Pergamon

Tetrahedron Letters 41 (2000) 7705–7710

TETRAHEDRON LETTERS

A novel multifunctional triterpene synthase from *Arabidopsis thaliana*

Tetsuo Kushiro,[a] Masaaki Shibuya,[a] Kazuo Masuda[b] and Yutaka Ebizuka[a,*]

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

[b]*Showa Pharmaceutical University, 3-3165 Higashi-Tamagawagakuen, Machida, Tokyo 194-8543, Japan*

Received 5 July 2000; revised 31 July 2000; accepted 4 August 2000

Abstract

The *Arabidopsis thaliana* genome sequencing project has identified several triterpene synthase homologues. One cDNA of these clones, YUP8H12R.43, was obtained and functionally expressed in yeast. At least nine triterpenes have been identified as its products based on NMR and LCMS analysis. The products include extensively migrated triterpenes, multiflorenol and bauerenol. Several chimeric clones were constructed between YUP8H12R.43 and *A. thaliana* lupeol synthase LUP1, to reveal that a C-terminal half and a part of N-terminus is important for such product multiplicity. The presence of such multifunctional triterpene synthase in plants is noteworthy from both a mechanistic and a physiological point of view.

Keywords: Arabidopsis thaliana; YUP8H12R.43; oxidosqualene cyclase; triterpene synthase.

Triterpenes are a structurally diverse family of natural products mainly found in plants. They exhibit a variety of pharmacological activities. In particular, their glycosides (saponins) are often active constituents of important traditional medicines.[1] In most cases the physiological role of these triterpenoids in plants is not clear. The spectrum of triterpene constituents of one plant differs from another, and it is an intriguing question as to why and how each plant produces a certain spectrum of these compounds.

All these triterpenes are biosynthesized from a common acyclic precursor 2,3-oxidosqualene by the action of oxidosqualene cyclases (OSCs).[2] More than 90 different carbon skeletons are known, which suggest that a corresponding number of product-specific triterpene synthases may exist. Up to now, eight different triterpene synthase cDNAs have been cloned from various plants. These are β-amyrin synthases from *Panax ginseng* (*PNY* and *PNY2*),[3] *Pisum sativum* (*PSY*),[4] and *Glycyrrhiza glabra* (*GgbAS1*),[5] lupeol synthases from *Olea europaea* (*OEW*),[6]

* Corresponding author. Tel/fax: +81-3-5841-4744; e-mail: yebiz@mol.f.u-tokyo.ac.jp

0040-4039/00/$ - see front matter
PII: S0040-4039(00)01347-2

Taraxacum officinale (*TRW*),[6] and *Arabidopsis thaliana* (*LUP1*),[7] and α-amyrin producing mixed amyrin synthase from *P. sativum* (*PSM*).[4] Among them, *P. sativum* PSM and *A. thaliana* LUP1 were shown to produce more than one product when expressed in yeast. This has led to the conclusion that a corresponding number of product-specific triterpene synthases might not exist in nature.

The genome sequencing project of *A. thaliana* has almost come to completion, and many of the sequences are open on the database. A database search for triterpene synthase homologues revealed the presence of several clones as candidates for OSCs. Among them, the chromosome 1 YAC YUP8H12R sequence contains four OSC homologues, which are clustered within 26 kb on the genome.[8] The YUP8H12R.43 clone, which is located 9 kb upstream of an apparent LUP1 genomic clone (YUP8H12R.42), exhibits 79.4% amino acid sequence identity to LUP1, and therefore, this LUP1 homologue was thought to code for another lupeol synthase of *A. thaliana*.

It would be an important task to characterize these clones precisely, so as to understand the physiological role of triterpenes in this model plant. Therefore, we have set out functional studies on these putative triterpene synthase clones. Here we report on the functional expression of the YUP8H12R.43 clone in yeast and demonstrated that it codes for a multifunctional triterpene synthase catalyzing the formation of nine different triterpenes as products.

