Supporting Information

Product Profile of PEN3: The Last Unexamined Oxidosqualene Cyclase in *Arabidopsis thaliana*

Pietro Morlacchi, William K. Wilson, Quanbo Xiong, Aparna Bhaduri, Diana Sttivend, Mariya D. Kolesnikova, and Seiichi P. T. Matsuda*

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

**Email: matsuda@rice.edu*

This pdf file is best viewed with Acrobat 5.0 or greater on Windows XP, Vista, and Macintosh OS 10.4-10.5. Accurate visualization of this file is not assured with older operating systems.

Contents

Materials and Methods S1
Molecular Biology and Yeast Strain Construction S3
In Vivo Production of Triterpenes by PEN3 ..... S3
GC-MS and NMR Spectra of Tirucalla-7,24-dien-3β-ol (**2**) S7
NMR Signal Assignments for Tirucalla-7,24-dien-3β-ol (**2**) S14
Basis for the Structure Identification of the Major PEN3 Product **2** S20
Spectra and Structure Identification of Minor PEN3 Products S22
Estimation of the Ratio of PEN3 Products S25
Origin of 20*R* and 20*S* Triterpenes: Comparison of Two Mechanisms S25
Preliminary Characterization of LUP5 Products S28
Sequence Alignments of Oxidosqualene Cyclases S30
References and Notes S33

Materials and Methods

Materials. Restriction endonucleases and the Quick Ligation Kit were purchased from New England Biolabs (Beverly, MA, USA). GenePure^{TM} agarose LE was from ISC BioExpress (Kaysville, UT, USA). The Qiagen Gel Extraction Kit used for DNA recovery from agarose gels was from Qiagen (Valencia, CA, USA). The RETROscript RT-PCR kit was from Ambion (Austin, TX, USA), Ex-Taq polymerase was from Panvera (Madison, WI, USA), and the pGEM-T vector was from Promega (Madison, WI, USA). Bacterial and yeast media components were obtained from United States Biological (Swampscott, MA, USA).

Heme (in the form of hemin chloride), ergosterol, bis(trimethylsilyl)trifluoroacetamide (BSFTA), and Diaion HP-20 were from Sigma-Aldrich (St. Louis, MO, USA). Silica gel and silica gel TLC plates were from EMD Chemicals (Gibbstown, NJ, USA). Deuterated chloroform ($CDCl_3$) for NMR was obtained from Cambridge Isotope Laboratories (Andover, MA, USA), and 5-mm NMR tubes were from Wilmad (Buena, NJ, USA) or Shigemi (Allison Park, PA, USA). Organic solvents for extraction and saponification were from EMD Chemicals or Honeywell Burdick & Jackson (Muskegon, MI, USA). Pyridine and other chemical reagents were obtained from Fischer Scientific (Fair Lawn, NJ, USA).

Preparative TLC (PTLC). PTLC was performed on 20 × 20 cm glass-backed plates (250-μm layer of silica gel) from EMD Chemicals. Prior to use, the plates were washed by developing with 1:1 methanol-dichloromethane and activated in a 100 °C oven for ~15 h. Products were visualized by staining the plate with *p*-anisaldehyde and heating on a hot plate.

GC-MS. GC-MS analysis was performed on an Agilent 5973 MSD connected to an Agilent 6890 GC system. Samples were injected either splitless or in a 40:1 split with an injector temperature of 280 °C. The Restek Rtx-35ms GC column (30 m × 0.25 mm × 0.1 μm) was held isothermally at 260 °C. MS data were acquired with electron impact (EI) ionization at 70 eV in full-scan mode (50 to 650 amu) after a 3-min solvent delay.

Trimethylsilyl (TMS) ether derivatives of triterpene alcohols were prepared by dissolving each sample in 1:1 dry pyridine/BSTFA and keeping the mixture at 37 °C for 2 h.

NMR. ^{1}H, ^{13}C, and 2D NMR spectra were acquired at 25 °C in $CDCl_3$ on a Bruker Avance DRX 500, Varian Inova 600, or Varian Inova 800 spectrometer. Chemical shifts were referenced to tetramethylsilane at 0 ppm for ^{1}H or to $CDCl_3$, at 77.0 ppm for ^{13}C. Spectra were analyzed with Bruker xwinnmr 2.5 and Topspin 2.0. 1D spectra shown in the figures were transformed with very mild resolution enhancement, whereas coupling constants were measured from strongly resolution-enhanced spectra (e.g., LB -2 and GB ≥ 0.4). Fine couplings are visible only in strongly resolution-enhanced spectra and not in the figures shown herein.

HPLC Conditions. HPLC analysis was performed on an Agilent 1100 HPLC system equipped with a Rheodyne 7125 injector and ultraviolet (UV) detection at 210 nm. Separations were carried out using an ODS column and a mobile phase of methanol-water.

Molecular Modeling. Molecular mechanics calculations were done in PCMODEL version 8 (Serena Software; Bloomington, IN, USA), which was used for construction and initial optimization of triterpene geometries. PCMODEL was also used to perform conformer searches (GMMX module), view and analyze triterpene structures, and predict vicinal coupling constants from an extended Karplus relationship.

Quantum mechanical calculations were done with Gaussian 03 (Linux version C.01 or D.01).[1] Geometry optimizations were performed with B3LYP/6-31G*, and B3PW91/6-311G(2d,p) was used for single-point energies, which were done with SCF=tight. Energies are given as kcal/mol. All models were considered as closed-shell systems (restricted calculations) with the frozen-core approximation. NMR shieldings were calculated using the GIAO method at the B3PW91/6-311G(2d,p)//B3LYP/6-31G* level.

Molecular Biology and Yeast Strain Construction

mRNA obtained from 2-day old *Arabidopsis thaliana* seedlings was used to transcribe first strand cDNA with a RETROscript RT-PCR kit following the manufacturer's instructions. The *PEN3* (*At5g36150*) open reading frame (2285 bp) was amplified using primers 5'-atatGTCGACatgtggaggctgaggatcggagc-3' and 5'-aGCGGCCGCtcaaaggagcaaccgtatagcc-3' with a 40-cycle program: 95 °C, 30 s for denaturation, 56 °C, 30 s for annealing, and 59.5 °C, 3 min for extension, with Ex-Taq polymerase. The amplicons were purified with a Qiaquick Kit, ligated into the pGEM-T Easy vector with a Quick Ligation Kit, and used to transform *E. coli* strain DH5α.[2] The inserts were sequenced, and subclones were made to remove PCR errors. Restriction sites for subcloning were installed by amplifying the full-length *At5g36150* coding sequence using the same program and the primers 5'-ccGTCGACtataatgtggaggctgaggat-3' and 5'-gcccGCGGCCGCtatatcaaaaggagcaaccgtatagcc-3'. The purified *PEN3* coding sequence was excised with *Sal* I and *Not* I restriction endonucleases and cloned into the yeast expression vector pRS426GAL[3] using a Quick Ligation Kit. Sequencing established the clone to be error-free, and it was renamed pDS3.0. The insert was also subcloned into the integrative plasmid pRS305GAL[4] similarly to make pDS3.1.

Saccharomyces cerevisiae strain SMY8[5] was transformed with plasmid pDS3.0 using the lithium acetate method.[6] The transformant was selected on synthetic complete medium[2] lacking uracil and containing 2% dextrose, hemin chloride (13 μg/mL), ergosterol (20 μg/mL), and Tween 80 (5 mg/mL), and solidified with 1.5% agar. The plates were incubated at 30 °C until colonies formed. The resultant recombinant yeast strain was named SMY8[pDS3.0].

SMY8 was also transformed with pDS3.1, but SMY8[pDS3.1] showed no advantage over SMY8[pDS3.0]. Also, yeast strain RXY6,[7] which lacks both squalene epoxidase and lanosterol synthase, was transformed with pDS3.0 and pDS3.1. However, in vitro reactions of oxidosqualene with RXY6[pDS3.0] and RXY6[pDS3.1] homogenates appeared not to give any triterpene products. Consequently, RXY6[pDS3.0], RXY6[pDS3.1], and SMY8[pDS3.1] are not described further. In another attempt to improve the modest triterpene output of SMY8[pDS3.0], the GAC codon immediately upstream of the PEN3 start site, was replaced by ATA, a preferred codon for start sites in yeast.[8] Again, no improvement in triterpene production was observed.

In Vivo Production of Triterpenes by PEN3

Experiment 1 (large-scale culture). Cultures of SMY8[pDS3.0] totaling 24 L were grown in synthetic complete medium lacking uracil and supplemented with hemin chloride (13 μg/mL), ergosterol (20 μg/mL), and Tween 80 (5 mg/mL); 2% galactose was used as the carbon source. The cultures were allowed to grow to saturation at 30 ºC with shaking (250 rpm). Cells (103 g) were harvested by centrifugation (3800 rpm for 20 min) and saponified with 10% KOH in 80:20 ethanol-water at 70 °C for 2 h. After saponification, products were extracted with hexanes (3 × 250 mL). The combined extracts were washed with brine (3 × 50 mL) and rotary evaporated to dryness, yielding 60 mg of nonsaponifiable lipids (NSL). An aliquot (~2%) of the NSL was analyzed by GC-MS, which showed the presence of **2**, the dominant PEN3 product.

The NSL (~60 mg) from the SMY8[pDS3.0] cultures was loaded onto a silica gel column (6 g; 230-400 mesh) and eluted with gradients of ethyl ether in hexanes (2-100%). Nine fractions were collected and analyzed by GC-MS and ^{1}H NMR. Fraction 1 (4.8 mg; elution with 2% ether in hexanes) contained squalene. Fraction 2 (3.2 mg; further elution with 2% ether in hexanes) contained oxidosqualene. Fractions 3-6 (5.8 mg; elution with 4% ether in hexanes) contained PEN3 products, dioxidosqualene, and other material. Fractions 7-8 (15.6 mg; elution with 14-20% ether in hexanes) contained ergosterol. Fraction 9 (3.1 mg; elution with 100% ether) did not show the presence of any potential PEN3 products. Fractions 3-6 were further subjected to column chromatography (4 g silica gel, 230-400 mesh; elution with dichloromethane). Two fractions (A and B) were collected and analyzed by GC-MS and ^{1}H NMR. Fraction A (1.1 mg) contained PEN3 products and Fraction B (4.2 mg) contained mainly dioxidosqualene. Fraction A was spotted onto a 20 × 20 cm silica gel TLC plate (250-µm layer), which was developed with 1:1 ether-hexanes. The triterpene alcohol region was divided into bands A1 and A2, corresponding to monocyclic and polycyclic triterpene regions, respectively), which were scraped onto a small column and eluted with dichloromethane. An aliquot of each band was analyzed by GC-MS and ^{1}H NMR. A flowchart of these chromatographic procedures is shown in Figure S1.

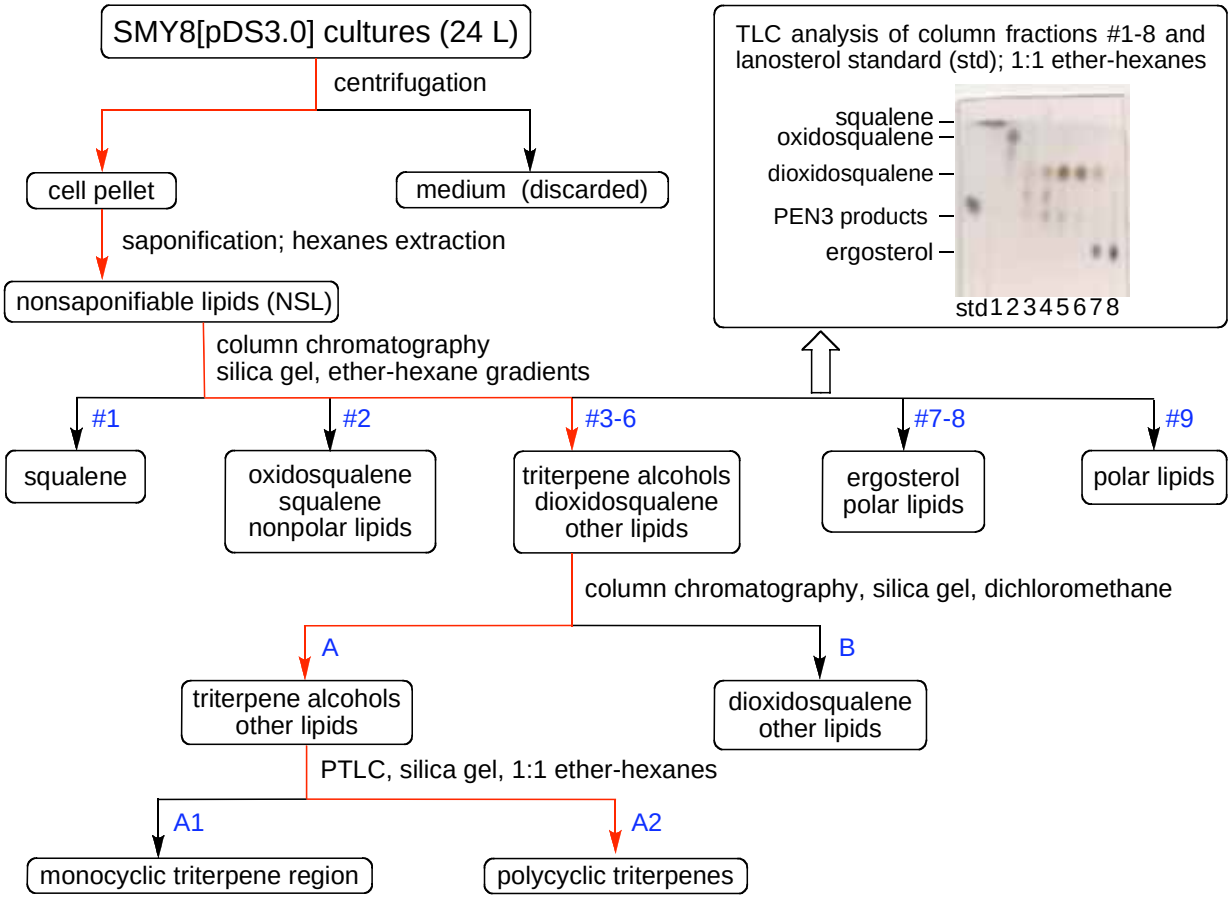


Figure S1. Flowchart summarizing the isolation of PEN3 products in Experiment 1. Chromatographic fractions (blue text) are given in order of decreasing mobility from left to right.

Experiment 2 (additional evidence for the PEN3 product profile; estimation of distortion from differential loss of triterpenes to the culture medium). In Experiment 2, SMY8[pDS3.0] cultures totaling 4 L were grown as in Experiment 1. In order to monitor loss of triterpenes to the culture medium, the PEN3 products were isolated by three different approaches. One workup (**2.1**) consisted of lyophilization of 0.5 L of the culture (without centrifugation), followed by saponification, chromatographic purification, and spectral analysis of PEN3 products. For the other workups, the remaining 3.5 L of culture was centrifuged to give a cell pellet and supernatant (culture medium). In the second workup (**2.2**), the cell pellet was saponified, followed by extraction and chromatographic isolation as described in Experiment 1. In the third workup (**2.3**), the supernatant was incubated with a polyaromatic adsorbent resin for hydrophobic compounds, followed by chromatographic isolation of triterpenes. A flowchart summarizing these procedures is given in Figure S2.

2.1 *Lyophilization workup.* A 500-mL aliquot of the 4-L culture of SMY8[pDS3.0] was grown using the experimental conditions previously described. Upon saturation, the culture was frozen in a dry ice-acetone bath and lyophilized. The lyophilized sample was saponified with 10% KOH in 80:20 ethanol-water at 70 ºC for 2 h, followed by extraction with hexanes (3 × 200 mL). The combined extracts were washed with brine (3 × 50 mL), dried over anhydrous sodium sulfate, and rotary evaporated to a residue (47 mg). Column chromatography (7 g silica gel, 230-400 mesh; elution with 2-100% ether in hexanes) gave four fractions (1-4), which were analyzed by GC-MS and ^{1}H NMR. PEN3 products were found in Fraction 2 (elution with 4% ether) together with dioxidosqualene and ergosterol. This fraction (1.4 mg) was further subjected to column chromatography (3 g silica gel, 230-400 mesh; elution with dichloromethane), giving fractions A and B. Both fractions were analyzed by GC-MS and ^{1}H NMR. Fraction A, containing the PEN3 products, was purified by PTLC (developed with 1:1 ether-hexanes). The silica layer was divided in four bands, A1-A4 in order of increasing polarity. Each band was scraped onto a small column, eluted with ether, and analyzed by GC-MS and ^{1}H NMR. PEN3 products were present in band A2 (0.5 mg).

