ChemPubSoc Europe

DOI: 10.1002/cbic.201500663

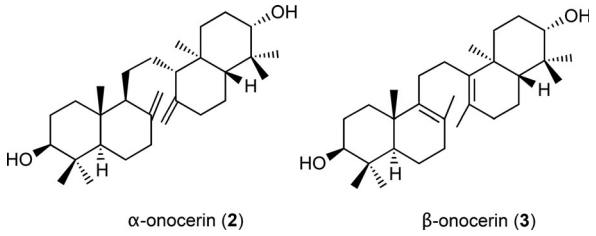
CHEMBIOCHEM Communications

Onocerin Biosynthesis Requires Two Highly Dedicated Triterpene Cyclases in a Fern *Lycopodium clavatum*

Takeshi Araki[+],[a] Yusuke Saga[+],[a] Momo Marugami,[a] Junnosuke Otaka,[a] Hiroshi Araya,[a] Kazuki Saito,[b] Mami Yamazaki,[b] Hideyuki Suzuki,[c] and Tetsuo Kushiro*[a]

Onocerin is known for its unusual structure among triterpenoids, with a symmetrical structure that is formed by cyclizations at the both termini of dioxidosqualene. The nature of the enzyme catalyzing these unusual cyclizations has remained elusive for decades. Here, we report the cloning of genes responsible for these reactions; they exhibited unprecedented substrate specificities among oxidosqualene cyclase family members. Two genes, LCC and LCD, were identified from the fern *Lycopodium clavatum*. Expression in yeast revealed that both were required to produce α-onocerin. LCC, the first dioxidosqualene cyclase, catalyzed the production of a novel intermediate pre-α-onocerin from only dioxidosqualene as a substrate; LCD catalyzed the second half of the cyclization, exclusively from pre-α-onocerin. These results demonstrated that these two most unusual oxidosqualene cyclases were involved in onocerin biosynthesis.

The enzymatic cyclization of 2,3-oxidosqualene (**1**) involved in triterpene biosynthesis has attracted much attention for decades, primarily because of its remarkable transformation of a simple linear substrate into complex polycyclic structures while controlling the stereochemistry of the multiple chiral centers generated.[1] Such reactions cannot (or can only partly) be mimicked by current organic chemistry and are thus still confined to enzymatic processes.[2] The enzymes responsible for catalyzing these reactions, oxidosqualene cyclases (OSCs), typically produce polycyclic structures comprised of four or five rings. These include sterol precursors such as lanosterol and cycloartenol, as well as precursors for various triterpenoids that are widely distributed among plants and are often bioactive constituents of a variety of medicinal plants.[1] In addition to the well-known triterpenes, there are several unusual types that are formed by irregular cyclization patterns and have attracted attention for a long time.[3] One of the most prominent of these, onocerin, is found in several ferns including *Lycopodium clavatum*[4] and some higher plants such as the legume *Ononis spinosa*.[5] Onocerin's unique symmetrical structure includes two bicycles connected by two methylene carbons, and it has been a target of intense synthetic studies (Scheme 1).[6a,b] Onocerins include α-onocerin (**2**) and β-onocerin (**3**). Since the discovery of onocerin in 1855 and its structure elucidation in 1955,[5] its biosynthesis has been proposed to involve cyclization from both termini of a squalene molecule. This was proven to be the case, as 2,3,22,23-dioxidosqualene (**4**), a bis-epoxidized squalene, was shown to be the precursor of **2** in *O. spinosa* cell-free extracts.[7,8] Despite onocerin's structure being discovered more than half a century ago, the nature of the synthase that catalyzes the unusual double cyclizations from both terminal epoxides remained elusive. Two possibilities have been envisioned: 1) a single enzyme responsible for catalyzing cyclizations at both terminal epoxides and 2) two enzymes, one catalyzing the cyclization at one terminus and another catalyzing the cyclization at the other. In both cases, the cyclase responsible for catalyzing the second cyclization should accommodate the unusual partially cyclized intermediate as a substrate.



