ORGANIC LETTERS

2007
Vol. 9, No. 25
5223−5226

Arabidopsis Camelliol C Synthase Evolved from Enzymes That Make Pentacycles

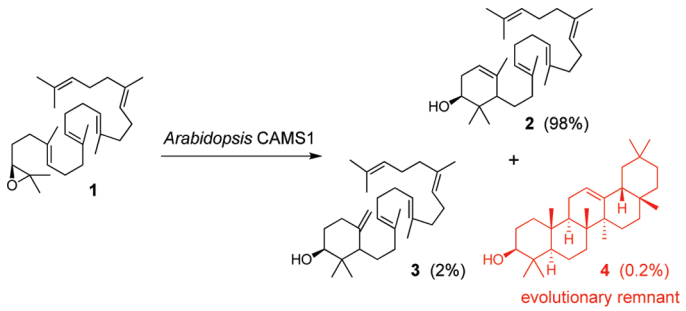
Mariya D. Kolesnikova,[†] William K. Wilson,[‡] David A. Lynch,[‡] Allie C. Obermeyer,[†] and Seiichi P. T. Matsuda*[,†,‡]

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

matsuda@rice.edu

Received October 2, 2007

ABSTRACT



We establish by heterologous expression that the *Arabidopsis thaliana* oxidosqualene cyclase At1g78955 (CAMS1) makes camelliol C (98%), achilleol A (2%), and β-amyrin (0.2%). CAMS1 is the first characterized cyclase that generates predominantly a monocyclic triterpene alcohol. Phylogenetic analysis shows that CAMS1 evolved from enzymes that make pentacycles, thus revealing that its pentacyclic β-amyrin byproduct is an evolutionary relic. Sequence alignments support prior suggestions that decreased steric bulk at a key active-site residue promotes monocycle formation.

Triterpene synthases[1] guide cationic cyclizations and rearrangements to convert squalene or oxidosqualene (**1**) to diverse skeletons, which usually contain four or five rings.[2] We show here that the *Arabidopsis thaliana* gene *At1g78955* (*LUP3*)[3] encodes an enzyme that converts oxidosqualene to camelliol C (**2**) and two minor byproducts. This is the first example of a native oxidosqualene cyclase that makes an A-ring monocycle as the dominant product.[1a,4] Our phylogenetic analysis indicates that this camelliol C synthase (CAMS1) is an evolutionary descendent of enzymes that make larger ring systems.

[†] Department of Chemistry.
[‡] Department of Biochemistry and Cell Biology.

(1) Review: (a) Phillips, D. R.; Rasbery, J. M.; Bartel, B.; Matsuda, S. P. T. *Curr. Opin. Plant Biol.* **2006**, *9*, 305−314. Subsequent reports on triterpene synthases: (b) Tansakul, P.; Shibuya, M.; Kushiro, T.; Ebizuka, Y. *FEBS Lett.* **2006**, *580*, 5143−5149. (c) Sawai, S.; Shindo, T.; Sato, S.; Kaneko, T.; Tabata, S.; Ayabe, S.-i.; Aoki, T. *Plant Sci.* **2006**, *170*, 247−257. (d) Basyuni, M.; Oku, H.; Inafuku, M.; Baba, S.; Iwasaki, H.; Oshiro, K.; Okabe, T.; Shibuya, M.; Ebizuka, Y. *Phytochemistry* **2006**, *67*, 2517−2524. (e) Shibuya, M.; Xiang, T.; Katsube, Y.; Otsuka, M.; Zhang, H.; Ebizuka, Y. *J. Am. Chem. Soc.* **2007**, *129*, 1450−1455. (f) Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835−2838. (g) Kolesnikova, M. D.; Obermeyer, A. C.; Wilson, W. K.; Lynch, D. A.; Xiong, Q.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 2183−2186. (h) 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.

(2) Reviews: (a) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2000**, *39*, 2812−2833. (b) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261−290.

(3) Husselstein-Muller, T.; Schaller, H.; Benveniste, P. *Plant Mol. Biol.* **2001**, *45*, 75−92.

