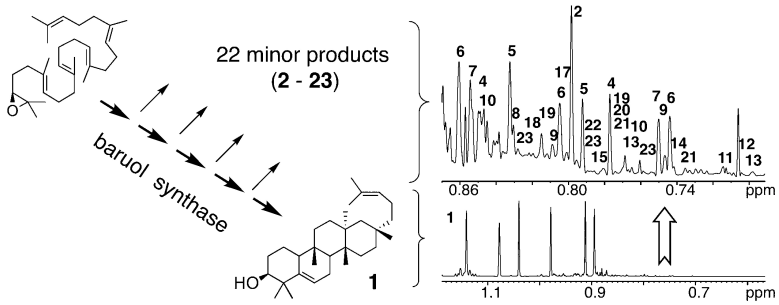
Article

An Oxidosqualene Cyclase Makes Numerous Products by Diverse Mechanisms: A Challenge to Prevailing Concepts of Triterpene Biosynthesis

Silvia Lodeiro, Quanbo Xiong, William K. Wilson, Mariya D. Kolesnikova, Carl S. Onak, and Seiichi P. T. Matsuda

J. Am. Chem. Soc., **2007**, 129 (36), 11213-11222• DOI: 10.1021/ja073133u • Publication Date (Web): 18 August 2007



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 5 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 08/18/2007

An Oxidosqualene Cyclase Makes Numerous Products by Diverse Mechanisms: A Challenge to Prevailing Concepts of Triterpene Biosynthesis

Silvia Lodeiro,[†] Quanbo Xiong,[‡,§] William K. Wilson,[‡] Mariya D. Kolesnikova,[†] Carl S. Onak,[†] and Seiichi P. T. Matsuda*[,†,‡]

Contribution from the Departments of Chemistry and of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005, and Department of Pharmaceutical Sciences, Texas Southern University, Houston, Texas 77004

Received May 3, 2007; E-mail: matsuda@rice.edu

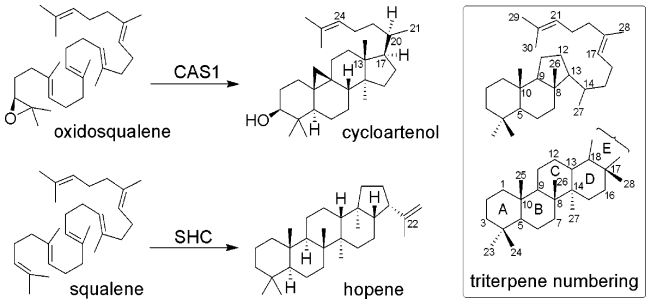
Abstract: The genome of the model plant *Arabidopsis thaliana* encodes 13 oxidosqualene cyclases, 9 of which have been characterized by heterologous expression in yeast. Here we describe another cyclase, baruol synthase (BARS1), which makes baruol (90%) and 22 minor products (0.02−3% each). This represents as many triterpenes as have been reported for all other *Arabidopsis* cyclases combined. By accessing an extraordinary repertoire of mechanistic pathways, BARS1 makes numerous skeletal types and deprotonates the carbocation intermediates at 14 different sites around rings A, B, C, D, and E. This undercurrent of structural and mechanistic diversity in a superficially accurate enzyme is incompatible with prevailing concepts of triterpene biosynthesis, which posit tight control over the mechanistic pathway through cation−π interactions, with a single proton acceptor in a hydrophobic active site. Our findings suggest that mechanistic diversity is the default for triterpene biosynthesis and that product accuracy results from exclusion of alternative pathways.

1. Introduction

Triterpene synthases convert oxidosqualene and squalene to polycyclic skeletons in one continuous reaction (Scheme 1).[1] Some cyclases perform the intricate transformation with apparent perfection, generating a single product out of thousands of conceivable isomers. Despite myriad attempts by chemists to tame this reaction, neither biomimetic synthesis nor enzyme engineering can routinely match the performance of cyclases created by natural selection.

The enzymatic mechanism has been studied intensively during the past 60 years.[2] The early focus on lanosterol synthase (LSS) and squalene-hopene cyclase (SHC) has expanded to include many cyclases from *Arabidopsis thaliana*[3,4,5] and other plants.[4] X-ray crystallography has revealed much about how the active-site residues guide the cyclization.[1a,b,6] Although crystal structures are available only for LSS and SHC,[1a,b,6] active-site geometries of mutants and other cyclases are inferred by homology modeling.[5k,7] Specific catalytic roles are now routinely attributed to individual residues.[8] The most commonly cited role involves the cation−π interaction,[8,9] in which an aromatic residue stabilizes a particular carbocationic intermediate and thereby directs the mechanism along a specific pathway. A constellation of such stabilizing residues in a hydrophobic active site is thought to guide a sequence of rearrangements to the final cationic intermediate, which is then deprotonated by a precisely positioned base. This logic is commonly used to rationalize the specificity of cyclases.

Scheme 1. Representative Cyclizations of (Oxido)squalene



Cyclases are often categorized as either multifunctional or accurate.[4] Multifunctional cyclases generate many products, none of which is dominant, whereas accurate cyclases seem to approach or achieve catalytic perfection. *Arabidopsis* cy-

[†] Department of Chemistry, Rice University.
[‡] Department of Biochemistry and Cell Biology, Rice University.
[§] Department of Pharmaceutical Sciences, Texas Southern University.

(1) Reviews: (a) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2000**, *39*, 2812−2833. (b) Wendt, K. U. *Angew. Chem., Int. Ed.* **2005**, *44*, 3966−3971. (c) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261−291. (d) Hoshino, T.; Sato, T. *Chem. Commun.* **2002**, 291−301. (e) Yoder, R. A.; Johnston, J. N. *Chem. Rev.* **2005**, *105*, 4730−4756. (f) Abe, I.; Rohmer, M.; Prestwich, G. D. *Chem. Rev.* **1993**, *93*, 2189−2206. (g) Segura, M. J. R.; Jackson, B. E.; Matsuda, S. P. T. *Nat. Prod. Rep.* **2003**, *20*, 304−317. (h) Cattel, L.; Ceruti, M. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 353−373.
(2) (a) Eschenmoser, A.; Ruzicka, L.; Jeger, O.; Arigoni, D. *Helv. Chim. Acta* **1955**, *38*, 1890−1904. (b) Stork, G.; Burgstahler, A. W. *J. Am. Chem. Soc.* **1955**, *77*, 5068−5077. (c) Eschenmoser, A.; Arigoni, D. *Helv. Chim. Acta* **2005**, *88*, 3011−3050.
(3) Husselstein-Muller, T.; Schaller, H.; Benveniste, P. *Plant Mol. Biol.* **2001**, *45*, 75−92.
(4) Review: Phillips, D. R.; Rasbery, J. M.; Bartel, B.; Matsuda, S. P. T. *Curr. Opin. Plant Biol.* **2006**, *9*, 305−314.

cloartenol synthase (CAS1)[10] and marneral synthase (MRN1)[5j] make only minor byproducts (<1% of total), and no byproducts from oxidosqualene have ever been reported for many cyclases, including LSS, cucurbitadienol synthase, five β-amyrin synthases, and four lupeol synthases.[11]

We now report that *At4g15370* of *A. thaliana* encodes a superficially accurate cyclase that makes a remarkable number of minor errors. We named the enzyme baruol synthase (BARS1) after its predominant product baruol. BARS1 makes 33 times as much baruol as any minor product, thus distinguishing itself from multifunctional cyclases and suggesting strong selective pressure toward a single product. However, BARS1 also produces over 20 minor products, which include monocycles, tricycles, tetracycles, and pentacycles. BARS1 accesses most of the known rearrangement pathways for cyclases,[1c,5j] performing a Grob fragmentation and making both C13 epimers of malabaricadienyl cations and both C20 epimers of tirucalla-7,24-dienols. Finally, cation deprotonation occurs from 14 sites widely distributed about the cyclic skeleton. Explaining this juxtaposition of nominal product accuracy, numerous minor products, and extraordinary mechanistic diversity is a challenge to the current concepts of triterpene biosynthesis. It is implausible that BARS1 possesses specific catalytic motifs to promote the individual pathway leading to each minor product, and we argue that mechanistic diversity is the default for triterpene cyclization and deprotonation.

The BARS1 study builds on previous efforts to mine the *Arabidopsis* genome for triterpenes. Although only two cyclases, CAS1[5a] and LUP2,[3,5f] are sufficient to produce all the C_{30} triterpene alcohols that have been detected in this plant,[12] the *Arabidopsis* genome encodes 13 homologs of triterpene synthases.[3,4] Nine of these have been at least partially characterized by cloning and heterologous expression in yeast,[3−5] and altogether 23 products have been found. BARS1 generates as many triterpenes as have been described for the 9 other characterized *Arabidopsis* cyclases combined (Figure 1). Described herein are the BARS1 products, a discussion of their mechanistic origin, and a reevaluation of factors that control product specificity in triterpene biosynthesis.

