ORGANIC LETTERS

2006
Vol. 8, No. 13
2835−2838

A New Triterpene Synthase from *Arabidopsis thaliana* Produces a Tricyclic Triterpene with Two Hydroxyl Groups

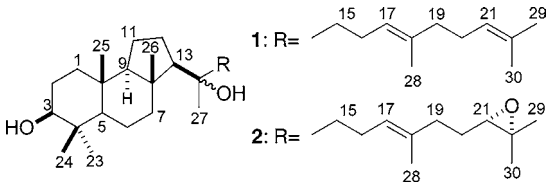
Ting Xiang,[†] Masaaki Shibuya,[†] Yuji Katsube,[‡] Takehiko Tsutsumi,[†] Miyuki Otsuka,[†] Hong Zhang,[†] Kazuo Masuda,[§] and Yutaka Ebizuka*,[†]

Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, Product Analysis Division, Research Laboratories, Kanebo Seiyaku, Ltd., 3-1, Kanebo-machi, Takaoka 933-0856, Japan, and Showa Pharmaceutical University, 3-3165 Higashi-Tamagawagakuen, Machida, Tokyo 194-8543, Japan

yebiz@mol.f.u-tokyo.ac.jp

Received April 24, 2006

ABSTRACT



Thirteen oxidosqualene cyclase homologues exist in the genome of *Arabidopsis thaliana*. One of these genes, At4g15340, was amplified by PCR and expressed in yeast. The yeast transformant accumulated tricyclic triterpene, (3*S*,13*R*)-malabarica-17,21-dien-3,14-diol (arabidiol), whose structure was determined by NMR and MS analyses. Its epoxide analogue, (3*S*,13*R*,21*S*)-malabarica-17-en-20,21-epoxy-3,14-diol (arabidiol 20,21-epoxide), was also isolated from the transformed yeast. This is the first example of a triterpene synthase that yields a tricyclic triterpene with two hydroxyl groups.

Triterpenes are a large and medicinally important group of natural products, many of which exhibit a variety of pharmacological activities including antiinflammatory, antimicrobial, antitumor, antiviral, cardiovascular system protective activity, etc.[1] More than 100 different types of triterpene skeletons from natural origin are known, which are derived from a common precursor, 2,3-oxidosqualene, via reaction catalyzed by oxidosqualene cyclase (OSC) (Figure 1).[2] In the previous studies, more than 20 OSC cDNAs have been cloned from various plant species, which included cycloartenol synthase,[3] lupeol synthase,[4] β-amyrin synthase,[3e,5] mixed-amyrin synthase,[5b] isomultiflorenol synthase,[6] cucurbitadienol synthase,[7] and multifunctional triterpene synthase.[3b,5d,8]

The completion of the genome sequencing project revealed the presence of 13 OSC homologues in *Arabidopsis thaliana*. Among them, At1g66960,[8a] At1g78500,[8a] At1g78960,[8b,c] and

[†] The University of Tokyo.
[‡] Kanebo Seiyaku Ltd.
[§] Showa Pharmaceutical University.

(1) (a) Mahato, S. B.; Sarkar, S. K.; Poddar, G. *Phytochemistry* **1988**, *27*, 3037−3067. (b) Mahato, S. B.; Nandy, A. K.; Roy, G. *Phytochemistry* **1992**, *31*, 2199−2249.

(2) (a) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261−291. (b) Abe, I.; Rohmer, M.; Prestwich, G. D. *Chem. Rev.* **1993**, *93*, 2189−2206.

(3) (a) Corey, E. J.; Matsuda, S. P. T.; Bartel, B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11628−11632. (b) Kawano, N.; Ichinose, K.; Ebizuka, Y. *Biol. Pharm. Bull.* **2002**, *25*, 477−482 (c) Hayashi, H.; Hiraoka, N.; Ikeshiro, Y.; Kushiro, T.; Morita, M.; Shibuya, M.; Ebizuka, Y. *Biol. Pharm. Bull.* **2000**, *23*, 231−234. (d) Morita, M.; Shibuya, M.; Lee, M. S.; Sankawa, U.; Ebizuka, Y. *Biol. Pharm. Bull.* **1997**, *20*, 770−775. (e) Kushiro, T.; Shibuya, M.; Ebizuka, Y.; *Eur. J. Biochem.* **1998**, *256*, 238−244.

