Pure Appl. Chem., Vol. 75, Nos. 2–3, pp. 369–374, 2003.

Functional genomics approach to the study of triterpene biosynthesis*

Yutaka Ebizuka‡, Yuji Katsube, Takehiko Tsutsumi, Tetsuo Kushiro, and Masaaki Shibuya

Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract: The *Arabidopsis thaliana* genome-sequencing project has identified the presence of 13 oxidosqualene cyclase homologs in this plant. In addition to the already identified clones, namely, CAS1 cycloartenol synthase, LUP1 lupeol synthase, and YUP8H12R.43 multifunctional triterpene synthase, two new cDNAs of the putative oxidosqualene cyclase genes, F1019.4 and T30F21.16, were obtained by polymerase chain reaction (PCR) and functionally expressed in yeast. Liquid chromatography/mass spectrometry (LC/MS) analysis led to the identification of some of their reaction products. Interestingly, except for CAS1 for sterol biosynthesis of primary metabolism, so-far-obtained all triterpene synthases of this plant are multifunctional, producing more than one cyclization product. A feeding experiment of ^{13}C-labeled acetate with LUP1 lupeol synthase transformant demonstrated the stereospecific water addition to lupenyl cation intermediate, yielding 3β,20-dihydroxylupane, which accounts for the multiproduct nature of this synthase.

INTRODUCTION

Recent international efforts have resulted in the complete genome sequencing of the model plant, *Arabidopsis thaliana* [1], and comparable efforts are currently directed toward the genome sequencing of more important crop plants, including rice, corn, soybeans, etc. These sequence data, together with those from EST (expressed sequence tag), facilitate cloning of cDNAs for biosynthetic enzymes much more easier than in the pre-genome era. For years, we have been studying the biosynthesis of plant triterpenoids, which represent one of the most structurally diverse groups of natural products. Many of these molecules show important and interesting biological activities such as antifungal, anti-HIV, antitumor promotion and metastasis, antioxidant, and so on [2]. However, a low level of production and difficulties associated with their separation from complex mixtures have hampered the development of these natural triterpenes for commercial products in the market. The ultimate goal of our study is to produce potentially useful triterpene compounds by recombinant technology of biosynthetic enzymes.

Thousands of triterpene products have been reported from the plant kingdom, and they arise from more than 80 different carbon skeletons. All of these carbon frameworks are elaborated from a common substrate, (*3S*)-oxidosqualene, by oxidosqualene cyclases (OSCs). Our interests have been focused on these enzymes, as the cyclization of oxidosqualene into various skeletal types is the primary origin of the diverse triterpene structures (Fig. 1).

To date, over 30 OSC cDNAs have been cloned in this laboratory by classical homology-based polymerase chain reaction (PCR) methods [3]. Their enzyme functions have been identified by het-

Pure Appl. Chem. **75**, 141–419 (2003). An issue of reviews and research papers based on lectures presented at the 23rd IUPAC International Symposium on the Chemistry of Natural Products, Florence, Italy, 28 July–2 August 2002 on the theme of natural products.
‡Corresponding author

