7-15-2019

Identification of RoCYP01 (CYP716A155) enables construction of engineered yeast for high-yield production of betulinic acid

Jiajian Huang
Jinan University - China

Wenlong Zha
Jinan University - China

Tianyue An
Jinan University - China

Hua Dong
Jinan University - China

Ying Huang
Chinese Academy of Sciences

Identification of RoCYP01 (CYP716A155) enables construction of engineered yeast for high-yield production of betulinic acid

Abstract
Betulinic acid (BA) and its derivatives possess potent pharmacological activity against cancer and HIV. As with many phytochemicals, access to BA is limited by the requirement for laborious extraction from plant biomass where it is found in low amounts. This might be alleviated by metabolically engineering production of BA into an industrially relevant microbe such as *Saccharomyces cerevisiae* (yeast), which requires complete elucidation of the corresponding biosynthetic pathway. However, while cytochrome P450 enzymes (CYPs) that can oxidize lupeol into BA have been previously identified from the CYP716A subfamily, these generally do not seem to be specific to such biosynthesis and, in any case, have not been shown to enable high-yielding metabolic engineering. Here RoCYP01 (CYP716A155) was identified from the BA-producing plant *Rosmarinus officinalis* (rosemary) and demonstrated to effectively convert lupeol into BA, with strong correlation of its expression and BA accumulation. This was further utilized to construct a yeast strain that yields > 1 g/L of BA, providing a viable route for biotechnological production of this valuable triterpenoid.

Keywords
betulinic acid, cytochrome P450, synthetic biology, yeast Introduction

Disciplines
Biochemistry, Biophysics, and Structural Biology | Biotechnology | Molecular Biology

Comments
This is a manuscript of an article published as Huang, Jiajian, Wenlong Zha, Tianyue An, Hua Dong, Ying Huang, Dong Wang, Rongmin Yu et al. "Identification of RoCYP01 (CYP716A155) enables construction of engineered yeast for high-yield production of betulinic acid." *Applied microbiology and biotechnology* (2019). doi: 10.1007/s00253-019-10004-z. Posted with permission.

Authors
Jiajian Huang, Wenlong Zha, Tianyue An, Hua Dong, Ying Huang, Dong Wang, Rongmin Yu, Lixin Duan, Xueli Zhang, Reuben J. Peters, Zhubo Dai, and Jiachen Zi

Identification of RoCYP01 (CYP716A155) enables construction of engineered yeast for high-yield production of betulinic acid

Jiajian Huang[1], Wenlong Zha[1], Tianyue An[1], Hua Dong[1], Ying Huang[2], Dong Wang[2], Rongmin Yu[1], Lixin Duan[3,4], Xueli Zhang[2], Reuben J. Peters[5], Zhubo Dai[2], Jiachen Zi[1]

[1] Biotechnological Institute of Chinese Materia Medic, Jinan University, 601 Huangpu Avenue West, Guangzhou 510632, China

[2] Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

[3] Guangdong Province Key Laboratory of Pharmacodynamic Constituents of Traditional Chinese Medicine and New Drugs Research, Jinan University, Guangzhou 510632, China

[4] Joint Laboratory for Translational Cancer Research of Chinese Medicine of the Ministry of Education of the People's Republic of China, International Institute for Translational Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

[5] Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, Iowa 50011, United States

Jiajian Huang, Wenlong Zha and Tianyue An. contributed equally to this work.

To whom correspondence may be addressed.

Reuben J. Peters:

rjpeters@iastate.edu

Zhubo Dai:

dai_zb@tib.cas.cn

Jiachen Zi:

jiachen_zi@163.com

Abstract

Betulinic acid (BA) and its derivatives possess potent pharmacological activity against cancer and HIV. As with many phytochemicals, access to BA is limited by the requirement for laborious extraction from plant biomass where it is found in low amounts. This might be alleviated by metabolically engineering production of BA into an industrially relevant microbe such as *Saccharomyces cerevisiae* (yeast), which requires complete elucidation of the corresponding biosynthetic pathway. However, while cytochrome P450 enzymes (CYPs) that can oxidize lupeol into BA have been previously identified from the CYP716A subfamily, these generally do not seem to be specific to such biosynthesis and, in any case, have not been shown to enable high-yielding metabolic engineering. Here RoCYP01 (CYP716A155) was identified from the BA-producing plant *Rosmarinus officinalis* (rosemary), and demonstrated to effectively convert lupeol into BA, with strong correlation of its expression and BA accumulation. This was further utilized to construct a yeast strain that yields > 1 g/L of BA, providing a viable route for biotechnological production of this valuable triterpenoid.

Keywords: betulinic acid, cytochrome P450, synthetic biology, yeast

Introduction

Betulinic acid (BA) and its derivatives have attracted great attention all over the world due to their promising pharmaceutical potential, especially anticancer and anti-HIV activity. In 1995, it was found that BA seems to serve as a melanoma-specific antitumor agent (Pisha et al. 1995). Indeed, BA is still considered as one of most promising natural products for developing new anticancer drugs, and numerous studies in the past ten years have been aimed at elucidating the efficacy and mechanism of BA and its derivatives in inhibiting cancer cell growth (Chintharlapalli et al. 2007; Liu et al. 2012; Mertens-Talcott et al. 2013; Majeed et al. 2014; Pal et al. 2015; Ye et al. 2017). Moreover, many BA derivatives possess potent inhibitory activity against HIV. 3-*O*-(3′,3′-Dimethylsuccinyl)betulinic acid (bevirimat, also known as PA-457) is capable of potently inhibiting replication of both wild type and drug-resistant HIV-1 isolates by blocking conversion of the capsid precursor p25 to mature capsid protein p24 (Li et al. 2003), with success completion of phase I and IIa clinical trials reported in 2009 (Qian et al. 2010). Recently, many new BA derivatives have been synthesized and found to exhibit inhibitory activity against

bevirimat-resistant HIV-1 variants (Dang et al. 2013; Zhao et al. 2016).

All the BA derivatives mentioned above, including bevirimat, are produced by semi-synthesis from BA. Unfortunately, the content of BA in the commercially used bark from birch (genus *Betula*) is quite low, 0.025% by weight, although this is alleviated to some degree by the alcoholic precursor betulin, which is found in higher amounts and can be converted to BA (Pezzuto and Kim 1998). Much effort has gone into increasing yields from phytochemical isolation of BA and betulin (Krasutsky et al. 2003; Siddiqui and Aeri 2016; Šiman et al. 2016), as well as conversion of betulin to BA (Pezzuto and Kim 1998; Csuk et al. 2006; Tulisalo et al. 2013; Ressmann et al. 2017). However, all these methods are based on isolation of BA and/or betulin from plant materials, which consumes large amounts of plant biomass, and requires a tedious and laborious extraction and purification process that is environmental unfriendly.

Synthetic biology approaches have recently succeeded in producing many important plant-derived natural products at industrial scales, such as artemisinic acid (Paddon et al. 2013) and resveratrol (Li et al. 2015). This requires complete characterization of genes involved in the biosynthesis of the target natural products. BA is produced by oxidation of C-28 in lupeol, which is derived from the common triterpenoid precursor 2,3-oxidosqualene via a polycyclization and rearrangement reaction catalyzed by lupeol synthase (Dewick 2002; Thimmappa et al. 2014). This oxidative conversion of C-28 from a methyl to carboxylic acid is catalyzed by cytochrome P450 enzymes (CYPs), specifically members of the CYP716A subfamily (Thimmappa et al. 2014; Ghosh 2017). Although it has been reported that CYP716A12 (Fukushima et al. 2011; Carelli et al. 2011), CYP716A15 (Fukushima et al. 2011), CYP716AL1 (Huang et al. 2012), CYP716A179 (Tamura et al. 2017), CYP716A80 (Khakimov et al. 2015) and CYP716A81 (Khakimov et al. 2015) are capable of catalyzing the conversion of lupeol to BA, these react more efficiently with α- and β-amyrin rather than lupeol. Recently, CYP716A180 was cloned from *B. platyphylla* and reported to selectively react with lupeol relative to α- and β-amyrin (Zhou et al. 2016). However, this has not yet led to construction of a high-yielding microbial strain suitable for biotechnological production of BA. Rosemary (*Rosmarinus officinalis*) has been reported to produce BA (Razboršek et al. 2007), and RNA-seq data for this has recently become available (Boachon et al. 2018), which can accessed from the corresponding website (http://medicinalplantgenomics.msu.edu). Here this resource was used to identify RoCYP01 (CYP716A155), which was found to preferentially react with lupeol to afford BA. This further enabled construction of an efficient BA biosynthetic pathway in yeast, with additional engineering leading to a strain that produces > 1 g/L.

Materials and Methods

Materials

Standard chemicals including betulinic acid (J&K Scientific, Ltd), betulin (J&K Scientific, Ltd), lupeol (Nanjing SenBeiJia Biological Technology Co., Ltd.), ursolic acid (Shanghai ZZbio Co., Ltd), oleanolic acid (Shanghai ZZbio Co., Ltd), α-amyrin (Shanghai ZZbio Co., Ltd) and β-amyrin (HePeng Biotech Co., Ltd.) were used. Amino acids were supplied from Sigma. Genes were amplified using Phanta Super-Fidelity DNA polymerase (Vazyme Biotech Co., Ltd) and Taq polymerase (TransGEN Biotech Co., Ltd). DNA gel purification and plasmid extraction kits were purchased from TIANGEN Biotech Co., Ltd. Restriction enzymes and T4-DNA polymerase from Thermo Fisher Scientific Co., Ltd., as well as pEASY-Blunt Cloning kit from TransGen Biotech Co., Ltd, were used for DNA cloning. The optimized gene sequences of *RoCYP01* and *AtLUP* were synthesized by Synbio Technologies.

S. cerevisiae strains BY4741 (ATCC 201388) and CEN.PK2-1D (EUROSCARF 30000B) were used as the hosts for functional elucidation and biosynthetic pathway assembly, respectively. *E. coli* DH5α was used for gene cloning. NZY medium was used for cultivation of *E. coli*. SD or SG media lack of corresponding nutrients (e.g. uracil, histidine, methionine, leucine or any combinations of them) were used for strain selection and fermentation of yeasts.

Identification and amplification of CYP and CPR candidates

CYP and CPR candidates were identified from rosemary transcriptome database (http://medicinalplantgenomics.msu.edu) and the RNA-Seq raw data of *B. pendula* and selfheal deposited in NCBI, using the corresponding reference sequences (listed in Supplementary Tables S1 and S2) as queries.

To clone CYP and CPR genes from plants, total RNAs of rosemary tissues were isolated using the TRIzol reagent (Invitrogen) following the manufacturer's instructions. cDNA of each RNA sample was synthesized using the SMARTer^{TM} RACE 5'/3' Kit (Takara Biomedical Technology Co., Ltd) according to product manual. Full-size sequences of truncated genes were obtained by rapid amplification of cDNA ends (RACE) using SMARTer^{TM} RACE 5'/3' Kit. The corresponding primers are listed in Supplementary Table S3.

Phylogenetic analysis

The phylogenetic tree was constructed, using the candidate CYPs of rosemary and the reference CYPs in Supplementary Table S1. Multiple sequence alignments were generated using the CLUSTAL W program. Phylogenetic analysis was conducted using the Maximum Likelihood method with MEGA7.0 software (Kumar et al. 2016). Bootstrap analysis with 1000 replicates was used to assess the strength of the nodes in the tree (Jones et al. 1992).

Construction of expression plasmids for functional elucidation

AtLUP and ATR1 genes were amplified from *Arabidopsis thaliana* cDNA prepared following the same method as preparation of rosemary cDNA.