2.2 *Analysis of triterpenes in the cell pellet.* This workup paralleled that of Experiment 1 and is summarized in Figure S2. All fractions were analyzed by GC-MS and ^{1}H NMR.

2.3 *Analysis of triterpenes in the culture medium.* After centrifugation of the 3.5-L culture at 3500 rpm for 20 min, the cell-free supernatant was incubated with methanol-washed Diaion HP-20 resin (60 g) at 30 ºC with shaking at 250 rpm. After 48 h, the suspension was loaded onto a column to filter the resin from the supernatant. Elution of the resin with ethanol (4 × 200 mL), followed by rotary evaporation gave a residue containing adsorbed triterpenes and detergent. The residue was dissolved in ether and passed through a silica-gel plug (3 g silica gel, 230-400 mesh) to remove highly polar material. To remove detergent, the sample was saponified with 10% KOH in 80:20 ethanol-water at 70 ºC for 2 h. After saponification, the sample was diluted with one volume of water, and triterpenes were extracted with hexanes (4 × 200 mL). The hexane extracts were washed with brine (3 × 50 mL), and the solvent was removed by rotary evaporation, yielding 82 mg of sample. The sample was then subjected to column chromatography and PTLC using the same experimental conditions described above. The purified PEN3 products were analyzed by GC-MS and ^{1}H NMR, which both provided product ratios (described on pages S25-S26).

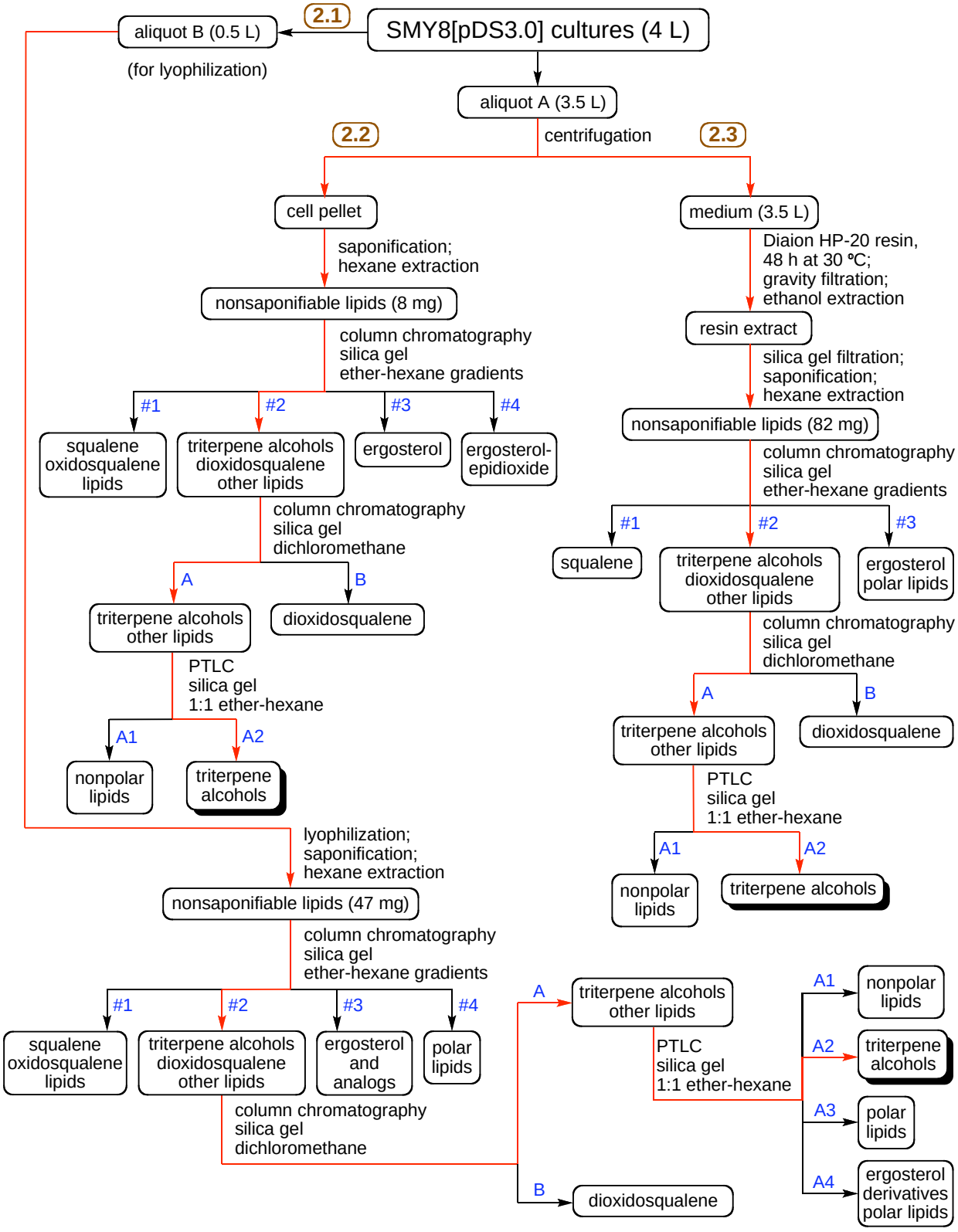


Figure S2. Flowchart summarizing the isolation of PEN3 products in Experiment 2. Chromatographic fractions (blue text) are given in order of decreasing mobility.

Experiment 3 (isolation of minor products). In another set of cultures totaling 10 L, the cell pellet was saponified, and the NSL was purified by a single long chromatography column (130 g silica gel; ether-hexane gradients). Selected fractions were subjected to reversed-phase HPLC (250 x 10 mm ODS column, 5μ particle size; gradient elution with 93:7 to 97:3 methanol-water) in order to isolate individual triterpene products. All column and HPLC fractions were analyzed by GC-MS and NMR.

GC-MS and NMR Spectra of Tirucalla-7,24-dien-3β-ol (2)

The major PEN3 product (**2**) was characterized by GC-MS (Figure S3), ^{1}H NMR (Figure S4), DEPT and ^{13}C NMR (Figure S5), COSYDEC (Figure S6), HSQC (Figure S7), and HMBC (Figures S8-S9). Except for the DEPT and ^{13}C NMR results (from Experiment 3), all GC-MS and NMR spectra of **2** are from Experiment 1, PTLC band A2. The 2D and ^{1}H NMR spectra were acquired using a cold probe.

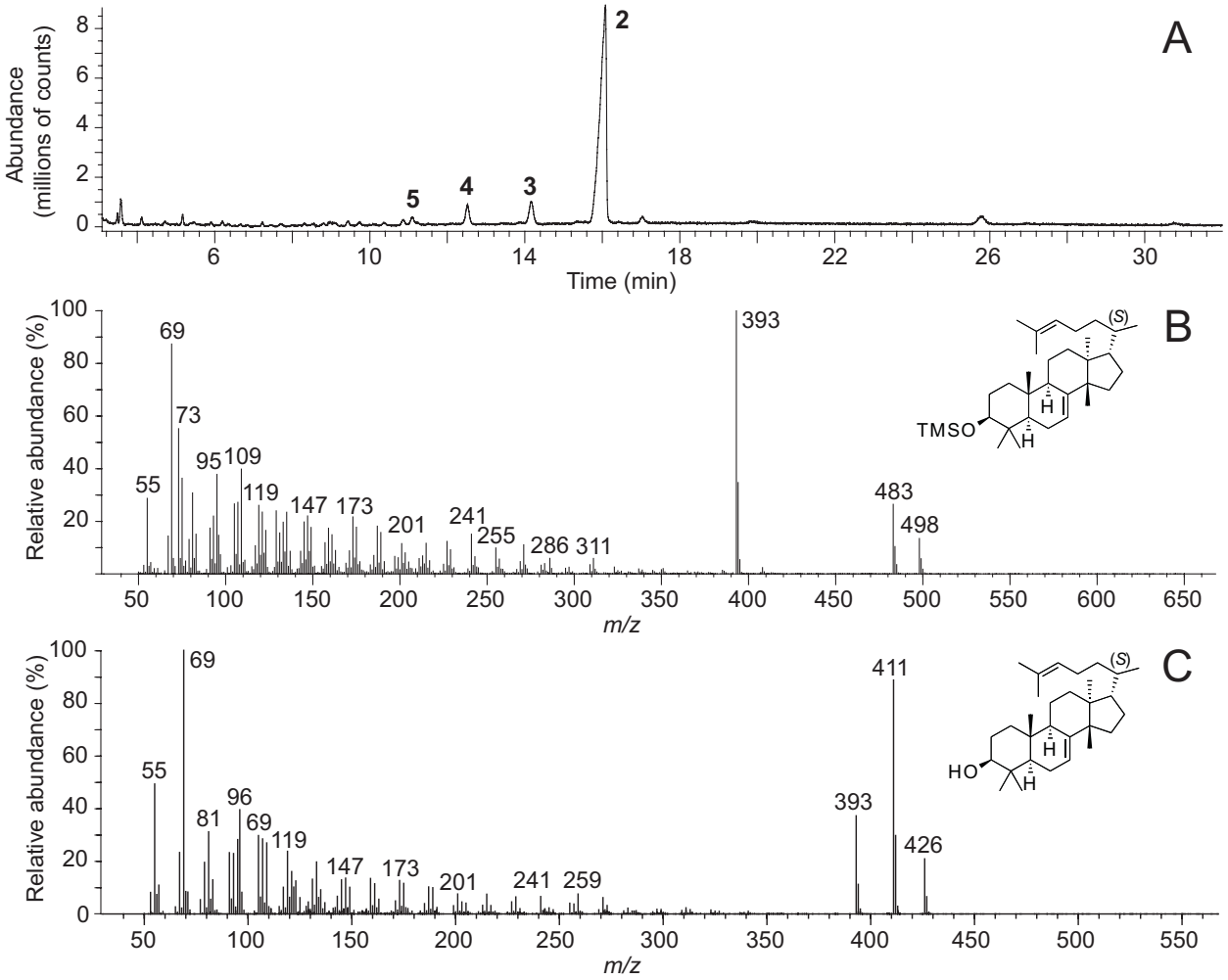


Figure S3. GC-MS analysis of Experiment 1, PTLC band A2: Total ion chromatogram (A) and mass spectra of tirucalla-7,24-dien-3β-ol (**2**) as its TMS ether (B) and without derivatization (C).

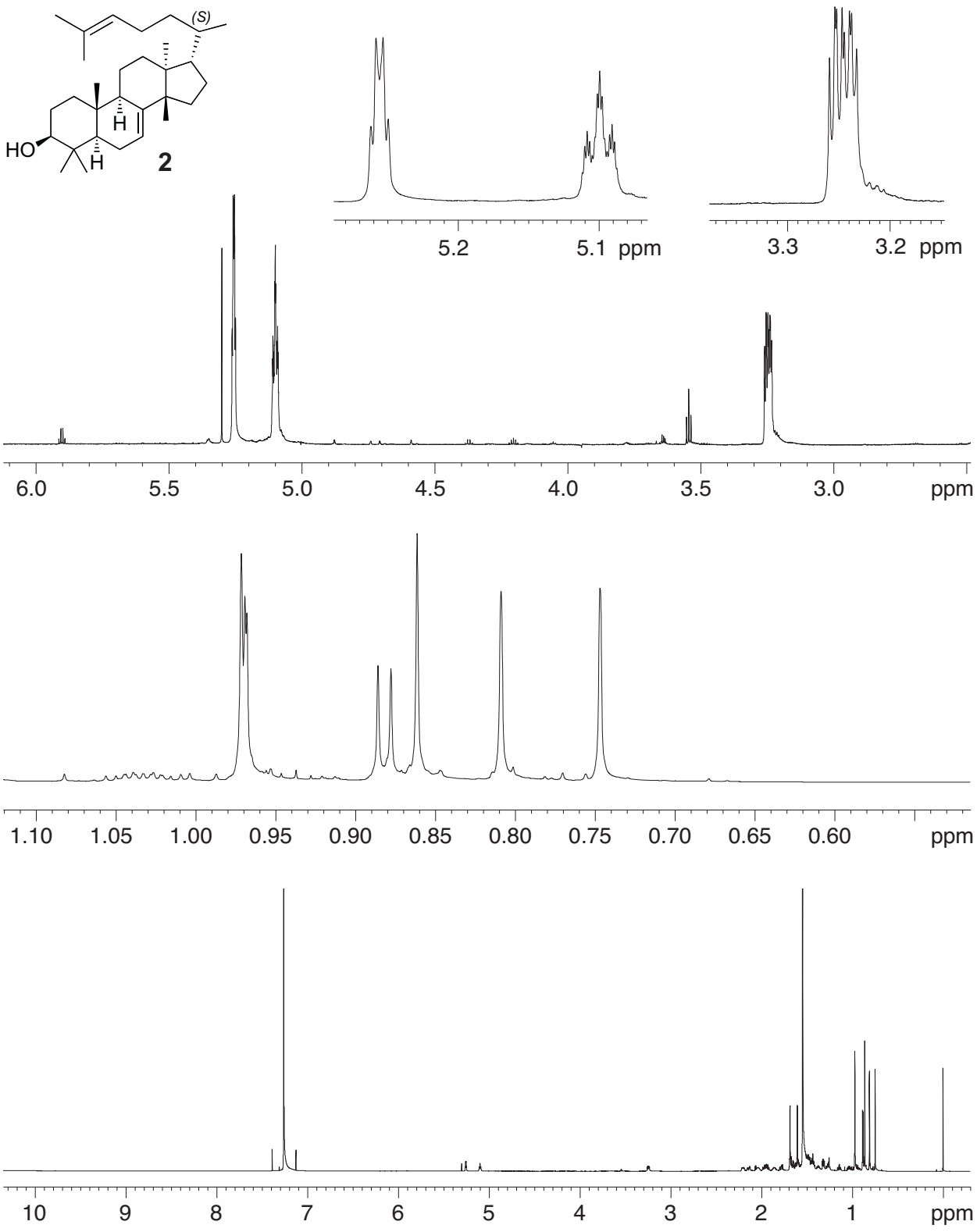


Figure S4. ^{1}H NMR spectrum of tirucalla-7,24-dien-3β-ol (**2**; Experiment 1, PTLC band A2): 800 MHz; 25 °C; ~2 mM solution of **2** in $CDCl_3$; referenced to $Si(CH_3)_4$ at 0 ppm.

