Biosynthesis

International Edition: DOI: 10.1002/anie.201700565
German Edition: DOI: 10.1002/ange.201700565

Discovery and Characterization of a New Family of Diterpene Cyclases in Bacteria and Fungi

*Yan-long Yang, Shasha Zhang, Ke Ma, Yuxing Xu, Qiaoqiao Tao, Yihua Chen, Juan Chen, Shunxing Guo, Jinwei Ren, Wenzhao Wang, Yong Tao, Wen-Bing Yin, and Hongwei Liu**

Abstract: *Diterpene cyclases from bacteria and basidiomycete fungi are seldom studied. Here, we presented the identification and verification of EriG, a member of the UbiA superfamily, as the enzyme responsible for the cyclization of the cyathane skeleton in the mushroom Hericium erinaceum. Genome mining using the EriG protein sequence as a probe led to the discovery of a new family of ubiquitous UbiA-related diterpene cyclases in bacteria and fungi. We successfully characterized seven new diterpene cyclases from bacteria or basidiomycete fungi with the help of an engineered Escherichia coli strain and determined the structures of their corresponding products. A new diterpene with an unusual skeleton was generated during this process. The discovery of this new family of diterpene cyclases provides new insight into the UbiA superfamily.*

Diterpenes, which are biosynthesized from the common isoprene precursor geranylgeranyl diphosphate, are widely distributed in plants, marine invertebrates, and microorganisms. They have attracted much attention from chemists and biologists owing to their unique structures and diverse bioactivities. The diverse carbon skeletons of diterpenes are formed through carbocation-mediated cyclization cascades catalyzed by diterpene cyclases. Compared with the plant diterpene cyclases, the diterpene cyclases from bacteria and basidiomycete fungi are seldom studied.[1] Specifically, the diterpene cyclases for the biosynthesis of dolabellane-, neodolabellane-, cyathane-, dolastane-, verrucosane-, isoagathane-, and neoverrucosane-type skeletons from bacteria or fungi have not been identified.[1c]

Fungi have been reported to produce a wide array of diterpenes. However, diterpenes from myxobacteria and actinobacteria are seldom reported.[2] Among fungal diterpenes, cyathane-type diterpenes, which have a 5-6-7 tricyclic skeleton, have attracted much interests owing to the great challenge of their chemical synthesis and their potential as drug candidates.[3] For example, erinacine A (Figure 1) from

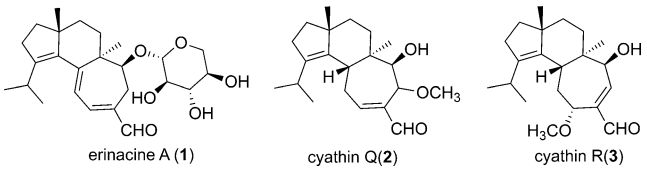


Figure 1. Representative cyathane-type diterpenes from mushrooms.

Hericium erinaceum was found to be a potent neurite outgrowth promoting agent that could be useful for the treatment of Alzheimer's disease;[4] and Cyathins Q (**2**) and R (**3**; Figure 1), which are the major secondary metabolites of *Cyathus africatus*, have been reported to be potent anticancer agents with novel action mechanisms.[5] Although the biosynthetic pathways of erinacines and cyathins have been investigated through isotope-labeling experiments, their corresponding genes or enzymes have not been identified so far.[6]

In general, the UbiA superfamily proteins catalyze the attachment of a prenyl chain or its derivatives in the production of ubiquinones, menaquinones, plastoquinones, hemes, chlorophylls, vitamin E, structural lipids, prenylated flavonoids, coumarin, and meroterpenoids.[7] However, two members from the UbiA family have recently been reported as terpene cyclases. PtmT1, a diterpene cyclase from *Streptomyces platensis*, catalyzes the formation of *ent*-atiserene (**6**) from *ent*-CPP in the biosynthesis of platencin.[8] FmaTC from *Aspergillus fumigatus* was revealed to be the sesquiterpene cyclase responsible for the biosynthesis of β-*trans*-bergamotene (**7**).[9] Herein, we present the characterization of seven UbiA-related diterpene cyclases from fungi and bacteria.