(4) The only other oxidosqualene cyclase reported to make fewer than three rings is marneral synthase (MRN1), which constructs a bicyclic cation that undergoes Grob fragmentation to the seco-A product (technically a monocycle): Xiong, Q.; Wilson, W. K.; Matsuda, S. P. T. *Angew. Chem., Int. Ed.* **2006**, *45*, 1285−1288.

10.1021/ol702399g CCC: $37.00
Published on Web 11/07/2007

CAMS1 was characterized by expression in *Saccharomyces cerevisiae* (for details, see the Supporting Information). The confirmed coding sequence was cloned into pRS426GAL[5] to generate plasmid pMDK4.8, which was expressed in the yeast strains SMY8[6] and RXY6.[7] The cell pellet from a 4 L culture of SMY8[pMDK4.8] was saponified, and the hexane extract of this in vivo experiment was analyzed by GC−MS and ^{1}H NMR, both of which displayed dominant signals consistent with data reported[8] for camelliol C (**2**). Minor signals corresponding to achilleol A (**3**)[9] and β-amyrin (**4**)[10] were also present. The bulk of the material was purified to discrete compounds, which were definitively characterized as **2**, **3**, and **4** by 800 MHz ^{1}H NMR. To establish the product ratios, we performed an in vitro reaction with synthetic oxidosqualene and enzyme present in the RXY6[pMDK4.8] homogenate.[11] GC−MS (Figure 1) and ^{1}H NMR analyses

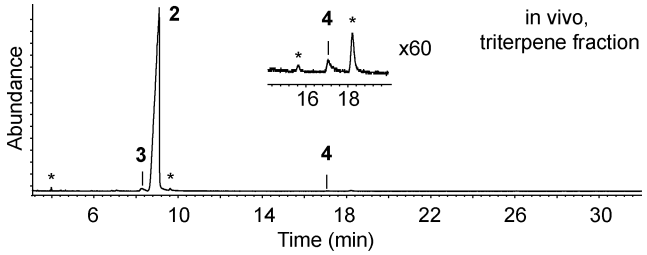


Figure 1. GC−MS total ion chromatogram of the underivatized triterpene fraction from SMY8[pMDK4.8]. Asterisks (*) denote non-triterpene components, as judged by their mass spectra.

of the in vitro reaction extract confirmed that these compounds were enzymatic cyclization products, and ^{1}H NMR provided quantitation for the relative amounts of **2** (98%), **3** (2%), and **4** (0.2%). Details of spectral analyses are given in the Supporting Information.

CAMS1 shows high product specificity, being substantially more accurate than many *Arabidopsis* cyclases that function in secondary metabolism, such as LUP1, BARS1,[1h] and PEN1[1g] (Table 1).[12] Although the enumeration of trace-level cyclase products is necessarily incomplete, the propensity

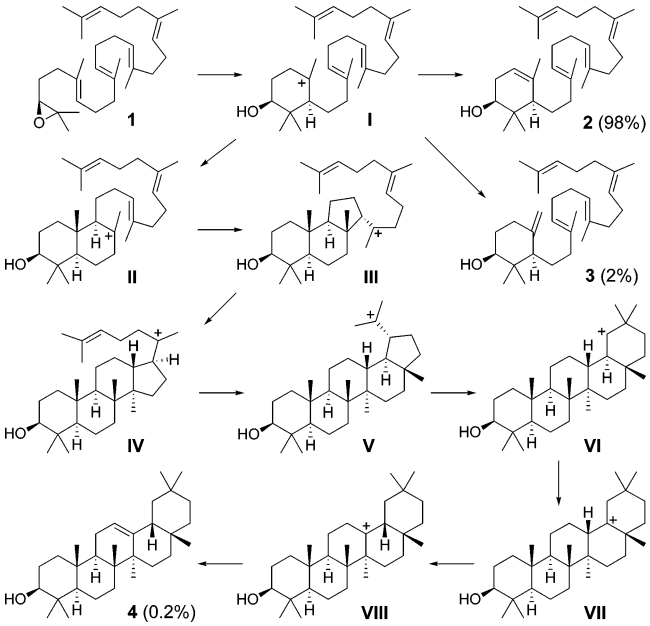
Table 1. Comparison of Product Accuracy among Cyclases[a]