All the gene fragments were amplified by PCR and ligated into the pEASY blunt cloning vector. The resulting plasmids were confirmed by gene sequencing. After digested, these gene fragments were respectively inserted into the corresponding MCSs of pESC-URA and pESC-LEU vectors to yield the plasmids pHJ01-pHJ07. The primers, resulting plasmids and yeast strains in this section are listed in Supplementary Tables S3-S5, respectively.

Feeding experiments

After a 24-hour fermentation, the cultures of the yeast stain HJ009 containing pHJ05 were suspended with authentic compounds (the final concentration of 4 mg/L) including lupeol, α-amyrin, β-amyrin and betulin, respectively, and then incubated at 30°C and 250 rpm for additional three days. The resulting cultures were extracted according to the method as described in the following section.

Construction and fermentation of high-yield BA-producing yeasts

To construct the expression cassettes, *optiRoCYP01*, *RoCPR01* and *optiAtLUP* were inserted into the SexAI/AscI site of pM4-AtSQE2, pM2-tHMG1 and pM3-SynPgPPD by cut-and-paste method to obtain plasmids pM4-optiRoCYP01, pM2-RoCPR01 and pM3-optiAtLUP. To integrate *optiRoCYP01*, *RoCPR01* and *optiAtLUP* into the NDT80 site of the strain WD2091 (Wang et al. 2018), five fragments were amplified from pM4-optiRoCYP01 (using primer set 3G-3-M-ADHt-TDH3-F/3G-3-M-TPI1t-TEF1-R), pM3-optiAtLUP (using primer set 3G-2-M-TPI1t-TEF1-F/M-CYC1t-pEASY-R), pM2-RoCPR01 (using primer set 1-M-pEASY-PGK1-F/3G-1-M-ADHt-TDH3-R) and pNDT80-HIS (using primer sets X1-M-pEASY-r-t-F/NDT80-interg-2 and NDT80-interg-1/X2-M-pEASY-r-t-R). These fragments were then transformed into strain WD2091 followed by selection on SD-HIS-URA-TRP plates to obtain the strain BA-1. Then, the second copy of the expression cassette of *optiRoCYP01*, *RoCPR01* and *optiAtLUP* was integrated into GAL80 site by transformation of BA-1 with the same three DNA fragments containing the corresponding expression cassettes as above and the other two fragments amplified from pGal80-LEU (using primer sets X1-M-pEASY-r-t-F/Gal80-interg-2 and Gal80-interg-1/X2-M-pEASY-r-t-R). The resulting strains were selected on SD-HIS-URA-LEU-TRP plates to obtain BA-2. All the primers, plasmids and strains in this section are listed in Supplementary Tables S3-S5.

Gene expression pattern analysis

The expression pattern of *RoCYP01* across various tissues of 4-month old rosemary plants was analyzed by quantitative reverse transcriptional polymerase chain reaction (qRT-PCR). Total RNA was isolated

from roots, stems and leaves according to the method as mentioned above. All of RNA samples were treated with HiScript II qRTSuperMix II at 50 °C. The qRT-PCR experiments were performed on a Roche thermal cycler (LightCyler 480 II) in three independent biological replicates. The rosemary actin gene *β-actin* was amplified as an internal standard. All primers for qRT-PCR are shown in Supplementary Table S3.

Extraction of triterpenoids from rosemary tissues and yeast culture*s*

To investigate the accumulation of BA in different tissues of rosemary, fresh plant materials, including roots, stems and leaves, were ground to powders. 1 g of each powder was extracted with 20 mL of THF and ethanol mixture (*v/v*, 1:1) by sonication for 30 min. The extracts were centrifuged at 1000 rpm for 10 min and the supernatants were collected. After repeating the extraction steps above 3 times, the supernatants were combined and concentrated to dryness by rotary evaporation. The residues were silylated for analysis by GC-MS.

Engineered yeast cultures were separated into cells and media by centrifugation. After crushed, cells were extracted with ethyl acetate twice by sonication for 30 min. Media were partitioned by ethyl acetate twice. The cell and medium extraction solutions were pooled together, concentrated to dryness by rotary evaporation and silylated for GC-MS analysis.

Cultivation in bioreactors

Single colonies were inoculated into 100 mL shake flasks containing 15 mL SD medium lacking leucine, uracil, tryptophan and histidine, and incubated at 30°C and 250 rpm for 24 h to an OD_{600} between 2.0 to 3.0. 50% Glycerol was added to culture media to a concentration of 25%, and 1 mL vials of cell suspension were stored at -80°C. Stored cells (1 ml) were thawed and transferred to a 250 mL shake flask containing 15 mL SD medium and incubated at 30°C and 250 rpm for 24 h, which was then transferred to three 1 L shake flasks containing 100 mL fermentation medium, respectively. Cells were cultivated at 30 °C and 250 rpm for about 36 h to 48 h to an OD_{600} between 8.0 to 10.0. Fermentation medium used in this work was based on the medium described previously (Lenihan et al. 2008). All 300 mL culture was inoculated into a 5-liter fermentor (New Brunswick Scientific BioFlo 310) with 3 L fermentation medium. Fermentation was maintained at 30°C and pH 5.0 with addition of NH_4OH. In an initial batch phase, dissolved oxygen (DO) was set to 30% and cascaded to agitation speed (from 300 to 1000 rpm) and air flow rate (from 3 to 20 L/min). After sugar and derived ethanol were consumed, the rise of DO upon 60% triggered pumping of feed medium (600 g/L glucose) to a final concentration of 5 g/L glucose in the fermentor in 3 minutes. As cell density increased, dissolved oxygen was allowed to reach 0% and the spacing between two feeds could be shorten to 30 minutes.

GC-MS analysis of triterpenoids

Each of the extracts above was treated with *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) at 80 °C for 30 min prior to GC-MS analysis. GC-MS analysis was performed in an Agilent 7890B GC instrument equipped with an HP-5MS column and 5977B single quadrupole mass spectrometer in electron ionization (70 eV) mode. Samples (generally 1 μL) were injected in splitless mode at 80 °C and, after a 1 min hold, the temperature raised at 20 °C/min to 300 °C, where it was held for 15 min.

To quantitative analysis of triterpenoid contents, a stock solution with a concentration of 1 mg/mL was accurately prepared for each authentic compound, and 100 μL stock solution was completely silylated. After dried, each silylated sample was resuspended by 1 mL hexane, from which 5-, 10-, 20-, 40- and 80-fold diluted solutions were prepared to construct an external calibration curve by GC-MS. Lupeol correlation coefficient: $y = 16615.5481x - 189.909$, $R^2 = 0.9986$; betulin correlation coefficient: $y = 18793.1521x - 567.7013$, $R^2 = 0.9981$; BA correlation coefficient: $y = 16247.8228x - 9223.9475$, $R^2 = 0.9995$.

Accession codes

The cDNA sequences of the genes involved in this work have been deposited in GenBank (*RoCYP01*: MK592859; *RoCYP02*: MK592860; *RoCYP03*: MK592861; partial *BpCYP01*: MK592862; *RoCPR01*: MK592863; *optiRoCYP01*: MK592864; *OptiAtLUP*: MK592865).

Results

Identification of candidate CYPs capable of oxidizing lupeol into BA

Birch species, such as *B*. *pubescens*, *B. platyphylla*, *B. pendula* and *B. papyrifera*, serve as the major sources of BA. Besides these, BA has been also found in many other plants (Pai and Joshi 2014; Ryu et al. 1992) – e.g. rosemary and selfheal (*Prunella vulgaris*). A well-assembled transcriptome database for rosemary has become available online (http://medicinalplantgenomics.msu.edu), and several versions of raw RNA-Seq datasets of *B. pendula* and selfheal have been released to NCBI as well. As the CYPs responsible for modifying triterpenoid backbones fall into CYP71, CYP72 and CYP85 clans (Thimmappa et al. 2014; Ghosh 2017), examples from each of the relevant (sub)families (Supplementary Table S1) were used as probes to identify homologs in the rosemary transcriptome and raw RNA-Seq data of *B. pendula* and selfheal. Two truncated (RoCYP01 and RoCYP02) and one full-length CYPs (RoCYP03) were found in the rosemary transcriptome; and a truncated CYP (BpCYP01) was assembled from the raw RNA-Seq data available for *B. pendula*. Rapid amplification of cDNA ends (RACE) was

used to obtain full-length clones for RoCYP01 and RoCYP02. Unfortunately, a full-length clone for BpCYP01 was not obtained.

Phylogenetic analysis of the obtained rosemary CYPs demonstrates that these all fall within the CYP85 clan (Fig. 1). More specifically, RoCYP01 clusters with the CYP716A subfamily that has been associated with BA biosynthesis and was assigned as CYP716A155 by Prof. David Nelson (Univ. Tenn.). By contrast, RoCYP02 and RoCYP03 are closer to CYPs that do not act in BA biosynthesis. Accordingly, RoCYP01 was hypothesized to be responsible for oxidizing lupeol into BA in rosemary.

RoCYP01 expression and BA accumulation are correlated across various rosemary tissues

To provide further support for the role of RoCYP01 as a physiologically relevant lupeol C-28 oxidase, its transcript levels were measured in leaves, stems and roots of rosemary by qRT-PCR. The highest *RoCYP01* transcript level was observed in leaves, whereas its transcript levels in stems and roots were about 60% and 80% lower, respectively (Fig. 2). BA accumulation in leaves, stems and roots was also investigated by GC-MS analysis and found to be 28.4, 22.1 and 1.7 μg/g (wet weight of tissues), respectively (Fig. 2). These results provide correlation between BA accumulation and the expression of *RoCYP01*, supporting a role for this in rosemary BA biosynthesis.

Functional characterization of RoCYP01

The utility of a synthetic biology approach towards clarifying enzymatic function in plant terpenoid biosynthesis has been previously demonstrated (Zi and Peters 2013; Zi et al. 2014; Ignea et al. 2016), in particular, via reconstruction of the upstream metabolic pathway to produce putative substrate(s) for the enzyme of interest. Here this approach was enabled by cloning of a lupeol synthase encoding gene from *Arabidopsis thaliana* (AtLUP) into the yeast expression vector pESC-LEU to yield pHJ01, transformation of which into *Saccharomyces cerevisiae* BY4741 led to the lupeol-producing strain HJ001. To enable CYP activity a cytochrome P450 reductase (CPR) encoding gene from *A. thaliana* (AtCPR1) was cloned into pESC-URA, along with RoCYP01 (as controls RoCYP02 and RoCYP03 were similarly cloned), leading to pHJ02 (as well as pHJ03 and pHJ04). These were transformed together with pHJ01 into *S. cerevisiae* BY4741 to obtain yeast strains HJ002-HJ004 (Supplementary Table S5). Unfortunately,

none of these produced BA. Given that CYP/CPR pairs from the same plant species have been shown to exhibit higher efficiency than non-native combinations (Guo et al. 2013), it was hypothesized that use of rosemary CPR might enable activity. Accordingly, characterized plant CPRs (Supplementary Table S2) were used as probes to identify homologs in the rosemary transcriptome. Only one full-length CPR candidate was identified. This RoCPR1 was cloned in place of AtCPR1 in pHJ02-pHJ04 to construct pHJ05-pHJ07 (Supplementary Table S4), respectively. These were co-transformed with pHJ01 into *S. cerevisiae* BY4741 to obtain yeast strains HJ005-HJ007 (Supplementary Table S5), respectively. BA was detected from the extract of the strain HJ005, in which AtLUP was co-expressed with RoCPR1 and RoCYP01 (Fig. 3).

