

## Supporting information for *Genome Mining to Identify New Plant Triterpenoids*

Gia C. Fazio, Ran Xu, and Seiichi P. T. Matsuda\*

Department of Chemistry and Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005

matsuda@rice.edu

**Construction of *S. cerevisiae* mutant RXY6:** The *erg1Δ* diploid strain (ATCC# 4024805, *MATa/α his3Δ1/his3Δ1 leu2Δ/leu2Δ MET15/met15Δ LYS2/lys2Δ ura3Δ/ura3Δ ERG1/erg1::KanMX4*) contains a native *ERG1* gene and the kanamycin resistance gene *KanMX4* at the second *ERG1* locus. Genomic DNA was extracted from ATCC# 4024805 and digested with *Pvu* II, which cuts the wild type *ERG1* open reading frame but not the deletion marker *KanMX4*. PCR was carried out using *Pvu* II digested genomic DNA as template and the primers ERG1pcrF1: 5'-CGAGTATGGGTTCGATGTGGCTATTCC-3' and ERG1pcrR1: 5'-GCACCTCCTCCGCCCTCACCTTAGC-3'. Because the native *ERG1* gene had been cut with *Pvu* II, only the deletion construct was amplified. PCR conditions were: 95°C for 5 min followed by the addition of *Ex Taq* polymerase; then 36 cycles of 94 °C for 15 sec, 65 °C for 30 sec, and 72 °C for 4 min. The 3.6 kb amplicon, containing (in sequence) the *ERG1* upstream flanking region, the *KanMX4* gene, and the *ERG1* downstream flanking region, was purified using a Qiaquick Gel Extraction Kit (Qiagen, Valencia, CA) and ligated into pGEM-T vector (Promega, Madison, WI) with Fastlink ligase (Epicentre Technologies, Madison, WI). The resulting plasmid, named pRX12.3, was confirmed by sequencing and then double digested with *Nde* I and *Sac* II to release the 3.6 kb insert. The digestion mixture was transformed into the yeast lanosterol synthase mutant SMY8<sup>1</sup> (*MATa erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ 63 leu2-3,112 his3-Δ200 ade2 Gal<sup>+</sup>*), using the lithium acetate method.<sup>2</sup> The transformants were selected on plates containing 1% yeast extract, 2% peptone, 2% dextrose, 0.75% agar, 200 μg/mL geneticin, 20 μg/mL ergosterol, and 13 μg/mL hemin chloride.<sup>1</sup> Two colonies were restreaked and named RXY6A and RXY6B, respectively.

The deletion of the *ERG1* gene was confirmed by PCR under conditions similar to those described above using RXY6A and RXY6B genomic DNA as templates for the following two primer pairs: ERG1cnfmF1 (5'-GCGTGCTTTCTTGAGCCCTG-3', upstream flanking the 3.6-kb deletion construct) and KANB1 (5'-TGTACGGGCGACAGTCACAT-3', within the *KanMX4* gene); KANC3 (5'-CCTCGACATCATCTGCCAGAT-3', within the *KanMX4* gene) and ERG1cnfmR1 (5'-CCCATCTACCAGAAGCGCCAC-3', downstream flanking the 3.6-kb deletion construct). RXY6A and RXY6B proved to be genetically identical, and RXY6A was named RXY6 (*MATa erg7::HIS3 erg1::KanMX4 hem1::TRP1 ura3-52 trp1-Δ 63 leu2-3,112 his3-Δ200 ade2 Gal<sup>+</sup>*). SMY8 genomic DNA was used as a negative control in the above PCRs.

The deletion of the *ERG1* gene was further confirmed by examination of the lipid components from RXY6, with SMY8 as a control. A sample of each strain was extracted by the Folch

method.<sup>3</sup> The extract was then loaded on a 5 cm × 5 mm i.d. silica gel column and eluted sequentially by hexane, hexane-acetone 1:1 (v/v) and acetone. <sup>1</sup>H NMR analyses showed that the hexane fraction of RXY6 contained squalene whereas that of SMY8 did not. The accumulation of squalene in RXY6 confirmed that the squalene epoxidase activity was abolished in this yeast strain.