2. Results and Discussion

2.1. *At4g15370* Encodes a Baruol Synthase. The *Arabidopsis* gene *At4g15370* (*PEN2*) encodes an amino acid sequence 75−84% identical to the other enzymes in the *Arabidopsis* PEN clade and 53−56% identical to those in the LUP clade. Both clades generate nonsteroidal triterpenes.[13] *At4g15370* cDNA from *A. thaliana* mRNA was amplified by RT-PCR and subcloned into the pRS426GAL yeast expression vector.[5d,14] The resultant plasmid pXQ13.2 was used to transform the yeast production hosts SMY8[15] and RXY6.[5i] SMY8 has a lanosterol synthase deletion, which abolishes native oxidosqualene cyclization so that any cyclization observed must derive from the foreign enzyme. RXY6 additionally has a squalene epoxidase deletion, which prevents in vivo biosynthesis of oxidosqualene so that no triterpenes are present prior to the in vitro incubation. The two mutations of RXY6 facilitate in vitro experiments,[16] and low NADPH levels during the incubation preclude further metabolism of the cyclization products.[17]

An in vitro reaction was performed by incubating synthetic oxidosqualene[18] with a homogenate from a 2-L culture of RXY6[pXQ13.2]. ^{1}H NMR and GC-MS analyses indicated one major product and numerous minor compounds (~80% yield from (3*S*)-2,3-oxidosqualene). In vivo experiments with SMY8-[pXQ13.2] gave comparable results. The major product was established by 2D NMR as the D:B-friedobaccharane alcohol **1**. While we were characterizing the minor products, **1** was isolated from root bark of the Panamanian tree *Maytenus blepharodes* and named baruol.[19] Hence, *At4g15370* encodes a baruol synthase.

(5) *Arabidopsis* cyclases: CAS1: (a) Corey, E. J.; Matsuda, S. P. T.; Bartel, B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11628−11632. LSS1: (b) Kolesnikova, M. D.; Xiong, Q.; Lodeiro, S.; Hua, L.; Matsuda, S. P. T. *Arch. Biochem. Biophys.* **2006**, *447*, 87−95. (c) Suzuki, M.; Xiang, T.; Ohyama, K.; Seki, H.; Saito, K.; Muranaka, T.; Hayashi, H.; Katsube, Y.; Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Plant Cell Physiol.* **2006**, 47, 565−571. LUP1: (d) Herrera, J. B. R.; Bartel, B.; Wilson, W. K.; Matsuda, S. P. T. *Phytochemistry* **1998**, *49*, 1905−1911. (e) Segura, M. J. R.; Meyer, M. M.; Matsuda, S. P. T. *Org. Lett.* **2000**, *2*, 2257−2259. LUP2: ref 3 and (f) Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. *Tetrahedron Lett.* **2000**, *41*, 7705−7710. LUP5: (g) Ebizuka, Y.; Katsube, Y.; Tsutsumi, T.; Kushiro, T.; Shibuya, M. *Pure Appl. Chem.* **2003**, *75*, 369−374. PEN6: ref 5g and (h) Shibuya, M.; Xiang, T.; Katsube, Y.; Otsuka, M.; Zhang, H.; Ebizuka, Y. *J. Am. Chem. Soc.* **2007**, *129*, 1450−1455. PEN4 (THAS1): (i) Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678−5679. PEN5 (MRN1): (j) Xiong, Q.; Wilson, W. K.; Matsuda, S. P. T. *Angew. Chem., Int. Ed.* **2006**, *45*, 1285−1288. PEN1: (k) Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835−2838. (m) Kolesnikova, M. D.; Obermeyer, A. C.; Wilson, W. K.; Lynch, D. A.; Xiong, Q.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 2183−2186.

(6) (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. (b) Lenhart, A.; Weihofen, W. A.; Pleschke, A. E. W.; Schulz, G. E. *Chem. Biol.* **2002**, *9*, 639−645. (c) Lenhart, A.; Reinert, D. J.; Aebi, J. D.; Dehmlow, H.; Morand, O. H.; Schulz, G. E. *J. Med. Chem.* **2003**, *46*, 2083−2092.

(7) (a) Wu, T.-K.; Griffin, J.-H. *Biochemistry* **2002**, *41*, 8238−8244. (b) Lodeiro, S.; Segura, M. J. R.; Stahl, M.; Schulz-Gasch, T.; Matsuda, S. P. T. *ChemBioChem* **2004**, *5*, 1581−1585. (c) Schulz-Gasch, T.; Stahl, M. *J. Comput. Chem.* **2003**, *24*, 741−753. (d) Wu, T.-K.; Chang, C.-H. *ChemBioChem* **2004**, *5*, 1712−1715. (e) Oliaro-Bosso, S.; Schulz-Gasch, T.; Taramino, S.; Scaldaferri, M.; Viola, F.; Balliano, G. *Biochem. Soc. Trans.* **2005**, *33*, 1202−1205. (f) Lodeiro, S.; Schulz-Gasch, T.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2005**, *127*, 14132−14133.

(8) (a) Morikubo, N.; Fukuda, Y.; Ohtake, K.; Shinya, N.; Kiga, D.; Sakamoto, K.; Asanuma, M.; Hirota, H.; Yokoyama, S.; Hoshino, T. *J. Am. Chem. Soc.* **2006**, *128*, 13184−13194. (b) Wu, T.-K.; Liu, Y.-T.; Chiu, F.-H.; Chang, C.-H. *Org. Lett.* **2006**, *8*, 4691−4694. (c) Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835−2838. (d) Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. *J. Am. Chem. Soc.* **2000**, *122*, 6816−6824. (e) Morita, M.; Shibuya, M.; Kushiro, T.; Masada, K.; Ebizuka, Y. *Eur. J. Biochem.* **2000**, *267*, 3453−3460. (f) Wu, T.-K.; Liu, Y.-T.; Yu, M.-T.; Wang, H.-J. *J. Am. Chem. Soc.* **2006**, *128*, 6414−6419.

(9) Reviews: (a) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303−1324. (b) Zacharias, N.; Dougherty, D. A. *Trends Pharmacol. Sci.* **2002**, *23*, 281−287. Early application to triterpene synthesis: (c) ref 1f. (d) Poralla, K.; Hewelt, A.; Prestwich, G. D.; Abe, I.; Reipen, I.; Sprenger, G. *Trends Biochem. Sci.* **1994**, *19*, 157−158.

(10) Herrera, J. B. R.; Wilson, W. K.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2000**, *122*, 6765−6766.

(11) For citations to the accurate cyclases, see ref 4.

(12) Apart from cycloartenol, only lupeol, α-amyrin, and β-amyrin have been reported in native *A. thaliana* plant tissue.[3,5c] α-Amyrin and β-amyrin were found in leaf wax: (a) Jenks, M. A.; Tuttle, H. A.; Eigenbrode, S. D.; Feldmann, K. A. *Plant Phys.* **1995**, *108*, 359−377. Lanosterol has been detected in hairy root culture and in plants overexpressing LSS1.[5c] Oxidized triterpenoid secondary metabolites are known in Brassicaceae but not in *Arabidopsis*: (b) D'Auria, J. C.; Gershenzon, J. *Curr. Opin. Plant Biol.* **2005**, *8*, 308−316.

(13) Although LSS was originally designated as PEN7,[3] phylogenetic analyses in refs 4 and 5b indicate that LSS is in neither the PEN nor LUP clade.

(14) (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.

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

(16) Lodeiro, S.; Wilson, W. K.; Shan, H.; Matsuda, S. P. T. *Org. Lett.* **2006**, *8*, 439−442.

(17) Joubert, B. M.; Hua, L.; Matsuda, S. P. T. *Org. Lett.* **2000**, *2*, 339−341.

(18) Nadeau, R. G.; Hanzlik, R. P. *Methods Enzymol.* **1969**, *15*, 346−351.

(19) Nunez, M. J.; Lopez, M. R.; Jimenez, I. A.; Moujir, L. M.; Ravelo, A. G.; Bazzocchi, I. L. *Tetrahedron Lett.* **2004**, *45*, 7367−7370.

