Supporting Information

Moses et al. 10.1073/pnas.1323369111

SI Materials and Methods

cDNA-AFLP Analysis. After appropriate sample preparation, cDNA-amplified fragment length polymorphism (AFLP)–based transcript profiling was performed with all 128 possible BstYI+1/MseI+2 primer combinations as described (1). Gel images were analyzed with the AFLP-Quantar Pro software (KeyGene) that allowed accurate quantification of all individual bands. The obtained raw expression data were corrected for lane variations (due to PCR or loading differences) by dividing the raw data by a correction factor. The correction factor was calculated by dividing the sum of the expression levels of all fragments within one lane by the highest sum of all lanes within a primer combination. Subsequently, the standard deviation (SD) and average for each band was calculated. Individual gene expression profiles were variance normalized by subtracting the calculated average from each data point, after which the obtained value was divided by the SD. A coefficient of variation (CV) was obtained by dividing the SD by the calculated average. Gene tags displaying expression values with a CV $\geq$ 0.5 were considered as differentially expressed. Based on this cutoff value and visual inspection of the cDNA-AFLP gels, differentially expressed gene tags were selected for further analysis. Cluster analysis, sequencing, and BLAST analysis were done as described (2).

Phylogenetic Analysis. The protein sequences were retrieved from GenBank and aligned with ClustalW. The phylogenetic tree was generated with MEGA 5.10 software (3), according to the Neighbor-Joining method and bootstrapping with 1,000 replicates. The evolutionary distances were computed with the Poisson correction method and all positions containing gaps and missing data were eliminated from the data set (complete deletion option).

Generation of Strain TM1. By means of a PCR-based gene targeting approach, P_{ERG7} of *Saccharomyces cerevisiae* S288c BY4742 was replaced with P_{MET3} as described for P_{ERG9} in *S. cerevisiae* CEN.PK 113–7D (4). Plasmid pIP007 was created by inserting P_{MET3}, amplified from genomic DNA of BY4742 with primers P1+P2, between SpeI and SacII of pUG6. Two overlapping fragments, a and b, were amplified from pIP007 with primers P3+P4 and P5+P6, respectively. Fragment c, corresponding to the genomic region 500 bp upstream of P_{ERG7}, and fragment d, corresponding to the first 500 bp of *ERG7*, were amplified from the BY4742 genome with primers P7+P8 and P9+P10, respectively. The fragments were combined in pairs a+c and b+d to generate the overlapping fragments e and f with the primers P7+P4 and P5+P10, respectively. A BY4742 wild-type strain was transformed with fragments e and f using the lithium acetate-mediated transformation method. Transformants were selected on yeast extract peptone dextrose medium (Clontech) supplemented with 200 μg/mL G-418 disulfate (Duchefa).

Generation of Plasmid Vectors. The truncated 3-hydroxy-3-methylglutaryl-CoA reductase (*tHMG1*) gene encoding amino acids 532–1054 of HMG1p (5) was amplified from the chromosomal *HMG1* gene with primers P13+P14 and inserted into MCS1 of pESC-URA (Agilent Technologies) to generate pESC-URA[*GAL10*/*tHMG1*]. The oxidosqualene cyclase (OSC) was inserted into MCS2 to create pESC-URA[*GAL10*/*tHMG1*;*GAL1*/*OSC*]. The β-amyrin synthase of *Medicago truncatula* and *Glycyrrhiza glabra* were amplified with primers P15+P16 and P17+P18, respectively, and *AtLUS1* and *CaDDS* with primers P31+P32 and P33+P34, respectively. All amplicons were cloned into the Gateway vector pDONR221 and sequence verified before insertion into pESC-URA.

The *AtATR1* (At4g24520), *M. truncatula CYP716A12*, and *Barbarea vulgaris UGT73C11* were amplified with primers P19+P20, P23+P24, and P25+P26, respectively and cloned into pDONR221 by Gateway recombination. The *AtATR1* was further recombined into the integrating, low-copy, and high-copy number yeast expression vectors (6) pAG305GAL-ccdB (plasmid 14137; Addgene), pAG415GAL-ccdB (plasmid 14145; Addgene), and pAG425GAL-ccdB (plasmid 14153; Addgene), respectively. *CYP716Y1* and *CYP716A12* were recombined into the high-copy number pAG423GAL-ccdB (plasmid 14149; Addgene) vector.

The constructs encoding the self-processing polyproteins of CYP716Y1 with CYP716A12 and AtATR1 with UGT73C11 were generated by amplifying *CYP716Y1* and *AtATR1* without a stop codon and with a 3′-overhang of the partial T2A sequence with primers P21+P27 and P19+P28, respectively. *CYP716A12* and *UGT73C11* were amplified with a 5′-overhang of the partial T2A sequence with primers P29+P24 and P30+P26, respectively. As primers P27 and P29 contain a uracil, the *CYP716Y1* and *CYP716A12* were amplified with the Pfu Turbo Cx polymerase (Stratagene). These fragments were used for USER Cloning (New England BioLabs) to generate complementary sticky ends that were ligated in vitro with the T4 DNA ligase (Invitrogen). The ligated fragment was amplified with primers P21+P24. The amplicons of *AtATR1* and *UGT73C11* were combined and reamplified with primers P19+P26. The resulting amplicons were Gateway cloned into pDONR221, sequence verified, and recombined to generate pAG423[*GAL1*/*CYP716Y1-T2A-CYP716A12*] and pAG425[*GAL1*/*AtATR1-T2A-UGT73C11*].

Gas Chromatography–Mass Spectrometry Analysis. Samples were trimethylsilylated with 20 μL pyridine and 100 μL *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich) that were added directly to the dried material, vortexed, and incubated for 15 min before transfer into gas chromatography (GC) vials. GC–mass spectrometry (MS) analysis was performed with the GC model 6890 and MS model 5973 (Agilent). A 1-μL aliquot was injected (splitless mode) into a VF-5MS capillary column (Varian CP9013; Agilent) and operated at a constant helium flow of 1 mL/min. The injector was set to 280 °C and the oven was held at 80 °C for 1 min postinjection, ramped to 280 °C at 20 °C/min, held at 280 °C for 45 min, ramped to 320 °C at 20 °C/min, held at 320 °C for 1 min, and finally cooled to 80 °C at 50 °C/min at the end of the run. The MS transfer line was set to 250 °C, the MS ion source to 230 °C, and the quadrupole to 150 °C, throughout. For identification of metabolites, full electron ionization (EI)–MS spectra were generated by scanning the *m*/*z* range of 60–800 with a solvent delay of 7.8 min. For quantification of β-amyrin and lupeol, a select ion mode was operated for the ions 498, 483, 393, 279, 218, 203, and 189. The areas of the peaks were calculated with the default settings of the AMDIS software (Version 2.6, NIST).