**Cloning of At5g48010 and yeast construction:** The At5g48010 open reading frame was PCR-amplified with the following two pairs of primers (enzyme restriction sites are underlined): At5g48010F1 (5'-GGGTCGACATTATATGTG GAGGCTGAGAACTGGACCG-3') and At5g48010R1 (5'-TTGCGGCCGCTTAAGGGAGGAGACGTCGCAGAGC -3'); PCR as above using 2 μL of *Arabidopsis* cDNA derived from young seedlings generated a 2.3 kbp fragment, which was gel-purified and ligated into pCR2.1-TOPO vectors with Fastlink ligase. The subclone was sequenced and compared to the genomic sequence to ensure that PCR errors had not been introduced. The insert was excised using *Not* I and *Sal* I and subcloned into the yeast expression vector pRS426GAL.<sup>4</sup> The resultant expression construct was named pGCF4.0 and was transformed into RXY6 and SMY8 using the lithium acetate method, to generate RXY6[pGCF4.0] and SMY8[pGCF4.0], respectively.

**In vitro reaction:** A 50-mL culture of RXY6[pGCF4.0] was grown in synthetic complete medium lacking uracil<sup>5</sup> with 2% galactose as a carbon source and supplemented with ergosterol (20 μg/mL), heme (13 μg/mL), and Tween 80 (5 mg/mL). The cells were obtained by centrifugation and resuspended as a 40% slurry in 100 mM sodium phosphate buffer at pH 7. Racemic oxidosqualene<sup>6</sup> was added from a 20 mg/mL solution in 20% Triton X-100 to a final concentration of 1 mg/mL. The cells were then broken by vortexing the suspension with glass beads for 6 min. The enzymatic reaction was allowed to incubate at room temperature for 24 h. Two volumes of ethanol were added to precipitate the protein, which was removed by filtration through diatomaceous earth. The solvent was removed by rotary evaporation, and the residue was suspended in ethyl acetate and passed through silica gel to remove the most polar cellular components and detergent. The solvent was removed under a nitrogen stream, and the residue was derivatized with 200 μL of 1:1 bis(trimethylsilyl)trifluoroacetamide/pyridine.

GC-MS analysis was performed on an Agilent 6890 gas chromatograph interfaced to a 5973 MSD; a Rtx-35ms capillary column (Restek, 30 m × 0.25 mm i.d., 0.10 μm film thickness) was used. A 5-μL aliquot of the derivatized product mixture was

<sup>3</sup> Folch, J.; Lees, M.; Stanley, G. H. S. *J. Biol. Chem.* **1957**, *226*, 497-509.

<sup>4</sup> Hart, E. A.; Hua, L.; Darr, L. B.; Wilson, W. K.; Pang, J.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **1999**, *121*, 9887-9888.

<sup>5</sup> Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K., Eds. *Current Protocols in Molecular Biology*; Wiley-Interscience: New York, 1999.

<sup>6</sup> Nadeau, R. G.; Hanzlik, R. P. *Methods Enzymol.* **1968**, *15*, 346-351.

<sup>1</sup> Corey, E. J.; Matsuda, S. P. T.; Baker, C. H.; Ting, A. Y.; Cheng, H. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 327-331.

<sup>2</sup> Schiestl, R. H.; Gietz, R. D. *Curr. Genet.* **1989**, *16*, 339-346.

injected at 280 °C with a split ratio of 40 : 1, 20 cm/s helium flow, and 270 °C column temperature. The reconstructed ion chromatogram for *m/z* 498 showed one peak at 4.9 min. No other component was visible within 2% of the major signal, indicating that a single C<sub>33</sub>H<sub>58</sub>OSi component (trimethylsilyl ether of a triterpene alcohol) was formed. The mass spectrum is shown in Figure S1. The only component with *m/z* 426 was oxidosqualene; no underivatized cyclized components were present.

**In vivo production:** A 1-L culture of SMY8[pGCF4.0] was grown to saturation in the same medium used to culture RXY6[pGCF4.0]. The cells were harvested by centrifugation and saponified using 200 mL 10% potassium hydroxide in 80% ethanol at 70 °C for 2 h. The reaction was allowed to cool to room temperature and 20 mL water was added. The suspension was extracted three times with 100 mL portions of hexane. The organic fractions were combined, and washed with three 5-mL portions of water and 5-mL of brine. The organic layer was dried over magnesium sulfate and concentrated to dryness. Silica gel thin-layer chromatography (ether/hexane, 1:1) showed two components that were absent in the host strain (*R<sub>f</sub>* 0.61 and 0.36). Silica gel column chromatography using a gradient from 10% ether in hexane to 20% ether in hexane provided about 1 mg of the less polar component and roughly 0.5 mg of the polar product. GC-MS of the TMS ethers established that the less polar component was the oxidosqualene cyclase product thalianol obtained from the in vitro reaction, and the more polar component had *m/z* 514, consistent with the TMS ether of a C<sub>30</sub>H<sub>50</sub>O<sub>2</sub> compound.