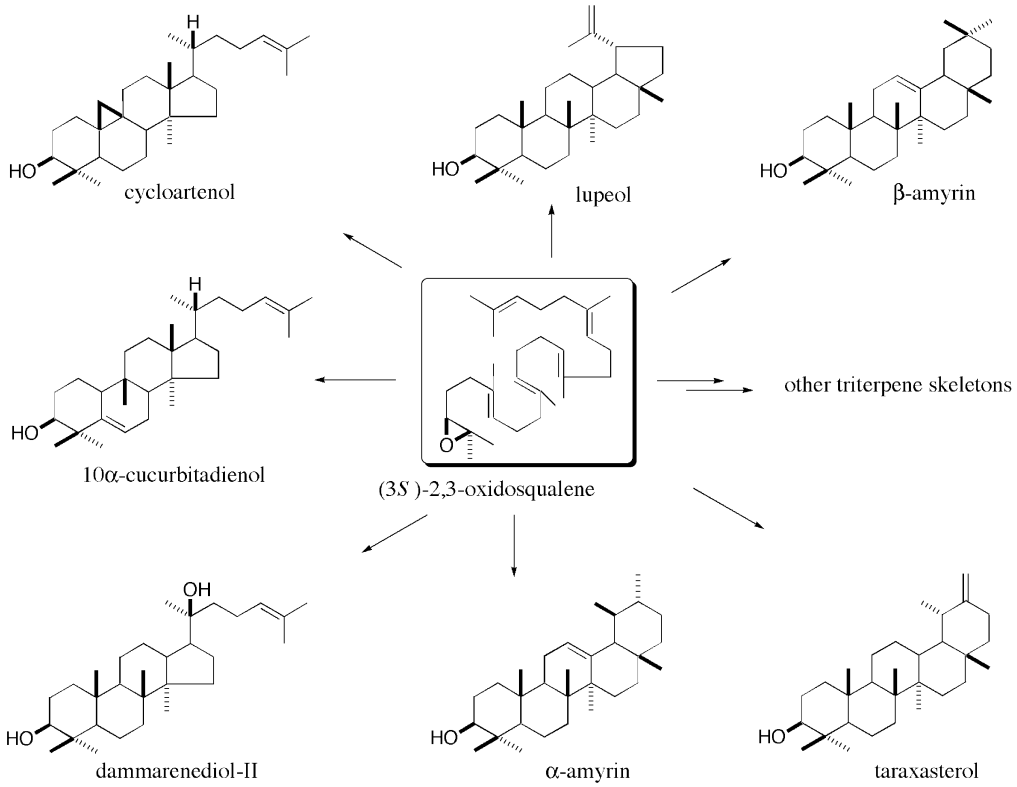

Fig. 1 Cyclization of oxidosqualene in higher plants.

erologous expression in yeast. Some enzymes show high fidelity to produce single cyclization products, suggesting the existence of more than 80 different OSCs. However, this estimate became more complicated by the existence of some multiproduct synthases. For example, one OSC clone (PSM) from *Pisum sativum* converted oxidosqualene into at least eight different cyclization products [4]. If such multiproduct enzymes occur commonly in the plant kingdom, the number of enzymes to generate diverse triterpene skeletons would be greatly reduced.

Arabidopsis genome has 13 distinct putative OSC genes, and so far only three of them have been identified for their enzyme functions [5–8]. In this paper, functional analysis of two additional *Arabidopsis* OSC clones and some mechanistic features of LUP1 lupeol synthase-catalyzed reaction are described.

cDNA CLONING OF OSCs FROM *Arabidopsis thaliana*

The cDNA clones that correspond to the genes F1019.4 and T30F21.16 [9], which show high sequence homology to the known OSCs (74 and 57 % homology to LUP1, respectively), were obtained from total RNA isolated from *A. thaliana* whole plant by essentially the same reverse transcriptase-polymerase chain reaction (RT-PCR) method as reported [7]. PCRs using primers specific to each nucleotide sequence in the database and engineered so as to incorporate appropriate restriction sites just up- and downstream of each start and stop codons of the respective genes gave 2.3 kb DNA fragments corresponding to the full length. Subcloning of these fragments into yeast expression vector pYES2 (Invitrogen) under the GAL1 promoter gave the plasmids pOSC-F1019.4 and pOSC-T30F21.16.

FUNCTIONAL EXPRESSION OF *ARABIDOPSIS* OSCs IN YEAST

Yeast mutant GIL77 (*gal2 hem3-6 erg7 ura3-167*) [10], which lacks lanosterol synthase activity, was used to transform with pOSC-F1019.4 and pOSC-T30F21.16. GIL77 harboring these plasmids were separately cultured (20 mL) and expression induced by galactose. Harvested cells were disrupted with 20 % KOH/50 % EtOH (aq.), and the cyclization products were extracted with hexane. Preparative silica gel thin-layer chromatography (TLC) separation followed by reverse-phase high-performance liquid chromatography (RP/HPLC) analysis showed the presence of several peaks in both of the transformants that are not found in the negative control with void vector. Liquid chromatography/atmospheric pressure chemical ionization mass spectrometry (LC/APCIMS) analysis in positive-ion mode [7] indicated that all three peaks eluted between 13–16 min (Fig. 2) from the transformant with pOSCF-1019.4 are triterpene monoalcohols as they gave an ion at *m/z* 409 $[M+H-H_2O]^+$. Comparison of retention time as well as MS/MS fragmentation pattern of *m/z* 409 with authentic sample clearly identified the last eluting component (with an arrow) to be tirucalla-7,21-diene-3β-ol (Fig. 2). Similar LC/APCIMS analysis identified the three peaks (with arrows in Fig. 3) of pOSC-T30F21.16 transformant to be lupeol, bauerenol, and α-amyrin (in the order of elution), respectively (Fig. 3). All peaks eluted between 9–20 min are triterpene monoalcohols as they gave an ion at *m/z* 409 $[M+H-H_2O]^+$. Retention times