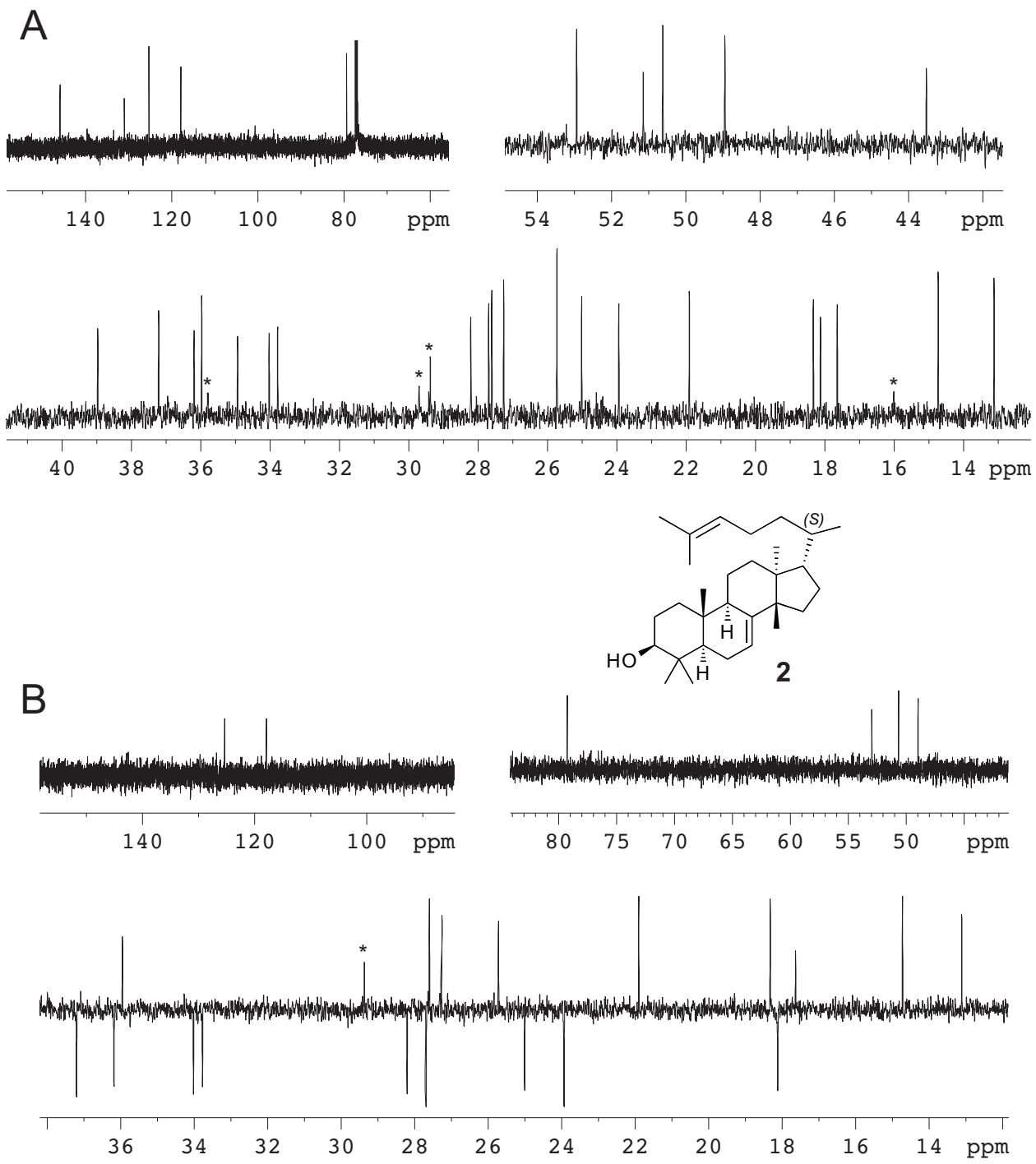


Figure S5. ^{13}C NMR (A) and DEPT (B) spectra of tirucalla-7,24-dien-3β-ol (**2**) from an HPLC fraction of Experiment 3: 125 MHz; 25 °C; ~5 mM solution of **2** in $CDCl_3$; inverse-gated decoupling (1-s acquisition time, 1.5-s relaxation delay). Chemical shifts were referenced to the central line of the $CDCl_3$ triplet at 77.0 ppm. Asterisks (*) denote minor impurities.

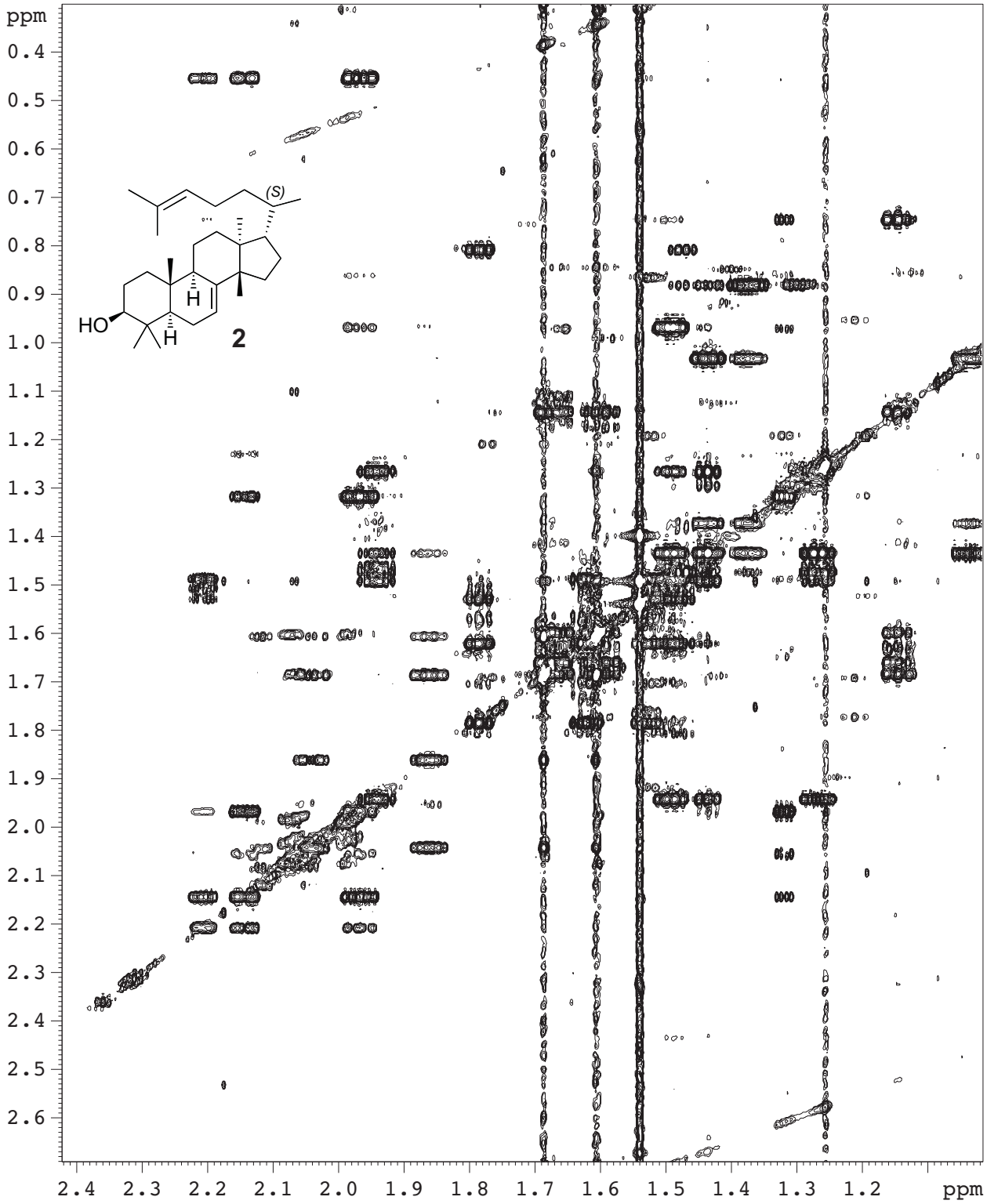


Figure S6. COSYDEC spectrum of tirucalla-7,24-dien-3β-ol (**2**; Experiment 1, PTLC band A2): 800 MHz; 25 °C; ~2 mM solution of **2** in $CDCl_3$; 276 t_1 increments; 200-ms constant time period; δ_H 0.3-2.7 window in f_1. Downfield signals for H3α, H7, and H24 are nominally aliased in f_1 by 2.4 or 4.8 ppm. Only the upfield region of the f_2 window (δ_H 7.9 to –0.1) is shown.

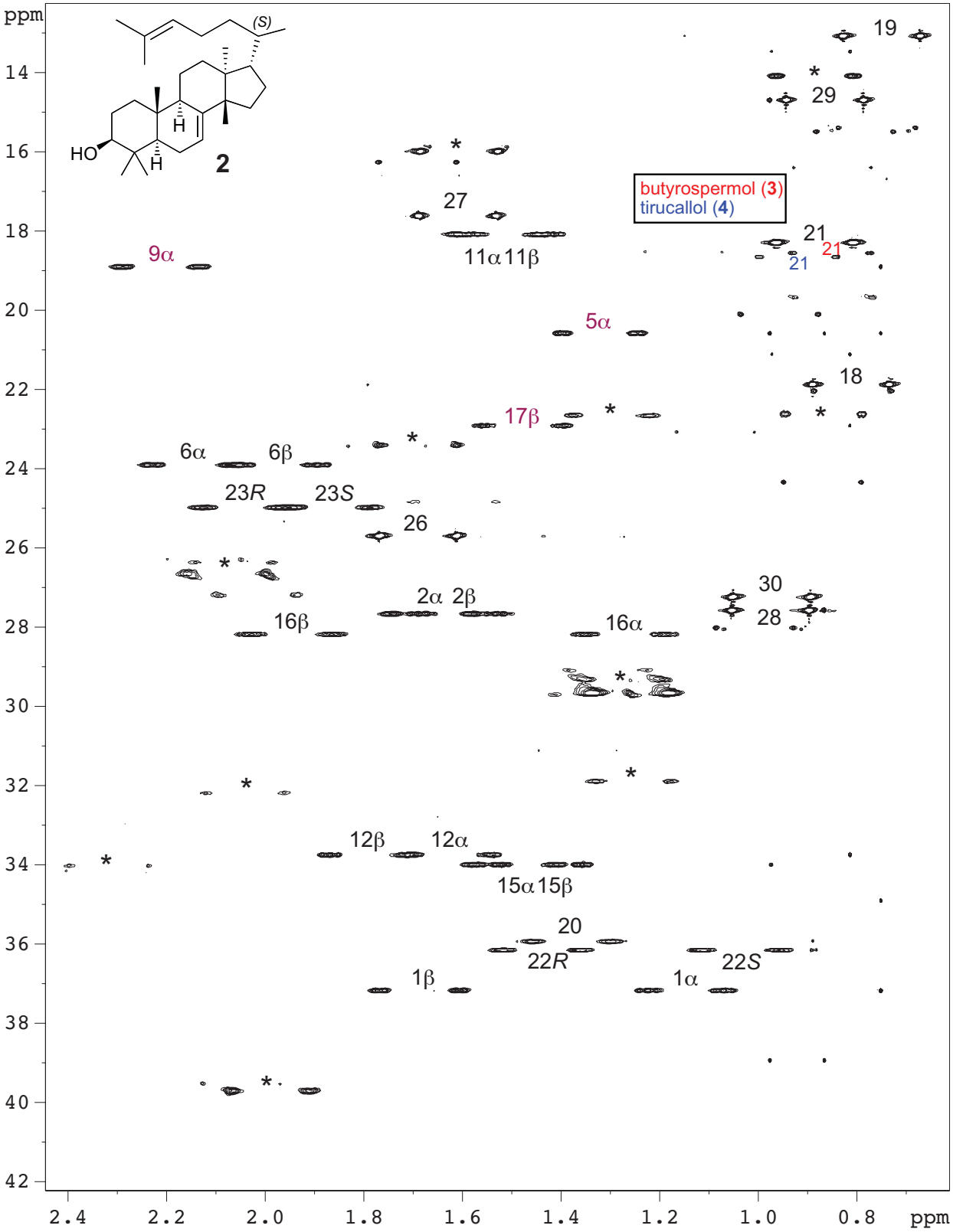


Figure S7. Coupled HSQC spectrum of tirucalla-7,24-dien-3β-ol (**2**): 800 MHz for ^{1}H; 25 °C; ~2 mM solution of **2** in $CDCl_3$; 348 complex points in t_1; δ_C 12.2-42.2 window in f_1. Assignments for signals aliased in f_1 (by 30.0 ppm) are shown in reddish brown. Asterisks denote impurities.

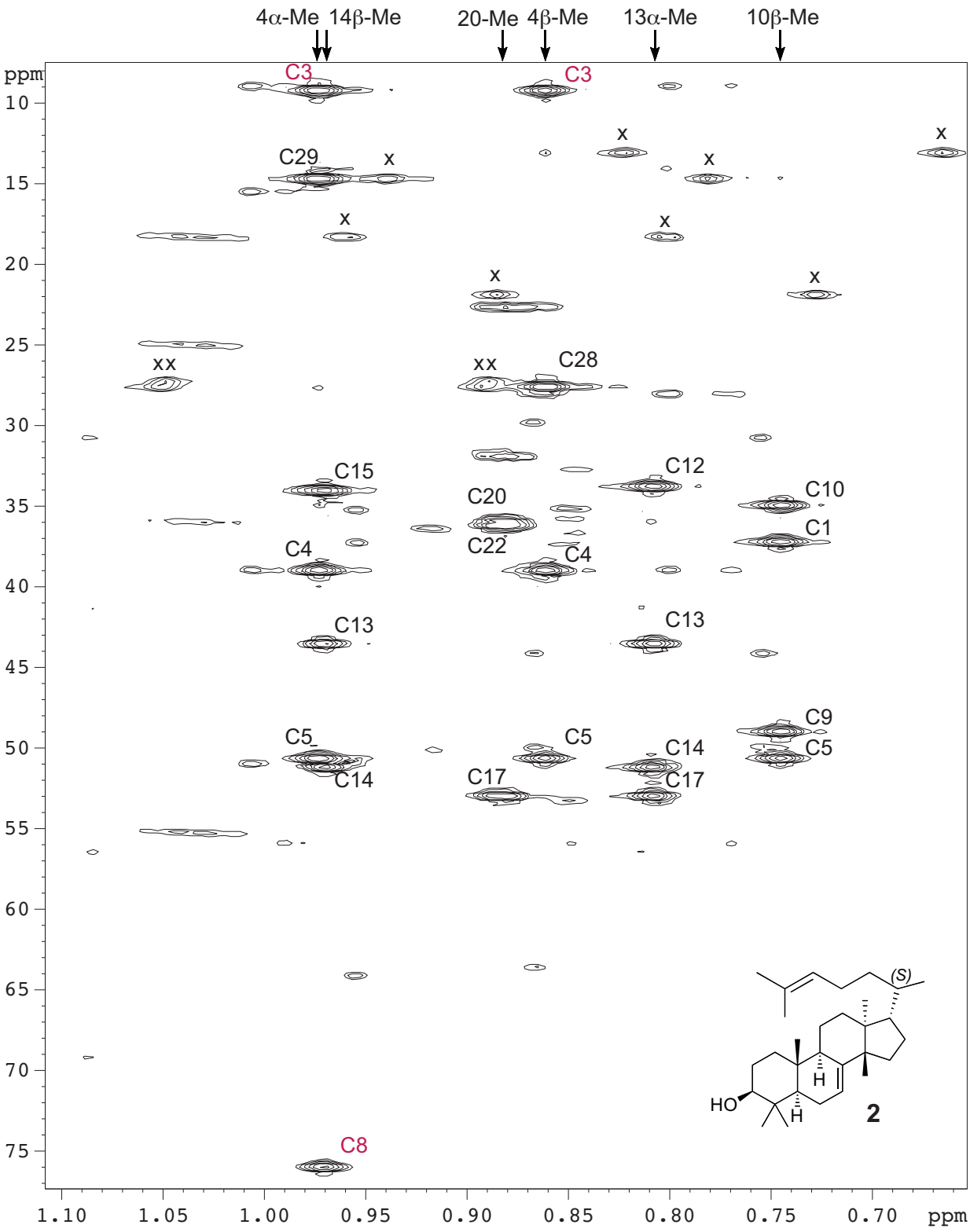


Figure S8. HMBC spectrum of tirucalla-7,24-dien-3β-ol (**2**; Experiment 1, PTLC band A2): upfield methyl region: 800 MHz for ^{1}H; 25 °C; ~2 mM solution of **2** in $CDCl_3$; 346 t_1 increments; δ_C 7.4-77.4 window in f_1. Assignments for signals aliased in f_1 (by 70.0 ppm) are shown in reddish brown. Artifactual signals from one-bond couplings are marked by "x".