As many CYPs involved in triterpenoid biosynthesis are promiscuous in terms of their substrate specificity (Thimmappa et al. 2014; Ghosh 2017), to further support a role for RoCYP01 its substrate preference also was investigated here. This was accomplished by transforming just pHJ05 for co-expression of RoCPR1 and RoCYP01 to obtain yeast strain HJ009 (Supplementary Table S5), which was then fed lupeol, α-amyrin, β-amyrin or betulin. The results indicated that although RoCYP01 can oxidize α-amyrin and β-amyrin into ursolic acid and oleanolic acid (Figs. 4c and 4d), respectively, it shows the highest efficiency in transformation of lupeol into BA (Fig. 4a). Interestingly, RoCYP01 prefers lupeol to betulin as well (Fig. 4b), suggesting that betulin might be a transient intermediate retained in the active site during the requisite series of reactions. Regardless, these results strongly indicate that RoCYP01 is the lupeol C-28 oxidase involved in rosemary BA biosynthesis.

Optimization of yeast BA production

It has previously been shown that high-level production of plant-derived terpenoids in a microbial host requires elucidation of the relevant enzymes (Paddon et al. 2013). Moreover, it also has been shown that modular pathway engineering is an efficient means to rapidly construct such high-yielding microbial strains for such natural products (Paddon et al. 2013; Wong et al. 2018; Zhou et al. 2012; Ajikumar et al. 2010). Thus, building on identification of RoCYP01 as a BA specific lupeol C-28 oxidase, this approach was taken here to increase yeast BA production levels.

As in previous studies (Dai et al. 2013; Dai et al. 2014; Wang et al. 2018), the upstream steps of the innate (tri)terpenoid biosynthetic pathway in yeast were arbitrarily divided into two modules (Fig. 5).

The first of these corresponds to the mevalonate-dependent (MVA) pathway for production of the universal isoprenoid/terpenoid precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) from the central metabolite acetyl-CoA. This module was optimized by overexpression of all the corresponding enzymes, although a truncated version of HMG (expressing just the catalytic domain) was used as this has been shown to increase flux (Donald et al 1997; Engels et al 2008), as previously described (Dai et al. 2013; Dai et al. 2014; Wang et al. 2018). The second module corresponds to the subsequent production of the common triterpenoid precursor 2,3-oxidosqualene (oxSQ). This module also was optimized as previously described (Wang et al. 2018). Briefly, by overexpression of a squalene synthase and squalene-epoxidase from *A. thaliana* (AtSQS2 and AtSQE2, respectively), as well as an FPP synthase from *Salvia miltiorrhiza* (SmFPS). Together, these optimized MVA and oxSQ modules provide significant increase of triterpenoid production levels (Wang et al. 2018).

For the production of BA specifically, the final module was targeted at improvement of the expression of RoCYP01 and AtLUP. For this purpose, synthetic genes, codon optimized for expression in yeast were obtained (optiRoCYP01 and optiAtLUP). Three expression cassettes, P_{PGK1}-*RoCPR1*-T_{ADH1}, P_{TDH3}-*optiRoCYP01*-T_{TPI1} and P_{TEF1}-*optiAtLUP*-T_{CYC1} (BA module) were constructed according to the previously reported method (Dai et al. 2013), and these were integrated into the *NDT80* site of a yeast strain in which the optimized MVA and oxSQ modules were already present (i.e., by integration into *HIS3* and *GAL7* sites of the yeast chromosome), to yield the strain BA-1 (Supplementary Table S5). After flask fermentation, BA-1 was found to produce 24.7, 99.3 and 21.7 mg/L of lupeol, betulin and BA, respectively (Fig. 6). We then inserted a second copy of the BA module into the *GAL80* site of BA-1 to construct the strain BA-2 (Supplementary Table S5). This led to an approximately ten-fold increase in BA yield (193.5 mg/L), with high yields of lupeol and betulin as well (23.6 and 59.5 mg/L), respectively; Fig. 6). Finally, the BA-2 strain of yeast was cultured in a fed-batch fermentation and, after a 168-h fermentation, titers of 0.1, 1.0 and 1.5 g/L of lupeol, betulin and BA (respectively) were observed. Notably, the co-production of betulin further increases the potential yield, as this can be relatively easily converted to BA (Pezzuto and Kim 1998; Csuk et al. 2006; Tulisalo et al. 2013; Ressmann et al. 2017).

Discussion

BA and its derivatives have exhibited strong potential for development into anticancer and anti-HIV

drugs. However, if this potential is realized it will be necessary to provide large-scale access to a more reliable and less resource intensive source of BA. One promising method for accessing such plant-derived terpenoid natural products is via a synthetic biology approach – i.e., biotechnological production via engineered microbes (Li and Pfeifer 2014; Bian et al 2017). However, as previously demonstrated for high-level production of arteminisic acid (Paddon et al. 2013), this requires identification of the relevant enzymes. While it has been shown that a number of CYP716A subfamily members can oxidize lupeol into trace amounts of BA, their major functions in planta seem to be conversion of α- and β-amyrin into ursolic acid and oleanolic acid, respectively (Fukushima et al. 2011; Huang et al. 2012). More recently, CYP716A180 was reported to selectively catalyze the oxidation of lupeol to BA (Zhou et al. 2016). However, this has not yet led to construction of a high-yielding microbial strain suitable for biotechnological production of BA.

Here publicly available transcriptome data for rosemary was used to rapidly identify RoCYP01 as a potential lupeol C-28 oxidase responsible for the observed production of BA by this plant, including not only phylogenetic analysis (Fig. 1), but also correlated expression (Fig. 2). Building on previous use of a synthetic biology approach to elucidation of plant terpenoid biosynthesis by upstream pathway reconstruction in *Escherichia coli* (Zi and Peters 2013; Zi et al. 2014; Boutanaev et al. 2015), RoCYP01 was investigated by a similar approach in yeast, which is more amenable to triterpenoid production. While requiring use of the rosemary RoCPR1, it was thus possible to demonstrate that RoCYP01 selectively reacts with lupeol, relative to α-amyrin, β-amyrin, or even betulin, which is the alcoholic intermediate in conversion of lupeol to BA (Figs. 3 and 4).

Identification of RoCYP01 as the rosemary lupeol C-28 oxidase for BA biosynthesis via this approach further encouraged construction of a high-yielding strain. This builds on previous work demonstrating not only how flux can be increased to isoprenoids/terpenoids (Dai et al. 2013; Dai et al. 2014; Wang et al. 2018), but triterpenoids more specifically (Wang et al. 2018). In particular, for production of the common triterpenoid precursor 2,3-oxidosequalene. This was coupled to optimization of the genes specific to BA biosynthesis, which led to titers of >1 g/L in batch-fed culture. To the best of our knowledge, this is the highest BA yield so far achieved by microbial fermentation. Accordingly, this work provides a good example of the rapid exploitation of the increasing amounts of sequence information available for plants to not only provides insights into biosynthetic pathways, but also facile synthetic biology approaches towards enabling biotechnological production of valuable natural products

such as BA and its derivatives.

Electronic supplementary material The online version of this article (https://doi.org/) contains supplementary material, which is available to authorized users

Acknowledgments We thank Prof. David R. Nelson for annotation of CYP716A155.

Funding information This work was supported by grants from the National Natural Science Foundation of China (NSFC, No. 81673530), China Postdoctoral Science Foundation (2018M633289) and Open Fund of Guangdong Province Key Laboratory of Pharmacodynamic Constituents of TCM and New Drugs Research (No. 2016B 030301004).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science 330: 70–74. doi: 10.1126/science.1191652

Bian G, Deng Z, Liu T (2017) Strategies for terpenoid overproduction and new terpenoid discovery. Curr Opin Biotechnol 48: 234–241. doi: 10.1016/j.copbio.2017.07.002

Boachon B, Buell CR, Crisovan E, Dudareva N, Garcia N, Godden G, Henry L, Kamileen MO, Kates

HR, Kilgore MB, Lichman BR, Mavrodiev EV, Newton L, Rodriguez-Lopez C, O'Connor SE, Soltis D, Soltis P, Vaillancourt B, Wiegert-Rininger K, Zhao D (2018) Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. Mol Plant 11: 1084–1096. doi: 10.1016/j.molp.2018.06.002

Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, Osbourn A (2015) Investigation of terpene diversification across multiple sequenced plant genomes. Proc Natl Acad Sci USA 112: E81–E88. doi: 10.1073/pnas.1419547112

Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, Graham N, Odoardi M, Piano E, Arcioni S, May S, Scotti C, Calderini O (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase invoved in the biosynthesis of hemolytic saponins. Plant Cell 23: 3070–3081. doi: 10.1105/tpc.111.087312

Chintharlapalli S, Papineni S, Ramaiah SK, Safe S (2007) Betulinic acid inhibits prostate cancer growth through inhibition of specificity protein transcription factors. Cancer Res 67: 2816–2823. doi: 10.1158/0008-5472.CAN-06-3735

Csuk R, Schmuck K, Schäfer R (2006) A practical synthesis of betulinic acid. Tetrahedron Lett. 47, 8769–8770. doi: org/10.1016/j.tetlet.2006.10.004

Dai Z, Liu Y, Zhang X, Shi M, Wang B, Wang D, Huang L, Zhang X (2013) Metabolic engineering of *Saccharomyces cerevisiae* for production of ginsenosides. Metab Eng 20: 146–156. doi: 10.1016/j.ymben.2013.10.004

Dai Z, Wang B, Liu Y, Shi M, Wang D, Zhang X, Liu T, Huang L, Zhang X (2014) Producing aglycons of ginsenosides in bakers' yeast. Sci Rep 4: 3698. doi: 10.1038/srep03698

Dang Z, Ho P, Zhu L, Qian K, Lee KH, Huang L, Chen CH (2013) New betulinic acid derivatives for bevirimat-resistant human immunodeficiency virus type-1. J Med Chem 56: 2029–2037. doi:

10.1021/jm3016969

Dewick PM (2002) Medicinal Natural Products: A Biosynthetic Approach, 2nd edn. John Wiley & Sons, New York, pp 167–290.

Donald KAG, Hampton RY, Fritz IB (1997) Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. Appl Environ Microbiol 63: 3341−3344.