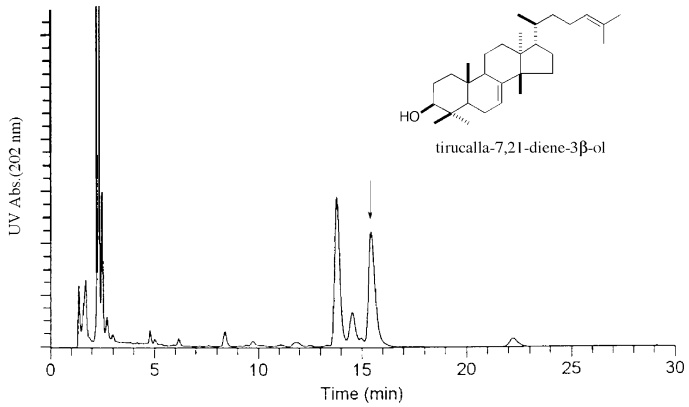

Fig. 2 HPLC analysis of clone-F1019.4 reaction products.

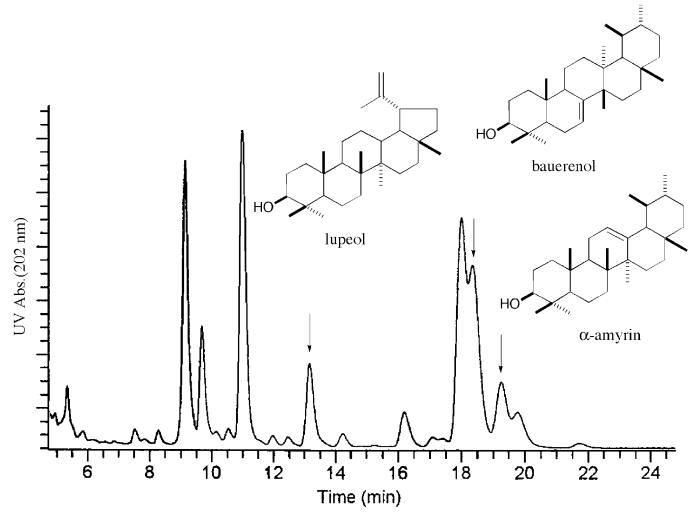

Fig. 3 HPLC analysis of clone-T30F21.16 reaction products.

and MS/MS patterns of these peaks, however, are different from those of the standard triterpene monoalcohols available in this laboratory, and rigorous identification of these products must await further large-scale cultures and separation of each component for spectroscopic structure analyses.

Previously known two triterpene synthases from *Arabidopsis*, LUP1 and YUP8H12R.43, are multifunctional. LUP1 lupeol synthase was first reported to produce lupeol as the major product with some minor triterpene monoalcohols including β-amyrin [6]. Later, it was shown to produce triterpene diol, 3β,20-dihydroxylupane, in comparable amount to lupeol [11,12]. The multiproduct nature of YUP8H12R.43 was more striking as it produced at least nine triterpene monoalcohols of different skeletal types in comparable level [7]. Two new clones, F1019.4 and T30F21.16, are again multifunctional. So far identified products of these clones are found among the products of YUP8H12R.43. However, the number of unidentified peaks in HPLC chromatograms suggests that F1019.4 and T30F21.16 produce at least two and six new triterpene skeletons, respectively, that have not been identified as OSC products of this plant. The total number of different triterpene skeletons produced by the characterized OSCs of this plant now exceeds 20. These triterpenes and their metabolites must be produced somewhere and at some timings in this tiny model plant, although phytochemical information on triterpene metabolites of this plant are still lacking. In the genome of *A. thaliana*, there remain eight putative OSC genes uncharacterized. One of such genes, d13715c, has been already obtained as a cDNA, however, it failed to show OSC activity in a similar expression system in yeast [8]. Further cDNA cloning and functional identification of the remaining genes are now in progress to know the magnitude of skeletal diversity of triterpenes in this model plant, although some of them might be pseudogenes.

SPECIFICITY OF LUP1-CATALYZED REACTION

In our previous study, feeding experiment of [1,2-$^{13}C_2$] acetate revealed that in LUP1-catalyzed reaction, deprotonation occurs on either of the two terminal methyl groups of lupenyl cation in equal ratio to produce lupeol [13]. Quenching of the same lupenyl cation by water would produce diol product, 3β,20-dihydroxylupane. Later identification of this diol as one of the major LUP1 products [11,12] led us to test whether this water addition is stereospecific or not, although the scrambling of methyl groups in deprotonation is suggestive of nonstereospecific water addition (Fig. 4).