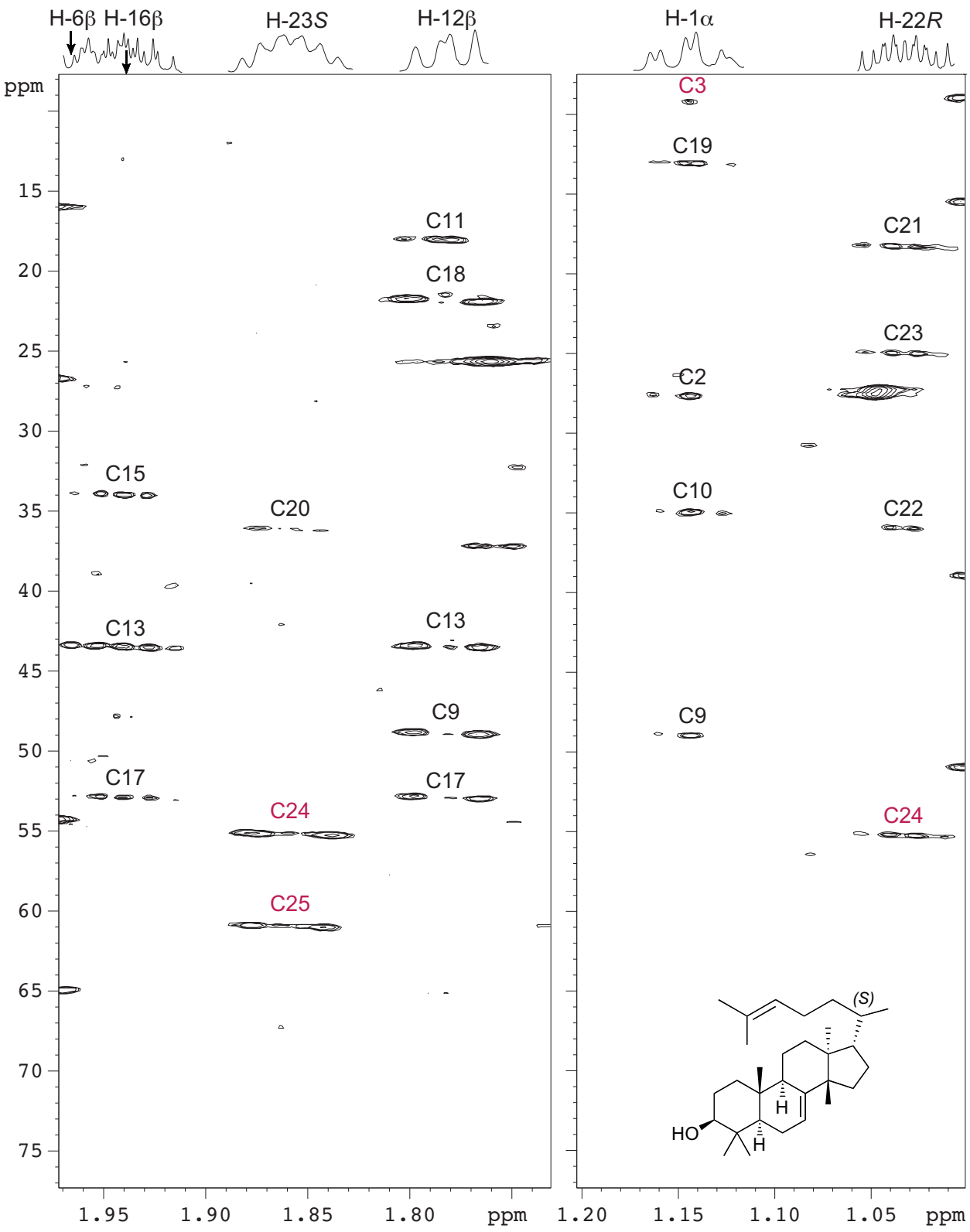


Figure S9. HMBC spectrum of tirucalla-7,24-dien-3β-ol (**2**), selected portions of the ^{1}H methylene envelope: same parameters as in Figure S8. Assignments for signals aliased in f_1 (by 70.0 ppm) are shown in reddish brown. The very weak correlations for H-6β are not visible here.

NMR Signal Assignments for Tirucalla-7,24-dien-3β-ol (2)

Reports of ^{1}H NMR signals for **2** appear to be limited to methyl, olefinic, and carbinol resonances.[9,10,11] We could not locate any ^{13}C NMR data for **2**, apart from a list of unassigned chemical shifts in a patent[11] and assigned values for the acetate derivative of **2**.[9] We assigned the ^{1}H and ^{13}C NMR signals for **2** from 1D and 2D NMR spectra (Figures S4-S9). The resulting assignments closely matched values predicted from quantum mechanical calculations (described below). Signal assignments are shown on the chemical structure in Figure S10, a format useful for interpreting 2D NMR results. The same assignments, together with ^{1}H-^{1}H coupling constants, are given in Table S1.

Most ^{13}C signals could be assigned from the HMBC spectrum (Figures S8-S9), and the attached protons were trivially assigned from the HSQC spectrum (Figure S7). The few remaining signals could be assigned from the COSYDEC spectrum (Figure S6). The stereochemistry of the diastereotopic protons was deduced from comparisons of observed and predicted coupling constants (Table S2) and/or by comparison with predicted chemical shifts.

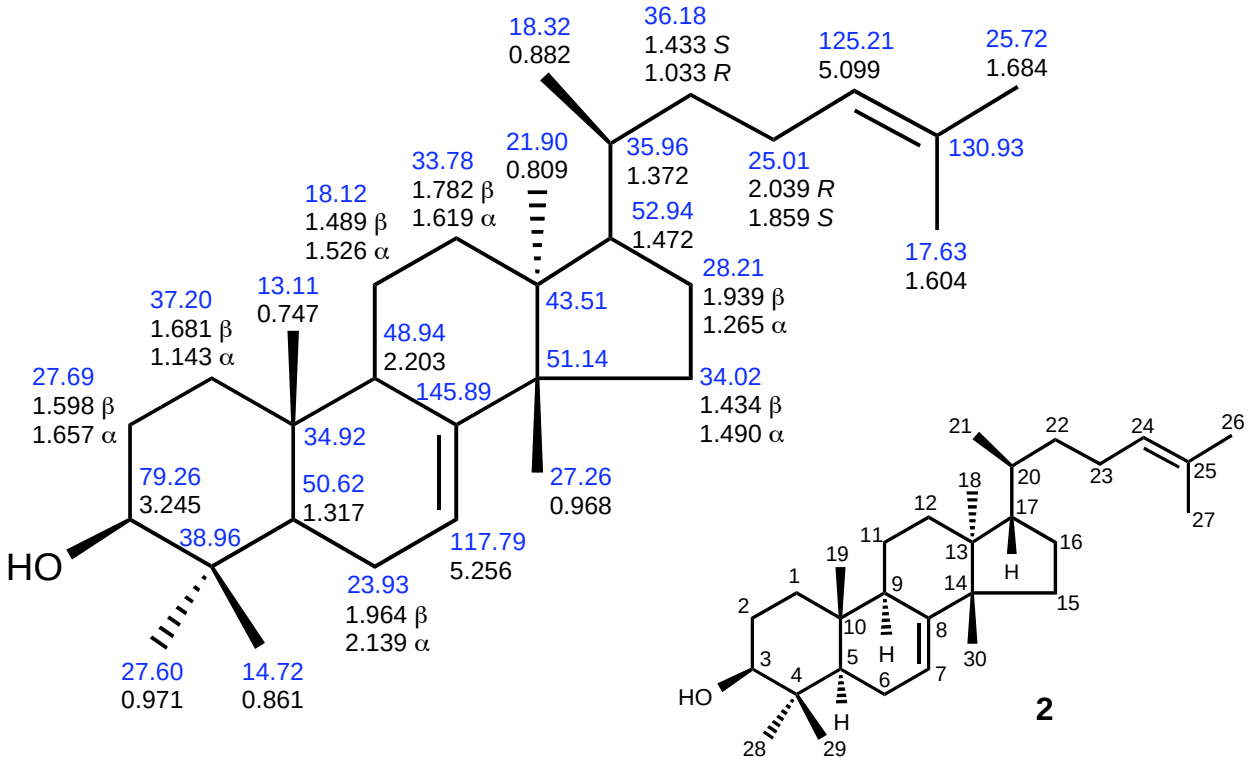


Figure S10. ^{1}H and ^{13}C NMR signal assignments and atom numbering for tirucalla-7,24-dien-3β-ol (**2**). ^{13}C NMR chemical shifts are in blue text.

Table S1. NMR chemical shifts and coupling constants for tirucalla-7,24-dien-3β-ol (**2**) [a]

^{13}C chemical shifts		^{1}H chemical shifts		Scalar ^{1}H-^{1}H couplings
atom	δ_C	atom	δ_H	(Hz)
C-1	37.20	H-1α	1.143	td, 13.3, 4.0
C-2	27.69	H-1β	1.681	dt, *~13*, 3.5
C-3	79.26	H-2α	*1.657*	dq, 13.0, 3.8
C-4	38.96	H-2β	*1.598*	dddd, 13.9, 13.0, 11.8, 3.7
C-5	50.62	H-3α	3.245	ddd, 11.6, 5.7, 4.0
C-6	23.93	H-5α	1.317	dd, 12.1, 5.6
C-7	117.79	H-6α	2.139	dddd, 17.8, 5.6, 4.3, 2.6
C-8	145.89	H-6β	1.964	dddd, 17.8, 12.1, 2.9, 1.1
C-9	48.94	H-7	5.256	dt, 4.3, 2.9
C-10	34.93	H-9α	2.203	dtq, *~13.4, 3.7, 2.8*
C-11	18.12	H-11α	*1.526*	m
C-12	33.78	H-11β	*1.489*	m
C-13	43.51	H-12α	1.619	ddd, 14.2, 10.3, 8.8
C-14	51.14	H-12β	1.782	br dd, 14.2, 9.9
C-15	34.02	H-15α	*1.490*	m
C-16	28.21	H-15β	*1.434*	ddd, 12.4, 9.3, 2.3
C-17	52.94	H-16α	1.265	dddd, 13.7, 11.0, 8.3, 2.5
C-18	21.90	H-16β	1.939	dtd, 13.6, 9.4, 7.3
C-19	13.11	H-17α	1.472	br q, *~9*
C-20	35.96	H-18	0.809	d, 0.7
C-21	18.32	H-19	0.747	d, 0.8
C-22	36.18	H-20	1.372	tqd, 9.1, 6.5, 2.7
C-23	25.01	H-21	0.882	d, 6.4
C-24	125.21	H-22*R*	1.033	dddd, 13.7, 10.0, 8.9, 5.0
C-25	130.93	H-22*S*	1.433	dddd, *~13.7, 10.4, 6.4, 2.7*
C-26	25.72	H-23*R*	2.039	m
C-27	17.63	H-23*S*	1.859	dq, *~14, 8*
C-28	27.60	H-24	5.099	t of septet, 7.1, 1.4
C-29	14.72	H-26	1.684	qd, 1.3, 0.5
C-30	27.26	H-27	1.604	m
		H-28	0.971	d, 0.4
		H-29	0.861	s
		H-30	0.968	d, 1.2

[a] Chemical shifts were measured at 800 MHz (^{1}H) or 125 MHz (^{13}C) in dilute $CDCl_3$ solution, referenced to $SiMe_4$ (^{1}H) or $CDCl_3$ (^{13}C, 77.0 ppm), and corrected for strong coupling effects. Solute concentrations were ~5 mM for ^{13}C and ~2 mM for ^{1}H. Chemical-shift accuracy for ^{13}C is ca. ±0.03 ppm and for ^{1}H is ±0.001 ppm (or ±0.03 ppm for values in blue italics). Designations 22*R*, 23*S*, etc. denote pro *R* and pro *S* hydrogens. Coupling constants have an estimated accuracy of ±0.5 Hz except for italicized values, which have elevated uncertainty (~1 Hz) owing to unresolved couplings, extraction of couplings from 2D experiments, or distortion from strong coupling.

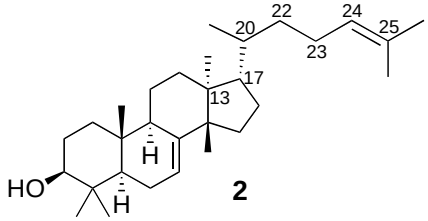
Table S2. Comparison of observed and predicted ^{1}H-^{1}H NMR coupling constants for **2** [a,b]

Atoms	Observed *J* (Hz)	Predicted *J* (Hz)	Deviation (Hz) [c]	Torsion angle (°)
1α-2α	4.0	3.6	–0.4	–57
1α-2β	13.8	13.5	–0.3	–174
1β-2α	3.5	3.2	–0.3	59
1β-2β	3.5	3.3	–0.2	–59
2α-3α	4.0	3.7	–0.3	63
2β-3α	11.7	11.5	–0.2	178
5α-6α	5.6	4.5	–1.1	–52
5α-6β	12.1	11.9	–0.2	–167
6α-7	4.3	4.3	0.0	–50
6β-7	2.9	3.4	0.5	64
9α-11α	3.7	3.7	0.0	56
9α-11β	13.4	12.2	–1.2	171
11α-12α	10.3	10.2	–0.1	–23
11α-12β	<1	0.3	<1	92
11β-12α	8.8	7.6	–1.2	–137
11β-12β	9.9	10.2	0.3	–23
15α-16α	11.0	11.3	0.3	–14
15α-16β	7.3	6.4	–0.9	–132
15β-16α	2.4	1.4	–1.0	105
15β-16β	9.4	11.4	2.0	–13
16α-17β	8.3	8.5	0.2	–142
16β-17β	9.4	8.7	–0.7	–25
17β-20	9.2	11.8	2.6	178
20-22*R*	9.0	10.5	1.5	
20-22*S*	2.7	2.3	–0.4	
22*R*-23*R*	5.0	3.4	–1.6	
22*R*-23*S*	10.0	10.5	0.5	
22*S*-23*R*	10.4	10.5	0.1	
22*S*-23*R*	6.4	4.5	–1.9	
23*R*-24	7.1	7.4	0.3	
23*S*-24	7.1	7.8	0.7	

[a] Observed coupling constants (*J* values; ±1 Hz) were derived from data in Table S1; observed geminal couplings: 1α-1β, 13.0 Hz; 2α-2β, 13.0 Hz; 6α-6β, 17.8 Hz; 12α-12β, 14.2 Hz; 15α-15β, 12.4 Hz; 16α-16β, 13.6 Hz; 22*R*-22*S*, 13.7 Hz. [b] Predicted *J* values for vicinal couplings (and the associated torsion angles in the ring system) were calculated from an extended Karplus relationship available in PCMODEL using B3LYP/6-31G* geometries optimized with Gaussian 03. The predicted *J* values shown represent the weighted average of couplings calculated for the eight most populated side-chain conformers of **2** (described below; comprising >98% of the population of conformers in Table S3). [c] Deviations correspond to predicted – observed values.

The ^{1}H and ^{13}C NMR chemical shift assignments for **2** were confirmed by comparison with shieldings predicted from quantum mechanical calculations using the GIAO method, as implemented in Gaussian 03. B3LYP/6-31G* geometries were optimized for the 15 side-chain conformers defined Table S3. Gas-phase NMR shieldings for the full $C_{30}H_{50}O$ structures were calculated at the B3PW91/6-311G(2d,p)//B3LYP/6-31G* level for each conformer. The Boltzmann distribution of conformers at 25 °C was calculated from B3PW91/6-311G(2d,p)//B3LYP/6-31G* energies and used to obtain a weighted average of the NMR shieldings. These averaged shieldings were further adjusted by applying correction factors for the conformational heterogeneity about the H-O-C3-C4 torsion angle. The corrected shieldings were converted to chemical shifts using empirical adjustments derived from comparisons between thousands of predicted and calculated chemical shifts in $CDCl_3$ solution for over 200 hydrophobic compounds. Observed ^{1}H and ^{13}C NMR chemical shifts are compared with predicted values in Table S4. Atomic coordinates are given for the most abundant conformer.

