

Natural Products

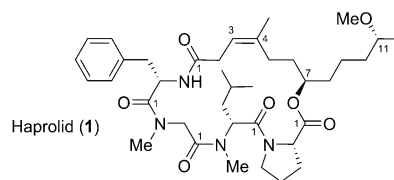
International Edition: DOI: 10.1002/anie.201603288
German Edition: DOI: 10.1002/ange.201603288

Isolation, Structure Elucidation, and (Bio)Synthesis of Haprolid, a Cell-Type-Specific Myxobacterial Cytotoxin

Heinrich Steinmetz, Jun Li, Chengzhang Fu, Nestor Zaburannyi, Birgitte Kunze, Kirsten Harmrolfs, Viktoria Schmitt, Jennifer Herrmann, Hans Reichenbach, Gerhard Höfle, Markus Kalesse,* and Rolf Müller*

Abstract: Myxobacteria are well-established sources for novel natural products exhibiting intriguing bioactivities. We here report on haprolid (**1**) isolated from *Byssovorax cruenta* Har1. The compound exhibits an unprecedented macrolactone comprising four modified amino acids and a polyketide fragment. As configurational assignment proved difficult, a bioinformatic analysis of the biosynthetic gene cluster was chosen to predict the configuration of each stereocenter. In-depth analysis of the corresponding biosynthetic proteins established a hybrid polyketide synthase/nonribosomal peptide synthetase origin of haprolid and allowed for stereochemical assignments. A subsequent total synthesis yielded haprolid and corroborated all predictions made. Intriguingly, haprolid showed cytotoxicity against several cell lines in the nanomolar range whereas other cells were almost unaffected by treatment with the compound.

Myxobacteria have shown a remarkable capacity to produce structurally diverse natural compounds exhibiting a broad spectrum of biological activities.^[1] During a screening campaign for novel metabolites, haprolid (**1**; Figure 1) was isolated from the myxobacterium *Byssovorax cruenta* strain Har1 at the HZI already in 2000 and its constitution was

Figure 1. Haprolid (**1**).

determined without defining the configuration.^[1d] We here report on the structure, the determination of its stereochemistry, the biosynthesis, and the synthesis of this compound and show haprolid to exhibit intriguing and selective cytotoxic activity against eukaryotic cells.

Haprolid (**1**) was obtained as a white powder and was assigned a molecular formula of $C_{38}H_{58}N_4O_7$ on the basis of HR-ESI-MS data. Analysis of the 1H and ^{13}C NMR spectra (Table S1) revealed the presence of one N-H amide proton ($\delta(H) = 8.28$ ppm), seven methyl groups (including one *O*-methyl and two *N*-methyl groups), one phenyl ring, seven methine groups (five of which were bound to heteroatoms), nine methylene groups, and five carbonyl-type carbon atoms. Interpretation of the 1H , ^{13}C , and 2D NMR data suggested a peptidic segment linked to a lipophilic C_{12} chain. The detailed analysis of COSY, HMQC, and HMBC correlations revealed one proline unit and three additional amino acids including NMe-leucine (*N*-Me-Leu), sarcosine (Sarc), and phenylalanine (Phe) identifying haprolid (**1**) as partially constructed as a tetrapeptide (Figures S4–S7). Specifically, the loss of Pro ($m/z = 115$), *N*-Me-Leu ($m/z = 127$), and Sarc ($m/z = 71$) observed in the ESI-MS/MS spectrum (Figure S11) confirmed the amino acid sequence. Attachment of the polyketide chain to Phe was confirmed by HMBC correlations between the Phe- α -Proton and C-1 of the polyketide chain. The constitution of the lipophilic dodecyl chain, including the position of unsaturation (C-3/4), was elucidated by combining the analyses of COSY and HMBC correlations. The geometry of the double bond between C-3 and C-4 was assigned as *Z* based on the strong ROESY correlation between H-3 and Me at C-4, and further supported by ROESY correlation between H-2b ($\delta(H) = 2.64$ ppm) and H-5b ($\delta(H) = 2.20$ – 2.14 ppm) (Figures S8 and S9). The cyclic structure of haprolid (**1**) was implied by strong HMBC correlation between H-7 ($\delta(H) = 4.62$ ppm) and Pro C-1. The absolute configuration of the amino acids was determined using a chemical degradation/derivatization procedure^[2] indicating the presence of L-Pro, D-NMe-Leu, and L-Phe.