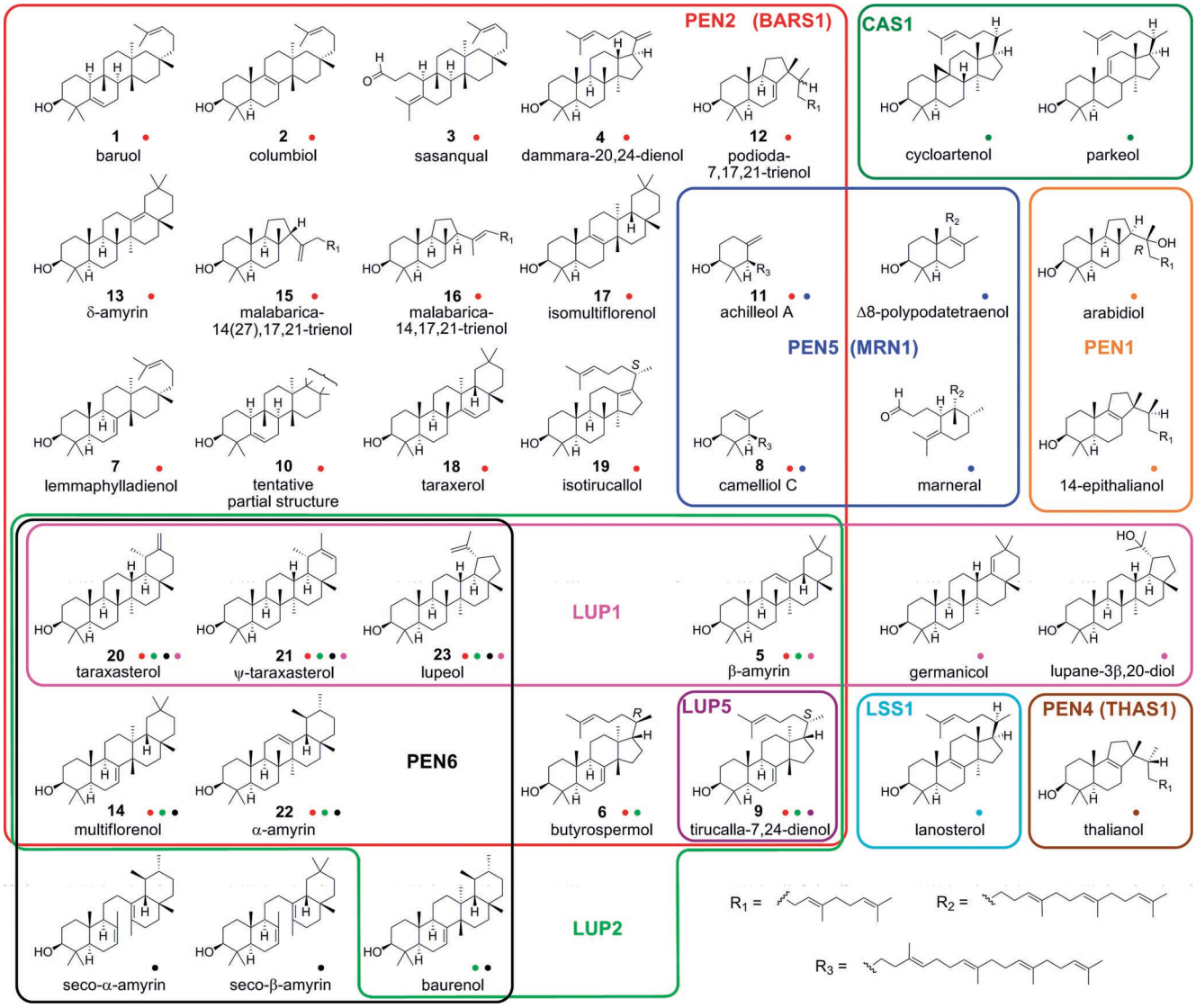


Figure 1. Triterpene structures from genome mining of *Arabidopsis* cyclases. Aldehyde **3** has not been isolated but is the presumed precursor of the observed alcohol **3a**. The structure of **10** has not been fully determined.

2.2. Identification of Minor BARS1 Products. The crude lipid extract from the in vitro reaction was subjected to a series of chromatographic separations (Figure 2). We avoided nonenzymatic cyclization of oxidosqualene during workup (see below) by immediately removing unreacted substrate on a short silica gel column. The combined triterpene eluate (fraction A) was separated by preparative TLC (PTLC) into fraction B_1 (mainly **1**) and fraction B_2 (most other triterpenes). Reversed-phase HPLC of fraction B_1 gave achilleol A (**11**),[20] camelliol C (**8**),[21] baruol (**1**),[19] and an unidentified Δ5 triterpene (**10**). Reversed-phase HPLC of fraction B_2 gave, in order of elution: (13*R*,17*E*)-malabarica-14(27),17,21-trien-3*β*-ol (**15**);[22] dammara-20,24-dien-3*β*-ol (**4**),[23] (13*R*,14ξ,17*E*)-podioda-7,17,21-trien-3*β*-ol (**12**),[24] (13*S*,14*E*,17*E*)-malabarica-14,17,21-trien-3*β*-ol (**16**); lupeol (**23**);[5d] tirucalla-7,24-dien-3*β*-ol (**9**),[25,26] lemmaphylla-7,21-dien-3*β*-ol (**7**);[27] *β*-amyrin (**5**),[28] butyrospermol (**6**),[8d] taraxerol (**18**),[29,30] taraxasterol (**20**),[28,30] columbiol (**2**); *δ*-amyrin (**13**);[30,31] multiflorenol (**14**),[5f,30] isomultiflorenol (**17**),[30,32] *ψ*-

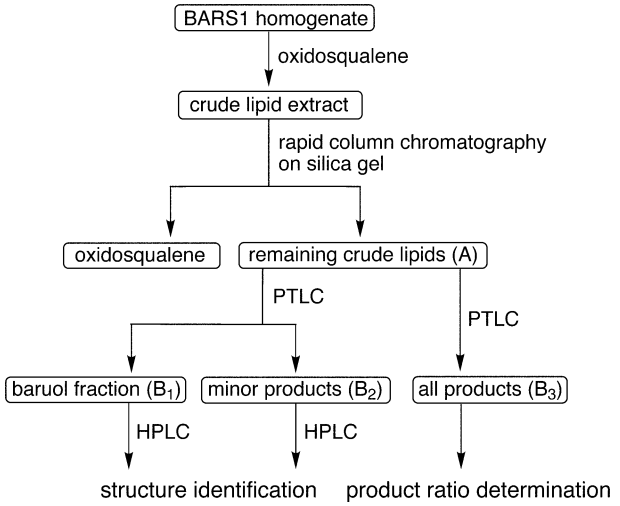


Figure 2. Chromatographic separations of the in vitro BARS1 products.

(20) Barrero, A. F.; Alvarez-Manzaneda, E. J.; Alvarez-Manzaneda, R. *Tetrahedron Lett.* **1989**, *30*, 3351−3352.
(21) Akihisa, T.; Arai, K.; Kimura, Y.; Koike, K.; Kokke, W. C. M. C.; Shibata, T.; Nikaido, T. *J. Nat. Prod.* **1999**, *62*, 265−268.
(22) Justicia, J.; Rosales, A.; Buñuel, E.; Oller-López, J. L.; Valdivia, M.; Haidour, A.; Oltra, J. E.; Barrero, A. F.; Cárdenas, D. J.; Cuerva, J. M. *Chem. Eur. J.* **2004**, *10*, 1778−1788.
(23) Leong, Y.-W.; Harrison, L. J. *Phytochemistry* **1999**, *50*, 849−857.

taraxasterol (**21**),[5f,28,30] and α-amyrin (**22**).[28,30] All products except **2**, **10**, **12**, and **16** were identified by electronically matching NMR and GC-MS spectra with spectra of authentic standards and by comparing spectral data with literature values.

The most abundant minor product **2** did not correspond to any triterpene standards. A large-scale culture of SMY8-[pXQ13.2] provided sufficient amounts of **2** for 2D NMR, which established the Δ8 D:C-friedobaccharane skeleton. We named **2** columbiol to recognize the organism that provided the genetic material (*A. thaliana* var Columbia). Structures of tricycles **12** and **16** were also elucidated by NMR, as described in the Supporting Information. The most abundant Δ5 triterpene besides baruol was **10**, which accounted for 0.3% of BARS1 products. The GC-MS and limited NMR data for **10** excluded known Δ5 triterpenes (see the Supporting Information), providing a partial structure determination. Additionally, several unidentified triterpene products were present at trace levels (<0.05%).

NMR and GC-MS analyses of the in vitro and in vivo samples showed the same minor products with a few exceptions. The large-scale culture showed isotirucallol (**19**) as a very minor product that was not detected in the in vitro experiment. A more abundant minor product found only in vivo was sasanquol (**3a**),[33] which is presumably derived from the postulated Grob fragmentation product sasanqual (**3**).[34] We have observed similar conversion of an enzymatic Grob fragmentation product (marneral) to its alcohol (marnerol) under in vivo conditions;[5j] this reduction may occur in yeast or during saponification.

2.3. Profile of BARS1 Products. PTLC of a second aliquot of fraction A of the in vitro reaction (Figure 2) afforded baruol and all minor triterpenes in a single fraction (B_3), which was analyzed by NMR and GC-MS (Figure 3). Whereas GC-MS suffered from extensive coelution of the enzymatic products, aliphatic methyl singlets in the δ_H 0.6−0.9 region of the 800 MHz NMR spectrum were sufficiently dispersed for preliminary quantitation of all BARS1 products except **19** and **3a**.[35] Several trace-level components were incompletely resolved in Figure 3, and additional quantitative information was obtained from relative signal intensities in spectra of PTLC and HPLC fractions. These complementary analyses were consolidated to provide the BARS1 product profile shown in Table 1.

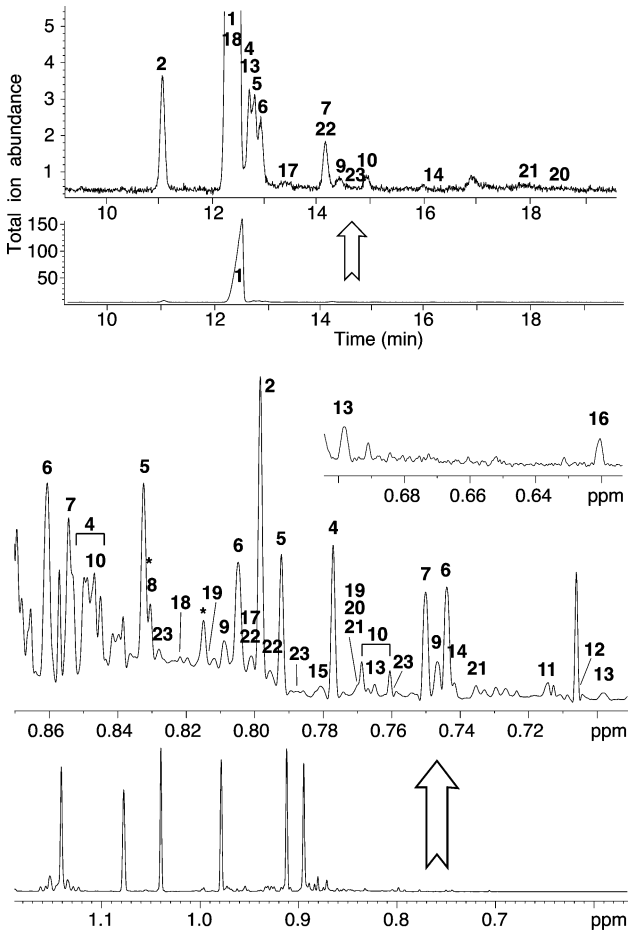


Figure 3. BARS1 products from the in vitro reaction (fraction B_3), as shown by selected portions of the total ion chromatogram from GC-MS (above) and the 800 MHz ^{1}H NMR spectrum (below). Asterisks (*) denote ^{13}C satellites from signals of **1**.