Given that a cluster-specific geranylgeranyl pyrophosphate synthase (GGPPS) is usually contained in the diterpene biosynthesis gene clusters of fungi and bacteria, we scanned all the GGPPS genes and their neighboring genes in the genomes of *H. erinaceum* and *C. striatus*. As a result, one gene

[*] Dr. Y.-L. Yang, K. Ma, Q. Tao, Dr. J. Ren, Dr. W. Wang, Prof. Dr. W.-B. Yin, Prof. Dr. H. Liu
State Key Laboratory of Mycology
Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101 (China)
E-mail: liuhw@im.ac.cn
K. Ma, Q. Tao, Prof. Dr. W.-B. Yin, Prof. Dr. H. Liu
Savaid Medical School, University of Chinese Academy of Sciences
Beijing, 100049 (China)
S. Zhang, Prof. Dr. Y. Tao
Chinese Academy of Sciences Key Laboratory of Microbial Physiological, and Metabolic Engineering, Institute of Microbiology
Chinese Academy of Sciences, Beijing, 100101 (China)
Y. Xu
Key Laboratory of Economic Plants and Biotechnology, Yunnan Key Laboratory for Wild Plant Resources, Kunming Institute of Botany
Chinese Academy of Sciences, Kunming 650201, Yunnan (China)
Prof. Dr. Y. Chen
State Key Laboratory of Microbial Resources
Yunnan Institute of Microbiology
Chinese Academy of Sciences, Beijing, 100101 (China)
Dr. J. Chen, Prof. Dr. S. Guo
Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College
Beijing, 100193 (China)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: http://dx.doi.org/10.1002/anie.201700565.

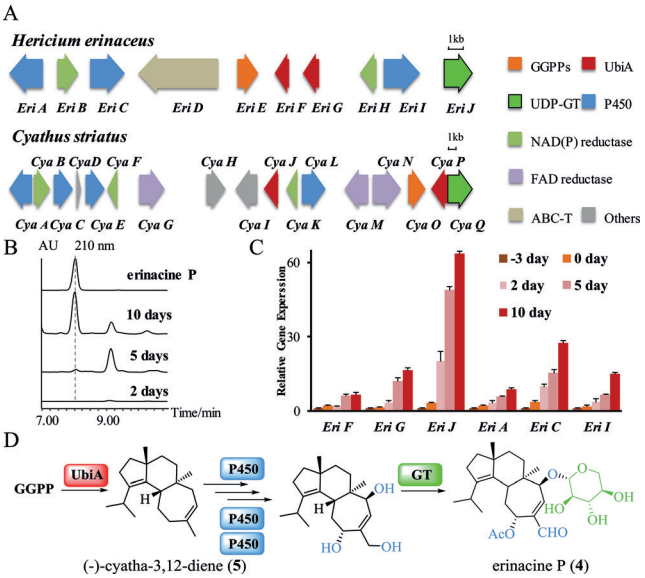


Figure 2. Identification of the gene clusters for cyathane diterpene. A) Organization and proposed function of the *Eri* and *Cya* gene clusters. B) HPLC metabolite profiling of the time-course cultures. C) qRT-PCR analysis of the putative erinacine biosynthesis genes in *H. erinaceus*. D) Proposed biosynthetic pathway for the erinacines.

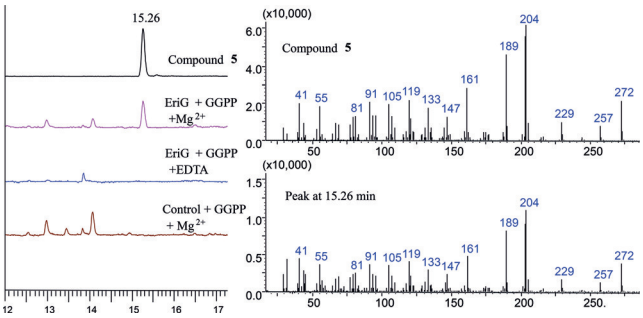


Figure 3. GC–MS results from the in vitro assay with GGPP and membrane fraction containing EriG.