[*] H. Steinmetz, Dr. B. Kunze
Helmholtz Center for Infection Research (HZI)
Department Microbial Drugs
Inhoffenstrasse 7, 38124 Braunschweig (Germany)

Dr. J. Li, Prof. Dr. M. Kalesse
Institute for Organic Chemistry
Gottfried Leibniz Universität Hannover
and
Centre of Biomolecular Drug Research (BMWZ)
Schneiderberg 1B, 30167 Hannover (Germany)
E-mail: markus.kalesse@oci.uni-hannover.de

Dr. J. Li, Prof. Dr. H. Reichenbach, Prof. Dr. G. Höfle,
Prof. Dr. M. Kalesse
Helmholtz Center for Infection Research (HZI)
Inhoffenstrasse 7, 38124 Braunschweig (Germany)

Dr. C. Fu, Dr. N. Zaburannyi, Dr. K. Harmrolfs, V. Schmitt,
Dr. J. Herrmann, Prof. Dr. R. Müller
Helmholtz Institute for Pharmaceutical Research Saarland (HIPS),
Helmholtz Center for Infection Research and Pharmaceutical
Biotechnology
Saarland University, Campus Building E8.1
66123 Saarbrücken (Germany)
E-mail: Rolf.Mueller@helmholtz-hzi.de

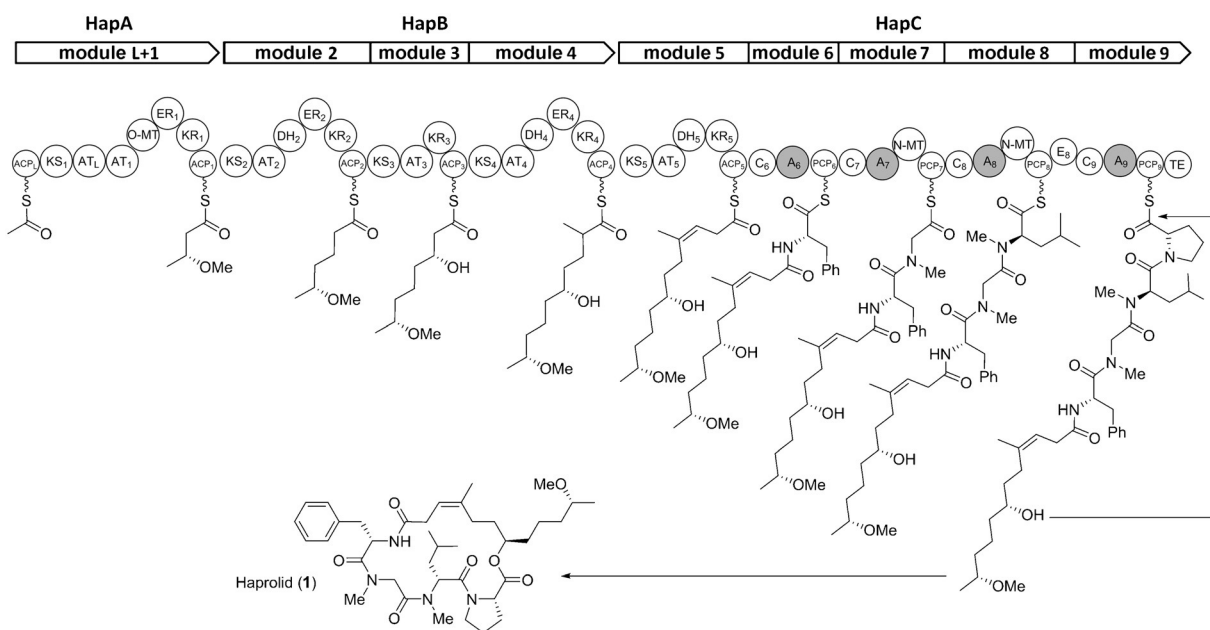
Supporting information and the ORCID identification number(s) for the article(s) of this article can be found under <http://dx.doi.org/10.1002/anie.201603288>.