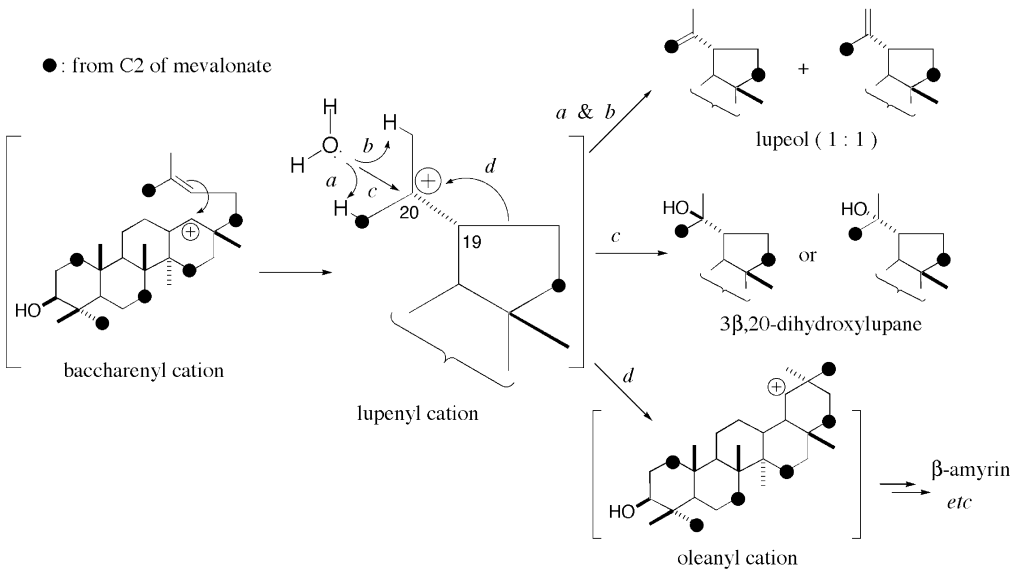

Fig. 4 Competition at lupenyl cation stage in LUP1-catalyzed reaction.

Two methyl groups on C20 of 3β,20-dihydroxylupane have different biosynthetic origin and could be distinguished by [1,2-$^{13}C_2$] acetate feeding. As C6 of mevalonate originates from intact incorporation of acetate, the ^{13}C NMR signal of any methyl group derived from this carbon should appear with an accompanying boublet, while one from C2 of mevalonate should appear as an enriched singlet. Relevant methyl groups of 3β,20-dihydroxylupane are magnetically nonequivalent and resonate at distinct chemical shifts (24.76 and 31.53 ppm in $CDCl_3$), thus allowing the application of the same methodology for specificity test. [1,2-$^{13}C_2$] Sodium acetate (90 % atom ^{13}C) was fed to the culture (1 L) of the yeast transformant, expressing LUP1 as reported [12]. 3β,20-Dihydroxylupane was extracted by hexane, separated and purified by silica gel column eluted with benzene/acetone mixture for NMR measurement. Contrary to our expectation, only one methyl signal in lower magnetic field at 31.53 ppm was accompanied by a boublet of J = 39.6 Hz, indicating that this methyl derived from C6 of mevalonate. On the other hand, a high-field resonating methyl carbon (24.76 ppm) appeared as an enriched singlet. Accordingly, a quaternary carbon of C20 at 73.52 ppm was accompanied by only one boublet of J = 39.6 Hz (Fig. 5).

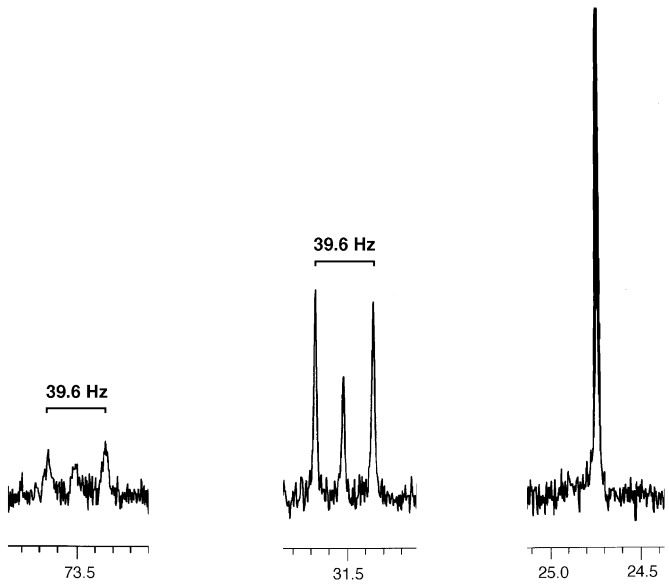

Fig. 5 Partial ^{13}C NMR spectrum of LUP1 dihydroxylupane fed with [1,2-$^{13}C_2$] acetate.