cluster (named *Eri*, Figure 2A and Table S3 in the Supporting Information), which includes one GGPPS, two annotated UbiA prenyltransferases (*EriF* and *EriG*), a UDP-glycosyltransferase, and three cytochrome P450 proteins, was found in the genome of *H. erinaceum*. A similar gene cluster (named *Cya*, Figure 2A and Table S4) is present in the genome of *C. striatus*. Taking the structures of erinacines and cyathins into consideration, we proposed that these two clusters may be responsible for the biosynthesis of cyathane diterpenes. Further analysis of the secondary metabolite profile and the transcription levels of the genes (detected by quantitative RT-PCR) in the *Eri* cluster revealed that the five putative erinacine biosynthetic genes are differentially expressed under erinacine P (**4**)-producing and non-producing conditions (Figures 2B,C). Based on these results, it was proposed that the UbiA-related genes in the *Eri* and *Cya* clusters account for the cyclization of the cyathane diterpenes.

To investigate the functions of *EriF* and *EriG* in the biosynthesis of erinacines (Figure 2D), the coding regions of *EriF* and *EriG* were synthesized and cloned into a pET-28a expression vector. The resulting plasmid was expressed in *Escherichia coli* BL21 (DE3), and a membrane fraction was prepared by ultracentrifugation at 100000 *g*.[10] When incubated with GGPP and Mg^{2+}, a new product at 15.26 min (retention time) was observed for *EriG* in the GC–MS spectrum. In contrast, no products were detected with *EriF* or empty vector (Figure 3 and Figure S4). The product peak at 15.26 min gave the dominant mass fragments at *m*/*z* 272, 257, 229, 204, 203, 189, and 161, which matches the reported data for (−)-cyatha-3,12-diene (**5**).[6b] To determine the structure of the enzyme product, *EriG* was expressed in the GGPP-engineered *E. coli* BW25113 (Table S2). After 24 h, we detected a highly hydrophobic compound with the molecular mass 272 in the hexane extract by GC–MS analysis (Figure S6). Larger-scale culturing enabled us to isolate the compound at a purified titer of 3 mg L^{-1}. The compound was further confirmed to be (−)-cyatha-3,12-diene (**5**) by NMR analysis (Table S8, Figures S13, S14).[6b] As a member of the UbiA family, the activity of EriG was strictly dependent on the presence of Mg^{2+}. The addition of 2.5 mM EDTA blocked the reaction completely (Figure 3). When geranyl diphosphate (GPP) or farnesyl diphosphate (FPP) were used as substrates, no enzymatic product was detected. The apparent K_m value for recombinant EriG is 99.6 μM with GGPP as a substrate. The maximum rate of this reaction is 4.0 nmol $min^{-1} mg^{-1}$. The K_m value for EriG is comparable to the reported values for DtcycA and DtcycB from *Streptomyces* sp., Cyc1 from *Kitasatospora griseola,* and SsCPS from *Streptomyces* sp. KO-3988 (Table S13).[11] Next, we attempted to identify the functions of *CyaJ* and *CyaP* in *C. striatus*. Unfortunately, no expected product was detected in the hexane extracts of *E. coli* BW25113 expressing *CyaJ* or *CyaP*. With the strain *C. africanus*, we cloned the complete coding sequence of *CyaTC*, a homologue gene of *CyaP*, from this strain. The deduced amino acid sequence of CyaTC shows 90% identity to CyaP and 68% identity to EriG. The crude recombinant CyaTC catalyzed the formation of (−)-cyatha-3,12-diene (**5**) from GGPP (Figure S5). These results confirmed that the membrane-bound EriG and CyaTC are indeed diterpene cyclases responsible for the biosynthesis of the cyathane skeleton.