Engels B, Dahm P, Jennewein S (2008) Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng 10: 201–206. doi: 10.1016/j.ymben.2008.03.001

Fukushima EO, Seki H, Ohyama K, Ono E, Umemoto N, Mizutani M, Saito K, Muranaka T (2011) CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. Plant Cell Physiol 52: 2050–2061. doi: 10.1093/pcp/pcr146

Ghosh S (2017) Triterpene structural diversification by plant cytochrome P450 enzymes. Front Plant Sci 8: 1886. doi: 10.3389/fpls.2017.01886

Guo J, Zhou YJ, Hillwig ML, Shen Y, Yang L, Wang Y, Zhang X, Liu W, Peters RJ, Chen X, Zhao ZK, Huang L (2013) CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. Proc Natl Acad Sci USA 110: 12108–12113. doi: 10.1073/pnas.1218061110

Huang L, Li J, Ye H, Li C, Wang H, Liu B, Zhang Y (2012) Molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in *Catharanthus roseus*. Planta 236: 1571−1581. doi: 10.1007/s00425-012-1712-0

Ignea C, Athanasakoglou A, Ioannou E, Georgantea P, Trikka FA, Loupassaki S, Roussis V, Makris AM,

Kampranis SC (2016) Carnosic acid biosynthesis elucidated by a synthetic biology platform. Proc Natl Acad Sci USA 113: 3681–3686. doi: 10.1073/pnas.1523787113

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275–282. doi: 10.1093/bioinformatics/8.3.275

Khakimov B, Kuzina V, Erthmann PØ, Fukushima EO, Augustin JM, Olsen CE, Scholtalbers J, Volpin H, Andersen SB, Hauser TP, Muranaka T, Bak S (2015) Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transient production of saponins in *Nicotiana benthamiana*. Plant J 84, 478–490. doi: 10.1111/tpj.13012

Krasutsky PA, Carlson RM, Nesterenko VV, Kolomitsyn IM, Edwardson CF (2003) Birch bark processing and the isolation of natural products from birch bark. US patent No. US 6,634,575 B2. 2003-10-21.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33: 1870–1874. doi: 10.1093/molbev/msw054

Lenihan JR, Tsuruta H, Diola D, Renninger NS, Regentin R (2008) Developing an industrial artemisinic acid fermentation process to support the cost-effective production of antimalarial artemisinin-based combination therapies. Biotechnol Prog 24: 1026–1032. doi: 10.1002/btpr.27

Li F, Goila-Gaur R, Salzwedel K, Kilgore NR, Reddick M, Matallana C, Castillo A, Zoumplis D, Martin DE, Orenstein JM, Allaway GP, Freed EO, Wild CT (2003) PA-457: A potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. Proc Natl Acad Sci USA 100: 13555–13560. doi: 10.1073/pnas.2234683100

Li M, Kildegaard KR, Chen Y, Rodriguez A, Borodina I, Nielsen J (2015) De novo production of resveratrol from glucose or ethanol by engineered *Saccharomyces cerevisiae*. Metab Eng 32: 1–11. doi: 10.1016/j.ymben.2015.08.007

Li Y, Pfeifer BA (2014) Heterologous production of plant-derived isoprenoid products in microbes and the application of metabolic engineering and synthetic biology. Curr Opin Plant Biol 19: 8–13. doi: 10.1016/j.pbi.2014.02.005

Liu X, Jutooru I, Lei P, Kim K, Lee S, Brents LK, Prather PL, Safe S (2012) Betulinic acid targets YY1 and ErbB2 through cannabinoid receptor-dependent disruption of microRNA-27a:ZBTB10 in breast cancer. Mol Cancer Ther 11: 1421–1431. doi: 10.1158/1535-7163.MCT-12-0026

Majeed R, Hamid A, Sangwan PL, Chinthakindi PK, Koul S, Rayees S, Singh G, Mondhe DM, Mintoo MJ, Singh SK, Rath SK, Saxena AK (2014) Inhibition of phosphotidylinositol-3 kinase pathway by a novel naphthol derivative of betulinic acid induces cell cycle arrest and apoptosis in cancer cells of different origin. Cell Death Dis 5: e1459. doi: 10.1038/cddis.2014.387

Mertens-Talcott SU, Noratto GD, Li X, Angel-Morales G, Bertoldi MC, Safe S (2013) Betulinic acid decreases ER-negative breast cancer cell growth in vitro and in vivo: role of Sp transcription factors and microRNA-27a:ZBTB10. Mol Carcinog 52: 591–602. doi: 10.1002/mc.21893

Pal A, Ganguly A, Chowdhuri S, Yousuf M, Ghosh A, Barui AK, Kotcherlakota R, Adhikari S, Banerjee R (2015) Bis-arylidene oxindole−betulinic acid conjugate: a fluorescent cancer cell detector with potent anticancer activity. ACS Med Chem Lett 6: 612–616. doi: 10.1021/acsmedchemlett.5b00095

Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Fleck M, Bajad S, Dang G, Dengrove D, Diola D, Dorin G, Ellens KW, Fickes S, Galazzo J, Gaucher SP, Geistlinger T, Henry R, Hepp M, Horning T, Iqbal T, Jiang H, Kizer L, Lieu B, Melis D, Moss N, Regentin R, Secrest S, Tsuruta H, Vazquez R, Westblade LF, Xu L, Yu M, Zhang Y, Zhao L, Lievense J, Covello PS, Keasling JD, Reiling KK, Renninger NS, Newman JD (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496: 528–532. doi: 10.1038/nature12051

Pai SR, Joshi RK (2014) Distribution of betulinic acid in plant kingdom. Plant Science Today 1, 103–107. doi: 10.14719/pst.2014.1.3.58

Pezzuto JM, Kim DSHL (1998) Improved methods of manufacturing betulinic acid. PCT patent No. WO 98/43936. 1998-10-08.

Pisha E, Chai H, Lee IS, Chagwedera TE, Farnsworth NR, Cordell GA, Beecher CW, Fong HH, Kinghorn AD, Brown DM, Wani MC, Wall ME, Hieken TJ, Das Gupta TK, Pezzuto JM (1995) Discovery of betulinic acid as a selective inhibitor of human melanoma that functions by induction of apoptosis. Nat Med 1: 1046–1051.

Qian K, Kuo RY, Chen CH, Huang L, Morris-Natschke SL, Lee KH (2010) Anti-AIDS agents 81. Design, synthesis, and structure-activity relationship study of betulinic acid and moronic acid derivatives as potent HIV maturation inhibitors. J Med Chem 53: 3133–3141. doi: 10.1021/jm901782m

Razboršek MI, Vončina DB, Doleček V, Vončina E (2007) Determination of major phenolic acids, phenolic diterpenes and triterpenes in rosemary (*Rosmarinus officinalis* L.) by gas chromatography and mass spectrometry. Acta Chim Slov 54: 60–67.

Ressmann A, Kremsmayr T, Gaertner P, Zirbs R, Bica K (2017) Toward a benign strategy for the manufacturing of betulinic acid. Green Chem 19: 1014–1022. doi: 10.1039/C6GC02641A

Ryu SY, Lee CK, Lee CO, Kim HS, Zee OP (1992) Antiviral triterpenes from *Prunella vulgaris*. Arch Pharm Res 15: 242–245.

Siddiqui N, Aeri V (2016) Optimization of betulinic acid extraction from *Tecomella undulata* bark using a Box-Behnken design and its densitometric validation. Molecules 21: 393–404. doi: 10.3390/molecules21040393

Šiman P, Filipová A, Tichá A, Niang M, Bezrouk A, Havelek R (2016) Effective method of purification

of betulin from birch bark: the importance of its purity for scientific and medicinal use. Plos one 11: e0154933. doi: 10.1371/journal.pone.0154933

Tamura K, Seki H, Suzuki H, Kojoma M, Saito K, Muranaka T (2017) CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of *Glycyrrhiza uralensis*. Plant Cell Rep 36: 437–445. doi: 10.1007/s00299-016-2092-x

Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A (2014) Triterpene biosynthesis in plants. Annu Rev Plant Biol 65: 225–257. doi: 10.1146/annurev-arplant-050312-120229

Tulisalo J, Pirttimaa M, Alakurtti S, Ylikauhaluoma J, Koskimies S (2013) Method for preparation of betulinic acid. PCT patent No. WO 2013/038314 A1. 2013-03-21.

Wang D, Liu Y, Xu J, Wang J, Dai Z, Zhang X, Huang L (2018) Construction of efficient yeast cell factories for production of ginsenosides precursor dammarenediol-II. Acta Pharm. Sin. 53: 1233−1241. doi: 10.16438/j.0513-4870.2018-0503

Wong J, de Rond T, d'Espaux L, van der Horst C, Dev I, Rios-Solis L, Kirby J, Scheller H, Keasling J (2018) High-titer production of lathyrane diterpenoids from sugar by engineered *Saccharomyces cerevisiae*. Metab Eng 45: 142–148. doi: 10.1016/j.ymben.2017.12.007

Ye Y, Zhang T, Yuan H, Li D, Lou H, Fan P (2017) Mitochondria-targeted lupane triterpenoid derivatives and their selective apoptosis-inducing anticancer mechanisms. J Med Chem 60: 6353–6363. doi: 10.1021/acs.jmedchem.7b00679

Zhao Y, Gu Q, Morris-Natschke SL, Chen CH, Lee KH (2016) Incorporation of privileged structures into bevirimat can improve activity against wild-type and bevirimat-resistant HIV-1. J Med Chem 59: 9262–9268. doi: 10.1021/acs.jmedchem.6b00461

Zhou C, Li J, Li C, Zhang Y (2016) Improvement of betulinic acid biosynthesis in yeast employing

multiple strategies. BMC Biotechnol 16: 59. doi: 10.1186/s12896-016-0290-9

Zhou YJ, Gao W, Rong Q, Jin G, Chu H, Liu W, Yang W, Zhu Z, Li G, Zhu G, Huang L, Zhao ZK (2012) Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. J Am Chem Soc 134: 3234–3241. doi: 10.1021/ja2114486

Zi J, Peters RJ (2013) Characterization of CYP76AH4 clarifies phenolic diterpenoid biosynthesis in the Lamiaceae. Org Biomol Chem 11: 7650–7652. doi: 10.1039/c3ob41885e

Zi J, Matsuba Y, Hong YJ, Jackson AJ, Tantillo DJ, Pichersky E, Peters RJ (2014) Biosynthesis of lycosantalonol, a *cis*-prenyl derived diterpenoid. J Am Chem Soc 136: 16951–16953. doi: 10.1021/ja508477e

Figure Legends

Figure 1. Phylogenetic analysis of the CYPs cloned from rosemary. RoCYP01, RoCYP02 and RoCYP03, as well as the CYPs involved in biosynthesis of plant triterpenoids (Thimmappa et al. 2014; Supplementary Table S1) were included to construct the tree via maximum likelihood analysis. The blue ones are the CYPs which can oxidize lupeol into BA, and the red ones are the CYPs identified from rosemary in this study. The bluish, blue, fuchsia and green shaded CYPs belong to CYP71, CYP72, CYP51 and CYP85 subfamilies, respectively.

Figure 2. Correlation between expression patterns of RoCYP01 and BA accumulation in rosemary tissues. Green bars: the mRNA levels of *RoCYP01* relative to those of the rosemary actin gene *β-actin* across leaves, stems and roots of rosemary, respectively; blue bars: BA accumulation across leaves, stems and roots of rosemary, respectively. The error bars indicate the SD from triplicate biological replicates

Figure 3. GC-MS analysis of the extracts of the engineered yeasts. (a) and (b) are the extracts of HJ005 expressing AtLUP, RoCYP01 and RoCPR1, and HJ008 expressing AtLUP and RoCPR1 (Supplementary Table S5), respectively. (c), (d) and (e) are the authentic compounds lupeol, betulin and BA. (f), (g) and (h) are the mass spectra of .lupeol, betulin and BA, respectively. All the samples are silylated before GC-MS analysis.

Figure 4. Investigation of substrate selectivity of RoCYP01. GC-MS analysis of the extracts of the yeast containing pHJ05 incubated with lupeol (a), betulin (b), α-amyrin (c) and β-amyrin (d), respectively. (e), (f), (g) and (h) are the mass spectra of α-amyrin, β-amyrin, ursolic acid (UA) and oleanolic acid (OA), respectively. All the samples are silylated before GC-MS analysis. (i) is the scheme which illustrates that RoCYP01 catalyzing oxidation reactions, using lupeol, α-amyrin and β-amyrin as substrates, respectively.

Figure 5. Biosynthetic pathway of BA in the engineered yeasts.

Figure 6. The yields of lupeol, betulin and BA of BA-1 and BA-2. BA-1 and BA-2 possess one and

two copies of P_{PGK1}-*RoCPR01*-T_{ADH1}-P_{TDH3}-*optiRoCYP01*-T_{TPI1}-P_{TEF1}-*optiAtLUP*-T_{CYC1}, respectively (Supplementary Table S5). The error bars indicate the SD from three independent fermentation experiments.