Table S3. Energies and relative populations for side-chain conformers of **2** [a]



2

Boltzmann	relative *E*	torsion angle (°)			
population (%)	kcal/mol	13-17-20-22	17-20-22-23	20-22-23-24	22-23-24-25
31.5	0.00	–178	171	–171	–116
23.3	0.18	–178	169	–179	116
16.1	0.40	–179	164	–67	126
10.5	0.65	–175	–64	178	117
8.2	0.80	–176	–63	–174	–114
6.8	0.91	–178	172	–59	–102
1.3	1.87	–179	163	79	–138
0.5	2.44	–89	159	175	116
0.4	2.54	–175	–57	–72	136
0.4	2.59	180	143	59	105
0.4	2.60	–176	–72	76	–133
0.3	2.70	–90	160	–71	121
0.2	3.02	–173	–72	71	93
0.1	3.37	–174	–57	–69	–92
0.1	3.75	65	164	–178	117

[a] Initial conformer geometries were from the GMMX module of PCMODEL. Torsion angles are from B3LYP/6-31G* geometries. The Boltzmann distribution and relative energies (*E*) are from gas-phase B3PW91/6-311G(2d,p) single-point energy calculations. All potential conformers are included except for the +*gauche* rotamer of 17-20-22-23 (which upon HF/3-21G minimization optimized to the *anti* conformer) and most of the high-energy ±*gauche* rotamers of 13-17-20-22.

Table S4. Observed and predicted NMR chemical shifts for tirucalla-7,24-dien-3β-ol (**2**) [a]

carbon	^{13}C chemical shifts			hydrogen	^{1}H chemical shifts		
atom	obsd	predicted	deviation	atom	obsd	predicted	deviation
C-1	37.2	37.3	0.1	H-1α	1.14	1.13	-0.02
C-2	27.7	28.6	0.9	H-1β	1.68	1.67	-0.01
C-3	79.3	78.6	-0.6	H-2α	1.66	1.67	0.01
C-4	39.0	38.5	-0.4	H-2β	1.60	1.59	-0.01
C-5	50.6	50.9	0.3	H-3α	3.25	3.23	-0.01
C-6	23.9	24.1	0.1	H-5α	1.32	1.37	0.05
C-7	117.8	119.2	1.4	H-6α	2.14	2.15	0.01
C-8	145.9	145.8	-0.1	H-6β	1.96	2.05	0.09
C-9	48.9	48.8	-0.1	H-7	5.26	5.31	0.06
C-10	34.9	35.1	0.2	H-9α	2.20	2.36	0.15
C-11	18.1	18.9	0.8	H-11α	1.53	1.46	-0.07
C-12	33.8	33.4	-0.4	H-11β	1.49	1.50	0.02
C-13	43.5	43.7	0.1	H-12α	1.62	1.53	-0.09
C-14	51.1	52.0	0.8	H-12β	1.78	1.78	0.00
C-15	34.0	33.8	-0.2	H-15α	1.49	1.56	0.07
C-16	28.2	28.6	0.4	H-15β	1.43	1.44	0.01
C-17	52.9	53.5	0.6	H-16α	1.27	1.28	0.02
C-18	21.9	22.7	0.8	H-16β	1.94	1.93	-0.01
C-19	13.1	13.4	0.3	H-17β	1.47	1.51	0.03
C-20	36.0	36.9	0.9	H-18	0.81	0.82	0.01
C-21	18.3	18.5	0.2	H-19	0.75	0.78	0.04
C-22	36.2	35.7	-0.5	H-20	1.37	1.37	0.00
C-23	25.0	24.8	-0.2	H-21	0.88	0.83	-0.05
C-24	125.2	125.1	-0.1	H-22*R*	1.03	0.90	-0.13
C-25	130.9	130.4	-0.5	H-22*S*	1.43	1.32	-0.11
C-26	25.7	26.6	0.9	H-23*R*	2.04	2.12	0.08
C-27	17.6	17.1	-0.5	H-23*S*	1.86	1.86	0.00
C-28	27.6	28.1	0.5	H-24	5.10	5.15	0.05
C-29	14.7	15.2	0.5	H-26	1.68	1.64	-0.05
C-30	27.3	28.4	1.1	H-27	1.60	1.57	-0.04
				H-28	0.97	1.00	0.03
				H-29	0.86	0.85	-0.01
				H-30	0.97	0.99	0.02
average deviation			0.23	average deviation			0.00
rms deviation [b]			0.58	rms deviation [b]			0.056

[a] Predicted chemical shifts were calculated at the B3PW91/6-311G(2d,p)//B3LYP/6-31G* level and reflect the Boltzmann distribution of gas-phase B3PW91/6-311G(2d,p) energies. Deviations correspond to predicted – observed (obsd) values. All results were calculated at full precision before rounding to 2-4 significant figures for readability. The calculation methodology is identical to that described in our previous work.[12] Designations 22*R*, 23*S*, etc. denote pro *R* and pro *S* hydrogens. [b] Root-mean-square deviation.

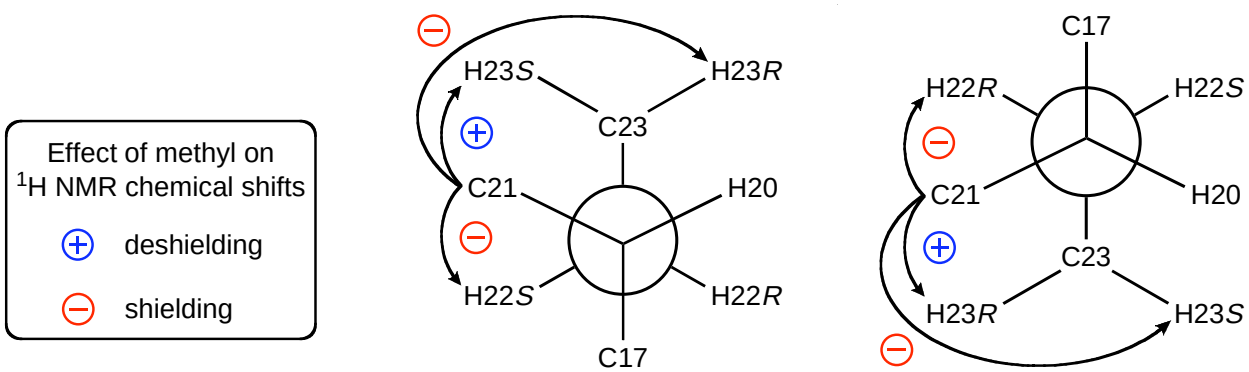
B3LYP/6-31G* atomic coordinates for the most abundant side-chain conformer of 2. The other conformers in Table S3 can be constructed from this prototype by simple bond rotations.

The condensed coordinate format used below is easily converted to tabular form by global find-and-replace routines available in most word processors. First, replace the paragraph mark with nothing; spaces might also need to be deleted; then replace "\" with the paragraph mark. If desired, commas can be replaced by the tab mark.

```
1\1\GINC-N114\FOpt\RB3LYP\6-31G(d)\C30H50O1\BILLW\15-Dec-2008\0\\# B3L
 YP/6-31G* Opt geom=allcheck guess=tcheck geom=(noang,nodist) pop=none
 test\\tirucalla-7,24-dienol, SC conformers\\0,1\C,4.9571297616,-1.6439
 560625,1.822342536\C,6.1096081121,-2.5773953822,2.2069346242\C,5.58609
 81437,-3.943566413,2.6499188091\C,4.7700862251,-4.6671257768,1.5475043
 611\C,3.6408770874,-3.6820139744,1.0654868146\C,2.7348805731,-4.288005
 7827,-0.0249454494\C,1.7017070069,-3.3254440881,-0.5460793323\C,1.6711
 514775,-2.0169438021,-0.2673460637\C,2.7009580373,-1.3967620838,0.6848
 136814\C,4.0483490734,-2.2073370037,0.7012790808\C,2.8472804883,0.1260
 639574,0.4670000093\C,1.4993407766,0.9122844251,0.4635240347\C,0.25040
 45406,0.0583854761,0.1170702752\C,0.6575902925,-1.0705646928,-0.900305
 9586\C,-0.7161775803,-1.6578786223,-1.295123488\C,-1.6622882748,-0.424
 7426799,-1.3378970448\C,-0.9137724257,0.7712749516,-0.6556816714\C,-0.
 3350586304,-0.5547446727,1.4176187854\C,4.7656020939,-2.0834688393,-0.
 6601225095\C,-1.8714971876,1.7178416959,0.1147519498\C,-1.1366115002,2
 .873116836,0.8148428612\C,-2.956275207,2.2697132942,-0.8423655095\C,-4
 .0930953917,3.0600234611,-0.1578551252\C,-5.2174678061,3.374067763,-1.
 109348767\C,-5.6533395972,4.5707004091,-1.5329616942\C,-6.8020677875,4
 .6722903124,-2.509329535\C,-5.0817415849,5.9014396972,-1.1067728062\C,
 4.1080719821,-5.9046943852,2.2008823735\C,5.7125321433,-5.1872619789,0
 .4391405721\C,1.2817149826,-0.5092783126,-2.2166802641\O,6.6517836164,
 -4.8066935354,3.0567530451\H,7.1378519866,-4.3526642149,3.7620530398\H
 ,5.360878348,-0.6686065767,1.525775904\H,4.3451114369,-1.4635132847,2.
 7195467072\H,6.8141636373,-2.7055744524,1.3770703843\H,6.6839920245,-2
 .1273368826,3.030614062\H,4.9032348969,-3.781411825,3.5041968584\H,3.0
 007450221,-3.5658469491,1.9547937169\H,2.2287801714,-5.1792560752,0.36
 86626922\H,3.3337777994,-4.6502588483,-0.8740441757\H,0.9594670062,-3.
 746750317,-1.2239110508\H,2.3048801151,-1.5070509827,1.7060573921\H,3.
 4851696127,0.5402703982,1.2543500816\H,3.3798094582,0.3129777588,-0.47
 02408327\H,1.5892452924,1.7470445513,-0.2424685095\H,1.3606817194,1.37
 14940455,1.4482066309\H,-0.6812191729,-2.1815998787,-2.2567249162\H,-1
 .0527679905,-2.3868504184,-0.5514441029\H,-2.5978927346,-0.6439182098,
 -0.8098586978\H,-1.9449170727,-0.1737649941,-2.3652993774\H,-0.4503128
 579,1.385918955,-1.4394490354\H,0.3401178853,-1.2716789669,1.889304198
 6\H,-0.5338780037,0.2354028205,2.1493545656\H,-1.2813516396,-1.0754220
 903,1.2423732865\H,5.7096949362,-2.6309717452,-0.6819838194\H,4.145090
 3228,-2.4534614765,-1.4810637278\H,5.0035564051,-1.0356717778,-0.87484
 86723\H,-2.3910195098,1.1279362691,0.8851807005\H,-0.571493222,3.47474
 09976,0.0906937047\H,-0.4318629102,2.5169738742,1.5702248178\H,-1.8390
 52662,3.5421023455,1.3225598148\H,-2.4813857198,2.9108972105,-1.599762
 2598\H,-3.4116537332,1.4385463162,-1.39384529\H,-3.701762051,3.9742964
 918,0.2974761241\H,-4.4882421634,2.4501166353,0.670250842\H,-5.7238510
 645,2.4924030375,-1.5087097692\H,-7.1903717674,3.6875003005,-2.7879766
 641\H,-6.4961679346,5.1900938888,-3.4297252114\H,-7.6313957815,5.25740
 81212,-2.0867557903\H,-4.2432742319,5.8118754246,-0.4122838291\H,-4.73
 3880167,6.4700372465,-1.9806386145\H,-5.8527158784,6.518731485,-0.6243
 54604\H,3.3494874979,-5.6100091001,2.9374923423\H,3.6242377729,-6.5474
 573058,1.4582569867\H,4.8663384356,-6.5006526306,2.716507104\H,6.32030
 22151,-4.4039720165,-0.018908234\H,6.4012644472,-5.9221649182,0.864086
 6686\H,5.1486726903,-5.6803278222,-0.3589885991\H,2.1774993096,0.09426
 21522,-2.0574595713\H,1.5660984354,-1.3552801562,-2.8514097258\H,0.569
 895947,0.0997193995,-2.7827821395\\Version=IA64L-G03RevD.01\HF=-1248.4
 9589\RMSD=1.460e-09\RMSF=3.965e-06\Thermal=0.\Dipole=-0.135794,-0.3628
 864,-0.5244266\PG=C01 [X(C30H50O1)]\\@
```

The signal assignments of the side-chain diastereotopic proton pairs in **2** (i.e. H22*R*, H22*S* and H23*R*, H23*S*) appear at first glance to disagree with the corresponding assignments in desmosterol, lanosterol, and other Δ24 sterols.[13] For example, H22*R* appears at δ ~1.4 in **2** but at δ ~1.0 in Δ24 sterols. However, a detailed analysis, outlined in Table S5, resolves this apparent discrepancy. In both **2** and Δ24 sterols, the 13-17-20-22 and 17-20-22-23 dihedral angles are mainly *anti*, as shown. The C22 proton *syn* to C21 is shielded (δ ~1.0) relative to the *anti* C22 proton (δ ~1.4), as noted previously.[14] The *anti* C22 proton is H22*R* in Δ24 sterols and H22*S* in **2**. The opposite substituent effects apply to the C23 protons, where the proton distal to C21 is shielded relative to its proximal partner. Again, the stereochemical labels for the C23 protons are reversed in **2** relative to Δ24 sterols. This simplified explanation is supported by an essentially complete conformation analysis (Table S3), which showed good agreement between prediction and observation for ^{1}H-^{1}H coupling constants and chemical shifts (Tables S2 and S4).

Table S5. Origin of the differences in NMR chemical shifts of diastereotopic side-chain protons in Δ24 sterols and tirucalla-7,24-dienol [a]



	Δ24-sterols	tirucalla-7,24-dienol
C17 stereochemistry	17β side chain	17α side chain
C20 stereochemistry	20*R*	20*S*
H22*R*	*anti* to C21	*syn* to C21
H22*S*	*syn* to C21	*anti* to C21
H22*R*	δ ~1.4	δ ~1.0
H22*S*	δ ~1.0	δ ~1.4
H23*R*	distal to C21	proximal to C21
H23*S*	proximal to C21	distal to C21
H23*R*	δ ~1.85	δ ~2.03
H23*S*	δ ~2.03	δ ~1.85

Basis for the Structure Identification of the Major PEN3 Product 2

Tirucalla-7,24-dien-3β-ol (2). The major PEN3 product was identified by comparison of the ^{1}H and ^{13}C NMR chemical shifts with the spectral data available for **2** in the literature[9-11] (Table

S6). Further support for the structure was derived from (a) de novo structure elucidation from 1D and 2D NMR spectra, (b) the identity of ^{1}H NMR chemical shifts with those we had observed previously for **2**,[12a] and (c) the good agreement of the NMR data with ^{1}H and ^{13}C chemical shifts predicted by quantum mechanical methods (Table S4) and ^{1}H-^{1}H coupling constants predicted by an extended Karplus relationship (Table S2).