To search for potential diterpene cyclase encoding sequences in the public database, bioinformatics analysis was conducted. A total of 1769 homologue sequences with considerable similarity ($< 10^{-5}$ e-value) to EriG were identified in the UbiA superfamily. Furthermore, a recently developed sequence clustering method was applied to predict the functions of these sequences in the UbiA superfamily.[7] At an e-value cutoff of 10^{-20}, the UbiA superfamily was segregated into 12 main clusters, each containing 10 or more members (Figure 4A). EriG, FmaTC, and their 989 homologues constitute a distinct sequence cluster, named cluster 11 (Figure 4A). We thus propose that cluster 11 represents a new family of terpene cyclases in the UbiA superfamily. To clarify their phylogenetic relationship, a phylogenetic tree of sequences in cluster 11 was constructed. The sequences of EriG and FmaTC were distinguished into two clades, Clade I and Clade II (Figure 4B and Figure S1). The sequences in Clade I are distributed widely in basidiomycetes, ascomy-

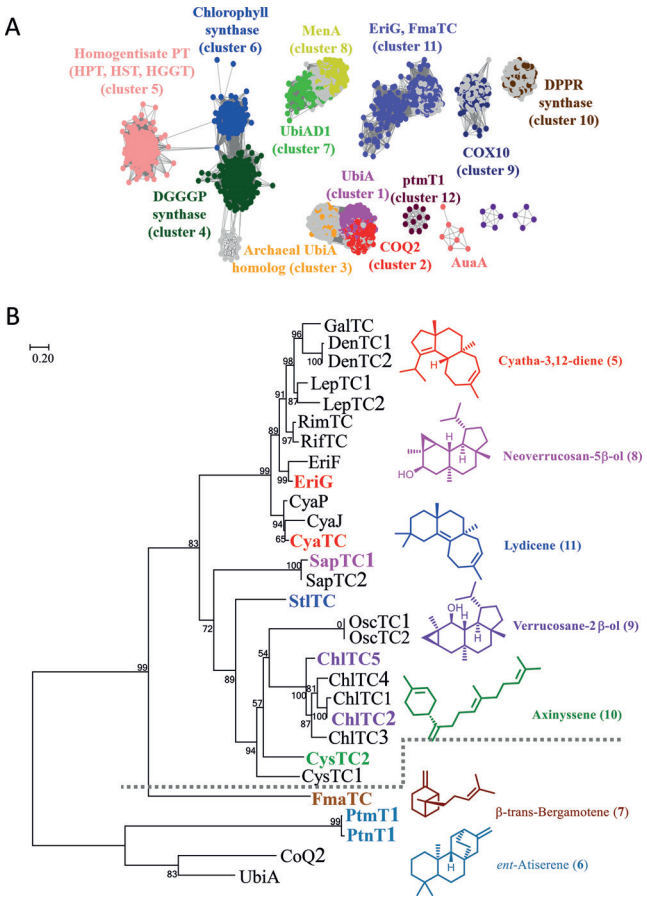


Figure 4. Sequence clustering and phylogenetic analysis of UbiA-like terpene cyclases. A) Sequence clustering of UbiA superfamily prenyltransferases. B) Phylogenetic analysis of UbiA-like terpene cyclases and the corresponding cyclized products.

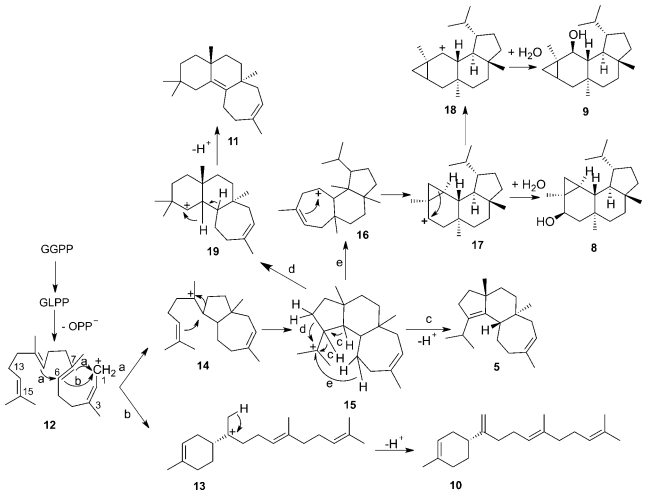
cetes, actinobacteria, cyanobacteria, and myxobacteria. Homologues of the GGPPS gene essential for the biosynthesis of diterpenes are also found in the vicinity of some *EriG*-like genes (Figure S2 and Table S5). On the other hand, diterpenes with various skeletons have been isolated from the bacteria and fungi clustered in Clade I and their related species, including verrucosane from the photosynthetic bacterium *Chloroflexus aurantiacus*,[12] neoverrucosane from the marine bacterium *Saprospira grandis*,[13] lepistal from the mushroom *Lepista sordida*.[14] We thus propose Clade I to be a new family of diterpene cyclases in bacteria and fungi.