Applied Microbiology and Biotechnology

Identification of RoCYP01 (CYP716A155) enables construction of engineered yeast for high-yield production of betulinic acid

Jiajian Huang[1], Wenlong Zha[1], Tianyue An[1], Hua Dong[1], Ying Huang[2], Dong Wang[2], Rongmin Yu[1], Lixin Duan[3,4], Xueli Zhang[2], Reuben J. Peters[5], Zhubo Dai[2], Jiachen Zi[1]

[1] Biotechnological Institute of Chinese Materia Medic, Jinan University, 601 Huangpu Avenue West, Guangzhou 510632, China

[2] Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

[3] Guangdong Province Key Laboratory of Pharmacodynamic Constituents of Traditional Chinese Medicine and New Drugs Research, Jinan University, Guangzhou 510632, China

[4] Joint Laboratory for Translational Cancer Research of Chinese Medicine of the Ministry of Education of the People's Republic of China, International Institute for Translational Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

[5] Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, Iowa 50011, United States

To whom correspondence may be addressed.

Reuben J. Peters:

rjpeters@iastate.edu

Zhubo Dai:

dai_zb@tib.cas.cn

Jiachen Zi:

jiachen_zi@163.com

Contents:

Table S1 The reference CYPs from plants

Table S2. The reference CPRs from plants

Table S3. The primers used in this study

Table S4. The plasmids involved in this study

Table S5. The yeast strains involved in this study

The sequences of RoCYP01, RoCYP02, RoCYP03, BpCYP01, RoCPR01, OptiRoCYP01 and OptiAtLUP

Table S1. The reference CYPs from plants

Clan	Gene	Species	GenBank ID	Substrate	Function
CYP71	CYP71A16	*Arabidopsis thaliana*	NP_199073	Marneral, marnerol	C-23 hydroxylation
	CYP71D353	*Lotus japonicus*	KF460438	Dihydrolupeol, 20-hydroxylupeol	C-20 hydroxylation, C-28 oxidation (three-step oxidation)
	CYP93E1	*Glycine max*	BAE94181	β-Amyrin, sophoradiol	C-24 hydroxylation
	CYP93E2	*Medicago truncatula*	ABC59085	β-Amyrin	C-24 hydroxylation
	CYP93E3	*Glycyrrhiza uralensis*	BAG68930	β-Amyrin	C-24 hydroxylation
	CYP705A1	*Arabidopsis thaliana*	NP_193268	Arabidiol	C-15–C-16 cleavage
	CYP705A5	*Arabidopsis thaliana*	Q9FI39	7β-Hydroxythalianol	C-15–C-16 desaturation
CYP72	CYP72A61v2	*Medicago truncatula*	BAL45199	24-Hydroxy-β-amyrin	C-22 hydroxylation
	CYP72A63	*Medicago truncatula*	BAL45200	β-Amyrin	C-30 hydroxylation, C-30 oxidation (three-step oxidation)
	CYP72A154	*Glycyrrhiza uralensis*	BAL45207	β-Amyrin, 11-oxo-β-amyrin	C-30 hydroxylation (β-amyrin); C-30,C-22, and C-29 Oxidation (11-oxo-β-amyrin)
	CYP72A67	*Medicago truncatula*	ABC59075	Oleanolic acid	C-2β hydroxylation
	CYP72A68v2	*Medicago truncatula*	BAL45204	Oleanolic acid	C-23 hydroxylation (three-step oxidation)
CYP51	CYP51H10	*Avena strigosa*	ABG88961	β-Amyrin	C-16 hydroxylation, C-12–C-13 epoxidation
CYP85	CYP88D6	*Glycyrrhiza uralensis*	BAG68929	β-Amyrin, 30-hydroxy-amyrin	C-11 oxidation (two-step oxidation)
	CYP87D16	*Maesa lanceolata*	AHF22090	β-Amyrin	C-16α Oxidation
	CYP708A2	*Arabidopsis thaliana*	Q8L7D5	Thalianol	C-7 hydroxylation
	CYP716A47	*Panax ginseng*	AEY75217	Dammarenediol II	C-12 hydroxylation
	CYP716Y1	*Bupleurum falcatum*	AHF45909	β-Amyrin, α-amyrin	C-16α Oxidation
	CYP716A141	*Platycodon grandiflorus*	BAX04008	β-Amyrin	C-16β Oxidation
	CYP716A53v2	*Panax ginseng*	AFO63031	Protopanaxadiol	C-6 hydroxylation
	CYP716A111	*Aquilegia coerulea*	APG38190	β-Amyrin	C-16β Oxidation
	CYP716A80	*Barbarea vulgaris*	ALR73782	β-Amyrin, α-amyrin, lupeol	C-28 oxidation (three-step oxidation)
	CYP716A81	*Barbarea vulgaris*	ALR73781	β-Amyrin, α-amyrin, lupeol	C-28 oxidation (three-step oxidation)

	CYP716A1	*Arabidopsis thaliana*	AED94045	β-Amyrin, α-amyrin, lupeol	C-28 oxidation
	CYP716A2	*Arabidopsis thaliana*	AED94048	β-Amyrin, α-amyrin, lupeol	C-16,C-22α,C-28 Oxidation
	CYP716A12	*Medicago truncatula*	ABC59076	β-Amyrin, α-amyrin, lupeol	C-28 oxidation (three-step oxidation)
	CYP716A179	*Glycyrrhiza uralensis*	BAW34647	β-Amyrin, α-amyrin, lupeol	C-28 oxidation (three-step oxidation)
	CYP716A14v2	*Artemisia annua*	AHF22083	β-amyrin, α-amyrin, lupeol	C-3 oxidation
	CYP716A140v2	*Platycodon grandiflorus*	BAX04007	β-amyrin	C-28 oxidation (three-step oxidation)
	CYP716A75	*Maesa lanceolata*	AHF22088	β-Amyrin	C-28 oxidation (three-step oxidation)
	CYP716A52v2	*Panax ginseng*	AFO63032	β-amyrin	C-28 oxidation (three-step oxidation)
	CYP716A244	*Eleutherococcus senticosus*	APZ88353	β-Amyrin	C-28 oxidation (three-step oxidation)
	CYP716AL1	*Catharanthus roseus*	AEX07773	β-Amyrin, α-amyrin, lupeol	C-28 oxidation (three-step oxidation)
	CYP716A78	*Chenopodium quinoa*	ANY30853	β-Amyrin	C-28 oxidation (three-step oxidation)
	CYP716A79	*Chenopodium quinoa*	ANY30854	β-Amyrin	C-28 oxidation (three-step oxidation)
	CYP716A15	*Vitis vinifera*	BAJ84106	β-Amyrin, α-amyrin, lupeol	C-28 oxidation (three-step oxidation)
	CYP716A17	*Vitis vinifera*	BAJ84107	β-Amyrin	C-28 oxidation (three-step oxidation)

Table S2. The reference CPRs from plants

CPRs	species	GenBank ID
Ob-CPR1	*Ocimum basilicum*	KX192385
Pc-CPR	*Perilla citriodora*	LC325999
Pf-CPR	*Perilla frutescens*	GQ120439
Sm-CPR1	*Salvia miltiorrhiza*	FR693803
Pk-CPR	*Picrorhiza kurrooa*	JN968968
Sd-CPR	*Scoparia dulcis*	KF306080
Sm-CPR2	*Salvia miltiorrhiza*	JX848592
Ob-CPR2	*Ocimum basilicum*	KX192386
Si-CPR1	*Sesamum indicum*	LC209223
AtCPR1	*Arabidopsis thaliana*	NM_118585

Table S3. The primers used in this study

Name	Sequence	Description
For RACE experiments		
RoCYP01-5'end-1	GGCCAGGTCGGCCTTCCGCTGCTTGATG	The gene specific primer of first round PCR for amplification of 5'-end of *RoCYP01*
RoCYP01-5'end-2	GGTCCATGATGCCAACGTAGCGGTGGAG	The gene specific primer of second round PCR for amplification of 5'-end of *RoCYP01*
RoCYP01-3'end-1	GGACGCCGTTCCACAAGGGGATCAAGGC	The gene specific primer of first round PCR for amplification of 3'-end of *RoCYP01*
RoCYP01-3'end-2	GCTCGTCGCCATCATCAAGCAGCGGAAG	The gene specific primer of second round PCR for amplification of 3'-end of *RoCYP01*
RoCYP02-5'end-1	CGTTCTCCAAGGCTCTTCAGCAGCATCG	The gene specific primer of first round PCR for amplification of 5'-end of *RoCYP02*
RoCYP02-5'end-2	GCAACGCTGGACGTGTCGTGGCCTGC	The gene specific primer of second round PCR for amplification of 5'-end of *RoCYP02*
RoCPR01-3'end-1	GCCACCTCCATTCCCCCCTTGCACTTTG	The gene specific primer of first round PCR for amplification of 3'-end of *RoCPR01*
RoCPR01-3'end-2	GTTCCCGTCCGCCAAGCCTTCTCTAGG	The gene specific primer of second round PCR for amplification of 3'-end of *RoCPR01*
For gene cloning		

EcoRI-RoCYP01-f	GAATTCATGGAGTTCTTCTATGCTTCCCTC	The forward primer for cloning of *RoCYP01*
NotI-RoCYP01-r	GCGGCCGCTTAAGAAGTGTGGGGGTAGAGTCG	The reverse primer for cloning of *RoCYP01*
EcoRI-RoCYP02-f	GAATTCATGGCTGAAATACTTTTGATGC	The forward primer for cloning of *RoCYP02*
NotI-RoCYP02-r	GCGGCCGCCTATGTCTGTGGTTGGATAAGAATG	The reverse primer for cloning of *RoCYP02*
EcoRI-RoCYP03-f	GAATTCATGGAGCTCCTAACGTTAGCTG	The forward primer for cloning of *RoCYP03*
ClaI-RoCYP03-r	ATCGATTCACTGTTTGTAAAGGCGGAC	The reverse primer for cloning of *RoCYP03*
BamHI-RoCPR01-f	GGATCCATGGAACCCTCGTCGCCGAAGCTC	The forward primer for cloning of *RoCPR01*
KpnI-RoCPR01-r	GGTACCTCACCATACATCGCGCAAATATCTTCCTG	The reverse primer for cloning of *RoCPR01*
SrfI-ATR1-f	GCCCGGGCATGACTTCTGCTTTGTATGCTTCCG	The forward primer for cloning of ATR1
KpnI-ATR1-r	GGTACCTCACCAGACATCTCTGAGGTATCTTCC	The reverse primer for cloning of *ATR1*
BamHI-AtLUP-f	GGATCCATGTGGAAGTTGAAGATAGGAAAGG	The forward primer for cloning of *AtLUP*
KpnI-AtLUP -r	GGTACCTTAATTAACGATAAACACAACTTTTCG	The reverse primer for cloning of *AtLUP*

For qRT-PCR

qRT-RoCYP01-f	CCTTCAGGGTGCATTCAGAGAAGC
qRT-RoCYP01-r	CGAACCTCGACGGATCGAACTTC
qRT-β-actin-f	GAAGGTTATGCACGCCCTCACG
qRT-β-actin-r	CAATTTCCCGCTCTGCAGTGG