Table S6. Comparison of ^{1}H and ^{13}C NMR chemical shifts for the major PEN3 product (this work) with literature values for tirucalla-7,24-dien-3β-ol (**2**) [a]

^{13}C NMR chemical shift comparisons			
atom	this work	ref 11	diff.
C-8	145.89	145.9	0.0
C-25	130.93	130.9	0.0
C-24	125.21	125.2	0.0
C-7	117.79	*b*	
C-3	79.26	79.3	0.0
C-17	52.94	52.9	0.0
C-14	51.14	51.1	0.0
C-5	50.62	50.6	0.0
C-9	48.94	48.9	0.0
C-13	43.51	43.5	0.0
C-4	38.96	39.0	0.0
C-1	37.20	37.2	0.0
C-22	36.18	36.2	0.0
C-20	35.96	35.9	-0.1
C-10	34.93	34.9	0.0
C-15	34.02	34.0	0.0
C-12	33.78	33.8	0.0
C-16	28.21	28.2	0.0
C-2	27.69	27.7	0.0
C-28	27.60	27.6	0.0
C-30	27.26	27.3	0.0
C-26	25.72	25.7	0.0
C-23	25.01	25.0	0.0
C-6	23.93	23.9	0.0
C-18	21.90	21.9	0.0
C-21	18.32	18.3	0.0
C-11	18.12	18.1	0.0
C-27	17.63	17.6	0.0
C-29	14.72	14.7	0.0
C-19	13.11	13.1	0.0

^{1}H NMR chemical shift comparisons					
atom	this work	ref 9	diff.	ref 10a,b	ref 11
H-7	5.256	5.253	-0.003	5.26	5.25
H-24*R*	5.099	5.098	-0.001	5.12	5.10
H-3α	3.245	3.253	0.008	3.22	3.24
H-26	1.684	1.683	-0.001	1.67	1.68
H-27	1.604	1.603	-0.001	1.62	1.60
H-28	0.971	0.968	-0.003	0.99	0.97
H-30	0.968	0.968	0.000	0.99	
H-21	0.882	0.880	-0.002	0.94	0.88
H-29	0.861	0.860	-0.001	0.87	0.86
H-18	0.809	0.806	-0.003	0.82	0.81
H-19	0.747	0.744	-0.003	0.76	0.75

[a] All chemical shifts are in ppm (δ); diff. denotes the chemical shift differences (literature – this work). The chemical shifts from references 10a and 10b are identical apart from minor discrepancies for H-26 and H-19. [b] ^{13}C NMR chemical shifts in ref 11 were not assigned; in an apparent oversight, the value δ 76.7 (probably corresponding to $CHCl_3$) was listed instead of the olefinic signal for C-7 at ca. δ 117.8.

Spectra and Structure Identification of Minor PEN3 Products

The most abundant minor products of PEN3 were isolated in Experiment 3 by HPLC separation on a C_{18} column using methanol-water gradients (order of elution: **6**, **5**, **2** and **4**, and **3**). Portions of ^{1}H NMR spectra of butyrospermol (**3**), tirucallol (**4**), isotirucallol (**5**), 13βH-malabarica-14(27),17,21-trien-3β-ol (**6**) are shown in Figure S11. Mass spectra of compounds **3**-**5** are shown in Figure S12. These minor compounds were identified by comparison with spectral data from the literature, as shown in Table S7. Dammara-20,24-dien-3β-ol (**7**) was identified from three distinctive signals resolved in the 800-MHz NMR spectrum of PEN3 products; the COSYDEC spectrum showed the olefinic signals of **7** correlating with each other. In addition to the literature comparisons, the ^{1}H NMR data for **2**-**7** matched fine coupling patterns and chemical shifts (±0.001 ppm) we have measured for authentic standards of these triterpenes.

Table S7. Comparison of ^{1}H NMR chemical shifts for the minor PEN3 products (this work) with literature values reported for **3**, **4**, **5**, **6**, and **7** in $CDCl_3$ solution [a]

NMR chemical shift comparisons			
atom	this work	lit.	diff.
butyrospermol (3)		ref 15 [b]	
H-18	0.805	0.804	0.001
H-19	0.744	0.744	0.000
H-21	0.849	0.848	0.001
H-28	0.974	0.970	0.004
H-29	0.860	0.854	0.006
H-30	0.971	0.970	0.001
H-3α	3.244	3.243	0.001
H-7	5.254	5.251	0.003
H-24	5.096	5.093	0.003
H-26	1.683	1.680	0.003
H-27	1.605	1.603	0.002
tirucallol (4)		ref 9 [c]	
H-18	0.756	0.751	0.005
H-19	0.953	0.950	0.003
H-21	0.917	0.917	0.000
H-28	1.003	0.995	0.008
H-29	0.801	0.791	0.010
H-30	0.867	0.864	0.003
H-3α	3.240		

NMR chemical shift comparisons			
atom	this work	lit.	diff.
isotirucallol (5)		ref 16	
H-18	1.082	1.08	0.00
H-19	0.847	0.85	0.00
H-21	0.960	0.96	0.00
H-28	0.987	0.99	0.00
H-29	0.770	0.77	0.00
H-30	0.815	0.82	0.00
malabaricatrienol (6)		ref 17	
H-24	0.781	0.77	0.01
H-25	0.855	0.84	0.02
H-26	0.959	0.94	0.02
H-23	0.978	0.96	0.02
H-3α	3.207	3.20	0.01
H-27	4.588	4.58	0.01
H-27	4.877	4.87	0.01
H-21	5.098	5.08	0.02
H-17	5.127	5.10	0.03
dammara-20,24-dienol (7)		ref 18	
H-29	0.777	0.78	0.00
H-21	4.706	4.71	0.00
H-21	4.740	4.74	0.00

[a] Some signal assignments for angular methyls are tentative. Not included here are ^{13}C NMR data for **3** and **4** from HSQC and HMBC spectra. [b] Similar chemical shifts for **3** are given in ref 9. [c] After application of acetylation shifts: 0.004, -0.024, 0.002, 0.122, –0.073, and 0.000 ppm.

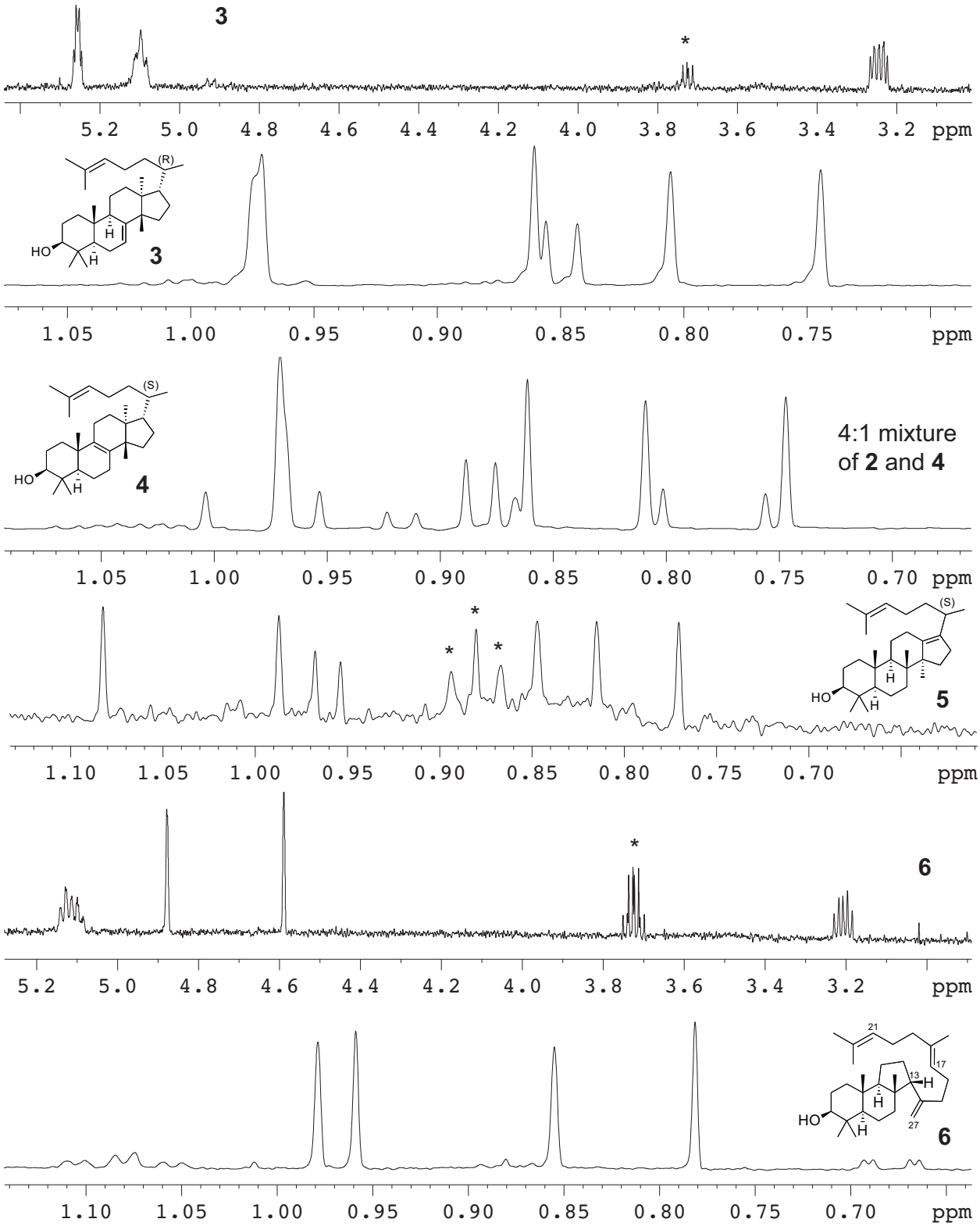


Figure S11. Partial ^{1}H NMR spectra of minor PEN3 products **3**-**6**: 500 MHz, ≤5 mM solution in $CDCl_3$. Signals from impurities are marked by asterisks (*).

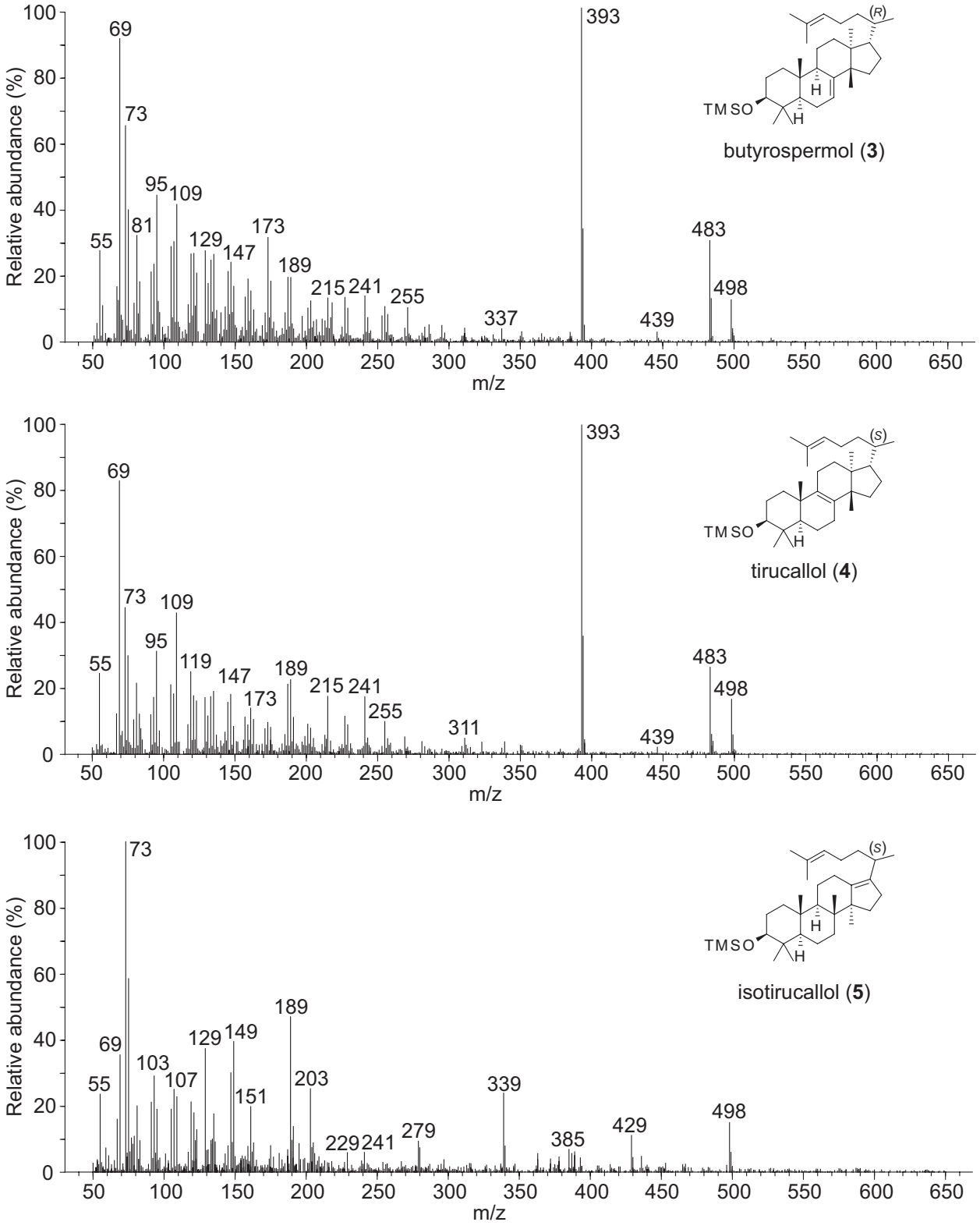


Figure S12. Mass spectra of minor PEN3 products **3**-**5**. These spectra correspond to the minor peaks shown in the total ion chromatogram of Figure 1 of the main text.

Estimation of the Ratio of PEN3 Products

The relative amounts of PEN3 products **2**-**7** were determined by integration of the upfield methyl signals in the 800-MHz NMR spectrum of PTLC band A2 (Figure 1A of the main text). Multiple signals of minor products **3**-**5** were resolved sufficiently for integration. Only the most upfield methyl singlets of **6** and **7** were resolved, and additional quantitation was obtained from integration of their olefinic signals. The following basic pattern of the PEN3 product profile was reproduced in several experiments involving different personnel over the course of four years:

tirucalla-7,24-dien-3β-ol	**2**	~85%
butyrospermol	**3**	6%
tirucallol	**4**	6%
isotirucallol	**5**	1.5%
13βH-malabarica-14(27),17,21-trien-3β-ol	**6**	1%
dammara-20,24-dien-3β-ol	**7**	0.5%

One potential source of error in the ratios is a neglect of possible diol, 3-keto, 3,10-epoxide, or Grob fragmentation products, which do not co-chromatograph with other triterpenes and would not be found in PTLC band A2. The many chromatographic fractions in Experiments 1 and 2 (Figures S1 and S2) were analyzed by GC-MS and ^{1}H NMR for possible formation of these compounds. The analyses suggested little or none of these unusual triterpenes. Friedo skeletons (3-keto products)[19] and 3,10-epoxides[20] were searched according to their distinctive ^{1}H NMR signals: δ 0.73 (s), 0.87 (s), 0.88 (d), 2.25 (q, 7 Hz), 2.29 (td, 13, 7 Hz), 2.39 (ddd, 14, 5, 2 Hz) and δ 3.75 (d, 5.5 Hz), respectively. Additional evidence against significant amounts of unusual triterpene skeletons was provided by Experiment 3, in which all fractions from a single large silica gel column of the NSL were analyzed by NMR and GC-MS. The first fraction (containing mainly oxidosqualene) showed no evidence of 3,10-epoxides or friedo skeletons, and the polar wash fraction after ergosterol showed no ^{1}H NMR signal at δ 3.19-3.25, where a diol H-3α resonance would be expected. (Such separation by a single silica gel column gives little protection against nonenzymatic cyclization of oxidosqualene and dioxidosqualene, and many known[12a] artifacts were observed.) Another signature of unusual triterpenes, including Grob fragmentation products, is a set of upfield methyl singlets that do not correspond to any known triterpene alcohols; no such patterns were noted among the various chromatographic fractions.

Another potential source of error in these in vivo experiments is yeast metabolism of the triterpene products via the sterol biosynthetic pathway. Analysis of the chromatographic fractions in Experiments 1–3 did not reveal any obvious triterpene metabolism, such as 24-methylation or 22-desaturation. Most early metabolites would contain a 3β-hydroxy-4,4-dimethyl moiety, and the few occurrences of the distinctive H-3α NMR signal at δ 3.19-3.25 were explained as known, unmetabolized triterpene alcohols.