Liquid Chromatography–MS Analysis. Reversed-phase liquid chromatography (LC) was performed with an Acquity UPLC system (Waters) fitted with an Acquity UPLC BEH C18 column (150 × 2.1 mm, 1.7 μm; Waters) and coupled to a LTQ Linear Ion Trap MS (IT-MS; Thermo Electron Corporation) via an electrospray ionization source. Of the sample, 15 μL was injected and a gradient program corresponding to 5% solvent B and 95% solvent A at 0 min, 50% solvent B and solvent A at 20 min, and 100% solvent B at 35 min was used to separate compounds. Solvents A

and B were water:acetonitrile (99:1, vol/vol) acidified with 0.1% (vol/vol) formic acid and acetonitrile:water (99:1, vol/vol) acidified with 0.1% (vol/vol) formic acid, respectively. The flow and column temperature were set to 350 μL/min and 40 °C, respectively. For negative ionization, the capillary temperature was set to 300 °C, sheath gas to 30 (arbitrary units), auxiliary gas to 5 (arbitrary units), and spray voltage to 5 kV. Full IT-MS spectra between *m/z* 150–1,400 were recorded. Extracted ion chromatograms were obtained by screening all chromatograms for the mass ranges 678.0–679.0 Da, 662.9–663.4 Da, and 648.0–649.0 Da, corresponding to the formate adducts of 3-*O*-Glc-echinocystic acid, 3-*O*-Glc-oleanolic acid, and 3-*O*-Glc-3β,16α-dihydroxyolean-12-ene, respectively.

Purification of 3-*O*-Glc-Echinocystic Acid. The 3-*O*-Glc-echinocystic acid was purified from cells of TM44 cultured for 72 h in synthetic defined medium with appropriate dropout supplements and without methyl-β-cyclodextrin. Yeast cells from a 3-L culture were collected and lyzed with CelLytic Y Cell Lysis Reagent (Sigma-Aldrich) and extracted thrice with ethyl acetate. The organic phase was evaporated to dryness and dissolved in 30 mL water. The 3-*O*-Glc-echinocystic acid was purified from this organic extract using a combination of flash chromatography and preparative HPLC. First, flash chromatography was performed with a Reveleris instrument (Grace) and a Reveleris 40g C18 cartridge (Grace). Solvent A was water:acetonitrile (99:1, vol/vol) acidified with 0.1% (vol/vol) formic acid and solvent B was methanol. A linear gradient from 10% to 80% methanol in 48 min was run using a flow of 40 mL/min and fractions of 15 mL were collected. The flash fractions containing 3-*O*-Glc-echinocystic acid were evaporated to dryness and subjected to two more purification rounds on a Waters 625 LC system. In the first round, a Luna 10 μm C18 column (Phenomenex) was used with water:acetonitrile (99:1, vol/vol) acidified with 0.1% (vol/vol) formic acid and acetonitrile:water (99:1, vol/vol) acidified with 0.1% (vol/vol) formic acid as solvents A and B, respectively. A linear gradient from 5% to 100% solvent B in 30 min was run using a flow of 6 mL/min and fractions of 1.2 mL were collected. The HPLC fractions containing 3-*O*-Glc-echinocystic acid were pooled, evaporated to dryness, dissolved in 1 mL water and used for a final purification round with a Platinum EPS C18 10-μm column and water:acetonitrile (99:1, vol/vol) adjusted to pH 7.0 with 0.1% triethyl ammonium acetate buffer and 0.05% trimethylamine as solvent A and acetonitrile:solvent A (99:1, vol/vol) as solvent B. A linear gradient from 5% to 100% solvent B in 30 min was run using a flow of 6 mL/min and fractions of 1.2 mL were collected. Flash and HPLC fractions were checked for the presence of 3-*O*-Glc-echinocystic acid using the same Acquity system (Waters) as for the LC-MS analysis. The final HPLC fractions containing 3-*O*-Glc-echinocystic acid were pooled and evaporated to dryness before NMR analysis.

NMR Measurements and Analysis. All NMR spectra were measured on an Avance II Bruker spectrometer operating at a ^{1}H and ^{13}C frequency of 700.13 and 176.05 MHz, respectively. A 1-mm $^{1}H/^{13}C/^{15}N$ TXI-z probe for mass-limited samples was used to measure the compound purified by HPLC as described in the previous step. Less than 1.0 mg (<1.5 μmol) dry sample was dissolved in 10 μL deuterated methanol (CD_3OD) with 99.96% atom-D to reduce interference of the solvent signal. The spectra recorded on the sample included 1D ^{1}H, 2D ^{1}H-{^{1}H} gradient selected correlation spectrometry (gCOSY), ^{1}H-{^{1}H} total correlation spectrometry (TOCSY) (100-ms spinlock), ^{1}H-{^{13}C} multiplicity-edited gradient heteronuclear single quantum correlation (gHSQC), and a ^{1}H-{^{13}C} heteronuclear multiple bond correlation (HMBC) with 8 Hz long-range proton-carbon coupling constant ($^{n}J_{CH}$), requiring a total of 80 h of measuring time. The available material was insufficient to record a 1D ^{13}C. The commercial standard was prepared by dissolving 5 mg in 550 μL CD_3OD and the same set of measurements (including an additional HMBC optimized for 5 Hz $^{n}J_{CH}$ coupling) were measured with a 5-mm $^{1}H/^{13}C/^{15}N$ TXI-z probe and were referenced to the residual solvent signals at 3.31 ppm for the ^{1}H frequency and 49.15 ppm for the ^{13}C frequency and were processed with TopSpin (version 3.2pl3; Bruker).