Scheme 1. Structures of onocerin.

We searched for genes encoding enzymes involved in the biosynthesis of onocerin, and intended to clone genes closely related to OSCs from *L. clavatum*. Because the substrate **4** closely resembles **1** and several OSCs have been shown to accept **4** as a substrate,[9–12] we speculated that genes related to OSCs might be involved in this biosynthetic process. Several degenerate oligonucleotide primers were designed based on conserved amino acid sequence motifs in plant OSCs. cDNA was prepared from total RNA extracted from aerial parts of *L. clavatum*. Nested PCR was carried out with several combinations of primers in order to amplify specific DNA fragments of

[a] *T. Araki,[+] Y. Saga,[+] M. Marugami, Dr. J. Otaka, Dr. H. Araya, Dr. T. Kushiro*
Graduate School of Agriculture, Meiji University
1-1-1 Higashimita, Tama-ku, Kawasaki
Kanagawa 214-8571 (Japan)
E-mail: kushiro@meiji.ac.jp

[b] *Prof. Dr. K. Saito, Dr. M. Yamazaki*
Graduate School of Pharmaceutical Sciences, Chiba University
1-8-1, Inohana, Chuo-ku, Chiba-shi
Chiba 260-8675 (Japan)

[c] *Dr. H. Suzuki*
Kazusa DNA Research Institute
2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818 (Japan)

[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201500663.

appropriate sizes. After the fragments had been sequenced, three were identified (LCA, LCB, and LCC; Figure S1 in the Supporting Information). Sequence comparison with OSCs indicated that LCA and LCB were more closely related to cycloartenol synthases (CASs) from ferns and higher plants (65–77% amino acid sequence identity),[13–15] and therefore were thought to be *L. clavatum* CAS genes. However, LCC was only distantly related to CAS (55% sequence identity), and thus was considered a candidate gene for onocerin synthase. The full-length sequence of LCC was obtained by RACE PCR, and the cDNA was introduced into the yeast expression plasmid pESC-Ura under control of the GAL1 promoter. The *Saccharomyces cerevisiae* ergosterol auxotrophic mutant strain GIL77, a mutant deficient in lanosterol synthase, was used as a host for the expression of LCC.[16] It accumulates **1** within the cells and, because of a squalene epoxidase that epoxidizes another terminal olefin, **4** also accumulates in these cells. After induction of the culture with galactose, cells were harvested and extracted with hexane. TLC, HPLC, and GC-MS analyses showed one major product derived from **4** (Figures 1 and S2). No product corresponding to onocerin or a typical triterpene monoalcohol was seen. NMR analyses (^{1}H, ^{13}C, COSY, HMBC, and HMQC) of a sample from a large-scale culture (4 L) indicated a bicyclic structure with an epoxide at C21–C22 (Figure S7). By comparing with published data of related compounds,[11,12,17] the data were determined to be in agreement with structure **5**. This corresponds to the expected intermediate product of onocerin biosynthesis where only the first cyclization of **4** has taken place. Therefore, the enzymatic product of LCC was identified as polypoda-8(26),13,17-trien-21,22-epoxy-3β-ol (**5**). The structure of this compound was previously proposed and tentatively named "pre-onocerin",[8] although it was not characterized or isolated from a natural source. We propose renaming it "pre-α-onocerin" (**5**). Notably, we were not able to find a trace of the non-epoxidized form of this compound (within the limit of detection). Expression of OSC genes in *S. cerevisiae* GIL77 should allow the expression of enzymes that use both **1** and **4** as substrates. The production of just the epoxidized product **5** indicated that LCC exclusively uses **4** as a substrate (not **1**). This is the first demonstration of an OSC family enzyme that specifically accepts **4** as a substrate, and thus can be termed "dioxidosqualene cyclase". Our finding is in good agreement with a previous observation with cell-free extracts of *O. spinosa*, which showed almost no conversion of radiolabeled **1**.[7] It is also the first report of an OSC that produces a bicyclic polypodane structure. Marneral synthase from *Arabidopsis thaliana* is the only other example where the cyclization stops at a bicyclic stage.[18] Therefore, LCC was demonstrated to be the pre-α-onocerin synthase that exclusively uses **4** as a substrate and catalyzes the formation of only the first two rings (corresponding to half of the onocerin molecule; Scheme 2).