triterpene synthase	P_1/P_2	$P_1/\Sigma P_i$	reference
LUP3 (CAMS1)	~50	0.98	this work
PEN5 (MRN1)	>99	0.99	4
LUP1	~1	0.4	1a
PEN1	20−25		1g
PEN2 (BARS1)	~33	0.90	1h

[a] Cyclase accuracy is quantified as the ratio of the primary product to the second most abundant product (P_1/P_2) or to total products ($P_1/\Sigma P_i$).[1h]

of CAMS1 to generate minor products in the 0.1−1% range is obviously much lower than that of BARS1 and PEN1. The high product accuracies of CAMS1 and MRN1[4] are partially attributable to their short biosynthetic pathways. For example, CAMS1 needs only to exclude rearrangement, further cyclization, and aberrant deprotonation of cation **I** (Scheme 1). In contrast, cyclases that make tetracycles and

Scheme 1. Mechanistic Pathway for the Formation of Camelliol C (**2**), Achilleol A (**3**), and β-Amyrin (**4**) from Oxidosqualene (**1**)



pentacycles generate many carbocationic intermediates, each of which is subject to premature deprotonation or rearrangement.

Phylogenetic analysis shows a close relationship between CAMS1 and enzymes that generate polycyclic triterpene alcohols (Figure 2). CAMS1 is most closely related to other oxidosqualene cyclases in the *Arabidopsis* LUP clade (70−78% identical), which form pentacycles or tetracycles. More distantly related are families of enzymes that make the

(5) (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.

(6) SMY8 is a lanosterol synthase deletion mutant that accumulates the substrate oxidosqualene and consequently supports in vivo biosynthesis: Corey, E. J.; Matsuda, S. P. T.; Baker, C. H.; Ting, A. Y.; Cheng, H. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 327−331.

(7) RXY6 is a squalene epoxidase/lanosterol synthase double deletion mutant that is useful for in vitro work because it cannot generate oxidosqualene: Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678−5679.

(8) Akihisa, T.; Arai, K.; Kimura, Y.; Koike, K.; Kokke, W. C. M. C.; Shibata, T.; Nikaido, T. *J. Nat. Prod.* **1999**, *62*, 265−268.

(9) Barrero, A. F.; Alvarez-Manzaneda, E. J.; Alvarez-Manzaneda, R. *Tetrahedron Lett.* **1989**, *30*, 3351−3352.

(10) Goad, L. J.; Akihisa, T. *Analysis of Sterols*; Blackie (Chapman & Hall): London, 1997; appendix 3, section (h).

(11) We favor in vitro yields because in vivo accumulation can distort the enzymatic product profile. See: (a) Joubert, B. M.; Hua, L.; Matsuda, S. P. T. *Org. Lett.* **2000**, *2*, 339−341. (b) Lodeiro, S.; Wilson, W. K.; Shan, H.; Matsuda, S. P. T. *Org. Lett.* **2006**, *8*, 439−442.

(12) Several other cyclases have qualitatively high product ratios,[1a] but because minor products were not quantitated and detection limits were not reported, quantitative measures of product specificity are unavailable.