**Determination of the thalianol structure:** The structure of thalianol was deduced by NMR and is shown in Figure S2, and the signal assignments are given in Figure S3. <sup>1</sup>H, <sup>13</sup>C, HSQC, HMBC, COSYDEC, and NOE difference spectra were obtained on a Bruker Avance 500 spectrometer at 25 °C in CDCl<sub>3</sub> solution. <sup>1</sup>H spectra were referenced to internal tetramethylsilane, and <sup>13</sup>C spectra were referenced to CDCl<sub>3</sub> at 77.0 ppm. Thalianol showed <sup>1</sup>H signals from two olefinic protons and three allylic methyl groups (Figure S5). This suggested that two of the olefins in oxidosqualene did not participate in the reaction and that thalianol probably was tricyclic. Signals in the 1D <sup>1</sup>H spectrum of the less polar compound were strikingly similar to those from the A and B rings of lanosterol and related 4,4-diMe-3β-OH-Δ8 sterols. HMBC, HSQC, and COSYDEC spectra were also consistent with the A-ring structure of lanosterol, and NMR signals, stereochemistry, and regiochemistry in the A ring were assigned accordingly. Signal assignments in rings B and C were made from HMBC (C8, C9, C12, C13), and COSYDEC (C6, C7, C11) spectra. Signal assignments and regiochemistry in side chain were established by HMBC and by analogy with NMR data for squalene, farnesol, and geranylgeraniol. The connectivity of the 13α-side chain was determined from HMBC signals of H25 and H27, and the C13 stereochemistry was established by NOE difference correlations of H26 and H25 to H-11β. The configuration of C14 was determined by NOE difference correlations between H26 and H27 (molecular modeling shows that each epimer has only one significant rotamer about the C13-C14 bond; in the R epimer, the two methyls are close, whereas in the S epimer the methyls are distant). Assignments are shown in Figure S3. 500 MHz <sup>1</sup>H NMR: δ 0.820 (br s, 3H, H-24), 0.835 (dd, *J*=6.7, 0.6 Hz, 3H, H-27), 0.874 (m, 1H, H-15), 0.934 (br s, 3H, H-26), 0.947 (br s, 3H, H-25), 1.017 (s, 3H, H-23), 1.066 (dd, *J*=12.4, 1.7 Hz, 1H, H-5), 1.251 (m, 1H, H-1α), 1.278 (m, 1H, H-