The structure elucidation and complete ^{1}H and ^{13}C assignment of the purified compound were not attempted because of the low quantity, incomplete removal of impurities, and the level of overlap caused by the limited structural variation of the β-amyrin skeleton, leading to considerable resonance overlap, even at 700 MHz. However, sufficient information could be extracted from the data to demonstrate the presence of 3-*O*-Glc-echinocystic acid. First, using the typical TOCSY pattern and matching approach (7), the glucose moiety was established and, through combination with the me-^{1}H-{^{13}C}-gHSQC, fully assigned, revealing the presence of two additional and as yet unaccounted CH units at (^{1}H;^{13}C) chemical shifts of 4.45;75.13 ppm and 3.20;90.66 ppm respectively. Using ^{13}C chemical-shift prediction (ChemDraw 12.0; PerkinElmer Informatics), the distinctly high ^{13}C shift of 90.66 was attributed to C-3 (92.5 predicted). This assignment proved an important step toward revealing key assignments of the β-amyrin structure. From the 1D ^{1}H and 2D me-^{1}H-{^{13}C}-gHSQC, the presence of seven individually resolved methyl groups corresponding to Me23–Me30 in the structure was easily established. Being threefold more intense, they provided a series of characteristic $^{n}J_{CH}$ cross-peaks with sufficient signal-to-noise ratio to explore and assign the surrounding carbon skeleton in the ^{1}H-{^{13}C}-gHMBC (Fig. S6). In short, Me23 and Me24 were assigned (nondiastereotopically) by establishing a contact with the previously assigned C-3 (A* in Fig. S6*B*). From this, Me25 could be assigned by noting a $^{n}J_{CH}$ correlation to a carbon also connected to Me23 and Me24, which can only be C-5 (B* in Fig. S6*B*). Furthermore, only Me27 is expected to connect to the quaternary alkene carbon C-13, allowing it to be unambiguously assigned (C* in Fig. S6*B*). Me26 was found through $^{n}J_{CH}$ correlations involving carbons in common with Me25 (to C-9) and Me27 (to C-8 and C-14), respectively. The remaining two methyl resonances should correspond to Me29 and Me30, which was proven by mutual $^{3}J_{CH}$ correlations and shared correlations to three carbons, necessarily representing C-19, C-20, and C-21 without definitive assignment of the latter three. Using the assigned methyl ^{1}H resonances as a starting point, the location of the hydroxylation site could be investigated by noting $^{n}J_{CH}$ correlations to the carbon at 75.13 ppm, referred to above. None of the methyls showed such correlation, eliminating all immediately surrounding carbon positions (5, 7, 9, 15, 19, and 21) in the β-amyrin skeleton as the location of the hydroxylation site, leaving C-16 as the most viable option. As the only methylene carbon close enough to Me27 to generate a $^{n}J_{CH}$, C-15 could be assigned at 36 ppm, a value considerably higher than the one expected from non-hydroxylated echinocystic acid, predicted at 29.0 ppm. By tracing both H15-methylene protons from the attached carbon (1.88 and 1.35 ppm), H16 could be identified within the COSY spectrum at 4.45 ppm and in the me-^{1}H-{^{13}C}-gHSQC, it connected to the carbon at 74.13 ppm, thus demonstrating C-16 as the hydroxylation site. The appearance of the H16 resonance is consistent with two similarly small $^{3}J_{HH}$ scalar couplings, which points to an α hydroxylation pattern, and establishes 3-*O*-Glc-echinocystic acid as the most probable structure for the purified compound.

With the possible structures narrowed down to a single one, confirmation could be attempted with a suitable reference compound. NMR data measured for a commercial sample of 3-*O*-Glc-echinocystic acid (Extrasynthese) proved highly similar to that of the purified compound. Comparison of the 1D ^{1}H spectra showed an

almost identical spectral pattern, both in terms of chemical shift and coupling in the HMBC and HSQC spectra (Fig. S7 *A*–*C*). Some small shifts in chemical shift between both could be noticed, which can probably be explained by concentration and/or solvent effects due to differences between the purified and commercial compounds. These effects influence the ^{13}C chemical shifts to a much lesser degree, and the respective carbon chemical shifts are identical within error. A superposition of the methyl region and analysis of the fine structure of H16 in both samples (Fig. S7*D*) shows identical correlation patterns, multiplicity, and line shape.

1. Vuylsteke M, Peleman JD, van Eijk MJT (2007) AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. *Nat Protoc* 2(6):1399–1413.
2. Rischer H, et al. (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *Proc Natl Acad Sci USA* 103(14):5614–5619.
3. Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.
4. Asadollahi MA, et al. (2008) Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: Effect of *ERG9* repression on sesquiterpene biosynthesis. *Biotechnol Bioeng* 99(3):666–677.
5. Polakowski T, Stahl U, Lang C (1998) Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. *Appl Microbiol Biotechnol* 49(1): 66–71.
6. Alberti S, Gitler AD, Lindquist S (2007) A suite of Gateway® cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* 24(10):913–919.
7. Gheysen K, Mihai C, Conrath K, Martins JC (2008) Rapid identification of common hexapyranose monosaccharide units by a simple TOCSY matching approach. *Chemistry* 14(29):8869–8878.