As the configurational assignment of the secondary alcohols in the polyketide side chain proved to be challenging, we decided to perform a bioinformatic ketoreductase analysis,^[3,4] for which the knowledge of the biosynthetic gene cluster sequence was required. To identify the biosynthetic gene cluster of haprolid the genome of the producer strain *Byssovorax cruenta* Har1 was sequenced with the aid of the Illumina sequencing technology and a draft genome comprising ten scaffolds was achieved. According to the chemical structure of haprolid, the biosynthetic gene cluster is supposed to include a hybrid polyketide synthase/nonribosomal peptide synthetase (PKS-NRPS) assembly line. After analyzing all predicted PKS-NRPS hybrid gene clusters by anti-SMASH analysis,^[5] we found a candidate gene cluster with two gaps which is in line with the “retro-biosynthesis” analysis of haprolid. The gaps of this gene cluster were closed by sequencing of the polymerase chain reaction (PCR)-amplified DNA fragments covering the gaps (Table S2). The function of every gene in the predicted gene cluster was analyzed by BLAST, indicating three giant genes (*hapA*, *hapB*, and *hapC*) which encode two PKS megaenzymes and one hybrid PKS-NRPS multienzyme complex. The detailed annotations of the upstream and downstream genes of *hapA*–*hapC* indicated that none of these adjacent genes have any obvious function in haprolid biosynthesis. Haprolid is synthesized starting from the multienzyme HapA containing a mixed-loading and chain-extending module frequently observed in biosynthesis of myxobacterial natural products like soraphen,^[6] melithiazol,^[7] thuggacin,^[8] and disciformycin.^[9] The polyketide chain is extended by three modules of the megaenzyme HapB. The hybrid PKS-NRPS protein HapC containing another PKS module and four NRPS modules not only extends the polyketide chain but also adds four amino acid residues.

To further confirm the functionalities encoded in the biosynthetic gene cluster, all the six acyltransferase (AT)

domains and four adenylation (A) domains were analyzed for substrate-activating specificity. The ATs were compared to the crystallized malonyl-CoA:acyl carrier protein transacylase from *Escherichia coli* (Table S3, Figure S12).^[10] The loading acyltransferase (AT_L) domain which is expected to transfer acetyl-CoA onto ACP_L, surprisingly, bears the arginine residue (R117), which is thought to be important for the activation of dicarboxylic acid extender units (Figure S12).^[11] However, there are already reported cases of this protein modification carrying an arginine residue which can shuttle either acetyl-CoA or malonyl-CoA.^[11,12] The substrate-specificity analysis of these AT domains revealed that the collinear assembly line loads acetyl-CoA as a starter unit, and then extends the polyketide chain by three subsequent malonyl-CoAs, followed by one methylmalonyl-CoA, and another malonyl-CoA incorporation, which is consistent with the haprolid chemical structure (Scheme 1, Figure S12, and Table S3). The A domains were analyzed for their putative amino acid specificity (Table S4, Figure S13)^[13] which was in agreement with the haprolid structure (Scheme 1).