Quantitation by NMR was always based on at least two signal intensities (olefinic or methyl) and was broadly consistent with intensities from GC-MS analyses. The relatively short T_1 relaxation times of the angular methyl groups (<1 s), the similarity of their T_1 values, and our use of a $5T_1$ interval between pulses assured signal intensities that are proportional to the molar ratios (and hence mass ratios) of each component in the mixture. Potential errors from signal overlap were monitored by quantifying multiple resonances for each component and by using resolution enhancement to check for nearly coincident signals. Interference from non-triterpene impurities was minimal because extraneous lipids were removed by PTLC. The many checks for consistency of the quantitation ensured against any major errors.

2.4. Avoiding Artifacts. We found that oxidosqualene undergoes nonenzymatic cyclization on silica gel under certain conditions. If unreacted oxidosqualene is not removed promptly at the beginning of workup, these artifacts may form and be incorrectly assigned as enzymatic products. Although silica-mediated cyclization is not efficient, the high level of sensitivity we employed mandated appropriate precautions to protect against artifacts.[36] Whereas isocamelliol (**24**),[24a] thalianol (**25**),[5i] 14-epithalianol (**26**),[5m] and polypodatetraenols **27**−**29** were undetectable (<0.02% of total triterpenes) in the protected in

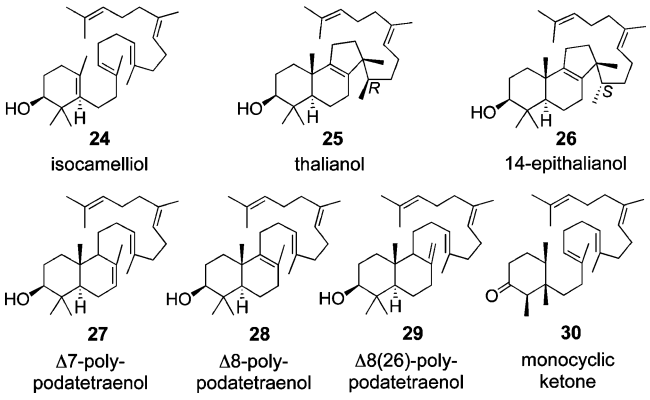
(24) (a) Kitahara, Y.; Kato, T.; Kishi, M. *Chem. Pharm. Bull.* **1968**, *16*, 2216−2222. (b) Arai, Y.; Hirohara, M.; Ageta, H. *Tetrahedron Lett.* **1989**, *30*, 7209−7212. (c) Hoshino, T.; Kouda, M.; Abe, T.; Sato, T. *Chem. Commun.* **2000**, 1485−1486.
(25) Itoh, T.; Tamura, T.; Matsumoto, T. *Lipids* **1976**, *11*, 434−441.
(26) Polonsky, J.; Varon, Z.; Rabanal, R. M.; Jacquemin, H. *Isr. J. Chem.* **1977**, *16*, 16−19.
(27) Akihisa, T.; Yasukawa, K.; Kimura, Y.; Takase, S.-I.; Yamanouchi, S.; Tamura, T. *Chem. Pharm. Bull.* **1997**, *45*, 2016−2023.
(28) Goad, L. J.; Akihisa, T. *Analysis of sterols*; Blackie (Chapman & Hall): London 1997, appendix 3, sections g and h.
(29) Bates, R. B.; Jacobsen, N. E.; Setzer, W. N.; Stessman, C. C. *Magn. Reson. Chem.* **1998**, *36*, 539−541.
(30) Bauer, S. Ph.D. dissertation, Westfälischen Wilhelms-Universität Münster, 2002.
(31) Tanaka, R.; Matsunaga, S. *Phytochemistry* **1988**, *27*, 3579−3584.
(32) Honda, C.; Suwa, K.; Takeyama, S.; Kamisako, W. *Chem. Pharm. Bull.* **2002**, *50*, 467−474.
(33) Akihisa, T.; Yasukawa, K.; Kimura, Y.; Yamanouchi, S.; Tamura, T. *Phytochemistry* **1998**, *48*, 301−305.
(34) Sasanqual has not been isolated but has implicitly been proposed as the precursor of sasanquol: Akihisa, T.; Koike, K.; Kimura, Y.; Sashida, N.; Matsumoto, T.; Ukiya, M.; Nikaido, T. *Lipids* **1999**, *34*, 1151−1157.
(35) The amount of sasanqual was estimated from the level of sasanquol from the in vivo experiment; **19** was also quantified from in vivo results. Although in vivo product profiles may be distorted through further metabolism, selective loss to the medium, and other factors, these effects were modest for BARS1. Nevertheless, our in vivo experiments were used mainly for obtaining standards and determining their structures rather than for quantifying product ratios.

Table 1. Yield and Characteristics of Baruol Synthase Products

product	% of total[a]	ring system	1,2-shifts[b]	proton loss[c]
1	89.7	6/6/6/6	6	C6
2	2.7	6/6/6/6	3	C9
3	1.8[d]	−/6/6/6[e]	6	O3
4	1.3	6/6/6/5	0	C28
5	1.2	6/6/6/6/6	2	C12
6	1.0	6/6/6/5	4	C7
7	0.9	6/6/6/6	3	C7
8	0.3	6	0	C1
9	0.3	6/6/6/5	4	C7
10	0.3	6/6/6/6/[f]	≥6[f]	C6
11	0.09	6	0	C25
12	0.07	6/6/5	2	C7
13	0.07	6/6/6/6/6	1	C13
14	0.06	6/6/6/6/6	4	C7
15	0.05	6/6/5	0	C27
16	0.05	6/6/5	0	C15
17	0.04	6/6/6/6/6	4	C9
18	0.04	6/6/6/6/6	3	C15
19	0.03	6/6/6/5	1	C13
20	0.03	6/6/6/6/6	1	C30
21	0.03	6/6/6/6/6	1	C21
22	0.02	6/6/6/6/6	4	C12
23	0.02	6/6/6/6/5	0	C29

[a] Yield from protected in vitro reaction. [b] Number of 1,2-hydride and methyl shifts after cyclization and any ring expansion. [c] Position of proton loss (triterpene numbering). [d] The yield of sasanqual (**3**) was estimated from that of sasanquol (**3a**) from in vivo reactions. [e] The −/6/6/6 ring system for **3** denotes a *seco*-A ring arising from Grob fragmentation. [f] The skeleton of **10** was not fully determined.

vitro reaction described above, **24**−**29** were present in unprotected workups of some early experiments. These artifacts were attributed to the PTLC conditions. More extensive nonenzymatic cyclization occurred when a crude product containing oxidosqualene was separated over 3 days on a large silica column. Tetracyclic and pentacyclic triterpenes were eluted with normal chromatographic behavior, but **15** and **24**−**28** were broadly distributed among the chromatographic fractions, indicating that they were formed on the column. Subsequent HPLC fractionation provided sufficient amounts of nonenzymatic products for structure identification by NMR, as described in the Supporting Information.



That oxidosqualene can cyclize on silica gel[37] was established by a control reaction in which oxidosqualene (shown by NMR

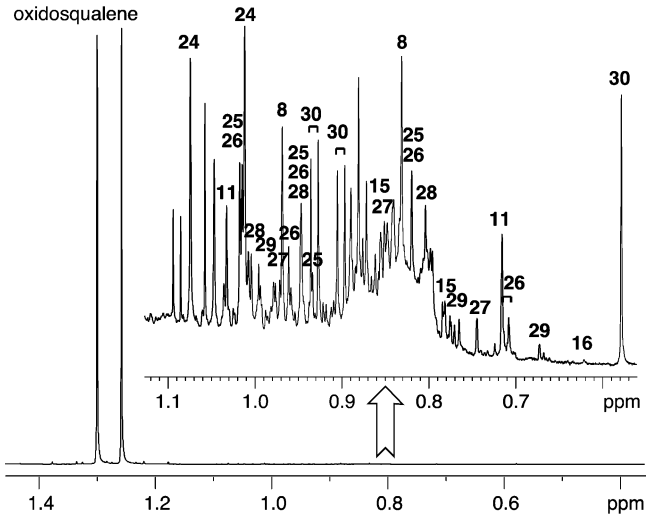


Figure 4. Partial 800 MHz ^{1}H NMR spectrum of a PTLC band containing oxidosqualene from the control reaction.

to contain at most 0.002% of any cyclic impurity) was incubated under our in vitro reaction conditions but without any yeast homogenate. NMR and GC-MS analysis of the crude product and fractions from a short silica gel column showed the absence of cyclized triterpenes (<0.02% relative to oxidosqualene). However, when the crude product was subjected to PTLC, mono-, bi-, and tricyclic triterpenes were found in several bands. Significant products were **24** and **30** (∼0.2% relative to oxidosqualene), with smaller amounts of **8**, **11**, **15**, and **25**−**29** (Figure 4). These results indicated the lack of nonenzymatic cyclization during incubation and rapid chromatography and demonstrated the ability of oxidosqualene to cyclize on silica gel under certain conditions.

In light of these findings, we reanalyzed the unprotected experiments with particular attention to monocycles, bicycles, and tricycles. Apart from **30**,[38] the most abundant artifact was **24**, which was present in several unprotected experiments, and its elevated levels correlated with elevated levels of **15**, **25**, and **26** but not **8**. We thus used the presence of **24** as a marker for nonenzymatic cyclization. The NMR methyl singlets for **24** are in a region that is usually free of interference and can be detected at levels as low as 0.01% of baruol.

