ORGANIC LETTERS

2009
Vol. 11, No. 12
2627−2630

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

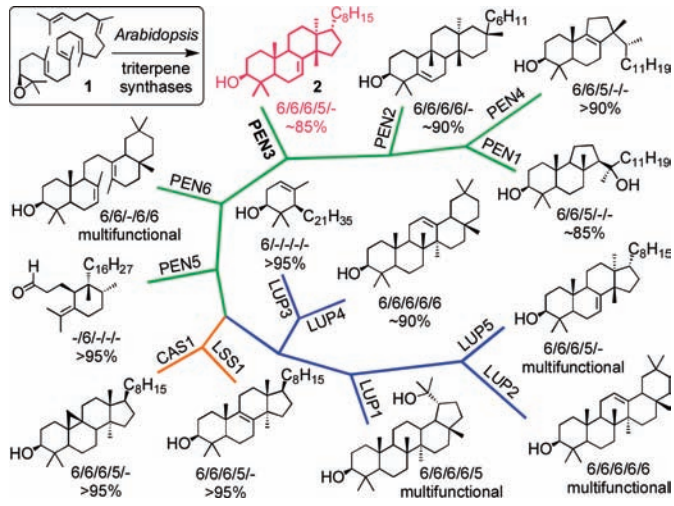
Pietro Morlacchi,[†] William K. Wilson,[‡] Quanbo Xiong,[‡] Aparna Bhaduri,[‡] Diana Sttivend,[‡] Mariya D. Kolesnikova,[†] and Seiichi P. T. Matsuda*[,†,‡]

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

matsuda@rice.edu

Received March 18, 2009

ABSTRACT



The triterpene product profile is reported for At5g36150 (PEN3), the last unexamined oxidosqualene cyclase in the reference plant *Arabidopsis thaliana*. PEN3 makes tirucalla-7,24-dien-3β-ol (~85%) and several minor products. Also discussed are the unexpectedly facile convergent evolution of another *Arabidopsis* tirucalladienol synthase (LUP5), mechanistic origins of the 20*S* configuration, and active-site remodeling necessary to accommodate the 17α side chain. This work marks the first completed functional characterization of all triterpene synthases in a higher plant.

Triterpenoids comprise a large and structurally diverse class of natural products. The key shared step in eukaryotic triterpenoid biosynthesis is the enzymatic cyclization of (3*S*)-oxidosqualene (**1**).[1] Regio- and stereochemical alternatives during the cyclization-rearrangement process can generate over 100 different skeletal types,[1d] which variously undergo oxidation, glycosylation, and other transformations to produce over 10 000 triterpenoids. Individual plant species exploit only a small fraction of this structural diversity, often with their own distinctive collection of triterpenoids. Some plants, such as licorice and ginseng, are renowned for their triterpenoid bouquet. In contrast, nonsterol triterpenoids were thought to be sparse[2] in the reference plant *Arabidopsis*

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

(1) Reviews: (a) Phillips, D. R.; Rasbery, J. M.; Bartel, B.; Matsuda, S. P. T. *Curr. Opin. Plant Biol.* **2006**, *9*, 305–314. (b) Abe, I. *Nat. Prod. Rep.* **2007**, *24*, 1311–1331. (c) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2000**, *39*, 2812–2833. (d) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261–290.

(2) D'Auria, J. C.; Gershenzon, J. *Curr. Opin. Plant Biol.* **2005**, *8*, 308–316.

10.1021/ol9005745 CCC: $40.75
Published on Web 05/15/2009

thaliana until recent evidence from genome mining revealed a potential abundance of triterpenes[1a] and triterpenoids.[3]

The *A. thaliana* genome encodes 13 oxidosqualene cyclases, 12 of which have received at least cursory functional characterization. Described herein is the product profile of the remaining uncharacterized cyclase, At5g36150 (PEN3), which generates predominantly tirucalla-7,24-dien-3β-ol. *A. thaliana* is now the first higher plant for which the catalytic function of all triterpene synthases is at least partially known.