For construction of expression cassettes

SexAI-optiRoCYP01-f	GCGACCTGGTAAAACAATGGAATTTTTCTATGCATC	For cloning *optiRoCYP01* into pM4-AtSQE2
optiRoCYP01-AscI-r	GCGGGCGCGCCTTAAGAAGTATGTGGGTATAATC	For cloning *optiRoCYP01* into pM4-AtSQE2
SexA1-RoCPR01-f	GCGACCTGGTAAAACAATGGAACCCTCGTCGCC	For cloning *RoCPR01* into pM2-tHMG1
RoCPR01-AscI-r	GCGGGCGCGCCTCACCATACATCGCGCAAATATC	For cloning *RoCPR01* into pM2-tHMG1
SexAI-optiAtLUP-f	GCGACCTGGTAAAACAATGTGGAAGTTGAAGATAGG	For cloning *optiRoAtLUP* into pM3-SynPgPPD
optiAtLUP-AscI-r	GCGGGCGCGCCTTAGTAGGAGTGAGCACATAAAAC	For cloning *optiRoAtLUP* into pM3-SynPgPPD

Primers used in DNA assembly

1-M-pEASY-PGK1-F	CTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCTTAAT

	TAAACGCACAGATATTATAA C
3G-1-M-ADHt-TDH3-R	CCTCCGCGTCATTAAACTTC TTGTTGTTGACGCTAACATT CAACGCTAGTATTCGGCAT GCCGGTAGAGGTGTGG
3G-3-M-ADHt-TDH3-F	CAGGTATAGCATGAGGTCG CTCTTATTGACCACACCTCT ACCGGCATGCCGAATACTA GCGTTGAATGTTAGCGTC
3G-3-M-TPI1t-TEF1-R	AGGAGTAGAAACATTTTGA AGCTATGGTGTGTGGGGGA TCACTTTAATTAA TCTATATAACAGTTGAAATT TGGA
3G-2-M-TPI1t-TEF1-F	GTCATTTTCGCGTTGAGAA GATGTTCTTATCCAAATTTC AACTGTTATATAGATTAATT AAAGTGATCCCCCACAC
M-CYC1t-pEASY-R	CGTATTACAATTCACTGGCC GTCGTTTTACAACGTCGTG ACTGGGAAAACCCTGGCGC GTTGGCCGATTCATTAATGC
X1-M-pEASY-r-t-F	CTTGCAAATGCCTATTGTGC AGATGTTATAATATCTGTGC GTTTAATTAAGGCTCGTATG TTGTGTGGAATTGT
X2-M-pEASY-r-t-R	CGAAGGCTTTAATTTGCAA GCTGCGGCCCTGCATTAAT GAATCGGCCAACGCGCCAG

	GGTTTTCCCAGTCACGACG TTG
NDT80-interg-1	CATCATAAGGAATTCCGGG ATTCTCCCCAT
NDT80-interg-2	CTGGCTTTAAAAAATGGAT AAAAAGGGATG
Gal80-1-interg-1	GGTTACCAGATCTACACCG TTCCCGATTTCA
Gal80-2-interg-2	ACAACATTTGGTCACTAAA TCGATATTTTAC
X1-M-pEASY-r-t-F	CTTGCAAATGCCTATTGTGC AGATGTTATAATATCTGTGC GTTTAATTAAGGCTCGTATG TTGTGTGGAATTGT
X2-M-pEASY-r-t-R	CGAAGGCTTTAATTTGCAA GCTGCGGCCCTGCATTAAT GAATCGGCCAACGCGCCAG GGTTTTCCCAGTCACGACG TTG

Table S4. The plasmids involved in this study

Name	Description	Source
pEASY-Blunt		TransGen Biotech
p-RoCYP01	Clone *EcoRI-RoCYP01-NotI* into pEASY-Blunt	This study
p-RoCYP02	Clone *EcoRI-RoCYP02-NotI* into pEASY-Blunt	This study
p-RoCYP03	Clone *EcoRI-RoCYP02-ClaI* into pEASY-Blunt	This study
p-RoCPR01	Clone *BamHI-RoCYP02-KpnI* into pEASY-Blunt	This study
p-ATR1	Clone *srfI-ATR1-KpnI* into pEASY-Blunt	This study
p-AtLUP	Clone *BamHI-AtLUP-KpnI* into pEASY-Blunt	This study
pM2-tHMG1	Clone P_{PGK1}-*SexAI-tHMG1-AscI-*T_{ADH1} into pEASY-Blunt	(Dai et al., 2013)
pM2-RoCPR01	Clone *RoCPR01* into the SexAI/AscI site of pM2-tHMG1	This study
pM3-synPgPPD	Clone P_{TEF1}-*SexAI-synPgPPD-AscI-*T_{CYC1} into pEASY-Blunt	(Dai et al., 2013)
pM3-optiAtLUP	Clone *optiAtLUP* into the SexAI/AscI site of pM3-synPgPPD	This study
pM4-AtSQE2	Clone P_{TDH3}-*SexAI-AtSQE2-AscI-*T_{TPI1} into pEASY-Blunt	(Wang et al., 2018)
pM4-optiRoCYP01	Clone *optiRoCYP01* into the SexAI/AscI site of pM4-AtSQE2	This study

pEASY-Blunt simple		TransGen Biotech
pNDT80-HIS	Clone *NDT80* and *HIS3* marker into pEASY-Blunt simple	(Wang et al., 2018)
pGal80-LEU	Cloning GAL80 and LEU2 marker into pEASY-Blunt simple	(Wang et al., 2018)
pESC-LEU		
pHJ01	Clone *AtLUP* into the BamHI/KpnI site of pESC-LEU	This study
pESC-URA		
pESC-URA-ATR1	ATR1 in the SrfI/KpnI site of pESC-URA	This study
pHJ02	*RoCYP01* is cloned into the EcoRI/NotI site of pESC-URA-ATR1	This study
pHJ03	*RoCYP02* is cloned into the EcoRI/NotI site of pESC-URA-ATR1	This study
pHJ04	*RoCYP03* is cloned into the EcoRI/ClaI site of pESC-URA-ATR1	This study
pESC-URA-RoCPR01	*RoCPR01* in the BamHI/KpnI site of pESC-URA	This study
pHJ05	*RoCYP01* is cloned into the EcoRI/NotI site of pESC-URA-RoCPR01	This study
pHJ06	*RoCYP02* is cloned into the EcoRI/NotI site of pESC-URA-RoCPR01	This study
pHJ07	*RoCYP03* is cloned into the EcoRI/ClaI site of pESC-URA-RoCPR01	This study

Table S5. The yeast strains involved in this study

Name	Description	Source
S. cerevisiae **BY4741**	*MATα, his3Δ1*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*	ATCC
S. cerevisiae **CEN.PK2-1D**	*MATα*, *ura3-52*, *leu2-3_112*, *trp1-289*, *his3Δ*, *MAL2-8c*, *SUC2*	This lab collection
HJ001	*S. cerevisiae* BY4741 with pHJ01	This study[a]
HJ002	*S. cerevisiae* BY4741 with pHJ01 and pHJ02	This study[a]
HJ003	*S. cerevisiae* BY4741 with pHJ01 and pHJ03	This study[a]
HJ004	*S. cerevisiae* BY4741 with pHJ01 and pHJ04	This study[a]
HJ005	*S. cerevisiae* BY4741 with pHJ01 and pHJ05	This study[a]
HJ006	*S. cerevisiae* BY4741 with pHJ01 and pHJ06	This study[a]
HJ007	*S. cerevisiae* BY4741 with pHJ01 and pHJ07	This study[a]
HJ008	*S. cerevisiae* BY4741 with pHJ01 and pESC-URA-RoCPR01	This study[a]
HJ009	*S. cerevisiae* BY4741 with pHJ05	This study[a]
WD-2091	*S. cerevisiae* CEN.PK2-1D, *HIS3::TRP1-*P_{PGK1}*-AtSQS2-*T_{ADH1}*-*P_{TDH3}*-AtSQE2-*T_{TPI1}*-*P_{TEF1}*-SmFPS-*T_{CYC1} *Gal7::URA3-*P_{PGK1}*-tHMG1-*T_{ADH1}*-*P_{PDC1}*-ERG12-*T_{ADH2}*-*P_{ENO2}*-IDI1-T-*$_{PDC1}$*-*P_{PYK1}*-ERG19-*T_{PGI1}*-*P_{FBA1}*-ERG13-*T_{TDH2}*-*P_{TDH3}*-ERG8-*T_{TPI1}*-*P_{TEF1}*-ERG10-*T_{CYC1}	(Wang et al., 2018)[b]
BA-1	WD-2091, *NDT80::HIS3-*P_{PGK1}*-RoCPR01-*T_{ADH1}*-*P_{TDH3}*-optiRoCYP01-*T_{TPI1}*-*P_{TEF1}*-optiAtLUP-*T_{CYC1}	This study[b]
BA-2	BA-1, *GAL80::LEU2-*P_{PGK1}*-RoCPR01-*T_{ADH1}*-*P_{TDH3}*-optiRoCYP01-*T_{TPI1}*-*P_{TEF1}*-optiAtLUP-*T_{CYC1}	This study[b]

[a] All these yeast strains are deposited in Microbiolagical Culture Collection Center of Jinan University, the collection numbers are JNU-JZ-HJ001, JNU-JZ-HJ002, JNU-JZ-HJ003, JNU-JZ-HJ004, JNU-JZ-HJ005, JNU-JZ-HJ006, JNU-JZ-HJ007, JNU-JZ-HJ008 and JNU-JZ-HJ009;
[b] All these yeast strains are deposited in Microbiolagical Culture Collection Center of Tianjin Institute of Industrial Biotechnology, the collection numbers are WD-2091, WD-2091-BA-1 and WD-2091-BA-2.

References

Dai, Z., Liu, Y., Zhang, X., Shi, M., Wang, B., Wang, D., Huang, L., Zhang, X., 2013. Metabolic engineering of *Saccharomyces cerevisiae* for production of ginsenosides. Metab. Eng. 20, 146–156.

Wang, D., Liu, Y., Xu, J., Wang, J., Dai, Z., Zhang, X., Huang, L., 2018. Construction of efficient yeast cell factories for production of ginsenosides precursor dammarenediol-II. Acta Pharm. Sin. 53, 1233−1241.

The sequence of *RoCYP01*

ATGGAGTTCTTCTATGCTTCCCTCCTCTGCCTCTTCGTCTCCCTGGTCTTCCTCTCCCT
CCACCTCCTCTTTTATAAGACGAAGACCGGCTCCCTCCCCCCGGGCAAGACCGGGTG
GCCGGTGATCGGGGAGAGCCTCGAGTTCCTCTCCACCGGCTGGAAGGGCCACCCAGA
GAAGTTCATCTTCGACCGGATGGCCCGCTACTCCTCCCACGTCTTCCGCACCCACCTC
CTCGGCGAGCCGGCCGCTGTCCTCTGCGGCTCCGCCGGCAACAAGTTCCTCTTCTCCA
ACGAGAACAAGCTCGTCCAAGCGTGGTGGCCGAGCTCCGTCGAGAAGATCTTCCCCA
ACGACAACGCCGAGACATCCTCGAAGGAAGAGTCCATCAAGATGCGCCGCATGCTC
CCCACCTTCTTCAAGCCGGAGGCCCTCCACCGCTACGTTGGCATCATGGACCACATT
GCCCGCCGCCACTTCGCCGACGGCTGGGACGGCAAACGGGAGGTCGTCGTCTTCCCC
TTGGCCAAGAACTACACCTTCTGGCTCGCCTGCCGCCTCTTTCTCAGCGTCGAGGATC
CCTCGCAGGTCGAAAAATTCGCGGCCCCGTTCAATCTTTTAGCATCGGGGCTGATCTC
GATCCCGATCGATCTGCCCGGGACGCCGTTCCACAAGGGGATCAAGGCCTCCGCCTA
CATCAGGAAGGAGCTCGTCGCCATCATCAAGCAGCGGAAGGCCGACCTGGCCGACG
GCACGGCGTCACCGACGCAGGACATCCTGTCGCACATGCTGCTCACCAGCAACGAAG
ATGGGAAATTCATGCAGGAGTCGGACATCGCGAACAAGATTCTGGGTTTGCTCATTG
GCGGCCATGACACTGCTAGCTCTGCTTGTACCTTCGTCGTCAAGTACCTCGCCGAGTT
GCCCCAGGTCTACGAGGGCGTCTACAAAGAGCAAATGGAGATCGCGAAATCAAAGG
CGGCCGGAGAGTTGCTGAACTGGGAGGACTTGCAGAAGATGAAGTACTCATGGAAC
GTTGCATGCGAAGTGCTGAGACTCGCACCGCCCCTTCAGGGTGCATTCAGAGAAGCC
CTTGCCGATTTCTCATTCAATGGCTTCTCCATCCCAAAGGGCTGGAAGTTATATTGGA
GTGCAAATTCGACGCACAAGAACTCGGAGTTCTTCCCGGAGCCGGAGAAGTTCGAT
CCGTCGAGGTTCGAGGGGTCGGGACCCGCCCCGTACACGTTCGTGCCGTTCGGCGGC
GGGCCGAGGATGTGCCCCGGCAAGGAATATGCCCGTTTGGAAATCCTAGTGTTCATG
CACCACCTTGTGAAGAGGTTCAAGTGGGAGAAGATGATTCCTGATGAGAAAATTGTG
GTTGATCCAATGCCCATTCCGGCTAATGGTCTCCCAGTTCGACTCTACCCCCACACTT
CTTAA