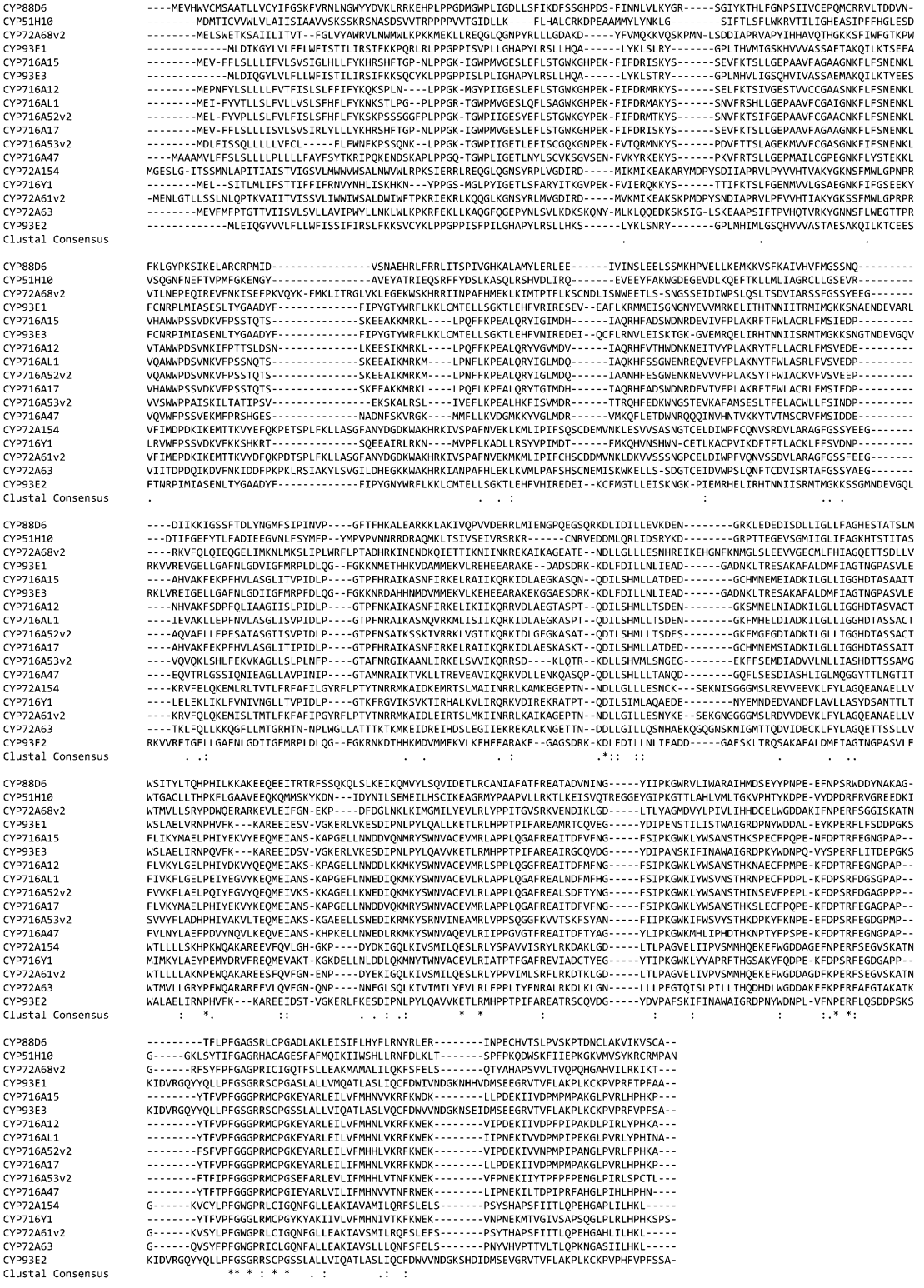

Fig. S1. Sequence alignment of the amino acid sequences of CYP716Y1 and other P450s involved in triterpene saponin biosynthesis used to generate the phylogenetic tree shown in Fig. 1*B*. Asterisk (*), identical amino acids; colon (:), conserved amino acid substitutions; full stop (.), semi-conserved amino acid substitutions.