At5g36150 cDNA was obtained by RT-PCR amplification of *A. thaliana* mRNA and subcloned into the yeast expression vector pRS426Gal.[4] The resultant plasmid pDS3.0 was used to transform SMY8,[5,6] a yeast lanosterol synthase deletion mutant unable to cyclize **1**. Cultures of SMY8[pDS3.0] totaling 24 L were grown to saturation in synthetic complete medium lacking uracil and supplemented with galactose, hemin chloride, and ergosterol. The resulting cell pellet was saponified, and the hexane extract was rapidly chromatographed to remove **1** and dioxidosqualene, which might otherwise generate artifacts of nonenzymatic cyclization.[7] Preparative TLC provided the triterpene alcohols. Each column and TLC fraction was analyzed by GC-MS and NMR to rule out a significant formation of diol, 3,10-epoxy, 3-keto, or Grob fragmentation products. These analyses and additional experiments suggested that the PEN3 product profile was not significantly distorted by metabolism or triterpene loss to the culture medium (see the Supporting Information).

The triterpene alcohol band from preparative TLC was analyzed by NMR and GC-MS (Figure 1). The major product

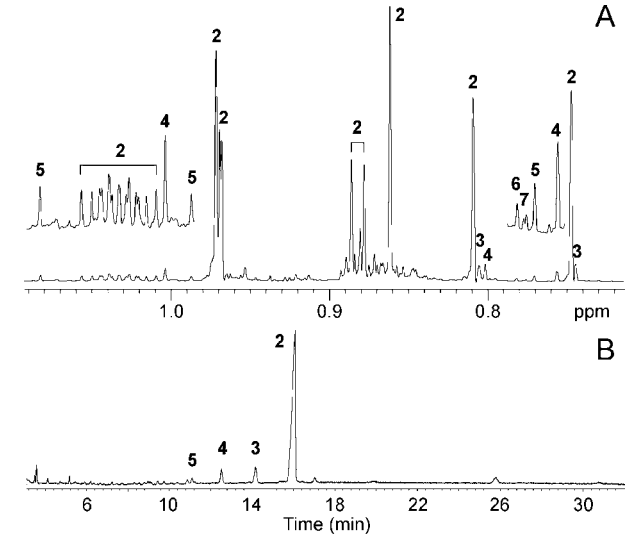


Figure 1. (A) ^{1}H NMR spectrum (800 MHz, upfield methyl region) and (B) GC-MS total ion chromatogram of the PEN3 triterpene band from preparative TLC. Minor unlabeled GC-MS peaks represent nontriterpene impurities.

was identified as tirucalla-7,24-dien-3β-ol (**2**, Figure 2) by comparison with reported proton[8] and carbon[9] NMR chemical shifts. Quantum mechanical predictions of NMR chemical shifts and connectivities from 2D NMR spectra were fully

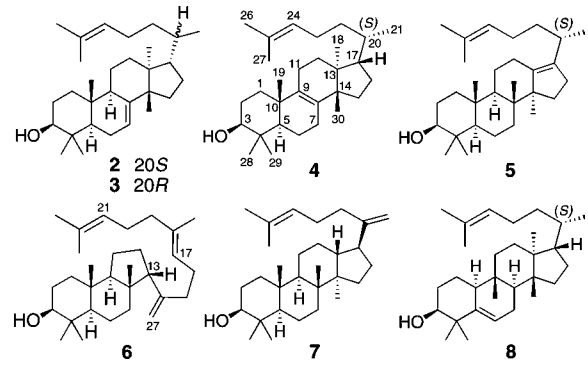


Figure 2. Products of PEN3 (**2**-**7**) and LUP5 (**2**-**6**, **8**).

compatible with the structure of **2**. NMR signal assignments for **2**, ^{1}H−^{1}H coupling constants, and predicted chemical shifts are given in the Supporting Information (Tables S1−S4).