In the genomes of *S. grandis* and *Chloroflexus* sp., genes encoding the UbiA-like proteins SapTC1 and SapTC2 (which show 26% identity to EriG at the amino acid level), and ChlTC1–ChlTC5 (which show 26–28% identity to EriG) were found. The genes and enzymes responsible for the biosynthesis of verrucosane and neoverrucosane in these bacteria remain unknown.[15] To verify the functions of *SapTC1* and *ChlTC1–ChlTC5*, these genes were synthesized and expressed in the engineered *E. coli* BW25113. Neoverrucosan-5β-ol (**8**; $R_t = 20.49$ min; Figure S7) and verrucosane-2β-ol (**9**; $R_t = 20.49$ min; Figures S8, S9) were detected in the mycelia of *E. coli* expressing the SapTC1 and ChlTC2 (ChlTC5) proteins by GC–MS and ^{1}H-NMR analysis, respectively. Furthermore, **8** was purified and confirmed to be neoverrucosan-5β-ol by 2D-NMR analysis (Table S9, Figures S15–S17).[13] The partially purified **9** was characterized by NMR, with some of its characteristic ^{1}H resonances matching the reported data for **9** [$\delta_H = 0.44$, dd ($J = 8.6$, 4.2) (H-4α); 0.18, t ($J = 4.4$) (H-4β); 3.57, d ($J = 9.5$) (H-2); 0.73 ppm, s (H-20); Table S10, Figures S18, S19].[12] The proteins ChlTC1, ChlTC3, and ChlTC4 did not function as diterpene cyclases. The gene *SapTC2* was not tested due to its 93% identity with *SapTC1*. Thus, SapTC1, ChlTC2, and ChlTC5 are concluded to be cyclases responsible for the biosynthesis of neoverrucosan-5β-ol and verrucosane-2β-ol.

To discover new diterpenes from myxobacteria and actinobacteria, we synthesized and expressed the potential diterpene cyclase genes *CysTC1* and *CysTC2* from *Cystobacter violaceus* and *StlTC* from *Streptomyces lydicus* in *E. coli* BW25113. Products for CysTC2 ($R_t = 20.49$ min; Figure S10) and StlTC ($R_t = 13.28$ min; Figure S11) were detected by GC–MS and ^{1}H-NMR analysis. No product was detected for CysTC1. After 3 L fermentation, the product of CysTC2 was isolated and determined to be axinyssene (**10**) by 2D NMR (Table S11, Figures S20–S22). Axinyssene (**10**) has been previously isolated from the Japanese marine Sponge *Axinyssa sp.*[16] The larger-scale culturing of *E. coli* BW25113 expressing *StlTC* afforded the new diterpene lydicene (**11**; Figure 4B). Compound **11** was isolated as a colorless oil. Its molecular formula was determined to be $C_{20}H_{32}$ (5 degrees of unsaturation) by the HRTOFMS data at m/z $[M+H]^+$ 273.2579 (calcd for $C_{20}H_{33}$ 273.2577). The structure of **11** was assigned by 1D and 2D NMR spectral elucidation (Table S12, Figures S12, S23–S29). Lydicene (**11**) possesses an unprecedented carbon skeleton of 2,2,4a,6a,9-pentamethyl-2,3,4,4a,5,6,6a,7,10,11-decahydro-1*H*-cyclhepta-[a]naphthalene.

Previous isotope-labeling experiments have revealed the possible cyclization mechanism for cyathane, verrucosane, and neoverrucosane diterpenes.[6a,15] In this study, we proposed that compounds **5**, **8**, **9**, and **11** share the common key intermediate **15** (Scheme 1). Compound **5** is synthesized through biosynthetic pathway c from **15**. Compounds **8** and **9** are formed via pathway e, with **16** and **17** as intermediates. A mechanism similar to that for the transformation of **17** into **18** through a cyclopropane rearrangement has been reported in the biosynthesis of cyclooctat-9-en-7-ol.[17] With **19** as the intermediate, **15** is converted into **11** by ring expansion followed by migration of the hydride at C-10. Axinyssene (**10**) is synthesized via the intermediate **13**, which is formed through a 1,6-cyclization of the pyrophosphate leaving product **12** (Scheme 1).