The YUP8H12R.43 cDNA clone was obtained from total RNA isolated from *A. thaliana* whole plant by RT-PCR.[9] Specific N- and C-terminal oligo DNA primers containing the *Kpn* I site immediately upstream of the start codon and the *Xba* I site immediately downstream of the stop codon were designed based on the database sequence (Kpn-YUP43-N and Xba-YUP43-C).[10] PCR using these primers and *A. thaliana* cDNA as a template gave a 2.3 kb DNA fragment corresponding to the full length. Subcloning of this fragment into the *Kpn* I and *Xba* I sites of yeast expression vector pYES2 (Invitrogen) under the *GAL1* promoter gave the plasmid $pOSC_{YUP43}$. Sequencing of this clone in both strands revealed three amino acid residues to be different from the reported genomic sequence.[11] Yeast mutant GIL77 (*gal2 hem3-6 erg7 ura3-167*), which lacks lanosterol synthase activity, was used to transform with this plasmid.[3] GIL77 harboring $pOSC_{YUP43}$ was cultured (20 mL), expression was induced by galactose and harvested. Cells were disrupted with 20% KOH/50% EtOH(aq) and the cyclization products were extracted with hexane. Preparative silica gel TLC separation followed by reverse-phase HPLC analysis showed a presence of nine major peaks (Fig. 1). A large scale culture (2000 mL) gave, after extraction and purification with silica gel column, 8 mg of triterpene alcohol mixture. ^{1}H and ^{13}C NMR analysis and comparison with authentic samples, especially around olefinic region, unambiguously identified seven triterpenes as lupeol (**1**), taraxasterol (**4**), β-amyrin (**5**), Ψ-taraxasterol (**6**), bauerenol (**7**), α-amyrin (**8**) and multiflorenol (**9**).[3,4,12] Additional signals were also observed in ^{1}H NMR which corresponded to either butyrospermol (**2**) or tirucalla-7,21-dien-3β-ol (**3**) or both.[13] Further analysis on LC-APCIMS in positive-ion mode, for comparison of retention time as well as selected ion monitoring MS/MS fragmentation patterns on m/z 409 $[M+H-H_2O]^+$ with authentic samples, clearly identified the presence of both **2** and **3** together with the above-mentioned products.[4,14] Therefore, YUP8H12R.43 was unambiguously demonstrated to be a multifunctional triterpene synthase producing at least nine triterpenes with **4** and **5** being the major products. This is the first triterpene synthase clone that gives multiflorenol and bauerenol. Other minor peaks were also observed on HPLC, which correspond to dammara-18(28),21-dien-3β-ol in retention time as well as in APCI MS/MS fragmentation patterns.[14] However, definitive identification of these products must await further rigorous analysis.

Multifunctional triterpene synthases such as PSM and LUP1 have already been reported; however, in the PSM case, α- and β-amyrin are the two major products with others being produced in only minor amounts,[4] while in the LUP1 case, lupeol dominates over other products.[7] From the present results, the product multiplicity in YUP8H12R.43 is overwhelming in that each of the nine triterpenes are produced in comparable levels. Formation of these triterpenes can be rationalized on the basis of cyclization and rearrangement reactions (Scheme 1). Thus, **1**, **2** and **3** are formed from intermediate carbocation en route to oleanyl cation. Two routes branch at this point as a methyl group shift from C-20 (route *a*) would give the taraxasteryl cation from which **4**, **6**, **7** and **8** are formed, while a hydride shift from C-18 (route *b*) would give **5** and **9**. The exact nature of such product multiplicity is not clear; however, flexible conformation in the active site of the enzyme might allow the cyclization reaction to terminate at various intermediate stages.