The NMR spectrum revealed five minor triterpene products (**3**-**7**), three of which were visible in the total ion chromatogram of the GC-MS (Figure 1). The byproducts, most of which were isolated by HPLC, were quantified by GC-MS and NMR and identified by comparing their resolved ^{1}H and 2D NMR signals with literature values[7,8,10] (see the Supporting Information): butyrospermol (**3**, 6%), tirucallol (**4**, 6%), isotirucallol (**5**, 1.5%), 13βH-malabarica-14(27),17,21-trien-3β-ol (**6**, 1%), and dammara-20,24-dien-3β-ol (**7**, 0.5%).[11] PEN3 is thus a moderately accurate tirucalla-7,24-dien-3β-ol synthase. Its accuracy ($P_1/P_2 \approx 14$, $P_1/\Sigma P_i \approx 0.85$)[12] is comparable to that of the *Arabidopsis* baruol,[7] arabidiol,[13] and β-amyrin[14] synthases (PEN2, PEN1, and LUP4).

(3) (a) Field, B.; Osbourn, A. E. *Science* **2008**, *320*, 543–547. (b) Shan, H.; Wilson, W. K.; Phillips, D. R.; Bartel, B.; Matsuda, S. P. T. *Org. Lett.* **2008**, *10*, 1897–1900.

(4) (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**, 76.

(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) Because metabolism can sometimes bias in vivo product profiles, in vitro reactions are generally preferred for quantitating cyclase products. We turned to in vivo studies because various in vitro experiments with PEN3 failed to generate detectable triterpene products.

(7) 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.

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

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

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

(11) Interferences in the NMR and GC-MS spectra may have obscured additional minor products (including possible monocycles and pentacycles) formed at a level of ~1% or less.

(12) Cyclase accuracy was assessed as the ratio of the primary product to the second most abundant product (P_1/P_2) or to total products ($P_1/\Sigma P_i$), as described in ref 7.

(13) (a) Kolesnikova, M. D.; Obermeyer, A. C.; Wilson, W. K.; Lynch, D. A.; Xiong, Q.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 2183–2186. (b) Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835–2838.

(14) Shibuya, M.; Katsube, Y.; Otsuka, M.; Zhang, H.; Tansakul, P.; Xiang, T.; Ebizuka, Y. *Plant Physiol. Biochem.* **2009**, *47*, 26–30.

PEN3 annulates **1** to tetracyclic intermediates **IIa** and/or **IIb** with ~99% success (Figure 3). Reorientation to a vertical cation via pathway A_1 or B_1, followed by a C17→C20 hydride shift, results in the 20*S* cation **IIIa**. Further 1,2-shifts give cation **IVa**, which is deprotonated at C7 to generate the dominant product **2**. Aberrant deprotonation from C9 and C13 leads to byproducts **4** and **5**. An intriguing catalytic error is the formation of 6% butyrospermol (**3**), the C20 epimer of **2**. This byproduct could arise from **IIb** via pathway B_2 (blue arrows) or from **IIa** via pathway A_2 (green arrows). These pathways differ according to the C17 configuration of the tetracyclic cation (**IIa** or **IIb**) and the direction of the subsequent side-chain rotation about the C17−C20 bond.