(4) (a) Shibuya, M.; Zhang, H.; Endo, A.; Shishikura, K.; Kushiro, T.; Ebizuka, Y. *Eur. J. Biochem.* **1999**, *266*, 302−307. (b) Herrera, J. B. R.; Bartel, B.; Wilson, W. K.; Matsuda, S. P. T. *Phytochemistry* **1998**, *49*, 1905−1911. (c) Segura, M. J. R.; Meyer, M. M.; Matsuda, S. P. T. *Org. Lett.* **2000**, *15*, 2257−2259. (d) Hayashi, H.; Huang, P.; Takada, S.; Obinata, M.; Inoue, K.; Shibuya, M.; Ebizuka, Y. *Biol. Pharm. Bull.* **2004**, *27*, 1086−1092.

10.1021/ol060973p CCC: $33.50
Published on Web 05/25/2006

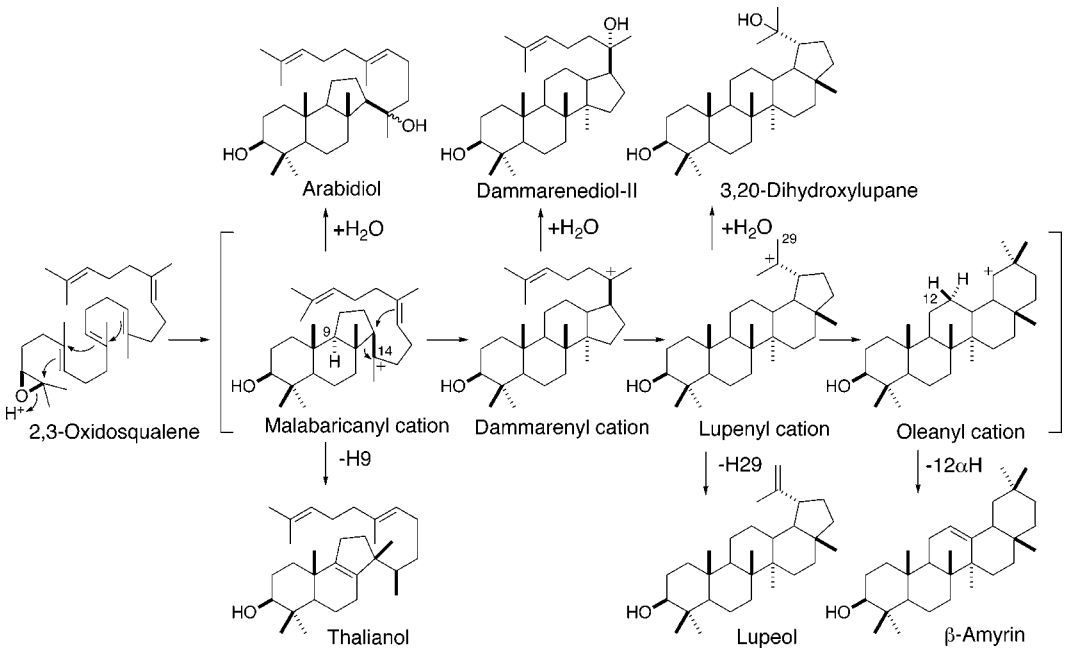


Figure 1. Proposed reaction mechanism of OSCs.

At1g78970 (LUP1)[4b,c] have been identified to encode multifunctional triterpene synthases with leaky product specificity yielding more than one product, while At2g07050,[3a] At5g48010,[9] and At5g42600[10] products are monofunctional, specifically yielding cycloartenol, thalianol, and marneral, respectively. Recently, At3g45130 product was identified as lanosterol synthase by Matsuda's and our groups.[11] Thalianol and marneral are new compounds which have not been isolated before from any plants including this plant. This is a good example of an alternative method to discover new natural products which may complement conventional phytochemical methods. Discovery of lanosterol synthase in higher plants, where cycloartenol synthase has been believed to be a unique enzyme that yields a sterol skeleton from oxidosqualene, provided a new aspect of phytosterol biosynthesis, in that two pathways might exist from oxidosqualene to phytosterols. Here, we report the identification of At4g15340 as a new triterpene synthase producing arabidiol, which has a 6,6,5-tricyclic ring system with two hydroxyl groups.

