Biosynthetic Gene Cluster of linaridin Peptides Contains Epimerase Gene

Wanlu Xiao,[a] Yasuharu Satoh,[b] Yasushi Ogasawara,*[b] and Tohru Dairi*[b]

[a]Graduate School of Chemical Sciences and Engineering and [b]Graduate School of Engineering, Hokkaido University

N13 W8, Kita-ku, Sapporo, Hokkaido 060-8628 (Japan)

*E-mail: yogasawa@eng.hokudai.ac.jp and dairi@eng.hokudai.ac.jp

Keywords: biosynthesis • D-amino acid • epimerase • linaridin • peptides

Abstract: Salinipeptins belong to the type-A linaridin class of ribosomally synthesized and post-translationally modified peptides (RiPPs) comprising 22 amino acid residues with multiple D-amino acids. Although chirality of other type-A linaridins, such as grisemycin and cypemycin, has not been reported, the biosynthetic gene clusters of type-A linaridins have an identical gene organization. Here, we report heterologous expression of the grisemycin biosynthetic gene cluster (*grm*) and show that grisemycin contains multiple D-amino acids, similar to salinipeptins. The heterologous expression experiments also confirmed involvement of a novel peptide epimerase in grisemycin biosynthesis. Gene-deletion experiments indicated that *grmL*, a sole gene of unknown function, was indispensable for grisemycin production. We also show that the presence of D-amino acids is likely a common feature of linaridin natural products by analyzing two other type-A linaridin clusters.

While most naturally occurring peptides consist of L-amino acids, D-amino acids are also widely identified both in primary and secondary metabolites. For example, peptidoglycan, a component of bacterial cell walls, contains D-Glu and D-Ala, which confer resistance to proteases.[1] In addition, vancomycin, gramicidins, and nisin are prominent examples of peptide natural products containing D-amino acids.[2] In peptide natural products, the presence of D-amino acids is generally a hallmark of peptides biosynthesized via non-ribosomal peptide synthetases, modular-type large enzyme complexes with multiple catalytic domains.[3] In this machinery, the epimerization domain specifically introduces D-amino acids into the growing peptide chain by epimerization at the α-carbon of a thioester.[4] By contrast, the presence of D-amino acids is rare in ribosomally synthesized and post-translationally modified peptides (RiPPs), because post-translational epimerization requires abstraction of an α-proton of an amide, which is less acidic than those of thioesters. To date, only three types of peptide epimerases have been identified in RiPP biosynthesis: radical *S*-adenosylmethionine (SAM)-dependent enzymes (PoyD, YydG),[5] α/β-hydrolase family enzyme (BotH)[6] that catalyzes epimerization of an Asp/Glu residue adjacent to a thiazolinyl group, and a C-terminal Trp epimerase (MslH)[7] involved in the biosynthesis of lasso peptide MS-271. In addition, a two-step process to introduce D-Ala from L-Ser via dehydroalanine was reported for lanthipeptide biosynthesis.[8]

Salinipeptins, produced by halotolerant *Streptomyces* sp. GSL-6C, are members of the linaridin class of natural products containing multiple D-amino acids (Figure 1).[9] Linaridins, <u>lin</u>ear peptides containing dehydrated (<u>arid</u>) amino acids, are a small but emerging group of RiPPs. In addition to salinipeptins, four linaridin family peptides have been characterized to date, namely grisemycin,[10] cypemycin,[11] legonaridin,[12] and cacaoidin.[13] Among these, salinipeptins, grisemycin, and cypemycin contain an N-terminal *N,N*-dimethylalanine moiety and a C-terminal aminovinyl-cysteine (aviCys) moiety and are categorized into the type-A subfamily. Analysis of the salinipeptin biosynthetic gene cluster (*sin*) revealed genes encoding a precursor