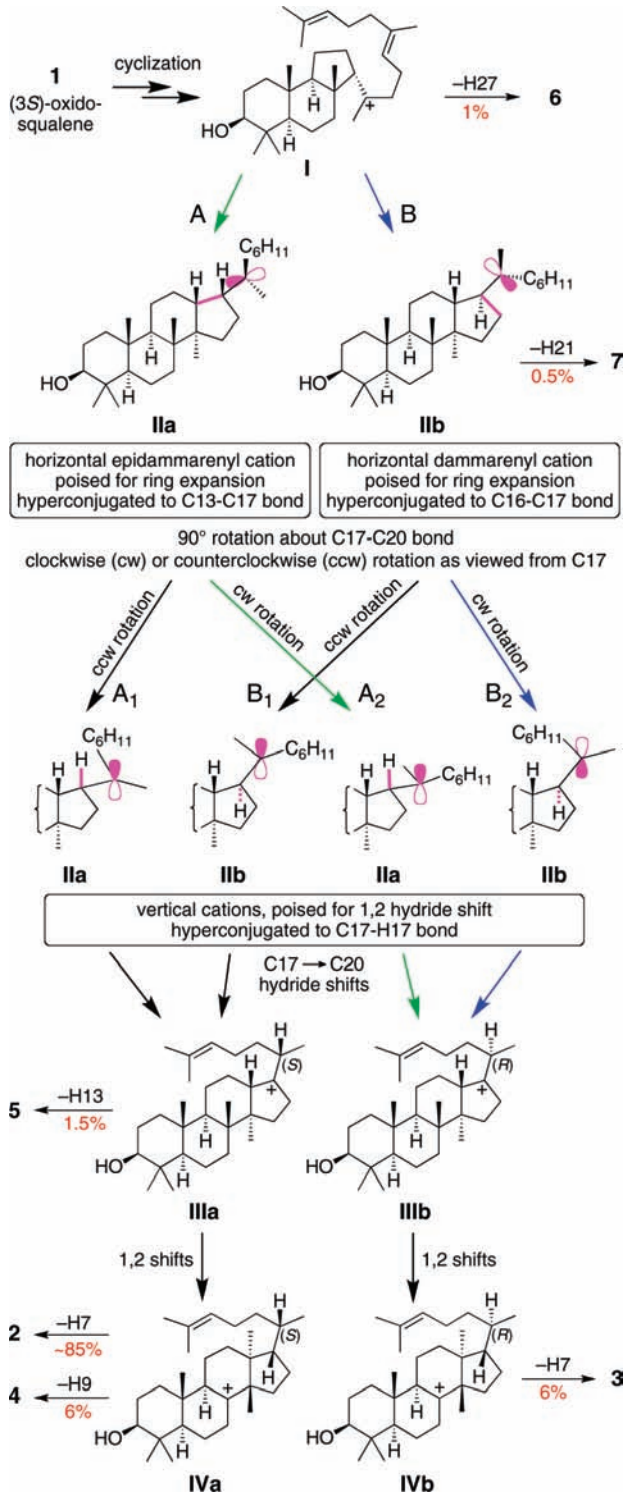


Figure 3. Possible mechanistic pathways to PEN3 products. Alternative pathways from **I** to vertical cations **IIa** and **IIb** are labeled as A_1, A_2, B_1, and B_2. Hyperconjugated bonds are shown in magenta. Side-chain rotation in **IIa** and **IIb** represents a conformational change from a horizontal to a vertical cation.[18] Ring D of **IIa** and **IIb** adopts a 13α envelope conformation.

The pathways for the major and minor PEN3 products **2**-**5** could thus access different tetracyclic cations or share the same tetracyclic cation and differ in side-chain rotation.[1d,15,16] If C20 epimers arise by opposite directions of C17−C20 rotation, PEN3 might generate extremes of 98.5% **IIa** (with 1% loss to **6** and 0.5% **IIb** leading to **7**) or 99% **IIb**. Alternatively, **2**, **4**, and **5** might arise from **IIa** via pathway A_1 and **3** from **IIb** via pathway B_2; such rotations would place the C_6H_{11} group *syn* to C13, a conformation that appears compatible with the L-shape of active-site cavities in oxidosqualene cyclases.[17]

The intermediacy of **IIb** has been established experimentally for LUP1 and theoretically for other cyclases that make nonhopanoid pentacycles.[15] Among tetracyclic products, dammarenyl cations with a 17β side chain have also been proposed on the basis of phylogenetic[15] and mechanistic[19] reasoning as the common precursor of tirucallol (20*S*) and euphol (20*R*). Nevertheless, elucidation of the mechanistic pathways to rearranged dammaranes (e.g., **2**-**5**) in PEN3 and other cyclases awaits definitive experimental evidence.