Based on the biosynthetic analysis, we proposed the absolute configurations of all chiral centers by analyzing ketoreductase (KR) domains, adenylation (A) domains, and condensation (C) domains of the Hap cluster.^[14] Configurational analyses of amino acid residues were in agreement with the results of chemical analysis (vide supra). The absolute configuration of both oxymethine groups in the polyketide chain was indicated by comparing the specific motif difference between the A-type and B-type stereospecificity KR domains (Figure S14).^[3,15] In addition, our statistical analysis of both ketoreductases gave a ScoreDiff value of 27.64 for the alcohol at C-18 and 47.39 for the C-22 methoxy group.^[3d] Both oxy functionalities at C-18 and at C-22 were therefore predicted to be D-configured as they are generated by B-type KR domains (KR1, KR3). This is only the second time that the fidelity of a biosynthetic structure prediction has been



Scheme 1. Biosynthesis of haprolid. For details see text.

quantified using the absolute value of a statistical analysis. Formation of the *Z*-configured double bond in haprolid is supposed to be catalyzed by module 5 which contains a B-type KR domain. Although most *cis* alkenes in polyketides are formed by dehydratase (DH) domain catalyzed elimination of water from (3*S*)-hydroxylacyl intermediates, which are in turn generated by A-type KR domains,^[16] exceptions to this rule were discovered such as the B-type KR domains found to be involved in the formation of *cis* double bonds in the biosyntheses of borrelidin,^[17] chivosazol,^[4b] difficidin,^[18] and mupirocin.^[19]

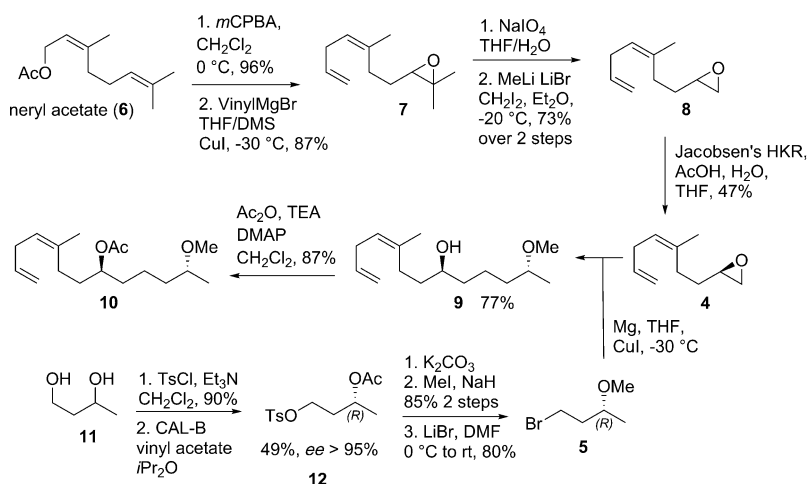
The combination of the C-domain-type determinations with the substrate-specificity prediction of the A domains leads to collinear biosynthesis, which starts from the activation of one L-Phe by domain A6 and the linking of the upstream polyketide acyl chain to the PCP-bound L-Phe by C6. Next, Gly is activated, incorporated into the growing hybrid PK-NRP chain, and N-methylated by the *N*-MT domain of module 7 followed by L-Leu insertion, *N*-methylation, and epimerization. Subsequently, the C9 domain acts as a ³C_L-type condensation domain constructing the D–L peptide linkage between the D-Leu and the L-Pro incorporated by module 9 (Scheme 1). Finally, the TE domain of module 9 performs the intramolecular macrocyclization yielding haprolid.

Based on the structural assignment we devised a retrosynthetic disconnection that joined the peptidic and polyketide portion through a peptide bond formation and either a Shiina^[20] or a Mitsunobu^[21] lactonization, potentially diverting one synthetic route into two diastereomers at the final stage of the synthesis.