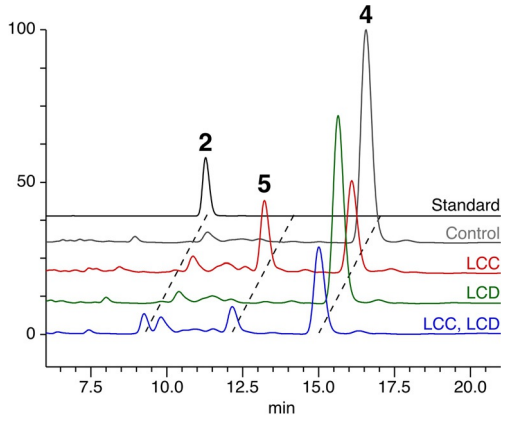
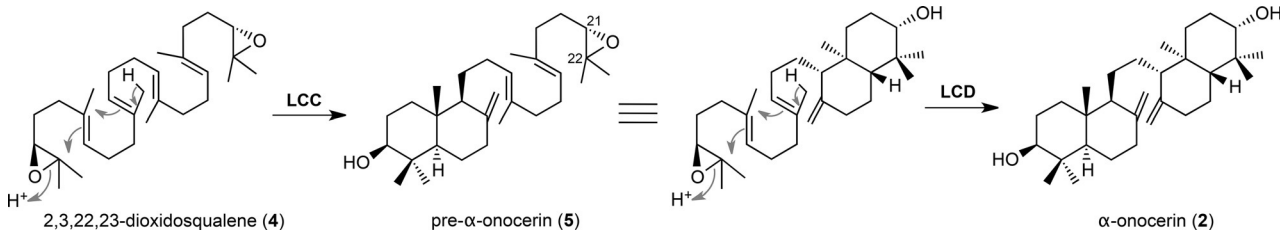


Figure 1. HPLC profiles of hexane extracts from yeast samples expressing empty vector (Control), LCC, LCD, both LCC and LCD, authentic **2** (Standard). Peaks are identified with dashed lines.

Identification of the function of LCC indicated that another enzyme was required for the formation of **2**. Our attention focused on identifying the second enzyme, which is different from LCA and LCB—our expression studies in yeast having shown that LCA catalyzed the formation of cycloartenol (Figure S21). We turned to the RNAseq data obtained from an RNA sample extracted from *L. clavatum* with an Illumina Hiseq 1000 sequencer. The data were searched for OSC homologues, and to our delight we were able to identify a new gene, named LCD, that exhibited 50% sequence identity to LCC. Phylogenetic analysis revealed that LCC and LCD are equally distantly related to CAS and other plant OSCs (Figure S27). The full-length LCD cDNA was amplified by PCR and introduced into the yeast expression plasmid pESC-Ura to obtain pESC-LCD. In addition, LCC and LCD were simultaneously introduced into pESC-Ura to obtain pESC-LCC,LCD. These plasmids were transformed into GIL77, the cells were cultured, and expression was induced with galactose. Inspection of the hexane extracts from these strains revealed a product that corresponded to authentic **2** by TLC, HPLC, and GC-MS analyses of a sample expressing both LCC and LCD (Figures 1 and S2). This product was absent in the sample expressing only LCD. Finally, a large-scale culture (4 L) was grown, and a sample was extracted and purified. All



Scheme 2. Cyclization of dioxidosqualene (**4**) into α-onocerin (**2**) catalyzed by LCC and LCD.