Diterpene cyclases are categorized into two classes, class I and class II, on the basis of their reaction mechanisms and catalytic domains. Class I cyclases have two conserved binding motifs [DDXXD and (N,D)DXX(S,T)XXXE] that directly bind the divalent metal ions to trigger the departure of the diphosphate group of substrate. Class II cyclases contain a DXDD motif that promotes carbocation formation by substrate protonation. Alignment of the amino acid sequences of the characterized diterpene cyclases with UbiA, CoQ2, and FmaTC showed the presence of two



Scheme 1. Proposed mechanism for the formation of **5** and **8–11** from GGPP as catalyzed by EriG-related diterpene cyclases.

highly conserved motifs, NXXX(G/A)XXXD and DXXXD, which correspond to motif I and motif II of UbiA (Figure S3). The NXXX(G/A)XXXD motif is similar to the NSE triad of class I terpene cyclases in bacteria and fungi, thus indicating that this motif is likely involved in binding of the Mg^{2+} cofactor.[1a] Based on the same motif and the similar catalytic mechanism between EriG and FmaTC, the EriG family was confirmed to represent class I diterpene cyclases. The catalytic mechanism and 3D structure of this new family of diterpene cyclases deserve further investigations.

In conclusion, we identified the membrane-bound diterpene cyclase EriG, which converts GGPP into (−)-cyatha-3,12-diene (**5**) in the biosynthesis of erinacines in *H. erinaceum*. This led to the discovery and characterization of a new family of widespread diterpene cyclases belonging to the UbiA superfamily. Three diterpene cyclases, SapTC1, ChlTC2, and ChlTC5, were characterized and shown to be involved in the biosynthesis of neoverrucosan-5β-ol and verrucosane-2β-ol. By expressing the potential diterpene cyclase gene *StlTC* from *S. lydicus* in engineered *E. coli*, we discovered a new diterpene with an unusual skeleton. Our study provides new opportunities for the applications of microbial diterpene synthases in biocatalysis and metabolic engineering.

Acknowledgements

This work was supported by the grants from the Ministry of Science and Technology of China (2014CB138304), the National Natural Science Foundation of China (81673334 and 21472233).

Conflict of interest

The authors declare no conflict of interest.

Keywords: biosynthesis · cyclases · diterpenes · natural products · UbiA

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 4749–4752
Angew. Chem. **2017**, *129*, 4827–4830