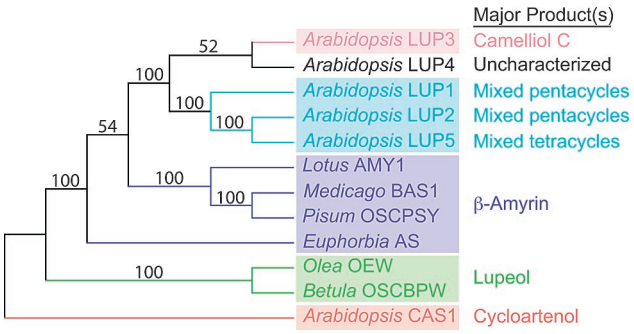


Figure 2. Computed phylogenetic relationships among known[1a,c] plant oxidosqualene cyclases closely related to CAMS1. The nucleotide sequences of these enzymes were aligned with the MegAlign program (DNASTAR, Inc., Madison, WI) by the Clustal W method. The alignment was used to create a phylogenetic tree rooted with the *A. thaliana* CAS1 using PAUP 4.05b. The tree was generated using the bootstrap method with 100 replicates with equal weight given to all the characters and maximum likelihood as the optimality criterion. Bootstrap values are listed at nodes. Cyclases are grouped by color according to phylogeny and product structure. For the full species names, see Figure 3.

	*	
Olea europa OEW	H G W Q V S D C T	485
Betula platyphylla OSCBPY	H G W Q V S D C T	487
Euphorbia tirucalli AS	H G W Q V S D C T	487
Medicago truncatula BAS1	H G W Q V S D C T	487
Glycyrrhiza glabra AS1	H G W Q V S D C T	487
Lotus japonicus AMY1	H G W Q V S D C T	475
Panax ginseng OSCPNY1	H G W Q V S D C T	488
Pisum sativum OSCPSY	H G W Q V S D C T	487
Arabidopsis thaliana PEN5 (MRN1)	Q G L P I S D G T	489
Arabidopsis thaliana LUP3 (CAMS1)	H G W Q A S D C T	488
Betula platyphylla OSCBPD	H G W Q A S D C T	460
Arabidopsis thaliana LUP4	H G W Q V S D C T	488
Arabidopsis thaliana LUP2	H G W Q V S D C T	488
Arabidopsis thaliana LUP1	H G W Q V S D C T	485
Arabidopsis thaliana LUP5	H G W Q V S D C T	488
Arabidopsis thaliana CAS1	H G W P I S D C T	485
Saccharomyces cerevisiae ERG7	Q G Y T V A D C T	458
Homo sapiens LSS	C G W I V S D C T	457

Figure 3. Partial amino acid alignment of oxidosqualene cyclases.[1a,c,15] An asterisk (*) designates position 484 in CAMS1, where steric bulk impacts B-ring formation.

pentacycles β-amyrin (72−75% identical) and lupeol (59−61% identical).[1a,13] One might imagine that cyclases that polycyclize oxidosqualene evolved from enzymes that form smaller ring systems by iterative addition of motifs that favor additional rings. However, phylogenetic analysis indicates the reverse evolutionary order, i.e., CAMS1 is a descendent of enzymes that form pentacyclic triterpenoids.

Sequence alignments show that nearly all plant cyclases contain valine or isoleucine at the position corresponding to Ala484 in CAMS1 (Figure 3). Previous mutational studies[11a,14] suggested that decreased steric bulk at this position (notably mutation to alanine or glycine) promotes the formation of monocycles. Our CAMS1 results strongly support this proposal. An uncharacterized cyclase from *Betula platyphylla* (OSCBPD)[15] also encodes alanine at this position and may likewise be compromised in B-ring formation.

The overwhelming dominance of **2** in the product profile of CAMS1 suggests that camelliol C or a metabolite thereof provides a competitive advantage that achilleol A does not replicate. The biological role of enzymes and their products can often be illuminated by microarray data. Unfortunately, the initial annotation of the *Arabidopsis* genome included *At1g78955* (*CAMS1*) as a fusion with *At1g78950* (*LUP4*), and these genes were not distinguished in early microarrays. The limited available nonarray expression data (Arabidopsis MPSS Plus: Gene Analysis) show elevated amounts of *At1g78955* mRNA in *Arabidopsis* inflorescence tissue.[16]

Monocyclic triterpenoids appear to be distributed sparsely across the vast diversity of higher plants. Compounds **2** and **3** have been found in a handful of asterids (*Camellia sasanqua*,[8,17a] *Camellia japonica*,[8,17a] *Achillea odorata*,[9] *Bupleurum spinosum*,[17b] and *Santolina elegans*[17c]), two eurosids (*Euphorbia antiquorum*[17d] and *Garcinia speciosa*[17e]), the monocots wheat and rice,[17a] and the fern *Polypodiodes formosana*.[17f] Because monocyclic and polycyclic triterpenes have rather different spectral and chromatographic signatures, most monocycles were described as components of plant oils (usually lacking polycyclic triterpenes) rather than in surveys of triterpene distribution. We suspect that the monocyclic triterpenes are more widespread than the literature suggests.