Most mono-, bi-, and tricyclic triterpenes can arise by both enzymatic and nonenzymatic processes. For example, the consistently high levels of camelliol (0.3% of total triterpenes) in both protected and unprotected workups indicated its enzymatic origin. The most abundant triterpene alcohol artifact was typically **24**, and we considered as enzymatic any mono-, bi-, and tricyclic products found at levels substantially higher than the detection limit for **24**. Compounds **8**, **11**, **12**, **15**, and **16** from the protected in vitro experiment met this criterion and were included in the list of BARS1 products. Because no tetracycles or pentacycles were detected in the control reaction

(36) The protected workup procedure was developed and used[7b,f,16] previously to avoid artifactual formation of monocycles and noncyclized derivatives from oxidosqualene. Previous reports of native cyclases are probably also unaffected by artifact formation because no minor monocyclic, bicyclic, or tricyclic products were reported except in PEN1[5m] and MRN1.[5j] Artifact formation was explicitly excluded in the PEN1 study,[5m] and we have reexamined the MRN1 fraction containing **8**, **11**, and **28** and confirmed the absence of isocamelliol.

(37) Although we are unaware of reports of oxidosqualene cyclization on silica gel, Sen et al. have described cyclization of epoxypolyenes on zeolites: (a) Sen, S. E.; Zhang, Y. z.; Roach, S. L. *J. Org. Chem.* **1996**, *61*, 9534−9537. (b) Sen, S. E.; Zhang, Y. z.; Smith, S. M.; Huffman, J. C. *J. Org. Chem.* **1998**, *63*, 4459−4465. Reaction on silica gel gave products of epoxide ring opening but no cyclization.

(38) (a) Abe, I.; Rohmer, M. *J. Chem. Soc., Perkin Trans. 1* **1994**, 783−791. (b) **30** has different chromatographic behavior from the other triterpenes and was usually excluded from the analyses.

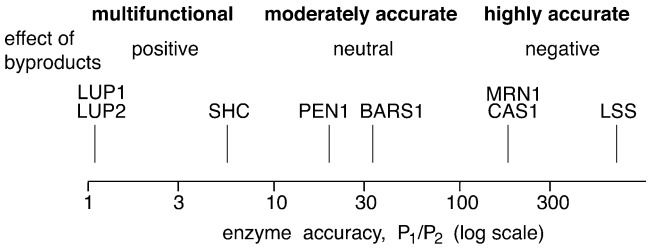


Figure 5. Range of product accuracy of (oxido)squalene cyclases. Improved analyses may result in substantial revisions to the P_1/P_2 values.

or in nonenzymatic cyclizations studied by others,[39] we regarded any amount of these triterpenes to be enzymatic.

We considered other sources of artifacts, including product isomerization, laboratory contamination (see the Supporting Information), and perturbation of the enzyme. The similarity of in vivo and in vitro results suggested that the environment of the enzyme had only modest effects on its product profile. Although heterologous expression might change protein structure to alter the nature of BARS1 catalysis, this caveat is no more pertinent to triterpene biosynthesis than to other recombinant systems. In another study, minor byproducts from in vitro reaction of purified SHC were shown not to be artifactual by detection of the same triterpenes in the native bacterial cell.[40]

2.5. Enzyme Accuracy.[41] Designating cyclases as either accurate or multifunctional is problematic. BARS1 obviously does not fit either canonical category. This dilemma prompted us to construct a continuum of cyclase accuracy based on the ratio of the primary product to the second most abundant product (P_1/P_2) or to total products ($P_1/\Sigma P_i$). By both criteria, BARS1 is less accurate ($P_1/P_2 \approx 33$; $P_1/\Sigma P_i \approx 0.90$) than the *Arabidopsis* CAS1 ($P_1/P_2 \approx 170$; $P_1/\Sigma P_i \approx 0.99$)[10] but considerably more accurate than *Arabidopsis* LUP1 ($P_1/P_2 \approx 1$; $P_1/\Sigma P_i \approx 0.4$)[5e] and the sesquiterpene cyclases δ-selinene synthase (34 products, $P_1/P_2 \approx 1.5$, $P_1/\Sigma P_i \approx 0.25$) and γ-humulene synthase (52 products, $P_1/P_2 \approx 1.9$, $P_1/\Sigma P_i \approx 0.29$).[42]

$P_1/\Sigma P_i$ requires a complete description of minor compounds. For partially characterized enzymes, we consequently prefer P_1/P_2, which necessitates reliable quantitation of only the two dominant components (Figure 5). Despite the merits of the P_1/P_2 continuum, some anomalies exist. For example, SHC might be presumed to be multifunctional based on its P_1/P_2 ratio of 5.[1d] However, a pioneering study of minor cyclase products by the groups of Rohmer and Poralla[40] found the next most abundant product at much lower levels ($P_1/P_3 \approx 50$) and showed that SHC cyclizes squalene to the final hopanyl cation with ~90−95% accuracy.

Problems with reporting an enzyme as accurate without quantifying minor products are illustrated by two investigations on the same enzyme. A single product was recently described for arabidiol synthase (PEN1),[5k] but more detailed analyses revealed numerous byproducts, with the most abundant at 4−5%.[5m] BARS1 would have similarly appeared to be fully accurate if it had been characterized at a 4% detection limit. Rigorous analyses that include in vitro assays[16] under documented conditions are essential. If the second most abundant product is not quantified, reporting the detection limit will allow an estimate, e.g., $P_1/P_2 > 10$ for a detection limit of 10%.

The accuracy of an enzyme is not readily understood without considering the nature of the selective pressure it has experienced. The most meaningful classifications of cyclase accuracy will describe whether byproducts generally have negative, neutral, or positive selective value. The selective pressure may be difficult to establish but can be roughly projected from P_1/P_2 values (Figure 5). The most accurate cyclases (LSS and CAS) are involved in sterol biosynthesis, where byproducts can have catastrophic effects. For example, genetic defects of mammalian sterol metabolism can generate elevated levels of sterol intermediates and byproducts that impair development, myelination, hormone synthesis, digestion, and other critical physiological functions.[43] In normal individuals, the same aberrant sterols are present at trace levels without adverse effects.[41b,44] The multiplicity of sterol byproducts is reminiscent of the BARS1 product profile, but selective pressures on sterol biosynthesis have evidently reduced the byproducts by an order of magnitude relative to those of BARS1.

Moderately and highly accurate cyclases are also found in secondary metabolism. The degree of accuracy appears to reflect a combination of selective pressure and catalytic constraints. High accuracy may be required for internal signaling functions and for further metabolism that entails high substrate specificity. In contrast, BARS1 may experience little selective pressure to increase $P_1/\Sigma P_i$ if the metabolic cost of its byproducts is modest and if their effects are neutral, e.g., because of excretion, segregation from the cell, or infrequent production in limited amounts. The moderate accuracy of PEN1 partially reflects the difficulty for a cyclase to hydroxylate a carbocation without some concomitant deprotonation.[5m] BARS1 catalysis accuracy could also be intrinsically constrained in excluding alternative pathways during the formation of the baccharenyl cation **VIa**.

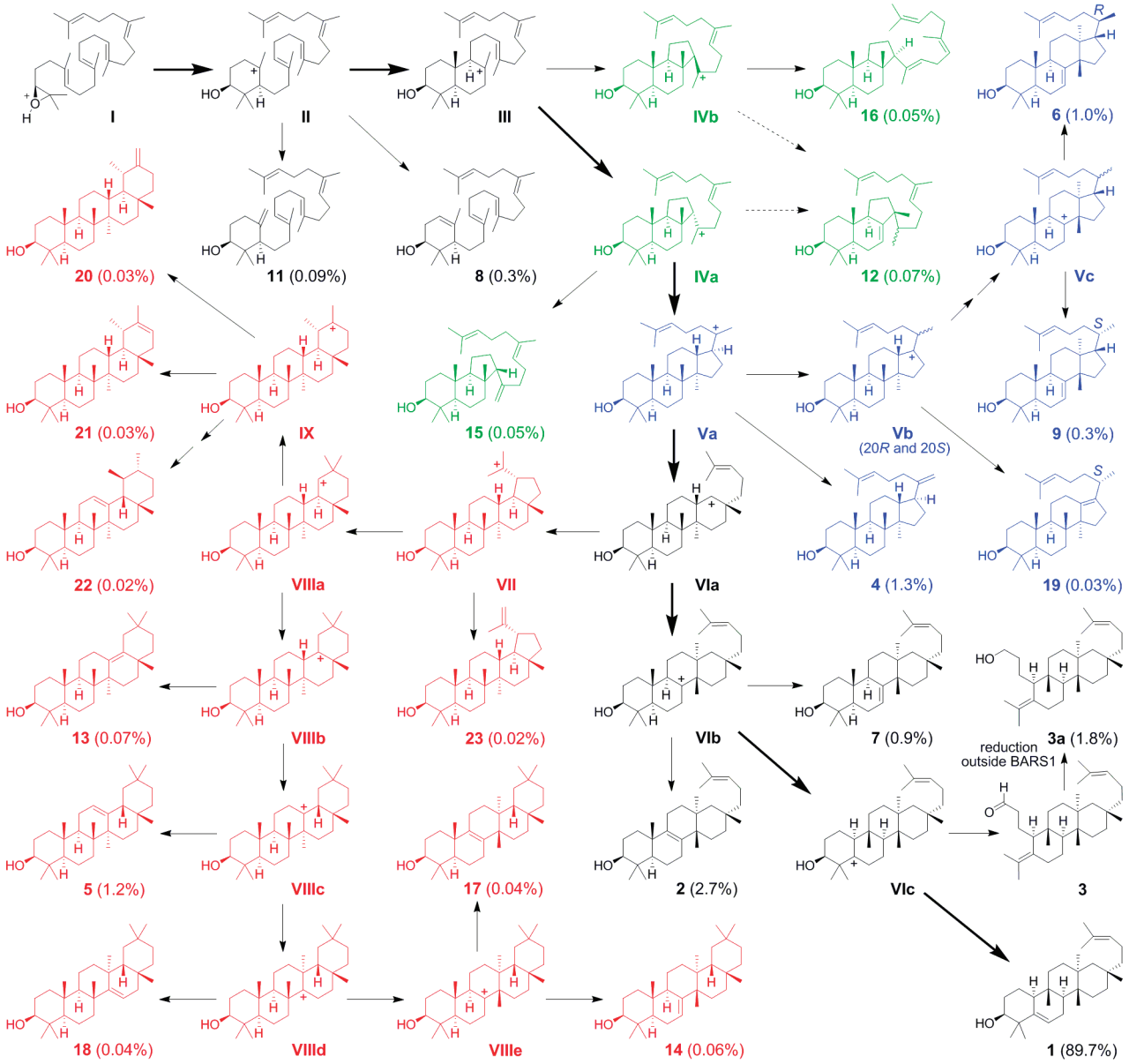
Multifunctional cyclases may have undergone selective pressure that favors multiple products for a shotgun approach to defense or for optimizing the physical properties of waxes and other materials. If the ultimate products of multifunctional cyclases are segregated and excreted (as in mono- and sesquiterpene biosynthesis), the byproducts would not disrupt cellular functions.