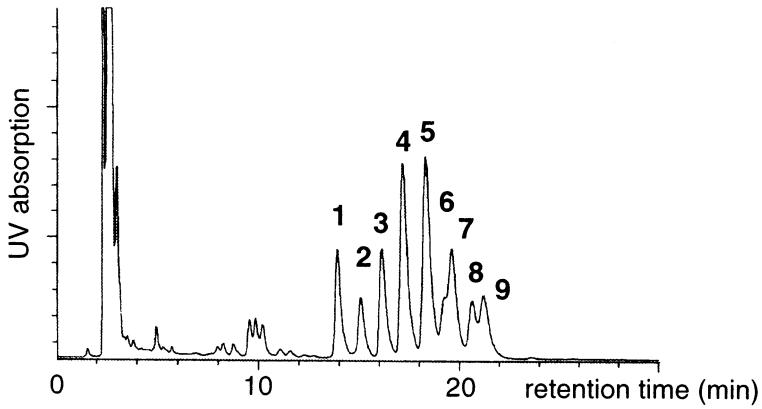
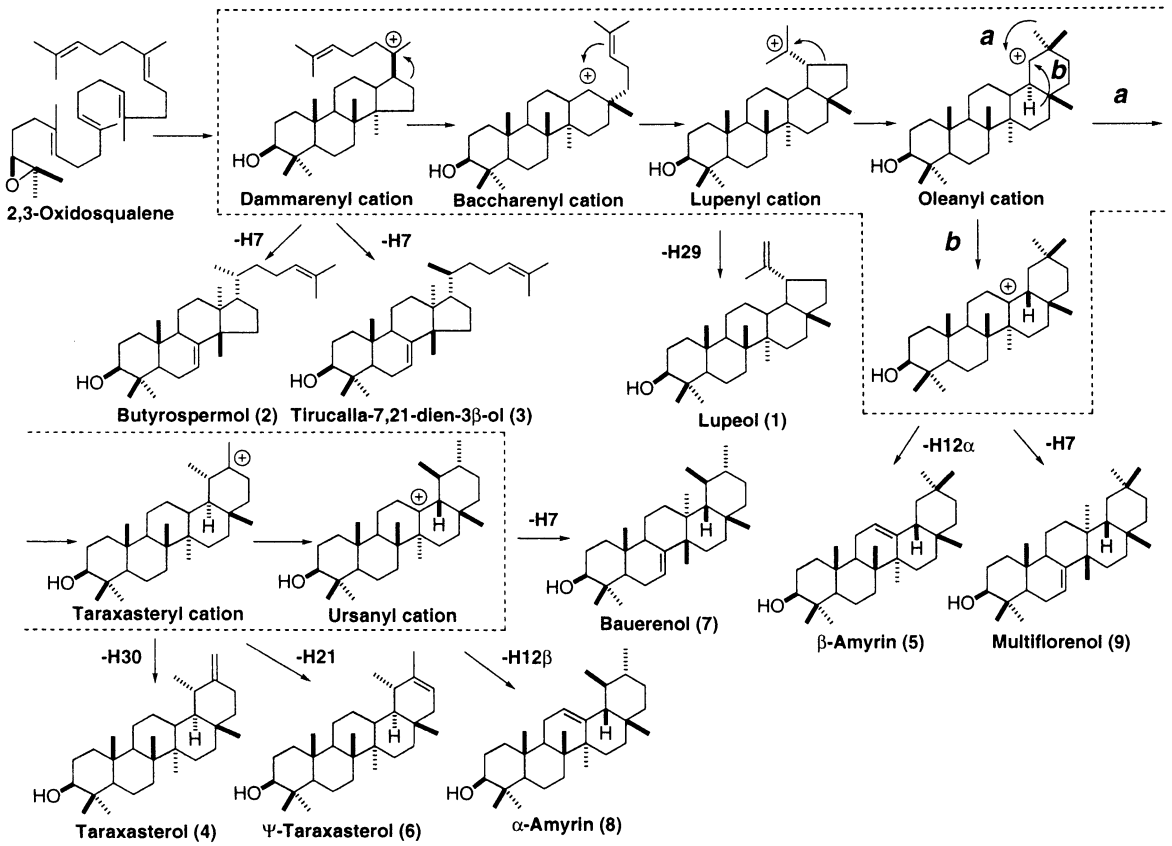


Figure 1. HPLC profile of products from YUP8H12R.43 clone. The numbers on peaks correspond to products on Scheme 1

Since YUP8H12R.43 shares 79.4% amino acid identity with LUP1 from the same *A. thaliana*, the remaining 20% sequence should contain amino acid residues responsible for such product multiplicity. In particular, YUP8H12R.43 should contain a residue responsible for the formation of bauerenol and multiflorenol, two extensively migrated triterpenes. In order to search for a sequence region containing such amino acid residues, several chimeric clones were constructed. *Nde* I site (nucleotide number 549, amino acid number 184 in YUP8H12R.43) and *Eco* RI site (nucleotide number 1190, amino acid number 398 in YUP8H12R.43) were common to both clones, and were used to construct the chimeras Chi 1–4 (Fig. 2). These clones were similarly expressed in yeast, and triterpene product patterns observed with LC-APCIMS. Chi 1 gave a product pattern similar to native LUP1 with increased production of minor compounds especially those which elute faster than lupeol in reverse-phase HPLC. Chi 2 gave an almost identical product pattern with native LUP1. On the other hand, Chi 3 gave increased production of **2**, **3** and **4** together with lupeol and β-amyrin exhibiting a similar product multiplicity to that observed in YUP8H12R.43. Importantly, this clone produced a minor amount of multiflorenol (Fig. 3). Chi 4 was inactive showing no production of triterpene. These results pointed out that a C-terminal half and a part of N-terminus of YUP8H12R.43 should contain the region important for observed product multiplicity and the formation of multiflorenol. Our previous chimeric and site-directed mutagenesis studies showed that the second quarter from the N-terminus contains an important residue for the formation of β-amyrin and lupeol.[9,13] The



Scheme 1. Cyclization mechanism for the formation of nine triterpenes by YUP8H12R.43 clone

present results suggest that residues governing the product specificities are not only restricted to this region, but are also scattered around on the whole sequence. In fact, formation of α-amyrin and bauerenol seem to require the whole native sequence of YUP8H12R.43. Studies toward identifying these residues for product multiplicity are now in progress.