A 2.3 kb full length cDNA corresponding to At4g15340 was amplified by RT-PCR with specific primers[12] designed from the database sequence and RNA from one-week-old seedlings of *A. thaliana* as a template. The obtained cDNA was subcloned into an expression vector pYES2 and expressed in *Saccharomyces cerevisiae* GIL77, a lanosterol synthase deficient strain.[3e] The transformant cells were extracted with hexane. TLC analysis showed two cyclization products, which were further separated by silica gel column chromatography and preparative HPLC. Structure of the products was analyzed by NMR (^{1}H and ^{13}C NMR, DEPT, DQF-COSY, HMQC, HMBC, NOESY, and differential NOE) and high-resolution EIMS (HREIMS).

HREIMS of the major product **1** gave a molecular ion peak at *m*/*z* 444.4015, indicating its molecular formula to be $C_{30}H_{52}O_2$ with five degrees of unsaturation. NMR analysis indicated the presence of two double bonds and two hydroxyl groups. Detailed analyses of DQF-COSY, HMQC, and HMBC gave a possible planar structure of **1** as malabarica-

(5) (a) Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Towards Natural Medicine Research in the 21st Century*; Excerpta Medica International Congress Series; Ageta, H., Aimi, N., Ebizuka, Y., Fujita, T., Honda, G., Eds; Elsevier Science: Amsterdam, The Netherlands, 1998; pp 421−427. (b) Morita, M.; Shibuya, M.; Kushiro, T.; Masuda, K.; Ebizuka, Y. *Eur. J. Biochem.* **2000**, *267*, 3453−3460. (c) Hayahsi, H.; Huang, P.; Kirakosyan, A.; Inoue, K.; Hiraoka, N.; Ikeshiro, Y.; Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Biol. Pharm. Bull.* **2001**, *24*, 912−916. (d) Iturbe-Ormaetxe, I.; Haralampidis, K.; Papadopoulou, K.; Osbourn, A. E. *Plant Mol. Biol.* **2003**, *51*, 731−743. (e) Suzuki, H.; Achnine, L.; Xu, R.; Matsuda, S. P. T.; Dixon, R. A. *Plant J.* **2002**, *32*, 1033−2048

(6) Hayashi, H.; Huang, P.; Inoue, K.; Hiraoka, N.; Ikeshiro, Y.; Yazaki, K.; Tanaka, S.; Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Eur. J. Biochem.* **2001**, *268*, 6311−6317.

(7) Shibuya, M.; Adachi, S.; Ebizuka, Y. *Tetrahedron* **2004**, *60*, 6995−7003.

(8) (a) Ebizuka, Y.; Katsube, Y.; Tsutsumi, T.; Kushiro, T.; Shibuya, M. *Pure Appl. Chem.* **2003**, *75*, 369−374. (b) Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. *Tetrahedron Lett.* **2000**, *41*, 7705−7710. (c) Husselstein-Muller, T.; Schaller, H.; Benveniste, P. *Plant Mol. Biol.* **2001**, *45*, 75−92.

(9) Fazio, G. C.; Xu, R.; Mstsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678−5679.

(10) Xiong, Q.; Wilson, W. K.; Matsuda. S. P. T. *Angew. Chem., Int. Ed.* **2006**, *45*, 1−5.

(11) (a) 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. (b) Kolesnikova, M. D.; Xiong, Q.; Lodeiro, S.; Hua, L.; Matsuda, S. P. T. *Arch. Biochem. Biophys.* **2006**, *447*, 87−95.

(12) First PCR primer: sense primer: 5′-AACTTGGAGGATTAATACAAA-3′, antisense primer: 5′-AGAATTGACCGTGCATTTTTC-3′, Second PCR primer: sense primer: 5′-TTAA**GGTACC**ATGTGGAGACTAAGAATTGGA-3′, antisense primer: 5′-CGTGCA**GAATTC** TCAAGGCTGAAGCCGCCGTAG-3′. Restriction sites are shown in bold face, start and stop codons are undelined.