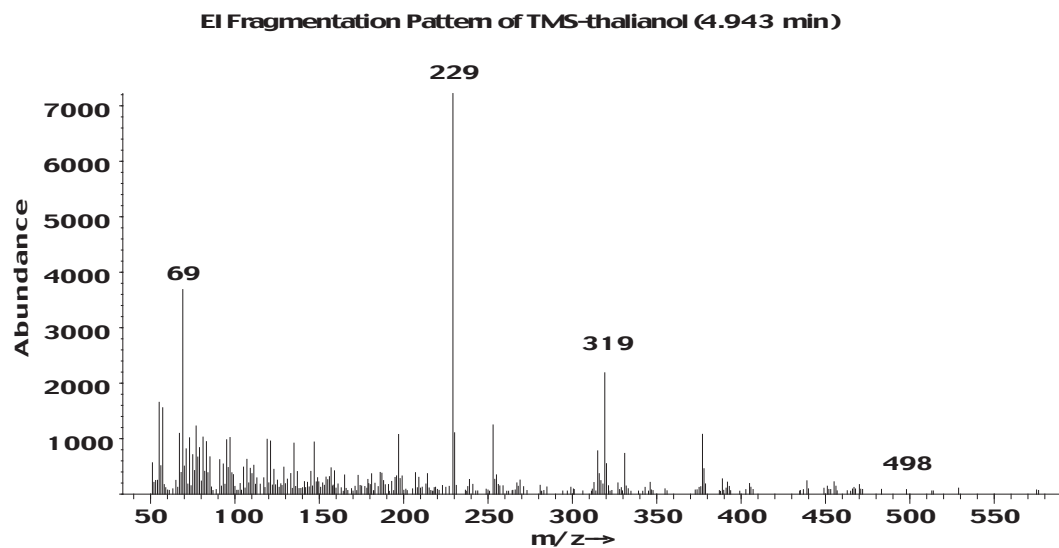
15), 1.287 (br ddd, *J*=13, 9, 5 Hz, 1H, H-12β), 1.300 (m, 1H, H-14), 1.441b (m, 1H, H-6β), 1.590 (br d, *J*=1.3 Hz, 3H, H-28), 1.602 (br d, *J*=1.4 Hz, 3H, H-29), 1.615 (m, 1H, H-2β), 1.660 (m, 1H, H-1β), 1.681 (br d, *J*=1.3 Hz, 3H, H-30), 1.705 (m, 1H, H-2α), 1.712, (m, 1H, H-12α), 1.732<sup>7</sup> (m, 1H, H-6α), 1.736<sup>7</sup> (m, 1H, H-7α), 1.858 (dq, *J*=14.6, 7.9 Hz, 1H, H-16), 1.941 (br dd, *J*=16, 7 Hz, 1H, H-7β), 1.98 (m, 2H, H-19), 2.03 (m, 1H, H-16), 2.036 (m, 1H, H-11α), 2.06 (m, 2H, H-20), 2.160 (m, 1H, H-11β), 3.252 (dd, *J*=11, 4 Hz, 1H, H-3), 5.097 (t of septet, *J*=7.1, 1.4 Hz, 1H, H-21), and 5.112 (t of septet, *J*=7, 1.3 Hz 1 H, H-17). 125 MHz <sup>13</sup>C NMR: δ 14.28 (C-27), 15.47 (C-24), 15.98 (C-28), 17.68 (C-29), 18.98 (C-6), 19.45 (C-25), 23.39 (C-7), 25.16 (C-26), 25.71 (C-30), 26.47 (C-16), 26.75 (C-20), 27.54 (C-11), 27.82 (C-2), 28.18 (C-23), 30.73, (C-12), 32.13 (C-15), 35.61 (C-1), 35.76 (C-10), 38.11 (C-14), 38.73 (C-4), 39.77 (C-19), 51.59 (C-5), 52.72 (C-13), 79.23 (C-3), 124.37 (C-21), 125.00 (C-17), 131.28 (C-22), 134.87 (C-18), 138.00 (C-8), and 143.38 (C-9).

**Determination of the thalianol epoxide structure:** The <sup>1</sup>H NMR spectrum of the more polar component (Figure S6) showed it to be closely related to thalianol. Signals were present within 0.002 ppm of the H-3 carbinol proton and each non-allylic methyl (H-23, H-24, H-25, H-26, H-27) of thalianol (Figure S4), indicating that the ring system and the atoms neighboring the non-allylic side-chain methyl (H-27) are the same as those in thalianol. However significant differences existed in other side chain signals. Although the H-17 olefinic proton (δ 5.162 ppm) and the H-28 allylic methyl protons (δ 1.609 ppm) paralleled those in thalianol, the absence of olefinic and allylic signals from the terminal olefin indicated that this portion of the more polar component was distinct from thalianol. New <sup>1</sup>H NMR signals at 5.162, 1.259, and 1.300 were consistent with the epoxide moiety of oxidosqualene as judged by comparisons with NMR data for oxidosqualene and 24,25-epoxylanosterol.<sup>8</sup> These results indicated the more polar component to be thalianol epoxide (Figure S2). Bis-oxidosqualene is known to accumulate in yeast lanosterol synthase mutants, and thalianol epoxide probably arose by normal thalianol synthase action on bis-oxidosqualene. Both stereocenters in bis-oxidosqualene generated by yeast are known to be *S*, and the C-21 stereocenter in thalianol should consequently also be *S*. Partial 500 MHz <sup>1</sup>H NMR: δ 0.820 (s, 3H, H-24), 0.836 (d, 6.6 Hz, 3H, H-27), 0.932 (s, 3H, H-26), 0.947 (s, 3H, H-25), 1.019 (s, 3H, H-23), 1.259 (s, 3H, H- 29), 1.300 (s, 3H, H- 30), 1.609 (br d, 1.2 Hz, 3H, H-28), 2.701 (t, 6.2 Hz, 1H, H-21), 3.252 (dd, 11, 4 Hz, H, H-3), 5.162 (t of sextet, 7, 1.3 Hz, 1H, H-17).