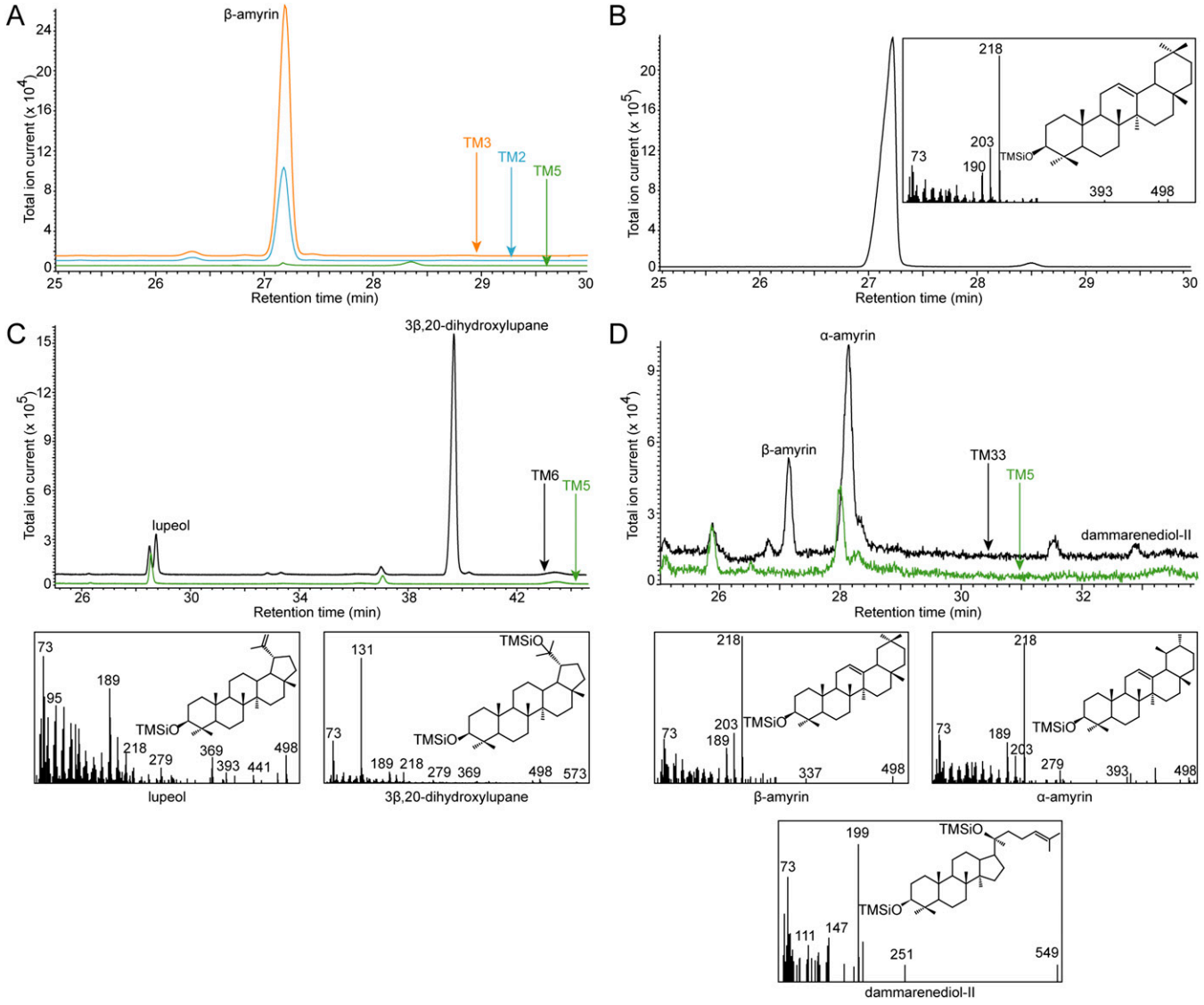

Fig. S2. Generation of yeast strains that produce different triterpene backbones. (*A* and *B*) Production of β-amyrin by yeast strains TM2 and TM3. Overlay of GC chromatograms (*A*) showing accumulation of β-amyrin in cells of strains TM2 and TM3 but not in control strain TM5. The GC retention time (t_R = 27.2 min) and EI-MS spectra of β-amyrin in strains TM2 and TM3 match those of an authentic standard (*B*). Inset shows the EI-MS spectrum and structure of trimethylsilylated β-amyrin. (*C* and *D*) Production of multiple triterpenes by yeast strains TM6 and TM33; 2,3-oxidosqualene is cyclized to multiple triterpene backbones in the yeast strains TM6 and TM33 expressing *AtLUS1* and *CaDDS*, respectively. Overlay of GC chromatograms showing accumulation of lupeol and 3β,20-dihydroxylupane in the cells of strain TM6 (*C*) and β-amyrin, α-amyrin, and dammarenediol-II in the cells of strain TM33 (*D*). The GC retention times of lupeol, 3β,20-dihydroxylupane, β-amyrin, α-amyrin, and dammarenediol-II were 28.9 min, 40.2 min, 27.2 min, 28.6 min, and 33.5 min, respectively. *Insets* show the EI-MS spectra and structure of all of the trimethylsilylated triterpenes produced by strains TM6 and TM33. TMSiO, trimethylsilylated oxygen.