peptide (SinA), dehydratase (SinH), decarboxylase (SinD), methyltransferase (SinM), transporters (SinT and SinP), and a protein of unknown function (SinL). Although salinipeptin possesses D-amino acid residues that were introduced via epimerization at C_α of amino acid residues of the precursor peptide, the *sin* cluster does not contain an obvious candidate gene for the epimerization.[9] Considering that SinH, SinD, and SinM are required for the formation of the dehydrobutyrine (Dhb), aviCys, and *N*,*N*-dimethylalanine moieties, respectively,[9, 14] the sole function of unknown protein SinL may be involved in the epimerization during salinipeptin biosynthesis. Interestingly, bioinformatics analysis revealed that gene clusters with similar gene organization were widely distributed in many different bacterial strains and that biosynthetic gene clusters of grisemycin (*grm*) and cypemycin (*cyp*) had orthologs organized similar to those of the *sin* cluster (Figure 1). While the absolute stereochemistry of amino acid residues in grisemycin and cypemycin has not been reported, cypemycin contained allo-Ile, an epimer of Ile, and we thus speculated that the presence of D-amino acid residues is a common feature of type-A linaridin natural products and that the allo-Ile residues in cypemycin are biosynthesized via epimerization at C_α of the Ile residues.

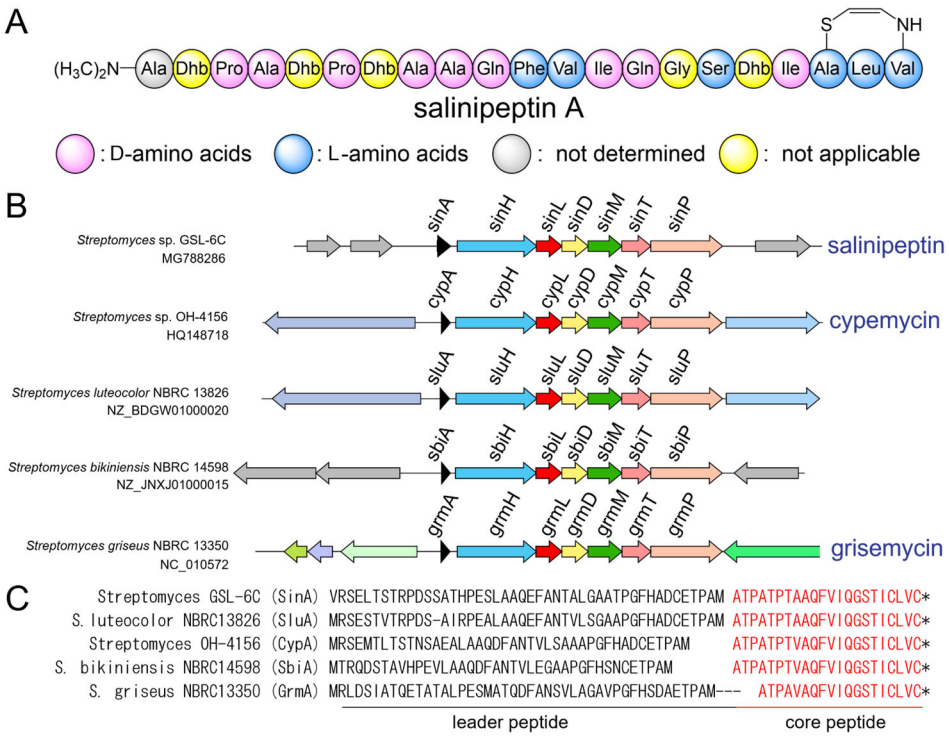

Figure 1. (A) Structure of salinipeptin A, (B) organization of the type-A linaridin biosynthetic gene clusters, and (C) deduced amino acid sequences of precursor peptides.

