engineered
552 *Escherichia coli*. *Biotechnology letters*, 2016. **38**(4): p. 603-609.
- 553 91. Yu, Y., et al., Productive Amyrin Synthases for Efficient α -Amyrin Synthesis in Engineered
554 *Saccharomyces cerevisiae*. *ACS synthetic biology*, 2018. **7**(10): p. 2391-2402.

- 555 92. Jin, C.-C., et al., Boosting the biosynthesis of betulinic acid and related triterpenoids in
556 *Yarrowia lipolytica* via multimodular metabolic engineering. *Microbial cell factories*, 2019.
557 **18**(1): p. 77.
- 558 93. Liu, X.-B., et al., Metabolic engineering of *Pichia pastoris* for the production of
559 dammarenediol-II. *Journal of biotechnology*, 2015. **216**: p. 47-55.
- 560 94. Zhang, G., et al., Refactoring β -amyrin synthesis in *Saccharomyces cerevisiae*. *AIChE Journal*,
561 2015. **61**(10): p. 3172-3179.
- 562 95. Milla, P., et al., Yeast oxidosqualene cyclase (Erg7p) is a major component of lipid particles.
563 *Journal of Biological Chemistry*, 2002. **277**(4): p. 2406-2412.
- 564 96. Mo, C. and M. Bard, A systematic study of yeast sterol biosynthetic protein–protein
565 interactions using the split-ubiquitin system. *Biochimica et Biophysica Acta (BBA)-Molecular*
566 *and Cell Biology of Lipids*, 2005. **1737**(2-3): p. 152-160.
- 567 97. Mo, C., et al., In yeast sterol biosynthesis the 3-keto reductase protein (Erg27p) is required for
568 oxidosqualene cyclase (Erg7p) activity. *Biochimica et Biophysica Acta (BBA)-Molecular and*
569 *Cell Biology of Lipids*, 2003. **1633**(1): p. 68-74.
- 570 98. Lutz, A.P., C. Renicke, and C. Taxis, Controlling protein activity and degradation using blue
571 light, in *Optogenetics*. 2016, Springer. p. 67-78.
- 572 99. Jungbluth, M., C. Renicke, and C. Taxis, Targeted protein depletion in *Saccharomyces*
573 *cerevisiae* by activation of a bidirectional degron. *BMC systems biology*, 2010. **4**(1): p. 176.
- 574 100. Zhang, X., et al., Bacterial cytochrome P450-catalyzed regio-and stereoselective steroid
575 hydroxylation enabled by directed evolution and rational design. *Bioresources and*
576 *Bioprocessing*, 2020. **7**(1): p. 2.

101. Milla, P., et al., Subcellular localization of oxidosqualene cyclases from *Arabidopsis thaliana*,
Trypanosoma cruzi, and Pneumocystis carinii expressed in yeast. Lipids, 2002. **37**(12): p.
1171-1176.
102. Farhi, M., et al., Harnessing yeast subcellular compartments for the production of plant
terpenoids. Metabolic engineering, 2011. **13**(5): p. 474-481.
103. Avalos, J.L., G.R. Fink, and G. Stephanopoulos, Compartmentalization of metabolic pathways
in yeast mitochondria improves the production of branched-chain alcohols. Nature
biotechnology, 2013. **31**(4): p. 335.

Figure1. Structures and application of terpenoids in nature.

Figure 2 2,3-Oxidosqualene biosynthetic pathway. AACT, acetyl-CoA C-acetyltransferase; HMGS,
3-hydroxy-3-methylglutaryl-CoA synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase;
MVD, mevalonate diphosphate decarboxylase; DXPS, 1-deoxy-D-xylulose-5-phosphate synthase;
DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; IPK, 4-hydroxy-3-methylbut-2-en-1-yl
diphosphate reductase; IDI, isopentenyl-diphosphate delta-isomerase; FPS, farnesyl diphosphate
synthase; SS, squalene synthase; SE, squalene epoxidase; HMGCsA,
3-hydroxy-3-methylglutaryl-CoA; MVA, mevalonate; MVPP, diphosphomevalonate; DXP,1-
deoxy-D-xylulose-5-phosphate; MEP, methylerythritol phosphate; HMBPP,
1-Hydroxy-2-methyl-2-butenyl 4-diphosphate; IPP, isopentenyl diphosphate; GPP, geranyl
diphosphate; FPP, farnesyl diphosphate; DMAPP, dimethylallyl diphosphate.

Figure 3 Triterpenoid saponins biosynthetic pathway. LPS, lupeol synthase; AAS, α -amyrin
synthase; DDS, dammarenediol II synthase; BAS, β -amyrin synthase; CDS, cucurbitadienol

598 synthase; TAS, tirucalla-7,24-3 β -ol synthase. CYP, cytochrome P450; CPR, cytochrome P450
 599 reductase; UGT, UDP-glycosyltransferase.

600 Figure 4 Site-specific mutagenesis of key amino acids in conserved regions of β -amyrin synthase
 601 (PNY).

602 Figure 5 Overview of artemisinic acid production pathway in *S. cerevisiae*. The genes in the
 603 pathway for biosynthesis of FPP are shown in blue, genes encoding the full three-step oxidation
 604 of amorphaadiene to artemisinic acid from *A. annua* are shown in green. DMAPP: dimethylallyl
 605 diphosphate; FPP: farnesyl diphosphate; IPP: isopentenyl diphosphate. tHMGR: truncated
 606 HMG-CoA reductase. CYP71AV1, CPR1 and CYB5 oxidize amorphaadiene to artemisinic alcohol,
 607 aldehyde, and aldehyde. ADH1 oxidizes artemisinic alcohol to artemisinic aldehyde; ALDH1
 608 oxidizes artemisinic aldehyde to artemisinic acid.