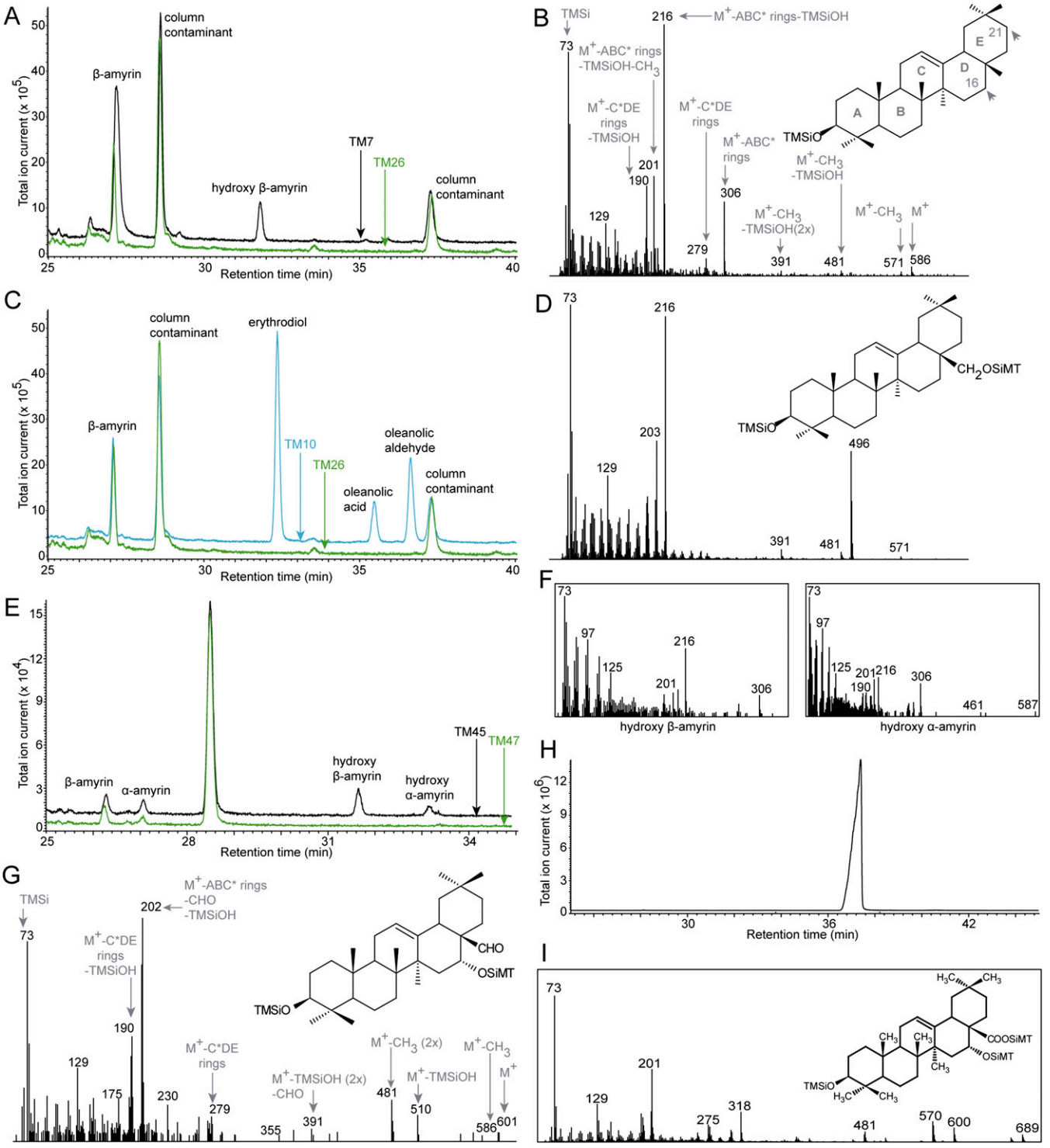

Fig. S3. CYP716Y1 hydroxylates β-amyrin on position 16. (*A*) Overlay of GC-MS chromatograms showing the conversion of β-amyrin to hydroxy-β-amyrin in strain TM7, but not in the control strain TM26. (*B*) EI-MS spectrum of trimethylsilylated hydroxy-β-amyrin accumulating in strain TM7. The possible hydroxylation positions of CYP716Y1 are indicated by arrows on the β-amyrin structure. The fragment ions corresponding to the loss of specific groups from the parent ion (M^+) during electron impact ionization are indicated. (*C*) Overlay of GC chromatograms showing the conversion of β-amyrin to erythrodiol, oleanolic aldehyde, and oleanolic acid by CYP716A12 in strain TM10 compared with the control strain TM26. (*D*) EI-MS spectrum and structure of an authentic trimethylsilylated erythrodiol standard. (*E*) Hydroxylation of β- and α-amyrin by CYP716Y1. Overlay of GC chromatograms showing the hydroxylation of β- and α-amyrin in strain TM45 but not in the control strain TM47. (*F*) EI-MS spectra of trimethylsilylated hydroxy-α- and hydroxy-β-amyrin in strain TM45. (*G*) Structure and EI-MS spectrum of trimethylsilylated 16α-hydroxy oleanolic aldehyde produced in strain TM30. Because of the lack of an authentic standard or NMR data of this compound, its identity could not be unambiguously established yet. Fragment ions corresponding to the loss of specific groups from the parent ion (M^+) are shown. (*H*) GC chromatogram showing elution of an authentic trimethylsilylated echinocystic acid standard at 37.3 min, differing from the 16α-hydroxy oleanolic aldehyde produced in strain TM30 that elutes at 40.5 min (Fig. 4*A*). (*I*) EI-MS spectrum and structure of the trimethylsilylated echinocystic acid standard. C*, partial C ring resulting from retro Diels–Alder fragmentation; CH_3, methyl group; CHO, aldehyde group; OSiMT, trimethylsilylated oxygen; TMSiOH, trimethylsilanol.

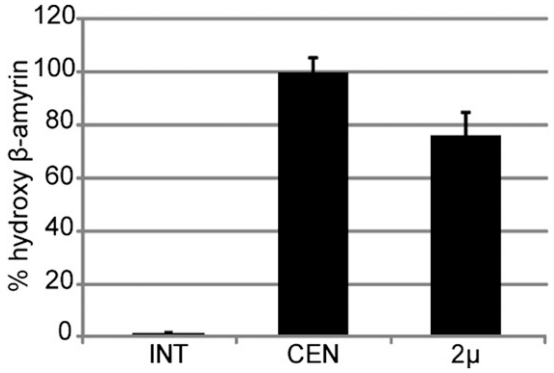

Fig. S4. Hydroxylase activity of CYP716Y1 enhanced by modulation of the CPR:P450 ratio. Relative comparison of the percentage of hydroxy-β-amyrin accumulating in strains TM8, TM9, and TM7 expressing *AtATR1* from an integrative (INT), centromeric (CEN), or 2-μm (2μ) vector, respectively.

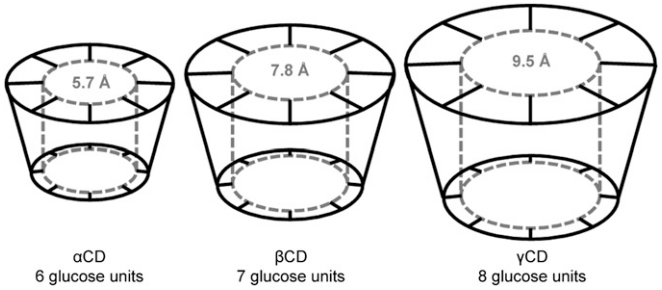

Fig. S5. Toroidal topology of the common cyclodextrin (CD) variants. The αCD, βCD, and γCD are cyclic oligosaccharides composed of 6, 7, and 8 glucose units, respectively (1). The CDs can sequester hydrophobic moieties because their interior is relatively more hydrophobic than their exterior. The interior diameter of each variant is indicated in ångströms and is highlighted by gray dashed lines.

1. Saenger W, et al. (1998) Structures of the common cyclodextrins and their larger analogues–Beyond the doughnut. *Chem Rev* 98(5):1787–1802.