In order to keep the synthesis as efficient as possible we started with neryl acetate, which had to be converted to segment **2** (Figure 2).^[22] The major challenges of this synthesis were the *Z*-configured double bond that had to be kept out of conjugation with respect to the acid moiety and the construction of two relatively isolated secondary alcohols. To overcome the first challenge we used a terminal olefin as a replacement for the acid moiety. Liberation of the acid was later achieved with Strukul's

catalyst which allowed selective epoxidation of the terminal double bond without affecting the *Z*-configured internal olefin. The construction of the secondary alcohols was achieved through enzymatic kinetic resolution and selective epoxide opening. For the transformation from **7** to **8**, an improved protocol for generating thermally stable iodomethyl lithium was employed which traces back to the seminal work of Köbrich.^[23]

The synthesis commenced with neryl acetate (Scheme 2), which was treated with *m*CPBA leading to epoxidation of the terminal double bond^[24] followed by treatment with vinylmagnesium bromide in the presence of copper(I) to provide compound **7**.^[25] The introduction of the monosubstituted double bond as a synthon for the carboxylate was the result of experimentation with other functional groups that failed to provide polyketide **2**. Consequently, compound **7** was first converted to the corresponding aldehyde by treatment with NaIO₄ and then treated with iodomethyl lithium generated in situ from CH₂I₂ and MeLi in the presence of LiBr.^[26] This improved protocol for using α -halogenated organolithium compounds (Köbrich reagents)^[27] allows reactions to take place at moderate temperatures (–20 to 0 °C) without



Scheme 2. Synthesis of segment **10**. CAL-B = lipase B from *Candida antarctica*, *m*CPBA = *m*-chloroperoxybenzoic acid, DMAP = 4-dimethylaminopyridine, DMS = dimethyl sulfide, HKR = hydrolytic kinetic resolution, TEA = triethylamine, Ts = tosylate.

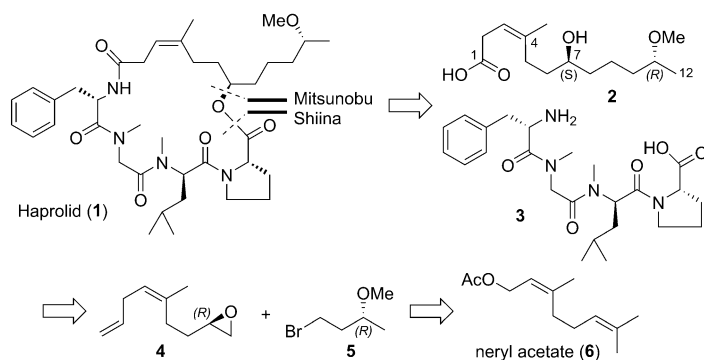
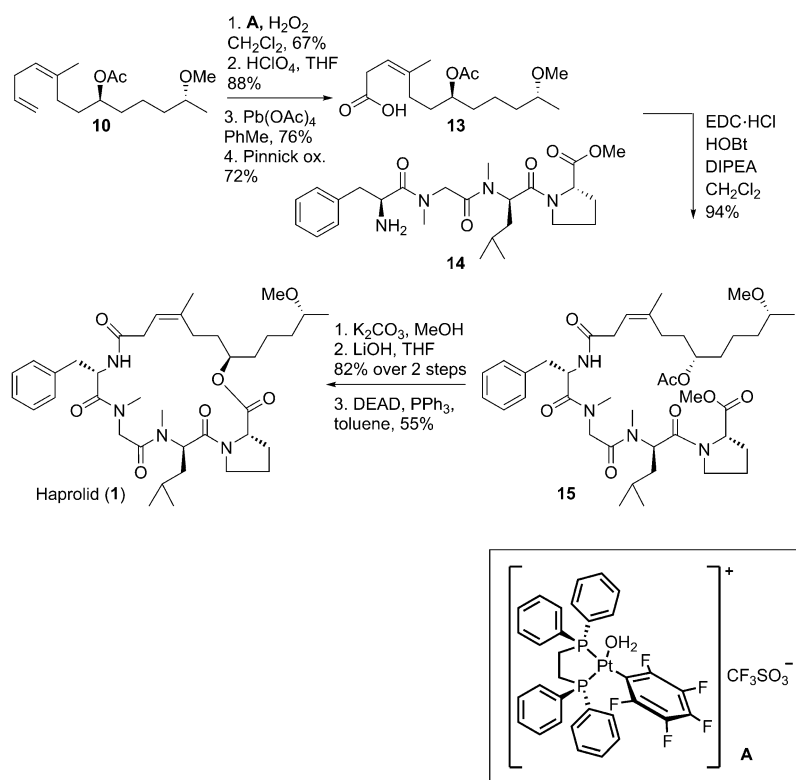