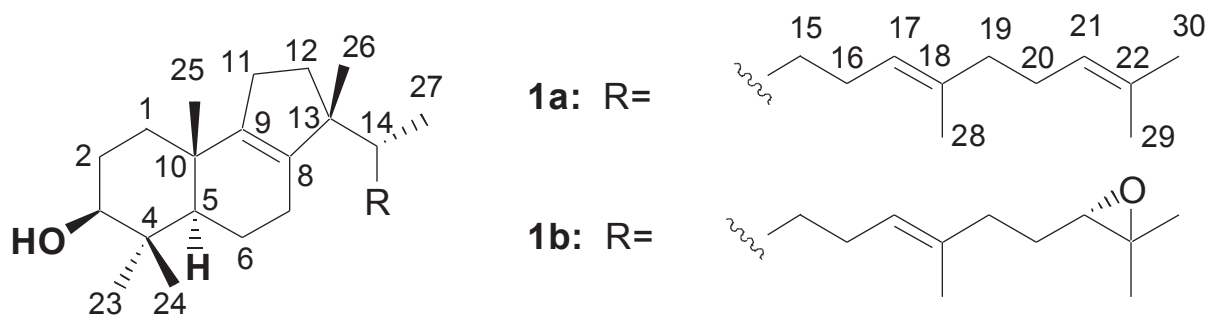
<sup>7</sup> The signals at 1.732 and 1.736 that were assigned to H-6α and H-7α might be exchanged.

<sup>8</sup> Emmons, G.T.; Wilson, W. K.; Schroepfer, G. J., Jr. *J. Lipid Res.* **1989**, *30*, 133-138.

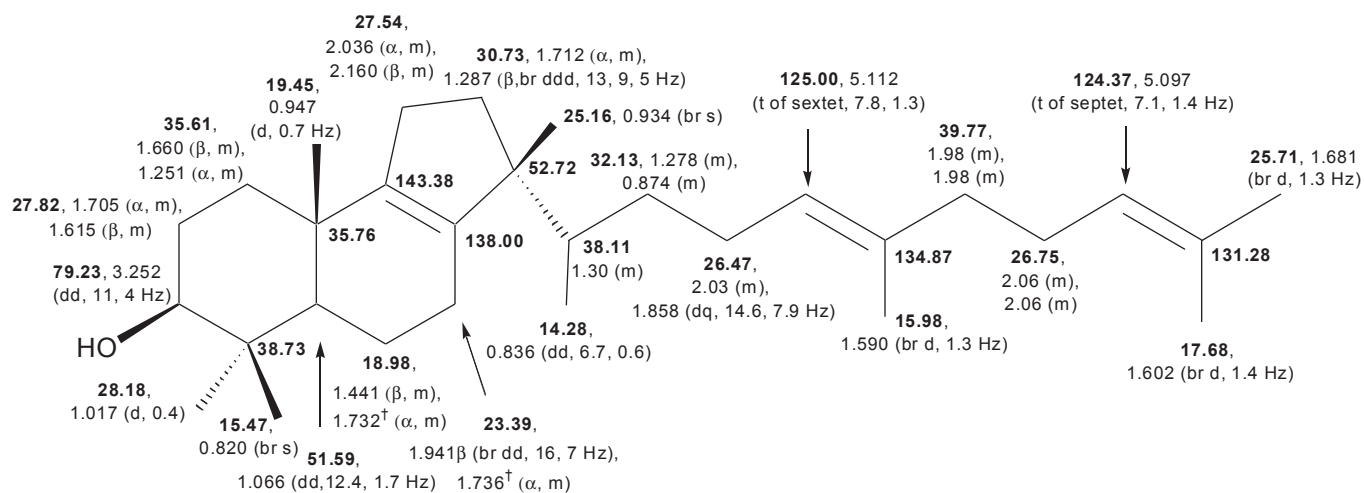
**Figure S1.** EI Spectrum of TMS-thalianol



**Figure S2.** Tricyclic triterpenoids produced by At5g48010.



**Figure S3.** <sup>13</sup>C and <sup>1</sup>H assignments for thalianol. <sup>13</sup>C signals are in bold and <sup>1</sup>H signals are in plain font.



**Figure S4.** <sup>1</sup>H assignments for thalianol epoxide. Signal assignments and stereochemistry and regiochemistry in ring A were deduced by analogy with thalianol. Signal assignments for H26 and H27 are based on data for 24,25-epoxylanosterol.<sup>8</sup>