The present work provides the most detailed product profile available for any oxidosqualene cyclase and illustrates the value of thorough analyses. A community standard on minor product analysis is essential for valid comparisons, and we suggest a detection limit of 0.5% or lower. Detailed product profiles of other triterpene synthases should provide additional insight into the cyclization mechanism, the role of individual residues in catalysis, the catalytic equivalence of orthologs,[45] and the rationale for byproduct formation.

2.6. Genome Mining. Conventional isolation approaches suggest that *Arabidopsis* is not a prolific producer of triterpe-

(39) van Tamelen, E. E. *Pure Appl. Chem.* **1981**, *53*, 1259−1270.
(40) Pale-Grosdemange, C.; Feil, C.; Rohmer, M.; Poralla, K. *Angew. Chem., Int. Ed.* **1998**, *37*, 2237−2240.
(41) Herein accuracy refers to product accuracy with respect to (3*S*)-oxidosqualene. Although "accurate" has the unwanted connotation of perfection and "multifunctional" suggests moonlighting functions,[41a] as exemplified by the emopamil binding protein,[41b] we continue to use these established terms. (a) Copley, S. D. *Curr. Opin. Chem. Biol.* **2003**, *7*, 265−272. (b) Wassif, C. A.; Brownson, K. E.; Sterner, A. L.; Forlino, A.; Zerfas, P. M.; Wilson, W. K.; Starost, M. F.; Porter, A. F. *Hum. Mol. Genet.* **2007**, *16*, 1176−1187.
(42) Steele, C. L.; Crock, J.; Bohlmann, J.; Croteau, R. *J. Biol. Chem.* **1998**, *273*, 2078−2089.
(43) Review: Porter, F. D. *Curr. Opin. Pediatr.* **2003**, *15*, 607−613.
(44) Ruan, B.; Wilson, W. K.; Pang, J.; Gerst, N.; Pinkerton, F. D.; Tsai, J.; Kelley, R. I.; Whitby, F. G.; Milewicz, D. M.; Garbern, J.; Schroepfer, G. J., Jr. *J. Lipid Res.* **2001**, *42*, 799−812.

Scheme 2. Proposed Mechanisms for the Conversion of Oxidosqualene to Cyclic Triterpenes by BARS1[a]



[a] Heavy arrows highlight the dominant route to baruol. Pentacycles, 6/6/6/5 tetracycles, and tricycles are shown in red, blue, and green, respectively.

noids; only four triterpenes and their metabolites have been found in *Arabidopsis* tissue.[12] However, genome mining is establishing a very different picture. BARS1 makes at least 23 products (Figure 1). In conjunction with the nine previously reported *Arabidopsis* oxidosqualene cyclases, this plant can form a prodigious array of 35 triterpenes (excluding **10**), i.e., a third of all known oxidosqualene cyclization products.[1c] Columbiol and tricycle **16** join thalianol,[5i] marneral,[5j] and seco-amyrins[5h] as novel structures obtained through mining the *Arabidopsis* genome. The set of known *Arabidopsis* oxidosqualene cyclase products is now more numerous and diverse than that reported for any organism, including plants in the Euphorbiaceae and Leguminosae families, which are renowned for triterpenoid production. Other plant genomes are revealing multiple oxidosqualene cyclases, suggesting that this level of discovery is attributable to the methodology rather than the specific plant. We anticipate that further genome mining will show that plants generally encode substantially more triterpenoid biosynthetic capacity than classical isolation approaches have revealed.

2.7. Mechanistic Pathways Used by BARS1. Baruol biosynthesis is readily understood by analogy to established mechanistic pathways of oxidosqualene cyclization.[1c,2a] As shown in Scheme 2, protonated oxidosqualene (**I**) undergoes a series of cation−olefin annulations and ring expansions, being transformed through monocyclic, bicyclic, tricyclic, and tetracyclic intermediates (**II**, **III**, **IVa**, and **Va**) to the baccharenyl cation (**VIa**). Rearrangement by 1,2-hydride and methyl shifts leads to the friedobaccharenyl cation **VIb** and then to **VIc**, which finally loses a proton from C6 to form the major product baruol. The intermediacy of the 13β-H tricyclic cation **IVa** and the

(45) Firn and Jones have noted the compromises inherent in categorizing enzymes by the structure of the dominant product when minor products are likely to be the catalytic distinction: Firn, R. D.; Jones, C. G. *Nat. Prod. Rep.* **2003**, *20*, 382−391.

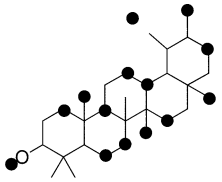


Figure 6. Deprotonation sites accessed by BARS1, shown as filled circles on a generic triterpene skeleton. Methyl deprotonation sites for tricyclic, tetracyclic, and lupanyl cations are approximated.

17α-H dammarenyl cation **Va** rather than their side-chain epimers is deduced on the basis of configurational transmission.[46]

Numerous minor products are shed by premature deprotonation en route to baruol. Rather than cyclize further through the horizontal[47] cation, 0.4% of the achilleyl cation (**II**) is deprotonated to monocyclic triterpenes **8** and **11**. An additional 0.2% of the intermediates are derailed to form tricycles **12**, **15**, and **16**, and 2.5% of the dammarenyl cation **Va** is similarly lost to side reactions that account for **4**, **6**, **9**, and **19**. A fraction (1.5%) of the baccharenyl cation **VIa** (or **Va**[46,47]) is further annulated to the lupanyl cation **VII**. Nearly all of **VII** is converted to the germanicyl cation **VIIIa**, which undergoes 1,2-shifts and deprotonation to rearranged oleananes **5**, **13**, **14**, **17**, and **18** (total 1.4%) or conversion via **IX** to (rearranged) taraxastanes **20**−**22** (total 0.1%). Conversion of **VIa** to baruol is accompanied by premature deprotonation to **2** and **7** (total 3.6%) and Grob fragmentation to **3** (1.8%).

2.8. The Profile of BARS1 Products Challenges Prevailing Concepts of Oxidosqualene Cyclization. The action of triterpene synthases is commonly thought to comprise the following processes: (a) folding the substrate, thereby predestining the ABC ring system, (b) protonating the substrate at C3 or O3, (c) guiding annulations by cation−π stabilization of intermediate carbocations, (d) similarly guiding 1,2-shifts by stabilizing the carbocation, (e) providing a hydrophobic active site that protects the cation from premature deprotonation, and (f) deprotonating the cation with a unique proton acceptor. However, cation−π stabilization, a truly hydrophobic active site,[48] and a unique proton acceptor are difficult to reconcile with the diversity of mechanistic pathways for BARS1. Although imperfect control of cyclization, rearrangement, and deprotonation can be rationalized as reflecting imprecise positioning of stabilizing residues and the proton acceptor, we describe several situations in which this line of argument is problematic.

2.8.1. Diversity of Deprotonation Sites. BARS1 mediates deprotonation at 14 different positions about rings A, B, C, D, and E (Figure 6). This diversity of deprotonation sites could not be accessed by a single proton acceptor in view of the limited mobility of the intermediate cations and active-site residues. We conclude that multiple proton acceptors are broadly dispersed around the active site cavity. We propose that deprotonation specificity in BARS1 is achieved not by minimizing the number of potential proton acceptors in the active site, but by constraining their access to the carbocation. Consistent with this hypothesis, many mutant cyclases generate products reflecting deprotonation at a variety of positions around the active site.[1d,g,8f]

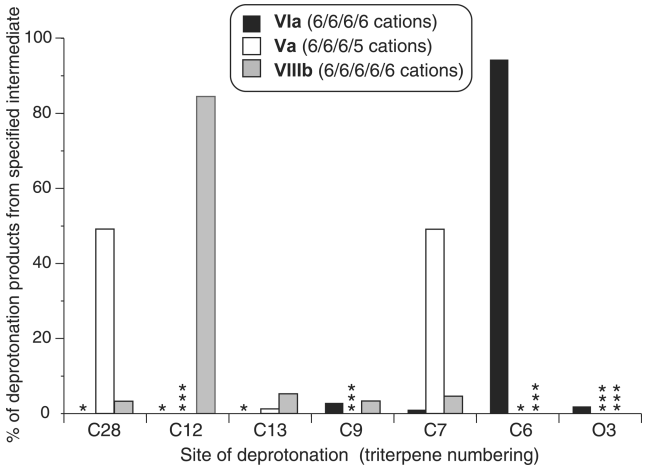


Figure 7. Deprotonation profiles for cations **Va**, **VIa**, and **VIIIb**. Asterisks (*) indicate the level of uncertainty for unobserved products. For **Va** and **VIIIb**, "C28" denotes any deprotonation beyond ring C.