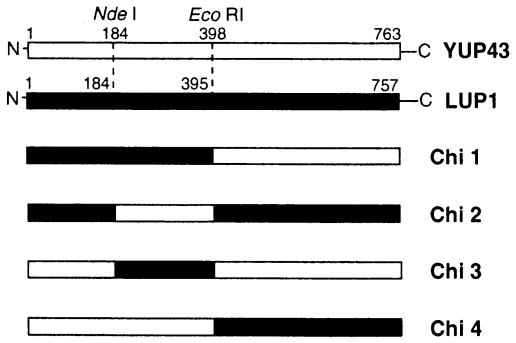


Figure 2. Chimera constructs

Up to now, studies on triterpene constituents of *A. thaliana* have not been reported. It would be interesting to see whether these products of YUP8H12R.43 can be found in the native plant.

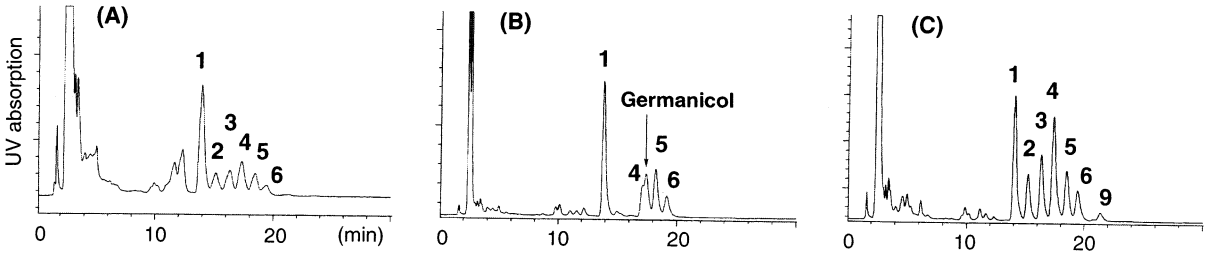


Figure 3. HPLC profiles of (A) Chi 1, (B) Chi 2 and (C) Chi 3

Identifying the function of the remaining homologous genomic clones and triterpene constituents are required to fully understand their physiological role in this model plant.

Acknowledgements

We are grateful to Professor S. Ayabe of Nihon University for kindly providing authentic taraxasterol. This work was in part supported by the Research for the Future Program (JSPS-RFTF96I00302) of The Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research ((A) 10358015) to Y.E. and Grant-in-Aid for JSPS Fellows to T.K. from the Ministry of Education, Science, Sports and Culture, Japan.