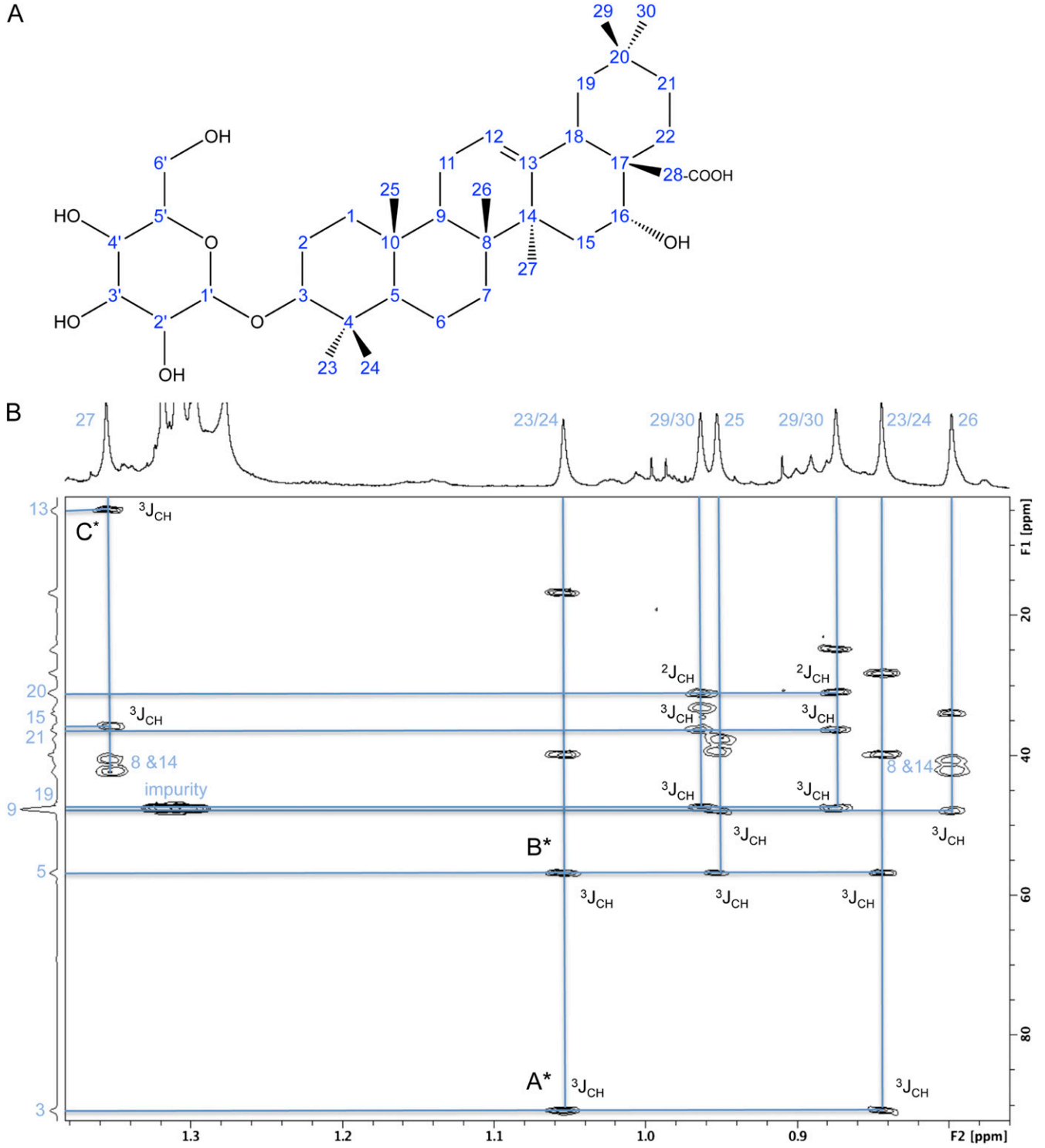

Fig. S6. NMR analysis of 3-*O*-Glc-echinocystic acid purified from yeast strain TM44. (*A*) Numbering scheme. (*B*) Methyl ^{1}H to carbon region of the 2D ^{1}H-{^{13}C} gHMBC spectrum of 3-*O*-Glc-echinocystic acid purified from yeast strain TM44. The spectrum was recorded at 700 MHz, 298 K, and optimized for $^nJ_{CH}$ = 8 Hz. The coupling pattern at C* corresponds with the alkene carbon C13 folded in at 5 ppm in the indirect ^{13}C dimension for higher resolution. The labels A*, B*, and C* refer to the NMR analysis in *SI Materials and Methods*, *NMR Measurements and Analysis*. $^2J_{CH}$, proton-carbon two bond scalar coupling constant; $^3J_{CH}$, proton-carbon three bond scalar coupling constant.

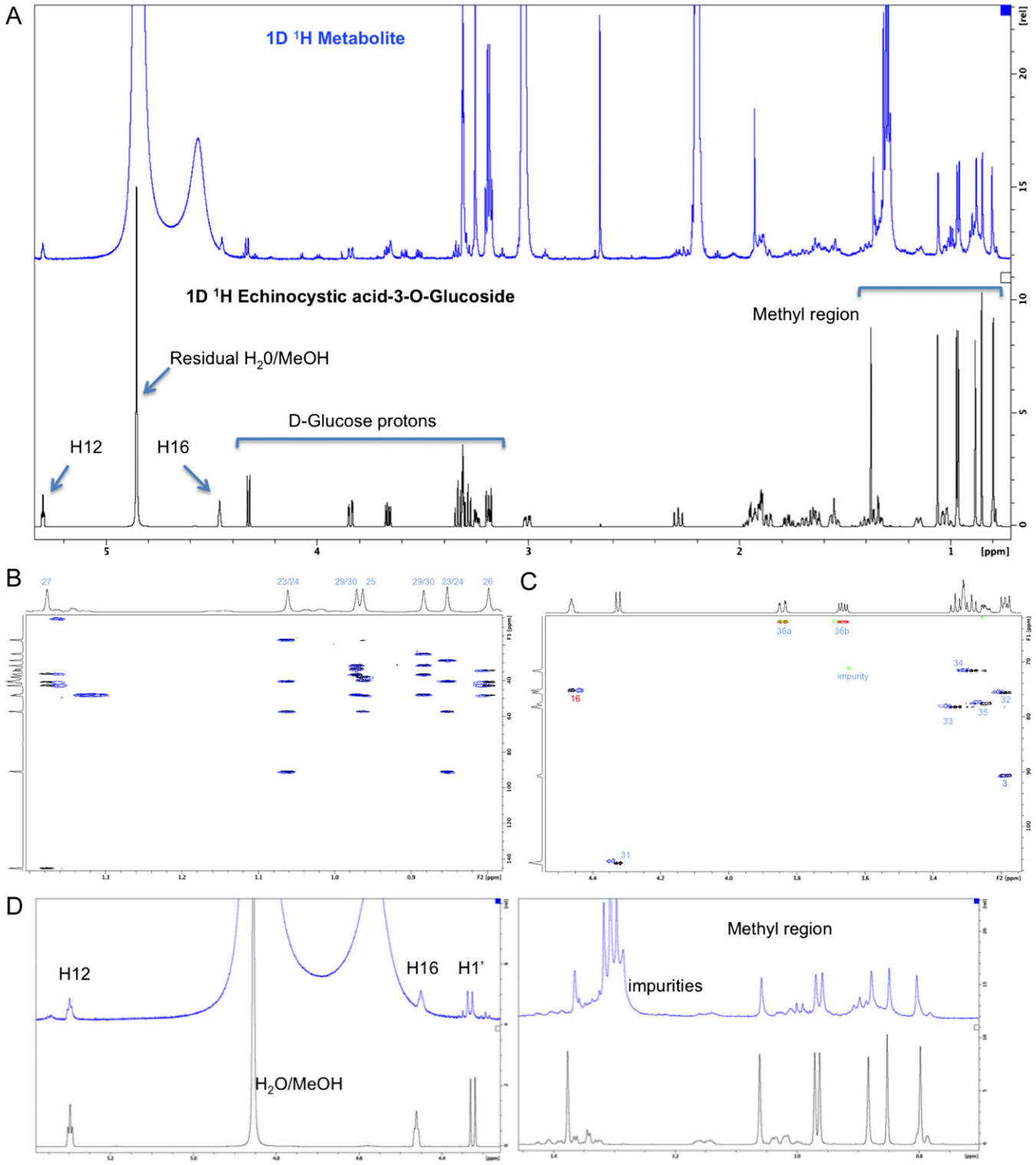

Fig. S7. NMR spectra of 3-*O*-Glc-echinocystic acid purified from yeast strain TM44 and the authentic commercial standard. (*A*) General comparison of the 1D ^{1}H spectra (blue and black represent metabolite and commercial standard, respectively). (*B*) Overlay of the ^{1}H-{^{13}C}-gHMBC spectra (8-Hz long-range coupling constant where blue represents the metabolite and black the commercial standard). (*C*) Overlay of the ^{1}H-{^{13}C} me-gHSQC spectra in the CH-O-R/H region (blue/green represents the metabolite and black/red the commercial standard). (*D*) Comparison of the two relevant 1D ^{1}H regions with comparison of the H16 fine structure (blue represents the metabolite and black the commercial standard). All spectra were recorded at 700 MHz and 298 K. MeOH, methanol.