[1] a) J. S. Dickschat, *Nat. Prod. Rep.* **2016**, *33*, 87–110; b) M. B. Quin, C. M. Flynn, C. Schmidt-Dannert, *Nat. Prod. Rep.* **2014**, *31*, 1449–1473; c) M. J. Smanski, R. M. Peterson, S. X. Huang, B. Shen, *Curr. Opin. Chem. Biol.* **2012**, *16*, 132–141.
[2] a) J. Herrmann, A. A. Fayad, R. Muller, *Nat. Prod. Rep.* **2017**, *34*, 135–160; b) P. Manivasagan, J. Venkatesan, K. Sivakumar, S. K. Kim, *Microbiol. Res.* **2014**, *169*, 262–278.
[3] a) J. W. Shen, Y. Ruan, B. J. Ma, *J. Basic Microbiol.* **2009**, *49*, 242–255; b) J. A. Enquist Jr., B. M. Stoltz, *Nat. Prod. Rep.* **2009**, *26*, 661–680.
[4] a) H. Kawagishi, A. Shimada, R. Shirai, K. Okamoto, F. Ojima, H. Sakamoto, Y. Ishiguro, S. Furukawa, *Tetrahedron Lett.* **1994**, *35*, 1569–1572; b) T. Tsai-Teng, C. Chin-Chu, L. Li-Ya, C. Wan-Ping, L. Chung-Kuang, S. Chien-Chang, H. F. Chi-Ying, C. Chien-Chih, Y. J. Shiao, *J. Biomed. Sci.* **2016**, *23*, 1–12; c) H. C. Kuo, C. C. Lu, C. H. Shen, S. Y. Tung, M. C. Hsieh, K. C. Lee, L. Y. Lee, C. C. Chen, C. C. Teng, W. S. Huang, T. C. Chen, K. F. Lee, *J. Transl. Med.* **2016**, *14*, 1–14.
[5] a) L. W. He, J. J. Han, B. W. Li, L. Huang, K. Ma, Q. Chen, X. Z. Liu, L. Bao, H. W. Liu, *Eur. J. Med. Chem.* **2016**, *111*, 183–192; b) L. Huang, J. J. Han, D. Ben-Hail, L. W. He, B. W. Li, Z. H. Chen, Y. Y. Wang, Y. L. Yang, L. Liu, Y. S. Zhu, V. Shoshan-Barmatz, H. W. Liu, Q. Chen, *J. Biol. Chem.* **2015**, *290*, 23563–23578.
[6] a) W. A. Ayer, S. P. Lee, T. T. Nakashima, *Can. J. Chem.* **1979**, *57*, 3338–3343; b) H. Kenmoku, N. Kato, M. Shimada, M. Omoto, A. Mori, W. Mitsuhashi, T. Sassa, *Tetrahedron Lett.* **2001**, *42*, 7439–7442; c) T. Anke, U. Rabe, P. Schu, T. Eizenhofer, M. Schrage, W. Steglich, *Z. Naturforsch. C* **2002**, *57*, 263–271.
[7] W. K. Li, *Trends Biochem. Sci.* **2016**, *41*, 356–370.
[8] M. J. Smanski, Z. G. Yu, J. Casper, S. J. Lin, R. M. Peterson, Y. H. Chen, E. Wendt-Pienkowski, S. R. Rajski, B. Shen, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13498–13503.
[9] H. C. Lin, Y. H. Chooi, S. Dhingra, W. Xu, A. M. Calvo, Y. Tang, *J. Am. Chem. Soc.* **2013**, *135*, 4616–4619.
[10] a) P. Zeyhle, J. S. Bauer, M. Steimle, F. Leipoldt, M. Rosch, J. Kalinowski, H. Gross, L. Heide, *ChemBioChem* **2014**, *15*, 2385–2392; b) P. Zeyhle, J. S. Bauer, J. Kalinowski, K. Shin-ya, H. Gross, L. Heide, *Plos One* **2014**, *9*, e99122.
[11] A. Meguro, T. Tomita, M. Nishiyama, T. Kuzuyama, *ChemBioChem* **2013**, *14*, 316–321.
[12] J. Hefter, H. H. Richnow, U. Fischer, J. M. Trendel, W. Michaelis, *J. Gen. Microbiol.* **1993**, *139*, 2757–2761.
[13] A. Spyere, D. C. Rowley, P. R. Jensen, W. Fenical, *J. Nat. Prod.* **2003**, *66*, 818–822.
[14] X. Mazur, U. Becker, T. Anke, O. Sterner, *Phytochemistry* **1996**, *43*, 405–407.
[15] a) C. Rieder, G. Strauss, G. Fuchs, D. Arigoni, A. Bacher, W. Eisenreich, *J. Biol. Chem.* **1998**, *273*, 18099–18108; b) W. Eisenreich, C. Rieder, C. Grammes, G. Hessler, K. P. Adam, H. Becker, D. Arigoni, A. Bacher, *J. Biol. Chem.* **1999**, *274*, 36312–36320.
[16] K. Kodama, R. Higuchi, T. Miyamoto, R. W. M. Van Soest, *Org. Lett.* **2003**, *5*, 169–171.
[17] A. Meguro, Y. Motoyoshi, K. Teramoto, S. Ueda, Y. Totsuka, Y. Ando, T. Tomita, S.-Y. Kim, T. Kimura, M. Igarashi, R. Sawa, T. Shinada, M. Nishiyama, T. Kuzuyama, *Angew. Chem. Int. Ed.* **2015**, *54*, 4353–4356; *Angew. Chem.* **2015**, *127*, 4427–4430.

Manuscript received: January 17, 2017
Revised: March 5, 2017
Final Article published: March 28, 2017