To test our hypothesis, we performed heterologous expression experiments of type-A linaridin clusters and examined whether the metabolites contained D-amino acid residues. Because we could not obtain producing strains of salinipeptin (*Streptomyces* sp. GSL-6C) and cypemycin (*Streptomyces* sp. OH4156), we first examined the grisemycin cluster. The *grm* cluster from *Streptomyces griseus* NBRC 13350, spanning a region of approximately 7 kbp from *grmA* to *grmP*, was cloned into pWHM3[15] to generate plasmid pWHM3-grm, which was expressed in *Streptomyces lividans* TK23. After cultivation of the resulting transformants, metabolites were extracted from 50 mL of culture broth with chloroform (50 mL × 3), concentrated *in vacuo*, and re-dissolved in 0.1% aqueous formic acid (500 μL). LC-MS analysis of the chloroform extract revealed the production of specific metabolite **1** at a retention time of 22.2 min (Figure 2, trace b). High-resolution electrospray ionization (HR-ESI) MS and MS/MS analysis of **1** showed that a doubly protonated molecular ion at *m*/*z* 917.0086 ($C_{86}H_{137}N_{21}O_{21}S$: *m*/*z* 917.0082 $[M+2H]^{2+}$) and the fragment ions (*b* and *y* ions) were fully consistent with the theoretical values for grisemycin (Figure S1). The planar structure of **1** was further confirmed by NMR analysis using 0.5 mg of **1** purified from large-scale cultures (50 mL × 100 flasks). The ^{1}H and ^{13}C-NMR data, NMR spectra (^{1}H, ^{13}C, COSY, TOCSY, HSQC, HSQC-TOCSY, HMBC, and ROESY), and 2D NMR correlation are summarized in Table S1 and Figures S2–11. The presence of 18 amino acid residues (Ala ×3, Dhb ×2, Pro ×1, Val ×3, Gln ×2, Phe ×1, Ile/allo-Ile ×2, Gly ×1, Ser ×1, Leu ×1, and aviCys ×1) were confirmed mainly by COSY, TOCSY, and HSQC-TOCSY analysis, and the sequence of these amino acids was established with HMBC and ROESY analysis. For nonproteinogenic amino acid residues, the N-terminal dimethyl group was confirmed with HMBC correlation from the *N*-CH_3 proton (δ_H 2.86, 6H) to C_α of Ala-1 (δ_C 64.05). In addition, the geometry of the double bond in AviCys-16 was deduced to be Z, based on the small coupling constant of 7.0 Hz between H-4 (δ_H 5.48) and H-5 (δ_H 7.11), as well as Nuclear Overhauser Effect (NOE) correlations between them. The geometry of the double bonds in both Dhb

residues was also deduced to be Z because NOE correlations between H-4 (δ_H 1.81) and NH (δ_H 9.90) in Dhb-2, as well as H-4 (δ_H 1.78) and NH (δ_H 10.00) in Dhb-14, were observed. It is worth noting that cypemycin contained (*E*)-Dhb residues, while salinipeptins contained (*Z*)-Dhb residues.[9,11]

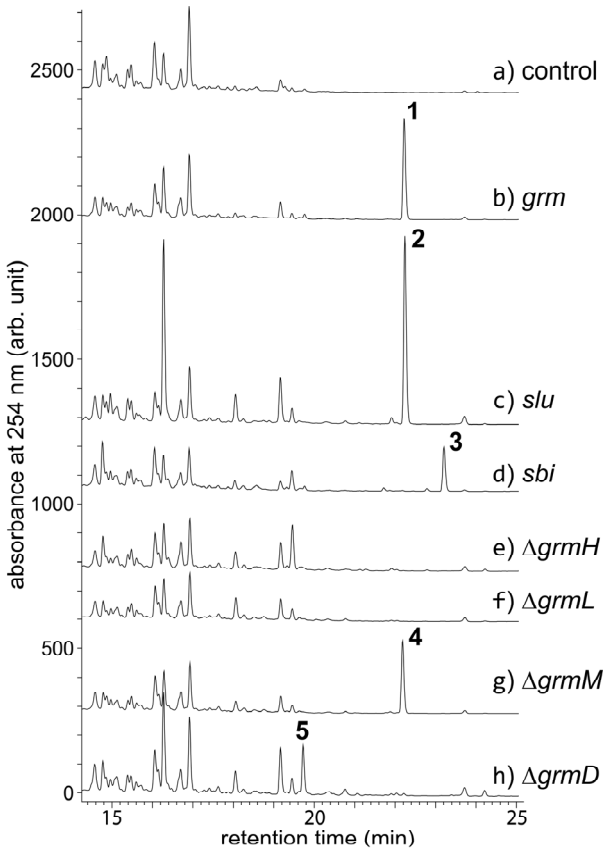

Figure 2. LC-MS analysis of the chloroform extracts of the culture broths. a) *S. lividans*/pWHM3 (empty vector control), b) *S. lividans*/pWHM3-grm, c) *S. lividans*/pWHM3-slu, d) *S. lividans*/pWHM3-sbi, e) *S. lividans*/pWHM3-grm-*ΔgrmH*, f) *S. lividans*/pWHM3-grm-*ΔgrmL*, g) *S. lividans*/pWHM3-grm-*ΔgrmM*, h) *S. lividans*/pWHM3-grm-*ΔgrmD*.