Another *Arabidopsis* cyclase, LUP5, makes a similar set of 6/6/6/5 tetracycles[20] (Figure 2). Despite the catalytic similarity of PEN3 and LUP5, their putative active-site residues are quite different (Figure 4A), consistent with convergent evolution. Notably, both enzymes have modified

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

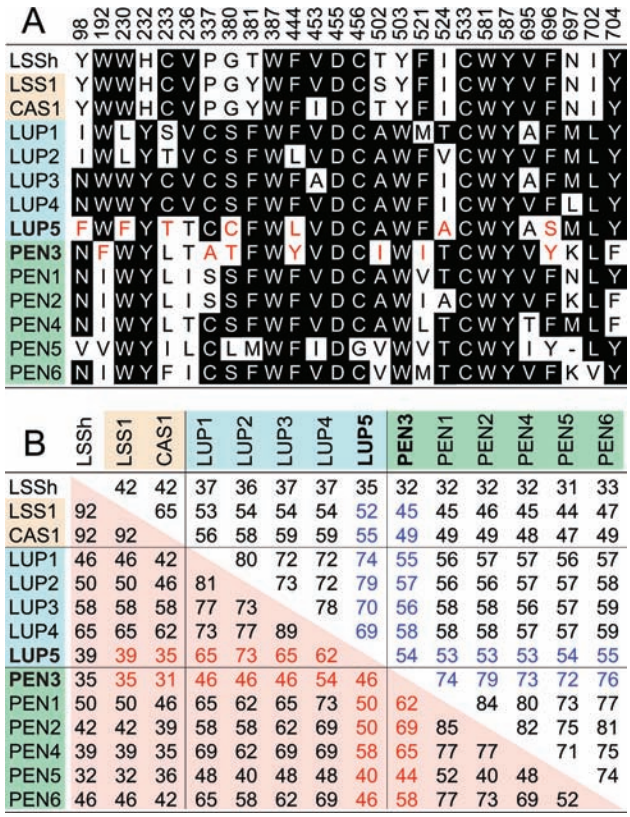
(16) These alternatives are exemplified in squalene cyclases and their mutants: (a) Pale-Grosdemange, C.; Feil, C.; Rohmer, M.; Poralla, K. *Angew. Chem., Int. Ed.* **1998**, *37*, 2237–2240. (b) Hoshino, T.; Sato, T. *Chem. Commun.* **2002**, 291–301. (c) Shinozaki, J.; Shibuya, M.; Masuda, K.; Ebizuka, Y. *Phytochemistry* **2008**, *69*, 2559–2564.

(17) Shape of the lanosterol substrate in the active-site cavity of a lanosterol synthase crystal structure: (a) Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. *Nature* **2004**, *432*, 118–122. Also, the 17*Z* geometry of (17*Z*)-protosta-17(20),24-dien-3β-ol implies a cationic precursor with C13 *syn* to the C_6H_{11} portion of the side chain: (b) 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.

(18) Matsuda, S. P. T.; Wilson, W. K.; Xiong, Q. *Org. Biomol. Chem.* **2006**, *4*, 530–543.

(19) (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. (c) In these papers, the pathway fork leading to 20*R* and 20*S* isomers (**IIIb** and **IIIa**) derives from different conformations of a 6-membered D-ring intermediate that incorporates a bridged C17-C20-C13 cation (chair form) or a C17-C20-C16 cation (boat form). The Supporting Information contains a side-by-side comparison of this mechanism with that of Figure 3.