the NMR data were in complete agreement with authentic **2** (Figure S23).[19a,b] Therefore, LCD was identified as the α-onocerin synthase catalyzing the second cyclization starting from **5** produced by LCC (Scheme 2). No product was seen when LCD was expressed alone, thus demonstrating that LCD accepts only **5** as a substrate (not **1** or **4**). The discovery of a dedicated OSC family enzyme that only accepts a partially cyclized isoprenoid as a substrate is unprecedented. Therefore, onocerin biosynthesis was demonstrated for the first time to involve two highly unusual OSC family enzymes: LCC, the first dioxidosqualene cyclase to be identified, and LCD, the α-onocerin synthase that exclusively accepts partially cyclized **5** as a substrate.

Although the reaction and the substrates were quite different from those of other OSCs, both LCC and LCD were remarkably similar to known OSCs in primary sequence. Both enzymes catalyze the same chemistry, although the substrate specificities were quite different from each other. The way LCC achieves substrate specificity towards **4** was intriguing, as LCC has to discriminate between two closely related substrates (differing only by the epoxide at C21–C22). According to the X-ray crystal structure of human lanosterol synthase (hLAS) in complex with lanosterol,[20] the hLAS motif ^{229}LWCHC^{233} is close to these carbons of the substrate, and notable differences were seen in this sequence. LCC has ^{254}MNIHG^{258} at the corresponding position (Figure S26), and this difference might explain the epoxide recognition upon substrate binding. LCC has ^{481}ISDCSAE^{487} at the conserved motif DCTAE of other OSCs. An isoleucine is two residues upstream of the critical aspartate of this motif in CAS,[21] cucurbitadienol synthase,[22] and marneral synthase,[18] all of which cyclize oxidosqualene folded in a *chair–boat* conformation; however **4** was folded in a *chair–chair* conformation during the formation of **5**. The role of this isoleucine in LCC is of particular interest. The sequence of LCD also showed a notable difference at the sequence motif ^{260}LPVLY^{264}. An unusual, partly cyclized decalin structure of **5** should be accommodated at the corresponding position in the active site, if the C21–C22 epoxide is bound near the critical aspartate of DCTAE.

Our results contrast sharply with the recently reported *Bacillus megaterium* enzyme that catalyzed cyclization of squalene from the both termini to produce onoceroid.[23] Therefore, in ferns, a strategy different from that in bacteria has evolved to construct an onoceroid molecule by two highly dedicated enzymes.

In conclusion, our studies uncovered the nature of the biosynthesis of **2** for the first time at the gene level. We demonstrated that two of the most unusual OSC family enzymes were responsible for the construction of two bicyclic structures cyclized from both termini of **4**, and we characterized intermediate **5**. With LCC and LCD in hand, we were able to achieve the first enzymatic synthesis of **2** in yeast. Identification of these OSC family enzymes opens the way to engineer and utilize them to prepare unusual kinds of triterpenes, such as by having partially cyclized structures that will contribute to structural diversity among triterpenoids.

Acknowledgements

*We thank Prof. Ushio Sankawa and Prof. Yutaka Ebizuka for providing the authentic sample of **2**, Sayaka Shinpo and Dr. Michimi Nakamura for technical support on RNAseq analysis, and Dr. Takahisa Nakane for measuring optical rotations. We are grateful to Meiji University for financial support.*

Keywords: biosynthesis · dioxidosqualene · onocerin · terpene cyclization · terpenoids