Figure 2. Haprolid and its retrosynthetic disconnection.

decomposition of the reagent. The resulting epoxide **8** was subjected to Jacobsen's hydrolytic kinetic resolution using the chiral (*R,R*)(salen)Co^{III} complex.^[28] With chiral epoxide **4** in hands treatment with the Grignard reagent derived from **5** provided the polyketide skeleton.

Transformation of the terminal double bond to the carboxylate was achieved by a four-step sequence involving an epoxidation of the terminal double bond with the platinum catalyst reported by Strukul et al (Scheme 3).^[29] Condensation with the corresponding tripeptide was done under standard peptide coupling conditions and generated compound **15**. Hydrolysis of both the acetate moiety and the methyl ester provided the seco acid, which was subjected to the Mitsunobu



Scheme 3. Endgame of the haprolid (**1**) synthesis. DEAD = diethyl azodicarboxylate, DIPEA = *N,N*-diisopropylethylamine, EDC = 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, HOBt = hydroxybenzotriazole.

protocol to provide haprolid in 55 % yield. All spectroscopic data of the synthetic material were in full agreement with those of the authentic sample (information on the organic synthesis in the Supporting Information).

Haprolid (natural and synthetic) was tested on a small panel of human cancer cell lines (Table 1). Intriguingly, the compound was highly active (IC_{50} values in low to mid nanomolar range) on the cervix carcinoma cells lines Hep2 and KB-3.1 (HeLa derivatives), and THP-1 and HL-60 leukemia cell lines. The IC_{50} values on Huh7.5 hepatocarcinoma and U-2 OS osteosarcoma cells were in the high nanomolar range. In contrast, haprolid was not active on A-549 lung carcinoma and SK-OV-3 ovarian adenocarcinoma cells. The reason for the selective cytotoxicity is currently unknown as is also the molecular target of the compound. We

Table 1: Cytotoxic activity of haprolid.

Cell line (accession no.)	IC ₅₀ [μM]	
	1 (from <i>B. cruenta</i>)	1 (synthetic)
A-549 (ACC-107)	> 100	> 100
Hep2 (CCL-240)	0.0042 ± 0.0032	0.01
HL-60 (ACC-3)	0.10 ± 0.11	n.d.
Huh7.5 (APC49)	0.29 ± 0.23	0.37
KB-3.1 (ACC-16)	0.014 ± 0.002	n.d.
SK-OV-3 (HTB-77)	> 100	> 100
THP-1 (ACC-16)	0.025 ± 0.014	0.06
U-2 OS (ACC-785)	1.73 ± 0.64	n.d.

did not observe specific effects on the cytoskeleton (actin, tubulin) or certain hallmarks of apoptosis, such as mitochondrial membrane depolarization, lysosomal rupture, and fragmentation of nuclei in KB-3.1 cells that were treated with haprolid (Figure S15). Future biochemical studies will address the intriguing question of the mode of action of the haprolid class of natural products.