Proton acceptors are typically polar residues or bound waters linked by hydrogen bonding to a basic residue or cytosolic water. The evident abundance and broad distribution of proton acceptors in BARS1 is incongruent with the concept that hydrophobicity of the active site protects the cationic intermediates from premature deprotonation. Indeed, crystal structures of other cyclases show many polar residues and bound waters lining the active site cavity.[6]

2.8.2. Differences in Deprotonation Profiles for Cations Va, **VIa**, **and VIIIb**. The dominant 6/6/6/6 product **1** is formed by deprotonation of rearranged cation **VIa** at C6, whereas deprotonation at other positions (C9, O3, and C7) is minimal (Figure 7). A conventional explanation for this product specificity would be that electrostatic stabilization from aromatic residues drives the rearrangement from the C18 cation **VIa** to the C5 cation **VIc**, from which an appropriately positioned base abstracts the C6 proton. If this were the case, one would expect 1,2 shifts of the 6/6/6/5 dammarenyl cation **Va** to produce a similar outcome. However, the 6/6/6/5 Δ5 product euferol was not detectable (<0.04% of **1**), and deprotonation of the 6/6/6/5 cations occurred mainly at C7 and C28 (Figure 7). Similarly, rearrangement of the 6/6/6/6/6 cation **VIIIb** led mainly to the Δ12 triterpene **5** rather than the Δ5 product.

The markedly different reactivity of analogous cations in the same active site is incompatible with the concept that electrostatic stabilization of cationic intermediates governs the extent of 1,2-shifts. If cation−π stabilization guides rearrangement of the 6/6/6/6 intermediates to the C5 cation, then the same stabilization should strongly favor the C5 cation among 6/6/6/5 and 6/6/6/6/6 intermediates. Some deviation in the product profiles might be expected owing to small geometric differences in the positioning of the substrate, but the magnitude of electrostatic interactions should be similar.[49]

(46) Xiong, Q.; Rocco, F.; Wilson, W. K.; Xu, R.; Ceruti, M.; Matsuda, S. P. T. *J. Org. Chem.* **2005**, *70*, 5362−5375.
(47) Matsuda, S. P. T.; Wilson, W. K.; Xiong, Q. *Org. Biomol. Chem.* **2006**, *4*, 530−543.
(48) Cyclase active sites exclude cytosolic water and are clearly hydrophobic from the viewpoint of protein folding. However, hydrophobic active sites of enzymes commonly contain a number of polar residues and bound waters linked by hydrogen bond networks. About 30−40% of the putative BARS1 active site residues (derived from LSS sequence alignments) have neutral polar side chains, many of which could quench a carbocationic intermediate. This environment is not hydrophobic from the viewpoint of an organic chemist.
(49) The magnitude of electrostatic stabilization is rather insensitive to the orientation and distance of the cation from the aromatic residue: (a) Matsuda, S. P. T.; Wilson, W. K.; Xiong, Q. Unpublished results.

The reason for the different deprotonation profiles for 6/6/6/6, 6/6/6/5, and 6/6/6/6/6 cations is not obvious but may be related to steric factors. All intermediates are tethered by the hydrogen bond between the 3β-hydroxyl and the protonating aspartate (D490), which is common to all oxidosqualene cyclases. The 6/6/6/6 cations are additionally constrained by their sterically demanding C17 substituents, i.e., an equatorial β-methyl and an axially attached side chain. This postulated rigidity of the 6/6/6/6 systems could allow the enzyme to protect these intermediates better from premature deprotonation than the presumably more mobile 6/6/6/5 cations.

Electrostatic effects do exist between aromatic residues and cationic intermediates[50] but appear to have a limited role in determining product specificity. This proposal is consistent with independent evidence from quantum mechanical modeling[49a] sesquiterpene mutagenesis studies,[51] and the nonenzymatic rearrangement of the 24,25-dihydroprotosteryl cation to form dihydroparkeol selectively without any specific cation stabilization.[52]

2.8.3. Products of Alternative Cyclization. Baruol synthase generates six classes of ring system from monocycles to 6/6/6/6/6 pentacycles (Table 1). If electrostatic stabilization is necessary to promote each annulation, ring expansion, and 1,2-shift, there should be a series of 10 stabilizing centers for cations linking **II** to **VIc** and perhaps less potent centers for minor cations **(IVb**, **Vb**, **VII**, etc.) and their premature rearrangement pathways. Although fringes of the major electrostatic fields might stabilize some minor intermediates, there is not enough space around the active site to accommodate so many stabilizing residues.

It is particularly difficult to rationalize the existence of overcyclized compounds like pentacycles if electrostatic stabilization of cations is necessary. BARS1 is a member of the PEN clade, which is notable for its dearth of cyclases that make pentacycles, yet BARS1 makes many minor pentacyclic products (**5**, **13**, **14**, **17**, **18**, and **20**−**23**). Phylogenetic analysis[4] and sequence comparisons[53] indicate little support for a model in which BARS1 inherited residual pentacyclization ability from an ancestor. BARS1 strongly favors baruol formation and has undergone substantial evolutionary optimization against other products. Specific residues necessary for pentacyclization should disappear quickly since they seem to provide no selective advantage. Because the presence of specific cation-stabilizing motifs to promote pentacyclization is doubtful, E-ring formation appears to be adventitious. The annulation is exothermic[47] and would proceed spontaneously in an enclosure that does not confine the substrate side chain to an extended conformation. We have reached similar conclusions from studies of epoxydammarane formation.[54]

2.8.4. Summary. It is not only implausible to postulate a specific motif to promote each of the annulations, ring enlargements, and rearrangements of triterpene biosynthesis; it is also unnecessary because these processes are energetically facile.[47,49a] Cyclases achieve accuracy more by deterring alternative possibilities than by lowering activation energies for the desired reaction.

These proposals are consistent with recent work in sesquiterpene biosynthesis, in which catalytic specificity was changed stepwise from one product to another, largely by mutagenesis of nonaromatic second-tier residues.[51] We[7b,f] and others[55] have also observed major changes in catalytic specificity through mutagenesis of second-tier residues. Such outer residues, which are too remote to provide significant electrostatic stabilization,[49a,50] are thought to alter catalytic specificity by creating subtle changes in the shape and dynamics of the active site.[51,55] These ideas are compatible with more general insights relating evolution, catalytic promiscuity, and enzyme engineering.[56]

3. Concluding Remarks

The study of oxidosqualene cyclization began in 1966 with a startling discovery: rat liver homogenate converts oxidosqualene to lanosterol.[57] As the first identified and most widely studied cyclase,[58] LSS came to be regarded as the prototypical model for oxidosqualene cyclization. Many of the seminal concepts[2] of triterpene biosynthesis were elaborated to explain the extreme accuracy of LSS relative to nonenzymatic reactions and were later used to rationalize the accuracy of other cyclases. Underlying this thinking has been the tacit assumption of a dichotomy between accurate and multifunctional cyclases. However, cyclase accuracy is a continuum, with BARS1 located about midway between multifunctional and highly accurate enzymes.

BARS1 makes one predominant product but also generates 22 minor products for no obvious biological advantage. These byproducts arise from aborted cyclization, overcyclization, Grob fragmentation, rearrangements, and aberrant deprotonation. BARS1 differs fundamentally from multifunctional cyclases, which make much larger amounts of byproducts but with less demonstrated structural variety. These observations suggested that cyclases, like organic chemists, have trouble controlling cationic rearrangements and that mechanistic diversity is the norm for (oxido)squalene cyclization.