[1] T. Kushiro, Y. Ebizuka in *Comprehensive Natural Products II, Vol. 1: Natural Products Structural Diversity—I. Secondary Metabolites: Organization and Biosynthesis* (Eds.: L. Mander, H. W. Liu), Elsevier, Amsterdam, **2010**, pp. 673 – 708.
[2] R. A. Yoder, J. N. Johnston, *Chem. Rev.* **2005**, *105*, 4730 – 4756.
[3] R. Xu, G. C. Fazio, S. P. T. Matsuda, *Phytochemistry* **2004**, *65*, 261 – 291.
[4] H. Ageta, K. Iwata, Y. Ootake, *Chem. Pharm. Bull.* **1962**, *10*, 637.
[5] D. H. R. Barton, K. H. Overton, *J. Chem. Soc.* **1955**, 2639 – 2652.
[6] a) G. Stork, J. E. Davies, A. Meisels, *J. Am. Chem. Soc.* **1959**, *81*, 5516 – 5517; b) Y. Mi, J. V. Schreiber, E. J. Corey, *J. Am. Chem. Soc.* **2002**, *124*, 11290 – 11291.
[7] M. G. Rowan, P. D. G. Dean, T. W. Goodwin, *FEBS Lett.* **1971**, *12*, 229 – 232.
[8] M. G. Rowan, P. D. G. Dean, *Phytochemistry* **1972**, *11*, 3111 – 3118.
[9] O. Boutaud, D. Dolis, F. Schuber, *Biochem. Biophys. Res. Commun.* **1992**, *188*, 898 – 904.
[10] H. Shan, M. J. R. Segura, W. K. Wilson, S. Lodeiro, S. P. T. Matsuda, *J. Am. Chem. Soc.* **2005**, *127*, 18008 – 18009.
[11] G. C. Fazio, R. Xu, S. P. T. Matsuda, *J. Am. Chem. Soc.* **2004**, *126*, 5678 – 5679.
[12] T. Xiang, M. Shibuya, Y. Katsube, T. Tsutsumi, M. Otsuka, H. Zhang, K. Masuda, Y. Ebizuka, *Org. Lett.* **2006**, *8*, 2835 – 2838.
[13] J. Shinozaki, M. Shibuya, K. Masuda, Y. Ebizuka, *FEBS Lett.* **2008**, *582*, 310 – 318.
[14] J. Shinozaki, M. Shibuya, Y. Takahata, K. Masuda, Y. Ebizuka, *ChemBioChem* **2010**, *11*, 426 – 433.
[15] M. Morita, M. Shibuya, M.-S. Lee, U. Sankawa, Y. Ebizuka, *Biol. Pharm. Bull.* **1997**, *20*, 770 – 775.
[16] E. G. Gollub, K.-P. Liu, J. Dayan, M. Adlersberg, D. B. Sprinson, *J. Biol. Chem.* **1977**, *252*, 2846 – 2854.
[17] G. J. Bennett, L. J. Harrison, G.-L. Sia, K.-Y. Sim, *Phytochemistry* **1993**, *32*, 1245 – 1251.
[18] Q. Xiong, W. K. Wilson, S. P. T. Matsuda, *Angew. Chem. Int. Ed.* **2006**, *45*, 1285 – 1288; *Angew. Chem.* **2006**, *118*, 1307 – 1310.
[19] a) S. Berger, D. Sicker, *Classics in Spectroscopy*, Wiley-VCH, Weinheim, **2009**, pp. 427 – 442; b) G. F. Pauli, *Planta Med.* **2000**, *66*, 299 – 302.
[20] R. Thoma, T. Schulz-Gasch, B. D'Arcy, J. Benz, J. Aebi, H. Dehmlow, M. Hennig, M. Stihle, A. Ruf, *Nature* **2004**, *432*, 118 – 122.
[21] E. A. Hart, L. Hua, L. B. Darr, W. K. Wilson, J. Pang, S. P. T. Matsuda, *J. Am. Chem. Soc.* **1999**, *121*, 9887 – 9888.
[22] M. Shibuya, S. Adachi, Y. Ebizuka, *Tetrahedron* **2004**, *60*, 6995 – 7003.
[23] D. Ueda, T. Hoshino, T. Sato, *J. Am. Chem. Soc.* **2013**, *135*, 18335 – 18338.

Manuscript received: December 10, 2015
Accepted article published: December 14, 2015
Final article published: January 28, 2016