Table S1. Generated yeast strains

Strain	Genotype
S288c BY4742	MATa *his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0*
TM1	S288c BY4742; P_{erg7}::P_{MET3}-*ERG7*
TM2	TM1; pESC-URA[*GAL10/tHMG1*;*GAL1/MtbAS*]
TM3	TM1; pESC-URA[*GAL10/tHMG1*;*GAL1/GgbAS*]
TM5	TM1; pESC-URA[*GAL10/tHMG1*]
TM6	TM1; pESC-URA[*GAL10/tHMG1*;*GAL1/AtLUS1*]
TM7	TM3; pAG423[*GAL1/CYP716Y1*], pAG425[*GAL1/AtATR1*]
TM8	TM3; pAG423[*GAL1/CYP716Y1*], pAG305[*GAL1/AtATR1*]
TM9	TM3; pAG423[*GAL1/CYP716Y1*], pAG415[*GAL1/AtATR1*]
TM10	TM3; pAG423[*GAL1/CYP716A12*], pAG425[*GAL1/AtATR1*]
TM17	TM3; pAG423[*GAL1/CYP716A12*], pAG415[*GAL1/AtATR1*]
TM26	TM3; pAG423, pAG425[*GAL1/AtATR1*]
TM28	TM6; pAG423, pAG425[*GAL1/AtATR1*]
TM30	TM3; pAG423[*GAL1/CYP716Y1-T2A-CYP716A12*], pAG415[*GAL1/AtATR1*]
TM33	TM1; pESC-URA[*GAL10/tHMG1*;*GAL1/CaDDS*]
TM44	TM3; pAG423[*GAL1/CYP716Y1-T2A-CYP716A12*], pAG425[*GAL1/AtATR1-T2A-UGT73C11*]
TM45	TM33; pAG423[*GAL1/CYP716Y1*], pAG425[*GAL1/AtATR1*]
TM46	TM6; pAG423[*GAL1/CYP716Y1*], pAG425[*GAL1/AtATR1*]
TM47	TM33; pAG423, pAG425[*GAL1/AtATR1*]
TM48	TM3; pAG423[*GAL1/CYP716Y1*], pAG425[*GAL1/AtATR1-T2A-UGT73C11*]
TM49	TM3; pAG423[*GAL1/CYP716A12*], pAG425[*GAL1/AtATR1-T2A-UGT73C11*]

Table S2. Primers used

Primer	Sequence (5′ to 3′)
P1	GGactagtCCTTGGTATAAGGTGAGGGGGTCCACAG
P2	TCCccgcggGGAGAATACCACCGTGAGGAGCAGGCATG
P3	GATCCCCGGGAATTGCCATGACGCTGCAGGTCGACAACCC
P4	CCATGAGTGACGACTGAATCCGG
P5	CTATCGATTGTATGGGAAGCCCG
P6	CGATTGTGTCAGAATAAAATTCTGTCATGTTAATTATACTTTATTCTTGTTATTATTATAC
P7	TAATACCCCTTGAGGAGAATGTCTTC
P8	CATGGCAATTCCCGGGGATCCCGGAAACATCGACTTTATCGAGG
P9	ATGACAGAATTTTATTCTGACACAATCG
P10	CCTTCCCATTTATACAAGTTTAGTGC
P11	CCTCCTTATACATTCGTTCCATTC
P12	TTAGGGTCTACTTTCTCCCATTTG
P13	TGGAAAgcggccgcTTTCTAGTACTATGGACCAATTGGTGAAAACTG
P14	GGCTTGCttaattaaGCGTGAACAGTCACATGGTGCTGTTGTGCTT
P15	GCAGGCTCCATATATctcgagATATATATGTGGAAGCTGAAGATTGGA
P16	GCTGGGTCATATATggtaccATATATTTAAACTGCAGTGGAAGGCAA
P17	GCAGGCTCCATATATctcgagATATATATGTGGAGGCTGAAGATAGCG
P18	GCTGGGTCATATATggtaccATATATTTAAGTTAAACAAACTGGAGTG
P19	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGACTTCTGCTTTGTATGC
P20	GGGGACCACTTTGTACAAGAAAGCTGGGTATCACCAGACATCTCTGAG
P21	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAACTTTCTATCACT
P22	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAGATGGAGATTTGTG
P23	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAATGGAGCCTAATTTCTATCTCTCCC
P24	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAGCTTTGTGTGGATAAAGGCG
P25	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGTTTCCGAAATCACCC
P26	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAATTATTAGATTGTGCTAGTTGC
P27	ACCGCA**U**GTTAGCAGACTTCCTCTGCCCTCAGATGGAGATTTGTGGGGAT
P28	GATTCTCCTCGACGTCACCGCATGTTAGCAGACTTCCTCTGCCCTCCCAGACATCTCTGAGGTATC
P29	ATGCGG**U**GACGTCGAGGAGAATCCTGGCCCAATGGAGCCTAATTTCTATC
P30	GGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCAATGGTTTCCGAAATCACCC
P31	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAggatccATGTGGAAGTTGAAG
P32	GGGGACCACTTTGTACAAGAAAGCTGGGTActcgagTTAATTAACGATAAAC
P33	GGGGACAAGTTTGTACAAAAAAGCAGGCTTActcgagATGTGGAAGCTGAAGATAGCA
P34	GGGGACCACTTTGTACAAGAAAGCTGGGTTgctagcTCAATTGGAGAGCCACAAGCG

The sequences in lowercase represent the restriction recognition sites used for restriction enzyme-mediated cloning. The underlined sequences correspond to partial T2A sequences. The uracil in the primer sequence is indicated in bold.