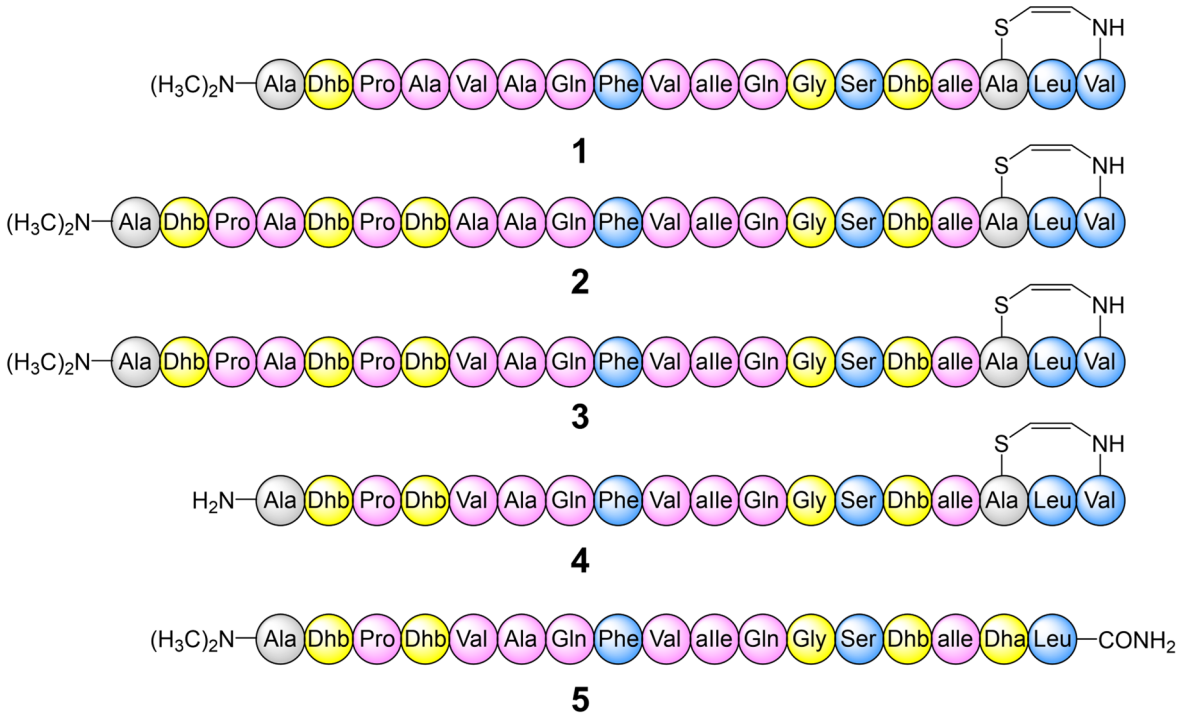

Figure 3. Structures of **1**–**5**. The color codes are the same as those indicated in Figure 1A.

We then examined whether **1** contained D-amino acid residues. Purified **1** was hydrolyzed in 6 M deuterium chloride at 110 °C for 6 h, and the released amino acids were analyzed by LC-MS after derivatization with D- and L-1-fluoro-2,4-dinitrophenyl-5-leucine amide (FDLA). In this method, the influence of unwanted racemization during acid hydrolysis can be excluded.[16] As shown in Figures S12 and S13, comparison with authentic standards of D-FDAA-derivatized amino acids revealed that all Ala, Gln, Pro, and allo-Ile (from Ile) residues and two Val residues in **1** are D-enantiomers. The presence of D-allo-Ile in **1** was confirmed by C_3 Marfey's method because Ile and allo-Ile could not be differentiated using the original Marfey's method.[17] By contrast, all Phe, Ser, and Leu residues and one Val residue were determined to be L-enantiomers. Although we were unable to determine which of the Val-5, -9, and -18 residues had an L-configuration, Val-18 is likely to be an L-enantiomer as predicted by the structure of compound **5** as discussed *vide infra*. Overall, these results clearly showed that **1** contained multiple D-amino acids in a similar manner to salinipeptins (Figure 3).