A third potential source of error is loss of triterpenes to the culture medium. As described on pages S5-S6, the spent culture medium from Experiment 2 was extracted with a hydrophobic

resin, followed by saponification, column chromatography, and PTLC. The ^{1}H NMR spectrum of the PTLC band containing tetracyclic triterpenes and some interfering impurities (Fraction A2) showed a ca. 9:1:1 mixture of tirucalla-7,24-dien-3β-ol (**2**), butyrospermol (**3**), and tirucallol (**4**), with ~1% each of isotirucallol (**5**) and 13βH-malabarica-14(27),17,21-trien-3β-ol (**6**). Lyophilization of a portion of the saturated culture of Experiment 2 (pages S5-S6), followed by chromatographic separations and spectra analysis, indicated a ca. 15:1:1 mixture of **2**, **3**, and **4**, with ~1% each of **5** and **6**. The limited NMR sensitivity and spectral inferferences from non-triterpene impurities precluded an accurate assessment of product ratios from the resin extraction and lyophilization experiments. This uncertainty limits our conclusions to the following: (a) the product profile estimated from the cell pellet is at most modestly distorted from the true product profile and (b) tirucalla-7,24-dien-3β-ol comprises ~80-90% of LUP5 products.

Origin of 20*R* and 20*S* Triterpenes: Comparison of Two Mechanisms

In Figure S13, the classical ETH mechanism[21] for the formation of the C20 epimers euphol (20*R*) and tirucallol (20*S*) is compared with the mechanism shown in Figure 3 of the main text. To facilitate the comparison, we have omitted the portion of Figure 3 pertaining to the dammarenyl cation with a 17α side chain and added a structure (**IIb-H$_1$**) corresponding to the horizontal C20 cation hyperconjugated to the C13-C17 bond.[22,23] To the ETH mechanism we added structures for the C17 cation (analogous to **IIIa** and **IIIb**). The ETH C17-H17-C20 bridged cation could be understood as either the transition state linking the C20 and C17 cations or as a C20 cation hyperconjugated to the C17-H17 bond (analogous to **IIb-V$_1$** and **IIb-V$_2$**).

One obvious difference between the mechanisms is trivial: the mechanism of Figure 3 is based on quantum mechanical calculations that show bond lengthening due to hyperconjugation; such calculations were virtually unavailable at the time of the ETH mechanism (1955), and the same phenomenon (a prescient insight) was denoted as a bridged cation.

A more fundamental difference between the mechanisms is the fork in the pathway leading to 20R and 20S isomers. In the ETH mechanism, this fork occurs at ETH intermediate **XIII**. This C13-C17-C20 bridged cation can reposition itself for a C17-C20 hydride shift either before or after undergoing a conformational (constellational) rearrangement to the C16-C17-C20 cation. The conformational change from D-ring boat to chair is presumably considered the decisive step that determines the fate of the nascent C20 stereocenter. Evidently, the direction of the inevitable rotation about the C17-C20 bond (shown only implicitly in the original mechanism) is fixed by the chair/boat D-ring conformation and is not energetically decisive in determining the ratio of C20 isomers. A possible interpretation of the ETH proposal would be that PEN3 accommodates the 6-membered D-ring intermediate in both chair and boat forms in a ratio of ~14:1.

In contrast, the pathway fork in Figure 3 for a dammarenyl cation with a 17β side chain derives from opposite directions of side-chain rotation as the horizontal cation (**IIb-H$_1$** or **IIb-H$_2$**) becomes a vertical cation (**IIb-V$_1$** or **IIb-V$_2$**) via pathways B$_1$ and B$_2$); **IIb-H$_1$** and **IIb-H$_2$** behave identically, cw (ccw) rotation resulting in 20*R* (20*S*) isomers. In the absence of substrate analog experiments[23] or relevant crystal structures, both mechanistic proposals are plausible.

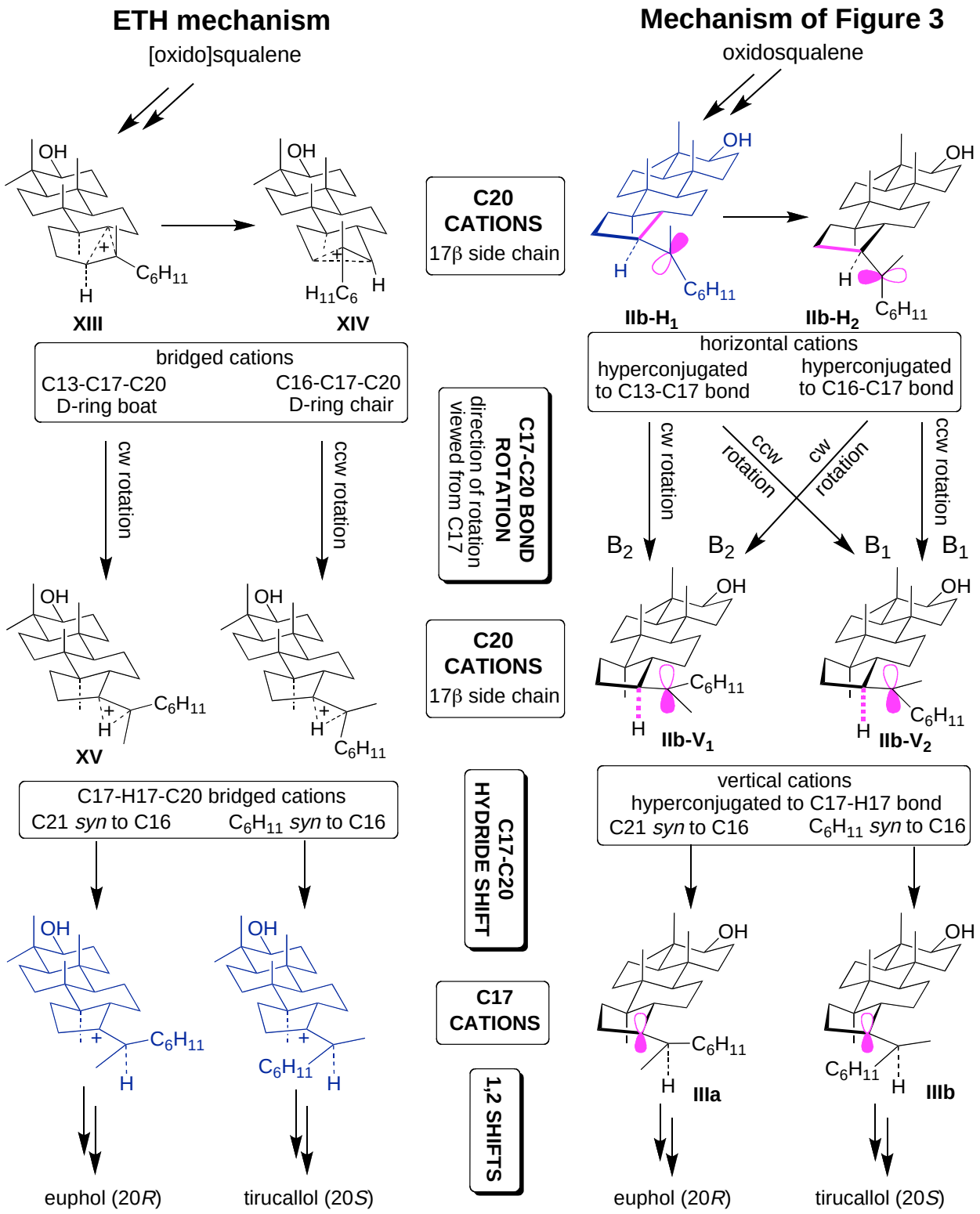


Figure S13. Comparison of the ETH mechanism for the enzymatic formation of euphol and tirucallol with the mechanism of Figure 3. Hyperconjugated bonds are shown in magenta. Structures in blue are additions not present in Figure 3 or in the published ETH mechanism. The Figure 3 structures are drawn in the style of the ETH mechanism but at a different perspective.

Preliminary Characterization of LUP5 Products

Strain construction. The *LUP5* (*At1g66960*) full length coding sequence was obtained from a cDNA bacterial pUNI vector library purchased from the Arabidopsis Biological Resource Center (http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm). The ~2.3 Kbp insert was subcloned into the integrative galactose-inducible yeast expression vector pRS305GAL[4] to give plasmid pPM2.1. Transformation of *S. cerevisiae* strain RXY6[7] and selection on synthetic complete medium[2] gave the recombinant yeast strain RXY6[pPM2.1].

Yeast culture. A single colony of yeast strain RXY6[pPM2.1] was used to inoculate synthetic complete medium lacking leucine (10 mL), supplemented with hemin chloride (13 mg/L), cholesterol (20 mg/L), and Tween 80 (5 g/L), with 2% glucose as the carbon source. The culture was grown to saturation at 30 °C with shaking, and 1-mL aliquots were used to inoculate two 100-mL portions of the same medium. Cultures were grown to saturation and used to inoculate two 1-L solutions of yeast peptone (YP) medium, supplemented with hemin chloride (13 mg/L), cholesterol (20 mg/L) Tween 80 (5 g/L), with 2% galactose as carbon source. Cultures were grown in a 30 °C incubator with shaking at 250 rpm to an OD_{600} of 6.5 (measured as 0.65 after 1:10 dilution). Cells were collected by centrifugation (3800 rpm for 20 min), and the supernatant was discarded. The cell pellet (22 g) was suspended in one volume of 100 mM phosphate buffer (pH 6.2), and cells were lysed by passing the suspension through a cell disruptor (Avestin Emulsiflex-C5 homogenizer).

In vitro reaction. The resultant slurry was centrifuged (3000 rpm for 5 min) to remove the cell debris, and oxidosqualene was added (from a 20 mg/L solution of (±)-2,3-oxidosqualene in 1:1 water-Tween 80) to a final concentration of 0.5 mg/mL. The suspension was incubated at ca. 23 °C, and the reaction progress was monitored by TLC analysis (using CH_2Cl_2 or 1:1 ether-hexane). After 36 h, the reaction was quenched with two volumes of ethanol, and the cell debris was removed by centrifugation (3000 rpm for 3 min). The ethanol was removed by rotary evaporation, and the remaining water phase was extracted with MTBE. The extracts were combined, washed with brine, and the organic solvent was evaporated to a residue. GC-MS analysis of an aliquot derivatized as the TMS ether indicated a yield of ~10% (based on (±)-oxidosqualene as judged by comparison of oxidosqualene and triterpene peak areas.

Chromatographic purification and spectral analysis. The above residue was rapidly chromatographed on a short silica gel column (5 g, 230-400 mesh) to remove unreacted oxidosqualene, which might otherwise generate non-enzymatic cyclization products.[12a] Squalene, oxidosqualene and other non-polar yeast components (10 mg) were eluted with 2% ether in hexane. Subsequent elution with 100% ether gave LUP5 products, cholesterol, and other polar yeast components (22 mg). ^{1}H NMR and GC-MS analysis confirmed that oxidosqualene was absent in the second fraction. This triterpene-containing fraction was divided into two batches (PTLC-1 and PTLC-2) of 9-11 mg, each of which was subjected to PTLC (developed with 1:1 ether-hexane). For PTLC-1, the TLC plate was divided into six bands in order of decreasing polarity (A-F) and analyzed by GC-MS and NMR to detect minor products; cyclic triterpenes (LUP5 products) were observed in bands D and E. For PTLC-2, the TLC plate was divided so

that a single band (DE) corresponded to bands D and E of PTLC-1; band DE of PTLC-2 was analyzed by GC-MS and NMR to determine the LUP5 product ratios. ^{1}H NMR spectra of PTLC-2 band DE and PTLC-1 band D are shown in Figure S14. GC-MS analysis gave similar results.

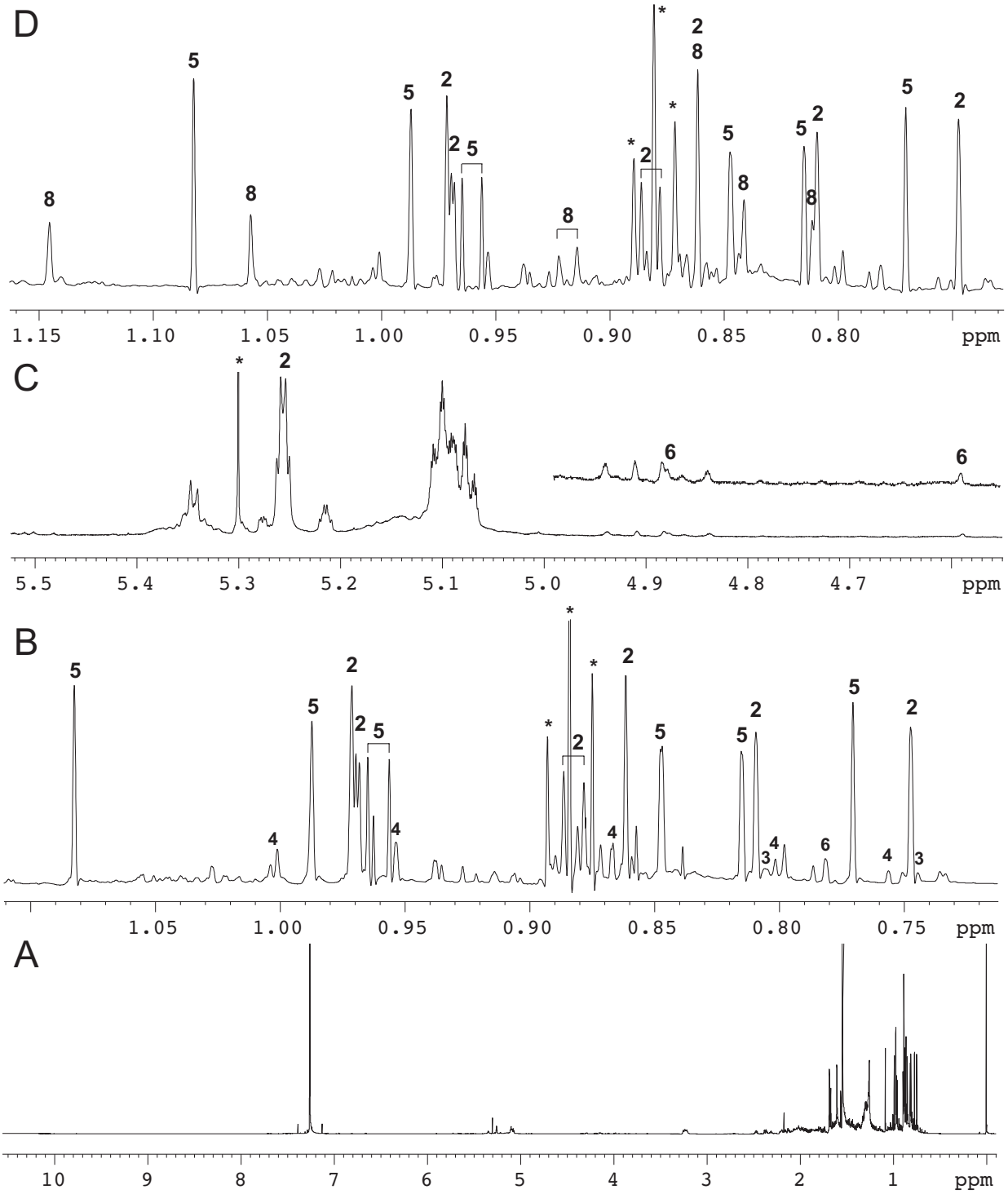


Figure S14. 800 MHz ^{1}H NMR spectra of LUP5 products: PTLC-1 band D, full spectrum (A), upfield methyl region (B), and olefinic region (C); PTLC-2 band DE, upfield methyl region (D).

The LUP5 product profile was estimated primarily from the signal intensities of resolved methyl singlets in the 800-MHz ^{1}H NMR spectrum. This characterization is considered preliminary because (1) the result is based on a single experiment, (2) some signal intensities were distorted by overlapping interferences, and (3) unidentified minor triterpene products appeared also to be present (Figure S14). The discrepancy between our estimate of a ~2:2:1 ratio of the Δ7,24, Δ13(17),24, and Δ5,24 isomers (**2**, **5**, and **8**) and the ~4:4:1 ratio apparent from the HPLC-UV chromatogram of ref 24 could be attributable to different UV responses of the three isomers at 202 nm.