The inherent inaccuracy of enzyme catalysis[59] has been recognized for DNA replication,[60] protein synthesis,[61] and other systems[62] but has rarely[40] been noted in triterpene synthesis. One reason for this neglect has been the difficulty of quantifying trace levels of numerous isomeric products. Advances in NMR

(50) Jenson, C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1997**, *119*, 10846−10854.
(51) Greenhagen, B. T.; O'Maille, P. E.; Noel, J. P.; Chappell, J. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9826−9833.
(52) (a) Corey, E. J.; Virgil, S. C. *J. Am. Chem. Soc.* **1990**, *112*, 429−431. For similar nonenzymatic rearrangements to dihydroparkeol, see: (b) van Tamelen, E. E.; Anderson, R. J. *J. Am. Chem. Soc.* **1972**, *94*, 8225−8228 and references therein.
(53) PEN6, the only characterized PEN cyclase that makes pentacycles, differs from BARS1 in seven putative active-site residues, four of which appear to be in the side-chain/D-ring region.
(54) Shan, H.; Segura, M. J. R.; Wilson, W. K.; Lodeiro, S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2005**, *127*, 18008−18009.
(55) Hyatt, D. C.; Croteau, R. *Arch. Biochem. Biophys.* **2005**, *439*, 222−233.
(56) (a) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. *Curr. Opin. Chem. Biol.* **2006**, *10*, 498−508. (b) O' Loughlin, T. L.; Patrick, W. M.; Matsumura, I. *Protein Eng. Des. Sel.* **2006**, *19*, 439−442. (c) Glasner, M. E.; Gerlt, J. A.; Babbitt, P. C. *Curr. Opin. Chem. Biol.* **2006**, *10*, 492−497.
(57) (a) Corey, E. J.; Russey, W. E.; Ortiz de Montellano, P. R. *J. Am. Chem. Soc.* **1966**, *88*, 4750−4751. (b) van Tamelen, E. E.; Willett, J. D.; Clayton, R. B.; Lord, K. E. *J. Am. Chem. Soc.* **1966**, *88*, 4752−4754.
(58) Being the sole cyclase in most animals and fungi, LSS could be studied by incubating crude preparations with oxidosqualene or an analogue. In contrast, plants contain multiple cyclases, whose individual study requires protein purification or cloning/heterologous expression.
(59) (a) Seminal review: Jensen, R. A. *Annu. Rev. Microbiol.* **1976**, *30*, 409−425. (b) We use "catalysis" in the broad sense of facilitating a reaction rather than the strict sense of lowering an activation energy barrier.
(60) Joyce, C. M.; Benkovic, S. J. *Biochemistry* **2004**, *43*, 14317−14324.
(61) Cochella, L.; Green, R. *Curr. Biol.* **2005**, *15*, R536−R540.
(62) Kazlauskas, R. J. *Curr. Opin. Chem. Biol.* **2005**, *9*, 195−201.

technology[63] now permit byproducts to be detected at trace levels and identified without extant authentic standards.[64] Our reanalysis[5m] of a reportedly[5k] accurate cyclase revealed many minor products, and we propose that the catalytic fidelity of cyclases is generally limited.

The product multiplicity of BARS1 derives largely from its ability to deprotonate cationic intermediates at numerous positions throughout the ring system. This ability mandates a broad distribution of polar residues and bound waters around the active site, as is observed in crystal structures of (oxido)-squalene cyclases.[6] This conclusion is inconsistent with the prevailing concept that the cationic cascade in triterpene biosynthesis is quenched by a single proton acceptor in a hydrophobic[48] active site. A major role for cation$-\pi$ stabilization of carbocation intermediates is also incompatible with the BARS1 product profile. Instead of stabilizing these highly reactive intermediates, a cyclase simply needs to protect them from unwanted reactions.[65] How cyclases perform this negative catalysis[66] may be a fruitful line of investigation.

4. Experimental Section

Instrumentation. NMR spectra were measured in dilute solution (<10 mM triterpenes) at 25 °C on a Bruker Avance DRX 500 or Varian Inova 600 or 800 NMR spectrometer and referenced to tetramethylsilane at 0 ppm (^{1}H) or $CDCl_3$ at 77.0 ppm (^{13}C). The Inova instruments were equipped with cold probes. GC-MS was performed on an Agilent 6890/5973 instrument using split injection (40:1), helium carrier gas (1 mL/min), and a Restek Rtx35MS column (30 m × 0.25 mm i.d., 0.1 μm film thickness) held isothermally at 260 °C. Mass spectra (*m*/*z* 50 to 650) were obtained at 70 eV. Separation of minor enzymatic products was carried out on an Agilent 1100 HPLC system using Phenomenex 5-μm ODS3 columns (250 × 4.6 mm i.d. or 250 × 21.2 mm i.d.), a mobile phase of methanol/water, and UV detection at 210 nm.

4.1. Cloning of Baruol Synthase. The whole length insert of the open reading frame *PEN2/At4g15370* (2292 bp) was RT-PCR amplified from RNA of 2-day old *Arabidopsis thaliana* seedlings using the two-step strategy described previously.[5j] The full length cDNA was recombined into the yeast expression vector pRS426GAL.[5d,14] The resulting plasmid, named pXQ13.2, was used to transform RXY6[5i] and SMY8[15] using the lithium acetate method[67] to generate yeast strains RXY6[pXQ13.2] and SMY8[pXQ13.2].

4.2. Large-Scale BARS1 Culture. SMY8[pXQ13.2] was cultured in synthetic complete medium (20 × 1 L) lacking uracil with 2% galactose as carbon source, supplemented with ergosterol (20 μg/mL), hemin chloride (13 μg/mL), and Tween 80 (5 μg/mL). The cultures were grown to saturation at 30 °C with shaking at 250 rpm. After centrifugation, the 195-g cell pellet was saponified with 10% KOH in 80% ethanol for 2 h at 70 °C to give 0.75 g of nonsaponifiable lipids (NSL). The NSL extract was loaded onto a silica gel column (130 g, 230−400 mesh) and eluted with CH_2Cl_2/hexanes (2:1). Five fractions were collected between oxidosqualene and ergosterol and subjected to preparative HPLC (elution with methanol/water (93:7)), affording sufficient material for 2D NMR structure determinations of many BARS1 products and artifacts.

4.3. In Vitro Reaction. The cell pellet (20 g) from a 2-L culture of RXY6[pXQ13.2] in the above medium was suspended in 0.1 M sodium phosphate buffer (20 mL, pH 6.2), and the cells were lysed in an Emulsiflex-C5 homogenizer. A solution of racemic oxidosqualene[18] and Triton X-100 was added to the homogenate to give a final concentration of 0.25 mg/mL substrate and 0.25% Triton X-100. An aliquot of the cell homogenate with no substrate added served as the control reaction. After 24-h incubation at room temperature, the reaction was quenched with two volumes of ethanol, and the cell debris was removed by centrifugation. The ethanol was evaporated, and the remaining aqueous phase was extracted with methyl *tert*-butyl ether (MTBE). The combined MTBE layers were washed with brine and concentrated to a crude extract, aliquots of which were analyzed by NMR and GC-MS.

4.4. Chromatographic Separation of Enzymatic Products from in Vitro Reaction. An aliquot of the MTBE crude extract was loaded onto a short silica gel column (7 g, 230−400 mesh) and separated with gradients of ethyl ether/hexanes. Four fractions were collected and analyzed by NMR and GC-MS. After fractions containing squalene and oxidosqualene, elution with ethyl ether/hexanes (1:9) gave the triterpene alcohols (Fraction A). An aliquot of fraction A was separated by PTLC on a 20 × 20 cm silica gel plate (250 μm layer) using CH_2Cl_2 as eluent. The plate was divided into 11 bands which were scraped onto small columns, eluted with CH_2Cl_2, and analyzed by NMR and GC-MS. Fraction B1 contained mainly baruol, and fraction B2 contained most of the minor triterpene alcohols. Fractions B1 and B2 were purified further by HPLC (elution with a linear gradient of methanol/water (9:1) to 100% methanol). HPLC fractions were analyzed by NMR and GC-MS.

4.5. Determination of the Ratios of BARS1 Products. A 10-mg aliquot of fraction A from the short silica-gel column of RXY6-[pXQ13.2] was further purified by PTLC as described above to remove extraneous lipids. The plate was divided into three bands, which were analyzed by NMR and GC-MS. The middle band (above ergosterol) contained baruol and minor triterpenes in a single fraction (B_3). The product profile was determined by the signal intensity of resolved methyl peaks in the 800 MHz NMR spectrum, in conjunction with other spectral information. All percentages and ratios were calculated at high precision prior to rounding for presentation.

4.6. Nonenzymatic Oxidosqualene Cyclization. A control reaction was performed in which an oxidosqualene/Triton X-100 solution was incubated with sodium phosphate buffer (in absence of cell homogenate) under the in vitro reaction conditions described above. The MTBE crude extract was analyzed by NMR and GC-MS. An aliquot of the MTBE crude extract was separated on a short silica gel column with gradients of ethyl ether/hexanes, yielding three fractions that were analyzed by NMR and GC-MS. A second aliquot of the crude extract was separated by PTLC (CH_2Cl_2) into five bands, which were analyzed by NMR and GC-MS.

Acknowledgment. We are grateful to Ella Bartel Matsuda for preparing the dandelion root extract. This work was supported in part by the National Science Foundation (MCB-0209769), The Robert A. Welch Foundation (C-1323), and the Herman Frasch Foundation. The 800 MHz NMR spectra were acquired on instrumentation purchased by the John S. Dunn, Sr. Gulf Coast Consortium for Magnetic Resonance. C.S.O. is grateful to the Arnold and Mabel Beckman Foundation for a Beckman Scholarship.

Supporting Information Available: Details of experimental procedures, structure elucidation, NMR and GC-MS spectra, and NMR assignments. This material is available free of charge via the Internet at http://pubs.acs.org.

JA073133U

(63) Our enhanced NMR sensitivity derived from a cold probe. Alternatively, capillary NMR offers comparable sensitivity: (a) Olson, D. L.; Norcross, J. A.; O'Neil-Johnson, M.; Molitor, P. F.; Detlefsen, D. J.; Wilson, A. G.; Peck, T. L. *Anal. Chem.* **2004**, *76*, 2966−2974. (b) Schroeder, F. C.; Gronquist, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 7122−7131.
(64) 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.
(65) (a) This idea was expressed almost 50 years ago by Robert Robinson: Wolstenholme, G. E. W.; O'Connor, M., Eds. *Ciba Foundation symposium on the biosynthesis of terpenes and sterols*; Little, Brown, and Co.: Boston, 1959; pp 299−300. (b) See also: Christianson, D. W. *Chem. Rev.* **2006**, *106*, 3412−3443.
(66) Rétey, J. *Angew. Chem., Int. Ed.* **1990**, *29*, 355−361.
(67) Schiestl, R. H.; Gietz, R. D. *Curr. Genet.* **1989**, *16*, 339−346.