How often have cyclases that generate A-ring monocycles evolved? This question can be addressed by considering gene divergences (Figure 2) in the context of organismal relationships. CAMS1 diverged relatively recently from LUP4 within the LUP clade. LUP enzymes appear to be unique to malvids (eurosids II). We conclude that CAMS1 arose within the malvids and is evolutionarily distinct from the unknown enzymes that generate monocyclic triterpenes in asterids, monocots, and ferns.[18] The phylogenetic analyses suggest

(13) (a) Still more distantly related are cyclases of the *Arabidopsis* PEN clade,[1a] three of which make fewer than four rings. (b) Protein sequence identities were calculated in MegAlign from pairwise alignments.

(14) Matsuda, S. P. T.; Darr, L. B.; Hart, E. A.; Herrera, J. B. R.; McCann, K. E.; Meyer, M. M.; Pang, J.; Schepmann, H. G. *Org. Lett.* **2000**, *2*, 2261−2263.

(15) Zhang, H.; Shibuya, M.; Yokota, S.; Ebizuka, Y. *Biol. Pharm. Bull.* **2003**, *26*, 642−650.

(16) Data available at http://mpss.udel.edu/at/.

(17) (a) Akihisa, T.; Koike, K.; Kimura, Y.; Sashida, N.; Matsumoto, T.; Ukiya, M.; Nikaido, T. *Lipids* **1999**, *34*, 1151−1157. (b) Barrero, A. F.; Haidour, A.; Munoz-Dorado, M.; Akssira, M.; Sedqui, A.; Mansour, I. *Phytochemistry* **1998**, *48*, 1237−1240. (c) Barrero, A. F.; Alvarez-Manzaneda, E. J.; Herrador, M. M.; Alvarez-Manzaneda, R.; Quilez, J.; Chahboun, R.; Linares, P.; Rivas, A. *Tetrahedron Lett.* **1999**, *40*, 8273−8276. (d) Akihisa, T.; Wijeratne, E. M. K.; Tokuda, H.; Enjo, F.; Toriumi, M.; Kimura, Y.; Koike, K.; Nikaido, T.; Tezuka, Y.; Nishino, H. *J. Nat. Prod.* **2002**, *65*, 158−162. (e) Rukachaisirikul, V.; Pailee, P.; Hiranrat, A.; Tuchinda, P.; Yoosook, C.; Kasisit, J.; Taylor, W. C.; Reutrakul, V. *Planta Med.* **2003**, *69*, 1141−1146. (f) Arai, Y.; Hirohara, M.; Ageta, H.; Hsu, H. Y. *Tetrahedron Lett.* **1992**, *33*, 1325−1328.

that natural selection can readily modify a cyclase that makes polycycles to instead deprotonate after A-ring formation. Frequent evolutionary events of this kind may underlie the punctate distribution of monocyclic triterpenoids across the plant kingdom.

Acknowledgment. The National Science Foundation (MCB-0209769), The Robert A. Welch Foundation (C-1323), and the Herman Frasch Foundation funded this research. The 800 MHz NMR spectra were acquired on instrumentation purchased by the John S. Dunn, Sr. Gulf Coast Consortium for Magnetic Resonance.

Supporting Information Available: Experimental procedures and NMR and GC−MS spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

OL702399G

(18) If the *Betula platyphylla* OSCBPD enzyme[15] (noted above) makes monocycles, this would represent a second evolution of an enzyme that forms monocycles within the eurosids because CAMS1 is too distant (63% identical) to be an ortholog.