References

1. Waller, G. R.; Yamasaki, K. *Saponins Used in Traditional and Modern Medicine*; Plenum Press: New York, 1996.
2. Abe, I.; Rohmer, M.; Prestwich, G. D. *Chem. Rev.* **1993**, *93*, 2189–2206.
3. (a) Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Eur. J. Biochem.* **1998**, *256*, 238–244. (b) Kushiro, T.; Shibuya, M.; Ebizuka, Y. In *Towards Natural Medicine Research in the 21st Century*, *Excerpta Medica International Congress Series 1157*; Ageta, H.; Aimi, N.; Ebizuka, Y.; Fujita, T.; Honda, G. Eds.; Elsevier Science: Amsterdam, 1998; pp. 421–428.
4. Morita, M.; Shibuya, M.; Kushiro, T.; Masuda, K.; Ebizuka, Y. *Eur. J. Biochem.* **2000**, *267*, 3453–3460.
5. Hayashi, H.; Huang, P.; Kobayashi, T.; Kirakosyan, A.; Inoue, K.; Hiraoka, N.; Ikeshiro, Y.; Kushiro, T.; Shibuya, M.; Ebizuka, Y., GenBank/EMBL data bank Acc. No. AB037203.
6. Shibuya, M.; Zhang, H.; Endo, A.; Shishikura, K.; Kushiro, T.; Ebizuka, Y. *Eur. J. Biochem.* **1999**, *266*, 302–307.
7. Herrera, J. R. B.; Bartel, B.; Wilson, W. K.; Matsuda, S. P. T. *Phytochemistry* **1998**, *49*, 1905–1911.
8. GenBank/EMBL data bank Acc. No. AC002986.
9. Kushiro, T.; Shibuya, M.; Ebizuka, Y. *J. Am. Chem. Soc.* **1999**, *121*, 1208–1216.
10. Kpn-YUP43-N: 5′-GACC**GGTACC**ATGTGGAAGTTGAAGATAGGAGAG-3′ (*Kpn* I site in bold face), Xba-YUP43-C: 5′-TATTAAG**TCTAGA**CTAAAGATCTTGATGAGTTGC-3′ (*Xba* I site in bold face).
11. Glycine instead of reported glutamate at position 88, methionine instead of isoleucine at position 90, and glycine instead of cystein at position 413.
12. Taraxasterol: ^{1}H NMR (500 MHz, $CDCl_3$) δ 0.767 (s), 0.852 (s), 0.855 (s), 0.931 (s), 0.971 (s), 1.018 (d, $J=6.7$ Hz), 1.021 (s), 3.203 (dd, $J=11.3$, 4.8 Hz), 4.600 (brs), 4.617 (brs). Ψ-Taraxasterol: ^{1}H NMR (500 MHz, $CDCl_3$) δ 0.734 (s), 0.767 (s), 0.852 (s), 0.950 (s), 0.974 (s), 0.987 (d, $J=6.2$ Hz), 1.043 (s), 1.634 (s), 3.208 (dd, $J=11.6$, 4.9 Hz), 5.262 (br). Bauerenol: ^{1}H NMR (500 MHz, $CDCl_3$) δ 0.745 (s), 0.857 (s), 0.903 (d, $J=5.8$ Hz), 0.943 (s), 0.967 (s), 0.995 (s), 1.036 (s), 1.048 (d, $J=7.0$ Hz), 3.246 (dd, $J=11.0$, 3.0 Hz), 5.416 (br). Multiflorenol: ^{1}H NMR (500 MHz, $CDCl_3$) δ 0.744 (s), 0.864 (s), 0.967 (s), 0.976 (s), 1.055 (s), 1.070 (s), 1.085 (s), 3.245 (dd, $J=11.0$, 3.4 Hz), 5.474 (br).
13. Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. *J. Am. Chem. Soc.* **2000**, *122*, 6816–6824.

14. LC-APCIMS was measured with LCQ (Thermo Quest). HPLC condition: SUPER-ODS column (ϕ 4.6×l 200 mm) (TOSOH) with 95% CH_3CN aq. (flow rate 1.0 mL/min, detection UV 202) at 40°C. The number in parentheses represents relative intensity of the peak in percentage. Tirucalla-7,21-dien-3β-ol: MS m/z 409 $[M+H-H_2O]^+$, MS/MS (precursor ion at m/z 409) 299 (20), 285 (20), 271 (25), 231 (15), 217 (100), 203 (30), 191 (50). Taraxasterol: MS m/z 409 $[M+H-H_2O]^+$, MS/MS (precursor ion at m/z 409) 339 (20), 325 (20), 315 (20), 313 (50), 299 (20), 285 (98), 271 (100), 259 (95), 257 (75), 231 (40), 229 (50), 217 (40), 215 (25), 203 (30), 177 (30). Ψ-Taraxasterol: MS m/z 409 $[M+H-H_2O]^+$, MS/MS (precursor ion at m/z 409) 339 (20), 325 (20), 313 (40), 299 (40), 285 (95), 271 (90), 259 (100), 257 (60), 231 (45), 229 (45), 217 (35), 215 (25), 177 (30). Bauerenol: MS m/z 409 $[M+H-H_2O]^+$, MS/MS (precursor ion at m/z 409) 325 (20), 313 (100), 299 (20), 285 (20), 271 (25), 231 (20), 217 (25), 191 (30). Multiflorenol: MS m/z 409 $[M+H-H_2O]^+$, MS/MS (precursor ion at m/z 409) 313 (10), 299 (20), 285 (30), 271 (50), 245 (30), 231 (35), 217 (60), 205 (30), 203 (30), 191 (100), 177 (20). Dammara-18(28),21-dien-3β-ol: MS m/z 409 $[M+H-H_2O]^+$, MS/MS (precursor ion at m/z 409) 353 (10), 299 (20), 285 (50), 271 (100), 257 (70), 245 (40), 231 (30), 217 (20), 203 (10), 191 (10).