Sequence Alignments of Oxidosqualene Cyclases

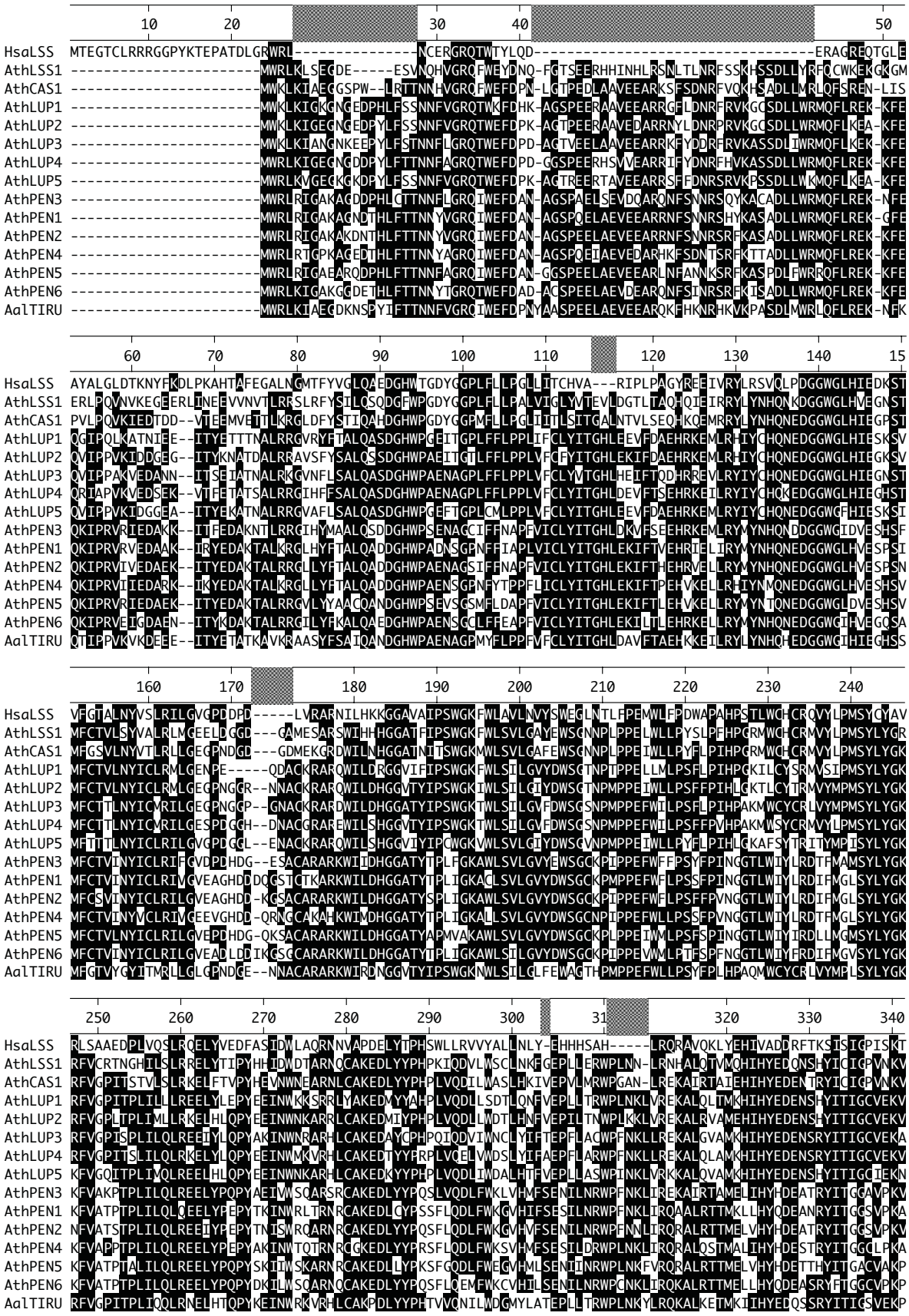
The nucleotide sequence for human lanosterol synthase (*LSSh*) and *Arabidopsis* cyclases used in Figure 4 is essentially sequence from GenBank (http://www.ncbi.nlm.nih.gov/Genbank) with modest corrections. Gene identifiers are given in Table S8. Sequence for the tirucalla-7,24-dienol synthase of *Ailanthus altissima*[11] was obtained from Chemical Abstracts.[25]

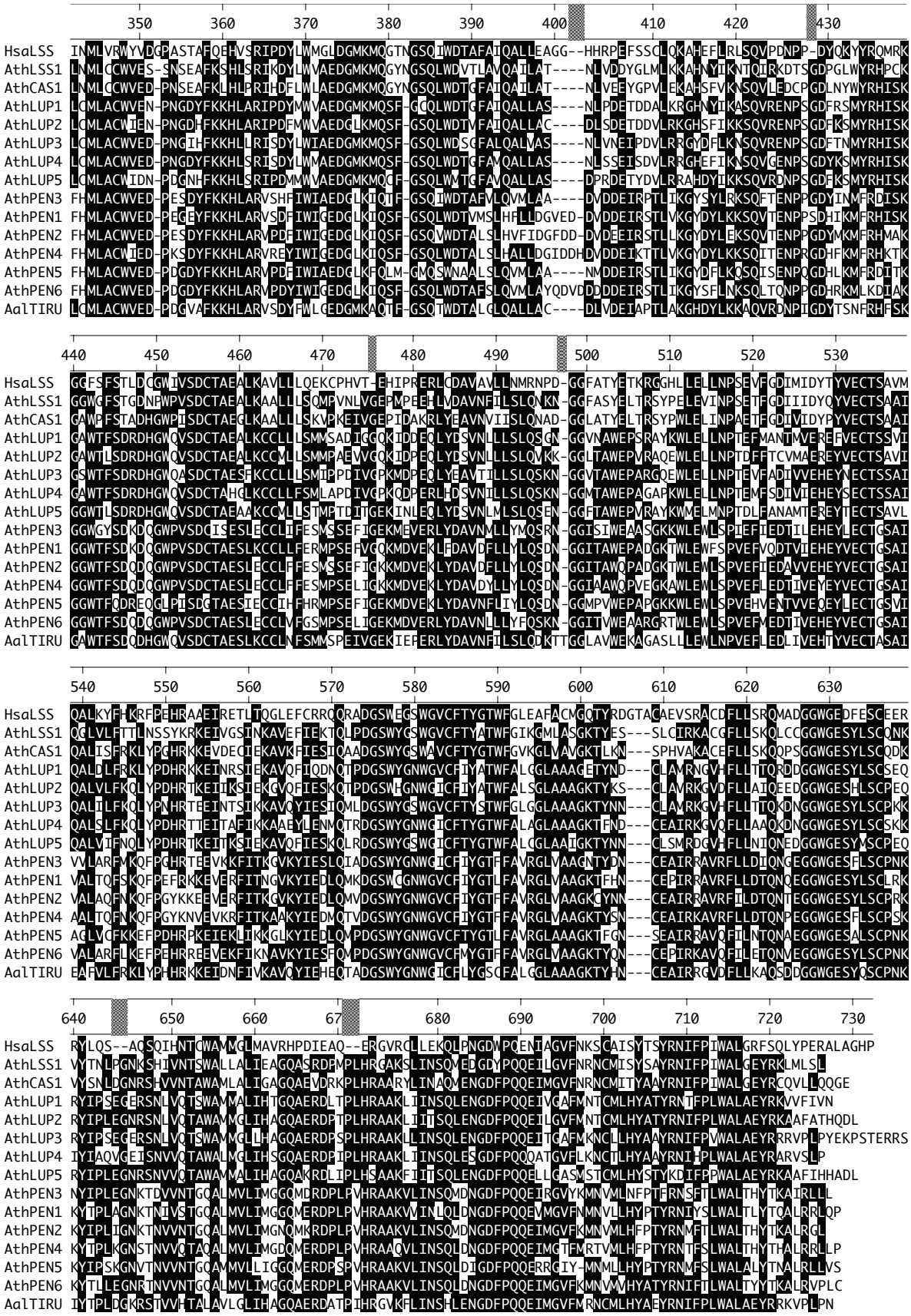
Table S8. Gene identifiers for human lanosterol synthase and *Arabidopsis thaliana* cyclases.

species	cyclase	GenBank locus tag or gene accession number
H. sapiens	Hsa LSSh	NM_001001438
A. thaliana	Ath CAS1	AT2G07050
A. thaliana	Ath LSS1	AT3G45130
A. thaliana	Ath LUP1	AT1G78970
A. thaliana	Ath LUP2	AT1G78960
A. thaliana	Ath LUP3	AT1G78955
A. thaliana	Ath LUP4	AT1G78950
A. thaliana	Ath LUP5	AT1G66960
A. thaliana	Ath PEN1	AT4G15340
A. thaliana	Ath PEN2	AT4G15370
A. thaliana	Ath PEN3	AT5G36150
A. thaliana	Ath PEN4	AT5G48010
A. thaliana	Ath PEN5	AT5G42600
A. thaliana	Ath PEN6	AT1G78500

The translated cyclase sequences were aligned using ClustalW, as implemented in the Megalign module of Lasergene 8.0 (DNASTAR, Madison, WI, USA), with minor manual editing. The amino acid alignment is shown in Figure S15. The nucleotide sequences were then aligned manually to match the protein alignment. Using the PHYLIP software package (available from http://evolution.gs.washington.edu/phylip.html) and a subset of the DNA and amino acid alignments containing only *Arabidopsis* cyclases, we calculated maximum likelihood and protein parsimony phylograms. These phylograms had the topology of the unrooted tree shown in the Graphical Abstract of the main text.

Figure S15 (pages S31-S32). Amino acid sequence alignment for *Arabidopsis* cyclases, human LSS, and a tirucalla-7,24-dienol synthase from *A. altissima*. The human LSS ruler is shown.


HsaLSS
AthLSS1
AthCAS1
AthLUP1
AthLUP2
AthLUP3
AthLUP4
AthLUP5
AthPEN3
AthPEN1
AthPEN2
AthPEN4
AthPEN5
AthPEN6
AalTIRU


HsaLSS
AthLSS1
AthCAS1
AthLUP1
AthLUP2
AthLUP3
AthLUP4
AthLUP5
AthPEN3
AthPEN1
AthPEN2
AthPEN4
AthPEN5
AthPEN6
AalTIRU

References and Notes

(1) Gaussian 03, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2005.

(2) 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.

(3) (a) 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. For details, see: (b) Hua, L. Ph.D. thesis, Rice University, 2000, p. 76.

(4) Sikorski, R. S.; Hieter, P. *Genetics* **1989**, *122*, 19-27.

(5) Corey, E. J.; Matsuda, S. P. T.; Baker, C. H.; Ting, A. Y.; Cheng, H. *Biochem. Biophys. Res. Commun*. **1996**, *219*, 327-331.

(6) Schiestl, R. H.; Gietz, R. D. *Curr. Genet.* **1989**, *16*, 339-346.

(7) Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678-5679.

(8) Yun, D.-F.; Laz, T. M; Clements, J. M.; Sherman, F. *Mol. Microbiol.* **1996**, *19*, 1225-1239.

(9) Goad, L. J.; Akihisa, T. *Analysis of Sterols*; Blackie (Chapman & Hall): London, 1997; appendix 3, pp. 406-410.

(10) (a) Itoh, T.; Tamura, T.; Matsumoto, T. *Steroids* **1976**, *27*, 275-285. (b) Itoh, T.; Matsumoto, T. *Lipids* **1976**, *11*, 434-441.

(11) Hidesaki, T.; Miyake, A.; Tabata, T. (Mitsui Chemicals) Jpn. Kokai Tokkyo Koho 2005052009, 2005 (*Chem. Abstr.* **2005**, *142*, 275994).

(12) (a) Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Kolesnikova M. D.; Onak, C. S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2007**, *129*, 11213-11222. (b) Kolesnikova, M. D.; Obermeyer, A. C.; Wilson, W. K.; Lynch, D. A.; Xiong, Q.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 2183-2186. (c) Shan, H.; Segura, M. J. R.; Wilson, W. K.; Lodeiro, S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2005**, *127*, 18008-18009. (d) Xiong, Q.; Wilson, W. K.; Pang, J. *Lipids* **2007**, *42*, 87-96. (e) Lodeiro, S.; Wilson, W. K.; Shan, H.; Matsuda, S. P. T. *Org. Lett.* **2006**, *8*, 439-442.

(f) Shan, H.; Wilson, W. K.; Phillips, D. R.; Bartel, B.; Matsuda, S. P. T. *Org. Lett.* **2008**, *10*, 1897-1900. (h) Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Ivanova, Y.; Smith, M. L.; May, G. S.; Matsuda, S. P. T. *Org. Lett.* **2009**, *11*, 1241-1244.

(13) (a) Wilson, W. K.; Sumpter, R. M.; Warren, J. J.; Rogers, P. S.; Ruan, B.; Schroepfer, G. J., Jr. *J. Lipid Res.* **1996**, *37*, 1529-1555. (b) Ruan, B.; Watanabe, S.; Eppig, J. J.; Kwoh, C.; Dzidic, N.; Pang, J.; Wilson, W. K.; Schroepfer, G. J., Jr. *J. Lipid Res.* **1998**, *39*, 2005-2020.

(14) Wilson, W. K.; Swaminathan, S.; Pinkerton, F. D.; Gerst, N.; Schroepfer, G. J., Jr. *Steroids* **1994**, *59*, 310-317.

(15) Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. *J. Am. Chem. Soc.* **2000**, *122*, 6816-6824.

(16) Akihisa, T.; Yasukawa, K.; Kimura, Y.; Takase, S.-i.; Yamanouchi, S.; Tamura, T. *Chem. Pharm. Bull.* **1997**, *45*, 2016-2022.

(17) Justicia, J.; Rosales, A.; Bunuel, E.; Oller-Lopez Juan, L.; Valdivia, M.; Haidour, A.; Oltra, J. E.; Barrero Alejandro, F.; Cardenas Diego, J.; Cuerva Juan, M. *Chem. Eur. J.* **2004**, *10*, 1778-1788.

(18) Leong, Y.; Harrison, L. J. *Phytochemistry* **1999**, *50*, 849-857.

(19) Akihisa, T.; Kimura, Y.; Koike, K.; Tai, T.; Yasukawa, K.; Arai, K.; Suzuki, Y.; Nikaido, T. *Chem. Pharm. Bull.* **1998**, *46*, 1824-1826.

(20) Shibuya, M.; Sagara, A.; Saitoh, A.; Kushiro, T.; Ebizuka, Y. *Org. Lett.* **2008**, *10*, 5071-5074.

(21) (a) Eschenmoser, A.; Ruzicka, L.; Jeger, O.; Arigoni, D. *Helv. Chim. Acta* **1955**, *38*, 1890-1904. (b) Eschenmoser, A.; Arigoni, D. *Helv. Chim. Acta* **2005**, *88*, 3011-3050. Both papers originated from the „ETH“, the Eidgenössische Technische Hochschule Zürich.

(22) To simplify Figure 3, we showed the dammarenyl cation **IIb** only as hyperconjugated to the C16-C17 bond. Quantum mechanical geometry optimizations of the enzyme-free substrate show the C16-C17 hyperconjugation to be energetically favored over C13-C17 hyperconjugation for dammerenyl cations with a 17β side chain, whereas C13-C17 hyperconjugation is favored for cations with a 17α side chain.[23] However, the energy differences between the C13-C17 and C16-C17 hyperconjugation are modest and may be markedly altered by effects of the surrounding enzyme. Thus, both forms should be considered in mechanistic proposals.

(23) Xiong, Q.; Rocco, F.; Wilson, W. K.; Xu, R.; Ceruti, M.; Matsuda, S. P. T. *J. Org. Chem.* **2005**, *70*, 5362-5375.

(24) Ebizuka, Y.; Katsube, Y.; Tsutsumi, T.; Kushiro, T.; Shibuya, M. *Pure Appl. Chem.* **2003**, *75*, 369-374.

(25) Two instances of “i” (roughly at positions 515 and 1854, perhaps the result of OCR errors) were replaced by “t”.