In summary, we have isolated a new secondary metabolite from myxobacteria, completed its biosynthetic analysis, and verified its structure by total synthesis. This biosynthetic structure elucidation took advantage of established protocols and a statistical refinement from which a quantitative assessment of the structure prediction was possible. The synthesis features under-represented protocols for the transformation of epoxides, namely the use of Strukul's catalyst and a thermally stable iodomethylolithium reagent. Finally, the remarkable biological activities in combination with its short and efficient synthesis (12 steps in the longest linear sequence) provide the background for potential biological applications of this metabolite.

Keywords: cell-line-selective cytotoxin · haprolid · polyketide (bio)synthesis · retrosynthesis · structure elucidation

How to cite: *Angew. Chem. Int. Ed.* **2016**, 55, 10113–10117
Angew. Chem. **2016**, 128, 10267–10271

- [1] a) H. Reichenbach, G. Höfle in *Drug Discovery from Nature* (Eds.: S. Grabley, R. Thiericke), Springer, Berlin, **1999**; b) K. J. Weissman, R. Müller, *Nat. Prod. Rep.* **2010**, 27, 1276–1295; c) T. F. Schäberle, F. Lohr, A. Schmitz, G. M. König, *Nat. Prod. Rep.* **2014**, 31, 953–972.
- [2] I. Abe, S. Kuramoto, S. Musha, *J. High Resolut. Chromatogr.* **1983**, 6, 366–370.
- [3] a) P. Caffrey, *ChemBioChem* **2003**, 4, 654–657; b) P. Caffrey, *Chem. Biol.* **2005**, 12, 1060–1062; c) R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry* **2003**, 42, 72–79; d) A. Kitsche, M. Kalesse, *ChemBioChem* **2013**, 14, 851–861; e) The online profile Hidden Markov Model analysis tool can be found at: https://akitsche.shinyapps.io/profileHMM_App.
- [4] a) S. C. Wenzel, R. Müller, *Nat. Prod. Rep.* **2009**, 26, 1385–1407; b) O. Perlova, K. Gerth, O. Kaiser, A. Hans, R. Müller, *J. Biotechnol.* **2006**, 121, 174–191.
- [5] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, B. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, et al., *Nucleic Acids Res.* **2015**, 43, W237–43.
- [6] C. J. Wilkinson, E. J. Frost, J. Staunton, P. F. Leadlay, *Chem. Biol.* **2001**, 8, 1197–1208.
- [7] S. Weinig, H.-J. Hecht, T. Mahmud, R. Müller, *Chem. Biol.* **2003**, 10, 939–952.
- [8] K. Buntin, H. Irschik, K. J. Weissman, E. Luxenburger, H. Blöcker, R. Müller, *Chem. Biol.* **2010**, 17, 342–356.
- [9] F. Surup, K. Viehrig, K. I. Mohr, J. Herrmann, R. Jansen, R. Müller, *Angew. Chem. Int. Ed.* **2014**, 53, 13588–13591; *Angew. Chem.* **2014**, 126, 13806–13809.