The sequence of *RoCYP02*

ATGGCTGAAATACTTTTGATGCTTGCAGCATTTTTGATAATCTTGGTCATAATAATTT
TCTCAAATAGGACGAGGTGTTCAAGAAACGGGGCCTTGAATCTGCCTCCGGGCAGCT
ATGGGTGGCCGGTTCTTGGGGAAACGGTTGAGTTCCTCCGTGCTGGATTGGATGGGA
CGCCGGAGAAATTTGTTAGAGAGAGAATGGAGAAATACAAGTCGCAAGTTTTCAAG
ACATGGATCATGGGGGAGCCCATGGCCGTTTTGTGTGGAGCTGCAGGTAACAAATTC
TTGTTCAGCAATGAAAAGAAGCTGGTGACAGTGTGGTGGCCGAGCTCCGTGAGGAAT
CTGCTGGGGCCGTGCTTGGCCACAAGCGGCGGCGATGAAGGCCGCCAGATGAGGAG
GATGGTGTCGTATTTCGTGAGCCCTGATGCTTTCACGAGGCTTTATATCAAGACAATG
GATATGGTCTCCCAGCAGCATATCAAGATTAACTGGCAAGGTAAAGAAGAGGTCAA
AGTGTTTCCGACGATCAAGTTGTACACGTTTGAACTTGCATGTAGACTGTTTATGAGC
CTTGAAGATGGCGAGCAGATTGGGAAGCTTGCTACTTTGTTTAATATTTTCTTGAAAG
GGATAATCTCAATCCCTATCAATTTGCCTGGAGCAAGATTCTACAGAGCAAAGAAAG
CTACATCTGCTATCAAGAATGAACTGCACAAGTTAGTGAGAGATAGACGAGGTGCTT
TGGAGCAGAAAACAGCTCTGCCCTCCCAAGATCTTCTCTCTCATCTGCTAGTAACCCC
TGATGAGAATGGCAAGTTCATGACTGAATCGGTAATCGTGAACAACATACTGATGCT
GCTCTTTGCAGGCCACGACACGTCCAGCGTTGCAATAACGATGCTGCTGAAGAGCCT
TGGAGAACGTCCTGATGTTTATGAGAAGGTGTTGAGAGAGCAGAAGGAGATAGCAT
CATCGGAAGGGGAGTTTCTGCAGTGGGAGGACATACAGAAGATGAAGTACTCGTGG
AGCGTGGTATGTGAAGTGACGAGGCTGTCACCGCCAGTCATTGGTGCTTTCAGAGAA
GCGTTAACGGACATAAGCTACGCGGGCTACCATATCCCCAAGGGGTGGAAGTTGTAT
TGGAGCTCGTCCAACACACATAGGGACCCAAGCTTGTTCACGAGCTACACCAGGTTT
GATCCATCGAGATTTGAAGAGTCACCGGGGCCAGCCCCATTTTCGTATGTGCCGTTT
GGTGGGGGGCCTAGGATGTGCCTGGGGAAAGAGTTTGCAAGGCTGGAGATACTCAT
ATTTTTACACAACGTGATTAAAAGATATAGGTGGAAGTTGGTGAATCCTGATGAGAA
AATTGCATATGATCCAATGCCCACACCTGTGGAAGGACTTCCCATTCTTATCCAACCA
CAGACATAG

The sequence of *RoCYP03*

ATGGAGCTCCTAACGTTAGCTGTTTCCTTACTCCTCATCGCCCTCACATTCTTC
CTCCTCCGCCGTAGCTCCGGCGGCGTCGATGCCCTCCCCCCGGGCAGCTTCGG
GTGGCCGTTGCTGGGAGAGAGCGTCGAGTTCCTGTTCGGCAAGCCGGAGAAA
TTCGTCGGCGACCGGATGAAGAAATACTCGCTGGACATCTTCAAAACTAAAA
TCCTCGGTGAGAAAACCGCCGTGATCTGCGGCCCCGCCGGCCACAAATTCCT
CTTCTCCAACGAGCAGAAGTACTTCACCGCATTCCGCCCCCACCCCATGCAGC
ACCTCTTCCGCTCCTACAAACCTAAATCCGCCTCCGCCGCCGAGAAGCCGGCT
GAAGCCGCCGCTCCGCCGACGACCAAAGCCACCGACGAGACCAAGGCCATA
CGCCAGCCCGGATTCCTCAAGCCCGAAGCCCTAATGCGATTTTTGGCGGGGA
TGGATTTAATCACTCAAAAGCAGCTGCAAATTCACTGTGCAGGTAAAAACGT
GGTCCAGTTTTACCCGCTGTCGAAAACAATTACCCTAACCATCGCGTGCCAGT
TTTTTATGGGTATTAATAACCCGGAGCGCATCGCGCGCTTGGTGAAAAACTTC
GACGATGTTACCGTGGGGATGCACTGCATCATGTTAAACTTACCGGGCACGA
TTTTTTACCGCGGTAACAAAGCTGCGGCGGCGCTCAGGAAGGAACTGATCGC
TGTCATCAAGGAGAAGAAGCAATCTATCGCCGGCGGGGCGCCGCTGCACGAC
ATCTTGTCGCACATGATCGTCGCCTCTGACCCGGCAGGGAGATCCATGCCCG
AACCTGAGATCGCAGATAAGATTATGGGGATTTTGACCGCCGGGTATAGTAC
GGTGGCGACCACGATGACCTTCTTGATGAAGTATCTAGGGTTGAATCCCGAT
ATCTACGAACGAGTTCGTGCAGAGCAATTGGAGATTGCGGCTGTGAAGAAAG
AAGGGGAACCATTGAATTGGGAAGATATGGGGAAAATGAAGTATTCATGGA
ATGTGATATGCGAGACAATGAGAATGGTTCCTCCATTGCAAGGGACATTTAG
GGAAGTGCTCGAGGAATTTACATATGCCGGCTACAATATTCCCAAGGGCTGG
AAGGTATATTGGACGGTGAGCACAACCAACATGAACCCACAGTATTTCAAG
AATCCAGAGAGATTCAATCCATCGAGATATGAGGAGGGAGAAGCGCCTCCGC
CGTACACCTACGTTCCATTCGGGGGAGGGCCGAGAATGTGCCCCGGAAAAGA
GTACGCTAGAATATCAATATTGGCTTTCGTTCACAATGTTGTGAAGAATTACC
GGTGGGAAGTTGTCGATCCAAATGAGAAGGTGGAAGGAGATATGATGCCGG
AGCCGCAGAAAGGTCTGCCGGTCCGCCTTTACAAACAGTGA

The truncated sequence of *BpCYP01*

……AAAACCGGGTACCCCATGATCGGCGAGAGCCTTGAGTTCTTGTCCACCGGATGG
AAAGGCCATCCGGAAAAGTTCATCTTCGACCGCATGACCAAGTTCTCATCGGAAGTC
TTCAAAACCTCCCTTCTTGGGCAGCCGGCCGCCGTCTTCTGTGGCGCGGCGTGTAACA
AGTTCTTGTTTTCCAATGAGAACAAGCTTGTAACTGCCTGGTGGCCGGACTCCGTCAA
C……ATCATAAAGCAGCGAAGGGTTGATTTGGGTGAGGGCAAGGCAAGCCCTACAC
AAGACATACTCTCTCACATGCTCTTGACTTCCGACGAGAGTGGGCAGTACATGACTG
AATTAGATATTGCCGACAAAATCCTCGGCCTGCTCATCGGCGGCCACGACACCGCCA
GCGCTGCTTGCACTTTCATTGTCAAGTATCTTGCTGAGCTTCCTCATATTTATGAGGG
TGTCTACAATGAGCAAATGGAGATAGCCAATTCCAAAGCACCCGGGGAGTTATTGAA
CTGGGAGGATATTCAAAAGATGAGGTATTCATGGAACGTAGCTTGTGAAGTCTTGAG
GCTTGCTCCACCGCTCCAAGGAGCTTTCCGGGAAGCCATTAATGACTTCATCTTCAAT
GGTTTCTCCATTCCCAAGGGCTGGAAGTTGTACTGGAGTGCAAACTCGACACACAGG
AGTGCGGAATACTTCCCAGAACCTGAGAAATTTGATCCGAGTAGATTTGAAGGAAGG
GGACCGGCACCGTACACATTTGTTCCATTTGGTGGAGGGCCGAGGATGTGCCCGGGA
AAAGAATATGCCCGATTGGAAATACTTGTTTTCATGCACAACTTAGTGAAAAGGTTC
AGGTGGGAGAAGATGATTCCTGATGAGAAAATCGTAGTTGATCCCATGCCCATGCCT
GCAAAGGGCCTGCCCGTTCGCCTTTACCCTCAC……