17,21-dien-3,14-diol. Recently, Hoshino et al. reported two diastereomers of malabarica-17,21-dien-3,14-diol, in a product mixture produced from oxidosqualene (nonphysiological substrate) by a bacterial squalene cyclase (SHC) mutant.[13] The mixture consisted of achilleol A (35%), (3*S*,13*S*)-malabarica-17,21-dien-3,14-diol (27%), (3*S*,13*R*)-malabarica-14(27),17,21-trien-3-ol (14%), thalianol (9%), the 9(11)-en-isomer of thalianol (9%), and (3*S*,13*R*)-malabarica-17,21-dien-3,14-diol (9%) (Figure 2B).[13]

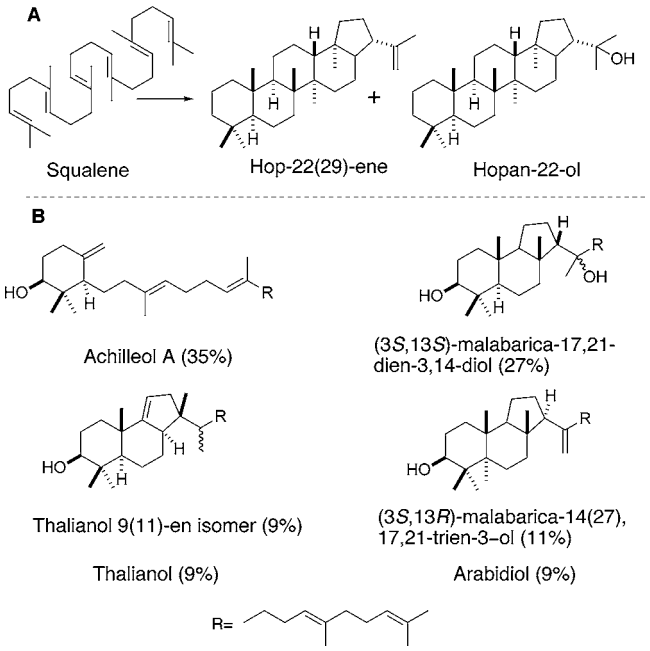


Figure 2. Multiple products of bacterial SHC (A) and its G600 deletion mutant (B).

Chemical shifts of C-9 and C-26 were distinct between (3*S*,13*S*)-malabarica-17,21-dien-3,14-diol and its 13*R* epimer. C-9 and C-26 of 13*S*-epimer resonated at δ 60.3 and 26.7, while those of the 13*R* epimer resonated at δ 63.2 and 15.9, respectively. Product **1** showed identical chemical shifts to those of the (3*S*,13*R*)-epimer including C-9 (δ63.2) and C-26 (δ 15.88 or δ15.92). Finally, NOESY analysis of **1** showed the correlations from H-9 to H-13, H-3, and H-5 and from H-3 to H-5 and 23-Me (Figure S2, Supporting Information), establishing the orientation of H-13 to be α. As the absolute configuration of C-3 is *S*, that of C-13 is determined as *R*. The configuration at C-14 was not determined in this study, although ^{13}C NMR showed the presence of only one diastereomer of **1**. From these results, **1** was determined as (*3S,13R*)-malabarica-17,21-dien-3,14-diol, which we named arabidiol (Figure 1).

HREIMS of minor product **2** gave a molecular ion peak at *m/z* 460.3898 indicating the molecular formula of $C_{30}H_{52}O_3$, which has one additional oxygen atom compared to arabidiol. The NMR data of **2** are very close to those of arabidiol except for the signals around terminal dimethyl group. **2** showed two oxygenated carbon signals at δ 64.2 (C-21) and 58.3 (C-22) in place of the terminal olefinic carbon signals at δ 124.8 (C-21) and 131.3 (C-22) observed in arabidiol, indicating the presence of epoxide between C-21 and C-22. The structure was finally determined as (3*S*,13*R*)-malabarica-17-en-20,21-epoxy-3,14-diol by the HMBC and NOESY analyses (Figure S2, Supporting Information), except for the configuration at C-14 and C-21.

Production of arabidiol epoxide can be explained by cyclization of dioxidosqualene accumulated in the host cells. Dioxidosqualene is synthesized in yeast from oxidosqualene by a squalene-epoxidase.[14] Lanosterol synthases from mammals[15] and yeast[16] cyclize dioxidosqualene to yield 24,25-oxidolanosterol, which is further metabolized to oxysterols.[17] Recently, it has been found that LUP1, which gives a pentacyclic product, can also cyclize dioxidosqualene to yield oxacyclic triterepenes.[18] These facts strongly suggest that OSCs can generally accept dioxidosqualene as the substrate to give the product with an oxirane, a pyran, or a furan ring at the other end. These oxacyclic triterpenes might be produced in the transformant cells in the previous OSC expression studies. If arabidiol epoxide is a direct cyclization product from dioxidosqualene by arabidiol synthase as in the case of thalianol epoxide,[9] the configuration at C-21 of **2** must be *S* since yeast has a unique squalene epoxidase that gives one diastereomer, (3*S*,22*S*)-dioxidosqualene. To the best of our knowledge, arabidiol epoxide has not been reported in the literature.