- [10] G. Yadav, R. S. Gokhale, D. Mohanty, *J. Mol. Biol.* **2003**, 328, 335–363.
- [11] L. Serre, E. C. Verbree, Z. Dauter, A. R. Stuitje, Z. S. Derewenda, *J. Biol. Chem.* **1995**, 270, 12961–12964.
- [12] S. F. Haydock, J. F. Aparicio, I. Molnár, T. Schwecke, L. E. Khaw, A. König, A. F. Marsden, I. S. Galloway, J. Staunton, P. F. Leadlay, *FEBS Lett.* **1995**, 374, 246–248.
- [13] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, 6, 493–505.
- [14] For other examples of structure prediction based on the biosynthetic analysis see: a) O. Hartmann, M. Kalesse, *Angew. Chem. Int. Ed.* **2014**, 53, 7335–7338; *Angew. Chem.* **2014**, 126, 7463–7466; b) C. Jahns, T. Hoffmann, S. Müller, K. Gerth, P. Washausen, G. Höfle, H. Reichenbach, M. Kalesse, T. Müller, *Angew. Chem. Int. Ed.* **2012**, 51, 5239–5243; *Angew. Chem.* **2012**, 124, 5330–5334; c) S. Essig, S. Bretzke, R. Müller, D. Menche, *J. Am. Chem. Soc.* **2012**, 134, 19362–19365; d) T. Brodmann, D. Janssen, M. Kalesse, *J. Am. Chem. Soc.* **2010**, 132, 13610–13611; e) M. Bock, K. Buntin, R. Müller, A. Kirschning, *Angew. Chem. Int. Ed.* **2008**, 47, 2308–2311; *Angew. Chem.* **2008**, 120, 2341–2344; f) D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik, R. Müller, *J. Am. Chem. Soc.* **2008**, 130, 14234–14243; g) D. Janssen, D. Albert, R. Jansen, M. Müller, M. Kalesse, *Angew. Chem. Int. Ed.* **2007**, 46, 4898–4901; *Angew. Chem.* **2007**, 119, 4985–4988.
- [15] A. P. Siskos, A. Baerga-Ortiz, S. Bali, V. Stein, H. Mamdani, D. Spiteller, B. Popovic, J. B. Spencer, J. Staunton, K. J. Weissman, et al., *Chem. Biol.* **2005**, 12, 1145–1153.
- [16] D. H. Kwan, F. Schulz, *Molecules* **2011**, 16, 6092–6115.
- [17] C. Olano, B. Wilkinson, C. Sanchez, S. J. Moss, R. Sheridan, V. Math, A. J. Weston, A. F. Brana, C. J. Martin, M. Oliynyk, et al., *Chem. Biol.* **2004**, 11, 87–97.
- [18] X. H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsis, G. Hitzeroth, N. Grammel, A. W. Strittmatter, et al., *J. Bacteriol.* **2006**, 188, 4024–4036.
- [19] A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas, *Chem. Biol.* **2003**, 10, 419–430.
- [20] I. Shiina, M. Kubota, H. Oshiumi, M. Hashizume, *J. Org. Chem.* **2004**, 69, 1822–1830.
- [21] O. Mitsunobu, *Synthesis* **1981**, 1–28.
- [22] For the use of terpenes in polyketide synthesis see: a) J. Swatschek, L. Grothues, J. O. Bauer, C. Strohmann, M. Christmann, *J. Org. Chem.* **2014**, 79, 976–983; b) P. Winter, J. Swatschek, M. Willot, L. Radtke, T. Olbrisch, A. Schäfer, M. Christmann, *Chem. Commun.* **2011**, 47, 12200–12202.
- [23] G. Köbrich, *Angew. Chem. Int. Ed. Engl.* **1967**, 6, 41–52; *Angew. Chem.* **1967**, 79, 15–27.
- [24] J. Germain, P. Deslongchamps, *J. Org. Chem.* **2002**, 67, 5269–5278.
- [25] P. Winter, C. Vaxelaire, C. Heinz, M. Christmann, *Chem. Commun.* **2011**, 47, 394–396.
- [26] J. M. Concellón, H. Cuervo, R. Fernández-Fano, *Tetrahedron* **2001**, 57, 8983–8987.
- [27] J. Aubé, *Comprehensive Organic Synthesis* B. M. Trost, I. Fleming, Eds. Pergamon: New York, **1999**, 1, p. 260.
- [28] a) M. Tokunaga, J. F. Larrow, F. Kakiuchi, E. N. Jacobsen, *Science* **1997**, 277, 936; b) L. P. C. Nielsen, C. P. Stevenson, D. G. Blackmond, E. N. Jacobsen, *J. Am. Chem. Soc.* **2004**, 126, 1360–1362.
- [29] M. Colladon, A. Scarso, P. Sgarbossa, R. A. Michelin, G. Strukul, *J. Am. Chem. Soc.* **2006**, 128, 14006–14007.

Received: April 4, 2016

Published online: July 12, 2016