The sequence of *RoCPR01*

ATGGAACCCTCGTCGCCGAAGCTCTCGCCGCTCGATTTCATAGCGGCGATCTTAAAG
GGCGATATTGAGGGGGCGGCGCCGCGGGGTGTGGCGGCGATGTTGATGGAGAACAG
AGACCTCGCGATGGTGCTCACTACATCCGTCGCGTTGCTCATAGGCTGCGTCGTTGTG
CTAGCGTGGCGGCGCACCGCCGGATCGGCGGGGAAGAAGCAGCTACAGCCGCCCAA
GCTGGTGGTGCCGAAGGCGGCGGTGGAGCCGGAGGAGGCGGAGGATGACAACACCA
AGGTTTCCGTCTTCTTCGGCACACAGACTGGTACAGCTGAAGGTTTCGCAAAGGCAT
TTGCTGAGGAAGCTAAAGCTAGATATCCACAGGCCAAGTTTAAAGTAATTGACTTAG
ATGATTATGCTGCCGATGATGATGAGTACGAGGATAAGTTGAAGAAGGAGAGTTTA
GCATTCTTCTTCCTGGCCTCATATGGAGATGGTGAGCCTACAGACAATGCTGCAAGG
TTCTACAAATGGTTTACTGAGGGAAAAGATAGGGAGGAATGGCTTAAGAATCTTCAG
TACGCTATATTCGGTCTTGGGAACAGACAATACGAGCATTTCAACAAGATTGCGATA
GTGGTAGATGACCTTATCACCGAGCAAGGAGGAAAGAAGCTTGTTCCAGTAGGCTTG
GGAGATGATGACCAATGCATTGAAGATGACTTTACTGCATGGCGTGAATTATTGTGG
CCTGAGTTGGATAAATTGCTCCGCAACGAGGACGATGCAACGGTTGCTACTCCATAC
ACTGCTGCGGTGTTGCAGTATCGTGTTGTGCTCCATGACCAAACAGATGGATTGATT
ACAGAGAATGGTTCACCAAATGGTCGTGCCAATGGTAACACTGTATATGATGCTCAA
CATCCCTGCAGGGCAAATGTTTCTGTAAAGAGAGAGCTGCACACTCCTGAATCAGAT
CGTTCTTGCACTCATTTGGAATTTGACATAGCTGGCACAGGACTTGTGTATGAAACG
GGGGACCATGTTGGTGTCTATTGTGAGAATTTGCTCAAGAATGTGGAGGAAGCGGAA
AAGTTACTAAATCTGTCCCCGCAAACATACTTTTCAGTTCATACTGATAACGAGGAT
GGCACTCCACTCAGTGGAAGCTCTTTGCCACCTCCATTCCCCCCTTGCACTTTGCGGA
CAGCACTAACTAAATACGCAGATCTTATGAGTATGCCCAAAAAGTCTGTGCTAGTTG
CATTAGCGGAATATGCTTCTGACCAAAGTGAAGCTGATCGACTCAGATATCTTGCAT
CCCCCGATGGAAAGGAGGAATATGCACAGTATGTAGTTGCAAGTCAGAGAAGCCTA
CTGGAGATCATGGCCGAGTTCCCGTCCGCCAAGCCTTCTCTAGGTGTTTTCTTTGCAG
CTGTTGCTCCTCGGCTCCAGCCCAGATTTTATTCTATCTCATCCTCCCCGAAAATCGC
ACCAACCAGAGTTCATGTGACTTGTGCTCTGGTTTATGACAAAACACCAACAGGACG
AATCCACAAGGGTATATGCTCTACATGGATAAAGAATGCTGTGCCTTTGGAGGAAAG
TAGCGATTGCAGTTGGGCACCAATTTTTATTAGAAGCTCTAACTTCAAACTCCCTACT
GATCCTAAAGTACCGGTAATAATGGTCGGCCCTGGTACTGGCTTGGCTCCATTTAGG
GGTTTCCTTCAGGAAAGGTTAGCTCTAAAGGAATCTGGAGCTGAACTTGGTCCTGCC
ATTTTATTTTTTGGTTGTAGGAACCGTAAAATGGATTTTATTTATGAAGATGAGTTGA
ATGGCTTTGTCAAAGCTGGAGCAATTTCCGAGCTCATAGTTGCTTTCTCACGTGAGGG
ACCTGCGAAGGAATACGTGCAACACAAGATGTCTCAAAGGGCTTCGGACGTCTGGA
AAATGATCTCCGATGGAGGTTATGTCTACGTCTGTGGTGATGCCAAGGGAATGGCGC

GTGATGTACACCGAACTCTCCACACAATCGCACAAAAACAGGGTTGTCTGAGCAGCT
CCGAAGCCGAAGGCATGGTCAAGAATCTGCAAACGACAGGAAGATATTTGCGCGAT
GTATGGTGA

The sequence of *OptiRoCYP01*

ATGGAATTTTTCTATGCATCTTTGTTGTGTTTGTTTGTTTCTTTAGTTTTCTTGTCATTG
CATTTGTTATTTTACAAGACTAAGACAGGTTCATTGCCTCCAGGTAAAACTGGTTGGC
CAGTTATTGGTGAATCTTTGGAATTTTTATCAACAGGTTGGAAGGGTCATCCAGAAA
AGTTTATTTTCGATAGAATGGCTAGATATTCTTCACATGTTTTTAGAACTCATTTGTTA
GGTGAACCAGCTGCAGTTTTATGTGGTTCTGCTGGTAATAAGTTCTTGTTTTCAAACG
AAAATAAGTTAGTTCAAGCATGGTGGCCATCTTCAGTTGAAAAGATTTTCCCAAACG
ATAACGCTGAAACTTCTTCAAAGGAAGAATCTATTAAAATGAGAAGAATGTTGCCAA
CTTTCTTTAAGCCAGAAGCATTGCATAGATACGTTGGTATCATGGATCATATCGCTAG
AAGACATTTTGCAGATGGTTGGGATGGTAAAAGAGAAGTTGTTGTTTTCCCATTGGC
TAAAAATTACACTTTCTGGTTGGCATGTAGATTGTTTTTATCTGTTGAAGATCCATCA
CAAGTTGAAAAGTTCGCTGCACCTTTTAATTTGTTGGCTTCTGGTTTGATCTCAATCC
CAATCGATTTGCCAGGTACACCATTCCATAAGGGTATTAAAGCTTCTGCATACATCA
GAAAGGAATTGGTTGCTATTATTAAGCAAAGAAAGGCTGATTTGGCAGATGGTACTG
CATCTCCAACACAAGATATCTTGTCTCACATGTTGTTGACATCAAACGAAGATGGTA
AATTCATGCAAGAATCAGATATCGCTAATAAGATCTTGGGTTTGTTGATTGGTGGTC
ATGATACTGCTTCTTCAGCATGTACATTCGTTGTTAAGTATTTGGCAGAATTGCCACA
AGTTTACGAAGGTGTTTACAAGGAACAAATGGAAATCGCTAAGTCTAAAGCTGCTGG
TGAATTGTTGAACTGGGAAGATTTGCAAAAGATGAAGTACTCATGGAATGTTGCATG
TGAAGTTTTGAGATTAGCTCCACCATTGCAAGGTGCTTTTAGAGAAGCTTTGGCAGA
TTTCTCTTTTAATGGTTTTTCAATCCCAAAGGGTTGGAAGTTGTACTGGTCTGCTAATT
CAACTCATAAAAATTCTGAATTTTTCCCAGAACCAGAAAAGTTCGATCCATCTAGAT
TCGAAGGTTCAGGTCCAGCACCATATACATTTGTTCCATTTGGTGGTGGTCCAAGAAT
GTGTCCTGGTAAAGAATACGCTAGATTGGAAATCTTGGTTTTTATGCATCATTTGGTT
AAGAGATTCAAATGGGAAAAGATGATCCCAGATGAAAAGATTGTTGTTGATCCAAT
GCCAATTCCAGCTAATGGTTTGCCAGTTAGATTATACCCACATACTTCTTAA

The sequence of *OptiAtLUP*

ATGTGGAAGTTGAAGATAGGTGAAGGTGGTGCTGGTTTGATTTCTGTAAACA
ATTTTATAGGTAGACAACACTGGGAATTTGACCCTAACGCTGGTACCCCTCAA
GAACATGCAGAAATCGAAAGATTGAGAAGAGAATTCACTAAAAATAGATTCT
CCATCAAGCAAAGTGCTGATTTGTTGATGAGAATGCAATTGAGAAAGGAAAA
CCACTACGGTACAAACAACAACATCCCAGCTGCAGTTAAATTGTCTGACGCC
GAAAATATAACCGTTGAAGCATTGGTCACTACAATAAGAAGAGCTATATCTT
TTTACTCTTCAATTCAAGCACATGATGGTCACTGGCCAGCAGAATCAGCCGGT
CCTTTATTTTTCTTGCAACCATTAGTTATGGCTTTATACATCACCGGTTCCTTG
GATGACGTCTTAGGTCCAGAACATAAAAAGGAAATCGTAAGATATTTGTACA
ACCACCAAAATGAAGATGGTGGTTGGGGTTTTCATATAGAAGGTCACTCTAC
TATGTTCGGTTCTGCATTATCATATGTCGCCTTGAGAATCTTAGGTGAAGGTC
CTGAAGATAAAGCTATGGCAAAAGGTAGAAAGTGGATCTTAGACCATGGTGG
TTTGGTTGCAATTCCATCCTGGGGTAAATTTTGGGTTACTGTCTTAGGTGCTTA
TGAATGGAGTGGTTGTAATCCTTTGCCACCTGAATTATGGTTGTTACCAAAGT
TTACTCCATTCCATCCTGGTAAAATGTTGTGTTACTGCAGATTGGTATACATG
CCTATGTCCTATTTGTACGGTAAAAAGTTCGTTGGTCCAATCACTGCTTTGAT
CAGAAGTTTAAGAGAAGAATTATACAACGAACCTTACAACCAAATTAACTGG
AATACAGCTAGAAACACCGTTGCAAAAGAAGATTTGTATTACCCACATCCTT
TGATACAAGACATGTTATGGGGTTTCTTGTATCACGTCGGTGAAAGATTTTTA
AACTGTTGGCCATTCTCTATGTTAAGAAGAAAAGCCTTGGAAATTGCTATAA
ACCATGTACACTACGAAGATGAAAATTCCAGATACTTGTGTATAGGTAGTGT
CGAAAAGGTATTATGCTTGATCGCAAGATGGGTTGAAGACCCTAATTCAGAA
GCATATAAATTGCATTTGGCCAGAATCCCAGATTACTTTTGGTTAGCTGAAGA
CGGTTTGAAGATTCAATCCTTTGGTTGTCAAATGTGGGATGCCGCTTTCGCTA
TCCAAGCAATTTTAGCCTGCAATGTTAGTGAAGAATATGGTCCTACATTGAGA
AAAGCTCATCACTTTGTAAAGGCATCTCAAGTTAGAGAAAACCCATCTGGTG
ACTTCAATGCCATGTACAGACATATTTCCAAAGGTGCTTGGACTTTTAGTATG
CATGATCACGGTTGGCAAGTCTCAGACTGTACAGCAGAAGGTTTAAAGGCAG
CCTTGTTATTGTCTGAAATGCCTTCAGAATTAGTAGGTGGTAAAATGGAAACT
GAAAGATTCTACGATGCTGTAAACGTTATCTTATCTTTGCAATCCAGTAATGG
TGGTTTCCCAGCCTGGGAACCTCAAAAAGCATATAGATGGTTGGAAAAGTTT
AATCCAACTGAATTTTTCGAAGATACAATGATCGAAAGAGAATACGTTGAA
TGTACAGGTTCTGCCATGCAAGGTTTAGCTTTGTTTAGAAAACAATACCCTCA
ACATAGATCAAAGGAAATAGATAGATGCATCGCCAAGGCTATCAGATACATC
GAAAACATGCAAAATCCAGACGGTTCTTGGTACGGTTGTTGGGGTATTTGCT

ATACCTACGGTACTTGGTTCGCAGTTGAAGGTTTGACAGCCTGTGGTAAAAA
CTGCCATAATTCCTTAAGTTTGAGAAAGGCATGTCAATTCTTGTTGTCTAAGC
AATTGCCAAATGCTGGTTGGGGTGAATCATATTTGTCTTCACAAAACAAGGTT
TACACCAATTTGGAGGGTAACAGAGCAAATTTGGTCCAATCCAGTTGGGCCT
TATTGTCTTTAACACATGCTGGTCAAGCAGAAATTGATCCAACCCCTATCCAC
CGTGGTATGAAATTGTTGATTAACTCACAAATGGAAGATGGTGACTTTCCAC
AACAAGAAATAACCGGTGTTTTCATGAGAAACTGCACTTTGAACTACTCTTCA
TACAGAAATATTTTTCCAATATGGGCAATGGGTGAATACAGAAGACAAGTTT
TATGTGCTCACTCCTACTAA