We next examined whether the presence of D-amino acid residues is the common feature of type-A linaridins by heterologous expression of two other gene clusters identified in publicly available genome database. As shown in Figure 1, the *slu* cluster identified in *Streptomyces luteocolor* NBRC 13826 is closely related to the cypemycin cluster in *Streptomyces* sp. OH-4156. We also selected the *sbi* cluster from *Streptomyces bikiniensis* NBRC 14598. Based on the deduced amino acid sequence of the precursor peptide, the metabolites of the *slu* and *sbi* clusters would be salinipeptin- and cypemycin-like linaridins, respectively. After heterologous expression of the *slu* (*sluA*–*sluP*) and *sbi* (*sbiA*–*sbiP*) clusters using a similar method to that described above, the specific metabolites were analyzed by LC-MS (Figure 2). As shown in Figures S14 and S15, HR-MS and MS/MS analyses revealed the production of salinipeptin-like linaridin **2** (obs.: *m*/*z* 1034.5568, calcd. for $C_{97}H_{150}N_{24}O_{24}S$, calcd: 1034.5561 $[M+2H]^{2+}$) and cypemycin-like linaridin **3** (obs.: *m*/*z* 1048.5722, calcd. for $C_{99}H_{154}N_{24}O_{24}S$: *m*/*z* 1048.5717 $[M+2H]^{2+}$) from the *slu* and *sbi* clusters, respectively. Furthermore, chiral analysis of the hydrolytic products of **2** and **3** revealed that both contained multiple D-amino acid residues in a similar manner to **1**. It is worth noting that there are stereochemical differences between the structures of **2** and salinipeptin A, which have two Val and two Ile residues. While salinipeptin A was reported to contain only L-enantiomers of the Val residues, both enantiomers of Val were observed in **2**. In addition, **2** contained D-allo-Ile residues instead of D-Ile residues in salinipeptin. Nonetheless, these results suggested that the presence of D-amino acid residues is a common feature of type-A linaridins. Furthermore, these results strongly indicated that the type-A linaridin cluster contains a cryptic peptide epimerase that catalyzes epimerization at C_{α} of the amino acid residues of the precursor peptide.

We next investigated the function of each gene by gene-deletion experiments of the *grm* cluster. After construction of pWHM3-*grm*-derived in-frame deletion plasmids lacking the *grmH*, *grmL*, *grmM*, or *grmD* gene, each plasmid was introduced into *S. lividans* and the metabolites were analyzed by LC-MS. When any one of the genes was removed, grisemycin production was

completely abolished (Figure 2, traces e–h). These results clearly indicated that all four genes are indispensable for grisemycin biosynthesis. Furthermore, inspection of the chromatograms revealed that the *grmM* and *grmD* mutants produced compounds **4** and **5**, respectively, while no putative intermediates were detected in the *grmH* and *grmL* mutants. To analyze the structures of these compounds, purified **4** and **5** were analyzed by LC-MS, MS/MS, and the modified Marfey's method described above. Compound **4** had a similar structure to **1** except for the absence of the N-terminal *N*,*N*-dimethyl groups (obs.: *m*/*z* 902.9930, calcd. for $C_{84}H_{133}N_{21}O_{21}S$: *m*/*z* 902.9926 $[M+2H]^{2+}$), as expected from the predicted function of GrmM (Figure S16). By contrast, the molecular formula of **5** deduced from HR-MS was $C_{79}H_{126}N_{20}O_{20}$ ($[M+2H]^{2+}$, obs.: *m*/*z* 838.4801, calcd.: *m*/*z* 838.4802), which was 203 mass units smaller than that of the expected non-decarboxylated peptide of the Δ*grmD* mutant (Figure S17). Considering that the N-terminal 16 amino acid residues of compound **5** were suggested to be the same as those of **1** by observation of a series of fragment ions (*b5*–*b16* ions), we proposed the structure of **5** as shown in Figure 3. Removal of two amino acid residues (Val-18 and Cys-19) from the C-terminus, presumably catalyzed by endogenous enzymes in the heterologous host, might occur to generate **5** from the non-decarboxylated intermediate. Importantly, both **4** and **5** contained multiple D-amino acids that correspond to **1** (Figures S12 and S13). These results suggested that the introduction of D-amino acids by the novel epimerase occurred prior to methylation and decarboxylation. Considering that GrmL was the only enzyme of unknown function among the proteins necessary for the biosynthesis of **1**, it is suggested that GrmL may be responsible for the introduction of multiple D-amino acid residues in the early stage of grisemycin biosynthesis.