A fused ring system such as 6,6,6,6,6-membered ring is one of the characteristic features of plant triterpene skeletons, produced by a carbocation mediated reaction by OSC as shown in Figure 1. Opening of epoxide by protonation, generation of preferred tertiary carbocation, and electrophilic addition to a double bond yield the first 6-membered ring intermediate. The second electrophilic addition yields a fused 6,6-membered ring, and the third one generates a 6,6,5-membered ring with the preferred tertiary carbocation (Markovnikov addition). Ring expansion yields a 6,6,6-membered ring. In a similar manner, 6,6,6,5- and 6,6,6,6,6-membered rings are formed. The carbocation could be quenched by deprotonation or addition of a water molecule at any intermediary step, producing a greater diversity of triterpene structures. Triterpenes with fused 6,6,6,5-, 6,6,6,6,5- or 6,6,6,6,6-membered rings are widely distributed in higher plants, but those with other ring systems occur only in limited plant families or species. Arabidiol and thalianol have a fused 6,6,5-membered ring. Addition of a water molecule to the 6,6,5-membered-ring-carbocation intermediate yields arabidiol. Hydride and methyl group shifts along the backbone of this intermediate yield migrated marabaricane skeleton, and quenching of carbocation by deprotonation from C-9

(13) Hoshino, T.; Shimizu, K.; Sato. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 6700−6703.
(14) Nagumo, A.; Kamei, T.; Sakakibara, J.; Ono, T. *J. Lipid Res.* **1995**, *36*, 1489−1497.
(15) Boutaud, O.; Dolis, D.; Schuber, F. *Biochem. Biophys. Res. Commun.* **1992**, *188*, 898−904.
(16) Field, R. B.; Holmlund, C. E. *Arch. Biochem. Biophys.* **1977**, *180*, 465−471.
(17) Spencer, T. A. *Acc. Chem. Res.* **1994**, *27*, 83−90.
(18) Shan, H.; Segura, M. J. R.; Wilson, W. K.; Lodeiro, S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2005**, *127*, 18008−18009.

yields thalianol. These reactions are precisely controlled by each enzyme (At4g15340 or At5g48010), and yield a single product, although many of the other *A. thaliana* triterpene synthases produce multiple triterpenes. It is noteworthy that *A. thaliana* has an ability to produce the rare triterpenes with 6,6,5-ring-skeleton, namely arabidiol and thalianol.

OSCs generally terminate the reaction by deprotonation to form a double bond. Some OSCs including LUP1 and arabidiol synthases terminate it by water addition to carbocation, to yield a product with two hydroxyl groups. LUP1 is a multiple-product OSC that yields dihydroxylupane and lupeol in nearly equal amount (Figure 1), together with minor triterpenes including β-amyrin, etc.[4c] Feeding of 1,2-[^{13}C] acetate to LUP1 transformant showed mixed-labeling patterns at the propene moiety in lupeol as the result of intact incorporation of the acetate unit to a single bond between C-20 and C-30, and a double bond between C-20 and C29,[22] while giving a unique labeling pattern at the hydroxypropane moiety in dihydroxylupane.[8a] Multiple productivity of LUP1 and these labeling patterns suggest that attack of a water to lupenyl cation is restricted from only one face in the quenching reaction.[8a] The presence of deprotonation and water addition products was also shown in bacterial SHC, which yields hop-22(29)-ene and hopan-22-ol in a ratio of 5:1 (SHC from *Bacillus acidocaldarius*)[19a] or 1:1 (SHC from *Acetobacter pasteurianum*)[19b] (Figure 2A), although diastereomeric purity regarding the prochiral center at C-22 was not determined. In contrast to multiple productivity of LUP1 and SHC, it is noteworthy that arabidiol synthase yields a single product. Arabidiol synthase is the first OSC yielding a single product with a hydroxyl group originating from a water molecule. Recently, we have succeeded in cloning of dammarerenediol-II synthase from *Panax ginseng*, which yields tetracyclic triterepene-diol as the sole product (to be published elsewhere). Arabidiol synthase and dammarenediol-II synthase yielded exclusively one diastereomeric product. These suggest that the position of the water molecule that terminates cyclization in the active site pocket of OSCs is precisely controlled.