(20) An HPLC-UV chromatogram of LUP5 products showed a ~4:4:1 ratio of three major triterpenes, one of which was identified as **2**: (a) Ebizuka, Y.; Katsube, Y.; Tsutsumi, T.; Kushiro, T.; Shibuya, M. *Pure Appl. Chem.* **2003**, *75*, 369–374. (b) Our preliminary NMR and GC-MS analyses of LUP5 products indicated a ~2:2:1 of **2**, **5**, and **8**, with minor products including **3**, **4**, and **6** (see the Supporting Information).



A

	98	192	230	232	233	236	337	380	381	387	444	453	455	456	502	503	521	524	533	581	587	695	696	697	702	704
LSSh	Y	W	W	H	C	V	P	G	T	W	F	V	D	C	T	Y	F	I	C	W	Y	V	F	N	I	Y
LSS1	Y	W	W	H	C	V	P	G	Y	W	F	V	D	C	S	Y	F	I	C	W	Y	V	F	N	I	Y
CAS1	Y	W	W	H	C	V	P	G	Y	W	F	I	D	C	T	Y	F	I	C	W	Y	V	F	N	I	Y
LUP1	I	W	L	Y	S	V	C	S	F	W	F	V	D	C	A	W	M	T	C	W	Y	A	F	M	L	Y
LUP2	I	W	L	Y	T	V	C	S	F	W	L	V	D	C	A	W	F	V	C	W	Y	V	F	M	L	Y
LUP3	N	W	W	Y	C	V	C	S	F	W	F	A	D	C	A	W	F	I	C	W	Y	A	F	M	L	Y
LUP4	N	W	W	Y	C	V	C	S	F	W	F	V	D	C	A	W	F	I	C	W	Y	V	F	L	L	Y
LUP5	F	W	F	Y	T	T	C	C	F	W	L	V	D	C	A	W	F	A	C	W	Y	A	S	M	L	Y
PEN3	N	F	W	Y	L	T	A	T	F	W	Y	V	D	C	I	W	I	T	C	W	Y	V	Y	K	L	F
PEN1	N	I	W	Y	L	I	S	S	F	W	F	V	D	C	A	W	V	T	C	W	Y	V	F	N	L	Y
PEN2	N	I	W	Y	L	I	S	S	F	W	F	V	D	C	A	W	I	A	C	W	Y	V	F	K	L	F
PEN4	N	I	W	Y	L	T	C	S	F	W	F	V	D	C	A	W	L	T	C	W	Y	T	F	M	L	F
PEN5	V	V	W	Y	I	L	C	L	M	W	F	I	D	G	V	W	V	T	C	W	Y	I	Y	-	L	Y
PEN6	N	I	W	Y	F	I	C	S	F	W	F	V	D	C	V	W	M	T	C	W	Y	V	F	K	V	Y

B

	LSSh	LSS1	CAS1	LUP1	LUP2	LUP3	LUP4	LUP5	PEN3	PEN1	PEN2	PEN4	PEN5	PEN6
LSSh		42	42	37	36	37	37	35	32	32	32	32	31	33
LSS1	92		65	53	54	54	54	52	45	45	46	45	44	47
CAS1	92	92		56	58	59	59	55	49	49	49	48	47	49
LUP1	46	46	42		80	72	72	74	55	56	57	57	56	57
LUP2	50	50	46	81		73	72	79	57	56	56	57	57	58
LUP3	58	58	58	77	73		78	70	56	58	58	56	57	59
LUP4	65	65	62	73	77	89		69	58	58	58	57	57	59
LUP5	39	39	35	65	73	65	62		54	53	53	53	54	55
PEN3	35	35	31	46	46	46	54	46		74	79	73	72	76
PEN1	50	50	46	65	62	65	73	50	62		84	80	73	77
PEN2	42	42	39	58	58	62	69	50	69	85		82	75	81
PEN4	39	39	35	69	62	69	69	58	65	77	77		71	75
PEN5	32	32	36	48	40	48	48	40	44	52	40	48		74
PEN6	46	46	42	65	58	62	69	46	58	77	73	69	52	