These results clearly indicate that the water addition to lupenyl cation is stereospecific and the rotation of C19–C20 bond of lupenyl cation is restricted before water addition. Simultaneous production of deprotonation product and water addition product by LUP1 implies that a water molecule in the active site takes both tasks of deprotonation to produce lupeol and direct quenching of carbocation to produce diol. If this is the case, the position of water in the active site must be flexible and pivotal to be a multiproduct synthase. When it is close enough to capture carbocation, diol is produced. And when it is a bit far from the tertiary cation center, it acts as a base to deprotonate from the methyl groups. Furthermore, to account for the stereospecific water addition and nondiscrimination of two methyl groups in deprotonation, water must be in one specified face only of the nascent lupenyl cation intermediate and at the same time in an equal proximity to both methyl groups. To know the stereochemistry of water addition, magnetically and biosynthetically nonequivalent two methyl groups have to be correlated. As has been reported, *m*-chloroperbenzoic acid oxidation of lupenyl acetate provided only one enantiomer of C20–C29 epoxide of unknown stereochemistry [14]. $LiAlD_4$ reduction of this epox-

ide afforded deuterated 3β,20-dihydroxylupane. ^{1}H-, ^{13}C-NMR and HMQC analyses established that deuterium was incorporated into the high-field resonating methyl group (^{1}H: 1.121 ppm, ^{13}C: 24.76 ppm). Establishment of the stereochemistry of epoxide, which is now underway by X-ray crystallography, will show the direction of water addition to lupenyl cation, and thus would provide an important insight for the active-site structure of this interesting multifunctional enzyme, whose 3D structure has not been obtained yet.

ACKNOWLEDGMENTS

The authors thank Dr. K. Masuda of Showa Pharmaceutical University for the generous gift of lupenyl acetate and Dr. Y. Fujimoto of Tokyo Institute of Technology for useful discussions. A part of this study was financially supported by a Grant-in-Aid for Scientific Research on Priority Area (A) (No. 12045213) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and a Regional R&D Consortium Project from the Ministry of Economy, Trade and Industry, Japan.

REFERENCES

1. *Arabidopsis* Genome Initiative. *Nature* **408**, 796–815 (2000).
2. G. R. Waller and K. Yamasaki (Eds.). *Saponins Used in Traditional Medicine*, Plenum, New York (1996).
3. N. Kawano, K. Ichinose, Y. Ebizuka. *Biol. Pharm. Bull*. **25**, 477–482 (2002) and references cited therein.
4. M. Morita, M. Shibuya, T. Kushiro, K. Masuda, Y. Ebizuka. *Eur. J. Biochem*. **267**, 3453–3460 (2000).
5. E. J. Corey, S. P. T. Matsuda, B. Bartel. *Proc. Natl. Acad. Sci. USA* **90**, 11628–11632 (1993).
6. J. B. R. Herreira, B. Bartel, W. K. Wilson, S. P. T. Matsuda. *Phytochemistry* **49**, 1905–1911 (1998).
7. T. Kushiro, M. Shibuya, K. Masuda, Y. Ebizuka. *Tetrahedron Lett*. **41**, 7705–7710 (2000).
8. T. H.-Muller, H. Schaller, P. Benveniste. *Plant Mol. Biol*. **45**, 75–92 (2001).
9. EMBL/GenBank/DDBJ data bank Acc. No. AC007152 for F1019.4 and AC007260 for T30F21.16.
10. T. Kushiro, M. Shibuya, Y. Ebizuka. *Eur. J. Biochem*. **256**, 238–244 (1998).
11. M. J. Segura, M. M. Meyer, S. P. T. Matsuda. *Org. Lett*. **2**, 2257–2259 (2000).
12. T. Kushiro, M. Shibuya, K. Masuda, Y. Ebizuka. *J. Am. Chem. Soc*. **122**, 6816–6824 (2000).
13. T. Kushiro, M. Shibuya, Y. Ebizuka. *J. Am. Chem. Soc*. **121**, 1208–1216 (1999).
14. W. H. Hui and M. M. Li. *Phytochemistry* **16**, 111–112 (1977).