As mentioned above, arabidiol and thalianol have a fused 6,6,5-membered ring. Quenching of the 6,6,5-membered-ring-carbocation by addition of water yields arabidiol, while deprotonation after hydride and methyl group shifts yields thalianol. Some particular amino acid residues in the active site of arabidiol synthase may keep a water molecule that reacts to carbocation at C-14 immediately after 6,6,5 ring formation. X-ray crystal structure analysis of human lanosterol synthase (hLAS)[20] revealed that His232 of the ^{229}LWCHCR^{234} motif is positioned to stabilize the anti-Markovnikov secondary cation formed at C-13 during C-ring formation. Homology modeling analysis of hLAS based on the crystal structure of bacterial SHC showed that Lys331 of the ^{330}TKSIS^{334} motif is situated directly above the two cationic centers at C-10 and C-15 during protosteryl cation formation.[21] These ^{229}LWCHCR^{234} and ^{330}TKSIS^{334} motives of lanosterol synthase correspond to ^{259}LWIYLR^{264} and ^{366}NRYIT^{370} of arabidiol synthase. If ^{259}LWIYLR^{264} and ^{366}NRYIT^{370} also locate at a similar position in the active site of arabidiol synthase, they may play an important role in stabilizing or quenching carbocation formed at the C-ring. ^{259}LWIYLR^{264} and ^{366}NRYIT^{370} motives are identical between thalianol and arabidiol synthases except for Asn366. The amino acid residue corresponding to Asn366 is Thr in thalianol synthase, and Ser or Thr in all other known triterpene synthases (Figure 3). Asn at this position in arabidiol synthase is quite unique compared to other triterpene synthases. We therefore speculate that Asn366 may take part in water addition to a carbocation at C-14 of the 6,6,5-membered-ring-intermediate because Asn is more polar than Thr and Ser, and can contribute to efficiently keep a water molecule in the active site.

At4g15340 (arabidiol synthase)	259LWIYLR	366NRYIT
At5g48010 (thalianol synthase)[9]	258LWIYLR	365TRYIT
At5g42600 (marneral synthase)[10]	258LWIYIR	365THYIT
At1g78500 (multifunctional synthase)[8a]	259LWIYFR	366SRYFT
PNY (β-amyrin synthase)[3e]	258MWCYCR	364SRYIT
OEW (lupeol synthase)[4a]	255MLCYCR	361SRYLC
At2g07050 (cycloartenol synthase)[3a]	254MWCHCR	359TRYIC
hLAS(lanosterol synthase)[20]	229LWCHCR	330TKSIS

Figure 3. Alignments of partial amino acid sequences of OSCs.

Nine out of thirteen OSCs in *A. thaliana* including arabidiol synthase (this study) have been identified for their enzyme function, demonstrating that *A. thaliana* has the ability to produce more than 15 triterpenes, although triterpene constituents of this plant itself have not been identified except for a few triterpenes such as β-amyrin, α-amyrin, lupeol, etc. (our unpublished data). It is noteworthy that *A. thaliana*, even though a small model plant, has the ability to produce nearly one-sixth of the total natural triterpene skeletons so far reported, owing to the contribution of both single-product and multiple-product OSCs.

Acknowledgment. We thank Prof. T. Hoshino (Niigata University) for providing us NMR data of (3*S*,13*S*)-malabarica-17,21-dien-3,14-diol. A part of this research was financially supported by a Grant-in-Aid for postdoctoral fellow to T.X., and a Grant-in-Aid for Scientific Research (S) (No. 15101007) to Y.E. from Japan Society for the Promotion of Science, Japan.

Supporting Information Available: Details of cloning, expression, and isolation procedures, ^{1}H and ^{13}C NMR data, NOESY correlations, and spectra of ^{1}H and ^{13}C NMR. This material is available free of charge via the Internet at http://pubs.acs.org.

OL060973P

(19) (a) Seckler, B.; Poralla, K. *Biochim. Biophys. Acta* **1986**, *881*, 356−363. (b) Rohmer, M.; Anding, C.; Ourisson, G. *Eur. J. Biochem.* **1980**, *112*, 541−547.

(20) 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.

(21) Schulz-gasch, T.; Stahl, M. *J. Comput. Chem.* **2003**, *24*, 741−753.

(22) Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Tetrahedron Lett.* **1999**, *40*, 5553−5556.