Figure 4. (A) Comparison of active-site residues (within 5 Å of lanosterol) in human lanosterol synthase[17a] (LSSh) with corresponding residues of triterpene synthases in *A. thaliana*. Nonconsensus PEN3 and LUP5 residues are shown in red. (B) Protein sequence identities for the full enzymes (upper right, unshaded) and active-site residues only (lower left, shaded).

the F696 residue,[21] which is highly conserved among lanosterol synthases and in the PEN and LUP clades.[22] In a human lanosterol synthase crystal structure,[17a] F696 lies closely beneath C17 and C20 of lanosterol and may interfere sterically with the formation of any intermediate bearing a 17α side chain, such as cations **IIa**, **IVa**, or **IVb**. Thus, oxidosqualene cyclases that make rearranged dammaranes may need to mutate or reposition[23] F696.

The significant remodeling of the cyclase active site to accommodate a 17α side chain is reflected in the sequence identities of PEN3 and LUP5 relative to other *Arabidopsis* cyclases (Figure 4B). The PEN3 and LUP5 identities are generally lower for putative active-site residues (red values) than for all residues (blue values). Moreover, the active sites of PEN3 and LUP5 include more nonconsensus and unique residues than other PEN and LUP cyclases except PEN5 (Figure 4A). These observations raise the credibility of **IIa** as the initial tetracyclic intermediate of PEN3 and LUP5 and challenge our postulate of a high evolutionary barrier between cyclases accessing **IIb** and those accessing **IIa**.[15]

With the characterization of PEN3, the triterpene products of all 13 oxidosqualene cyclases in *A. thaliana* have now been examined. As shown in the Graphical Abstract, these cyclases generate an impressive diversity of triterpene skeletal types. Two cyclases make sterol precursors. Only two cyclases make predominantly the usual pentacyclic triterpenes of secondary metabolism;[22] two others make 6/6/6/5 tetracyclic nonsteroidal triterpene alcohols. Two cyclases produce diols. Two more cyclases make either a bicyclic or pentacyclic cation and then cleave one ring through a Grob fragmentation. The remaining enzymes make a monocycle, tricycle, or 6/6/6/6 tetracycle. The many unconventional triterpene products are known in *Arabidopsis* mainly through genome mining and have been detected only infrequently in other plants. Much of the remarkable diversity of mechanism and product structure among the *Arabidopsis* cyclases was achieved with only limited modifications of the active-site residues (Figure 4). Ironically, no PEN cyclase makes mainly pentacycles, and no enzyme in the LUP clade makes predominantly lupeol.

Functional characterization of about half of the 13 *Arabidopsis* cyclases is essentially finalized, minor products having been identified and quantified down to the level of ~1% or better. However, product profiles and structure assignments are incomplete for most multifunctional cyclases. Further discussion of the collective functionality of *Arabidopsis* triterpene synthases awaits completion of these characterizations.

Acknowledgment. The National Science Foundation (MCB-0209769) and The Robert A. Welch Foundation (C-1323) 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, NMR and GC-MS spectra, NMR signal assignments, structure identification of **2**-**7**, and PEN3 and LUP5 product ratios. This material is available free of charge via the Internet at http://pubs.acs.org.

OL9005745

(21) Human lanosterol synthase residue numbering is used herein.

(22) Elucidation of the PEN and LUP clades: Husselstein-Muller, T.; Schaller, H.; Benveniste, P. *Plant Mol. Biol.* **2001**, *45*, 75–92.

(23) Some cyclases containing F696 (F601 in squalene-hopene cyclase) make tetracyclic intermediates or products with a 17α side chain, e.g., tirucalla-7,24-dienol synthase in *Ailanthus altissima* (ref 9) and squalene-hopene cyclase (ref 1c, Figure 20 and ref 16b, Figure 3).