In conclusion, we performed heterologous expression of three type-A linaridin biosynthetic gene clusters and demonstrated that all of the metabolites contained multiple D-amino acids in a similar manner to salinipeptin A. Importantly, these observations strongly indicated that type-

A linaridin clusters contained a gene encoding a novel peptide epimerase. Furthermore, gene-deletion experiments showed that the *grmL*, the sole gene of unknown function in the cluster, is indispensable for linaridin biosynthesis and that the epimerization preceded decarboxylation and methylation. Taken together, GrmL is suggested to catalyze epimerization in the early stage of biosynthesis. Because site-specific epimerization of peptides is chemically challenging, diverse strategies to enzymatically introduce D-amino acid residues into peptides have long been sought. Our results clearly reveal the involvement of a novel peptide epimerase responsible for linaridin biosynthesis. This work sets the stage for further biochemical investigations into the biosynthesis of the linaridin class of natural products.

Acknowledgments

This work was supported by Grants-in-Aid for Scientific Research on Innovative Areas from MEXT, Japan (JSPS KAKENHI Grant Number JP16H06452 to T.D.) and Grants-in-Aid for Scientific Research from JSPS (JP18H03937 and JP22H04976 to T.D. and JP18K05449 to Y.O.). We thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

The authors declare no competing financial interest.

References

[1] a) A. Gautam, R. Vyas, R. Tewari, *Crit. Rev. Biotechnol.* **2011**, *31*, 295-336; b) J. van Heijenoort, *Cell Mol. Life Sci.* **1998**, *54*, 300-304.

[2] S. Kapil, V. Sharma, *Can. J. Microbiol.* **2021**, *67*, 119-137.

[3] a) T. Stachelhaus, C. T. Walsh, *Biochemistry* **2000**, *39*, 5775-5787; b) C. T. Walsh, *Nat. Prod. Rep.* **2016**, *33*, 127-135.

[4] S. A. Samel, P. Czodrowski, L. O. Essen, *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 1442-1452.

[5] a) M. F. Freeman, C. Gurgui, M. J. Helf, B. I. Morinaka, A. R. Uria, N. J. Oldham, H. G. Sahl, S. Matsunaga, J. Piel, *Science* **2012**, *338*, 387-390; b) B. I. Morinaka, A. L. Vagstad, M. J. Helf, M. Gugger, C. Kegler, M. F. Freeman, H. B. Bode, J. Piel, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 8503-8507.

[6] A. Sikandar, L. Franz, S. Adam, J. Santos-Aberturas, L. Horbal, A. Luzhetskyy, A. W. Truman, O. V. Kalinina, J. Koehnke, *Nat. Chem. Biol.* **2020**, *16*, 1013-1018.

[7] a) Z. Feng, Y. Ogasawara, S. Nomura, T. Dairi, *Chembiochem* **2018**, *19*, 2045-2048; b) Z. Feng, Y. Ogasawara, T. Dairi, *Chem. Sci.* **2020**, *12*, 2567-2574.

[8] a) X. Yang, W. A. van der Donk, *J. Am. Chem. Soc.* **2015**, *137*, 12426-12429; b) L. Huo, W. A. van der Donk, *J. Am. Chem. Soc.* **2016**, *138*, 5254-5257; c) M. Xu, F. Zhang, Z. Cheng, G. Bashiri, J. Wang, J. Hong, Y. Wang, L. Xu, X. Chen, S. X. Huang, S. Lin, Z. Deng, M. Tao, *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 18029-18035.

[9] Z. Shang, J. M. Winter, C. A. Kauffman, I. Yang, W. Fenical, *ACS Chem. Biol.* **2019**, *14*, 415-425.

[10] J. Claesen, M. J. Bibb, *J. Bacteriol.* **2011**, *193*, 2510-2516.

[11] J. Claesen, M. Bibb, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16297-16302.

[12] M. E. Rateb, Y. Zhai, E. Ehrner, C. M. Rath, X. Wang, J. Tabudravu, R. Ebel, M. Bibb, K. Kyeremeh, P. C. Dorrestein, K. Hong, M. Jaspars, H. Deng, *Org. Biomol. Chem.* **2015**, *13*, 9585-9592.

[13] F. J. Ortiz-Lopez, D. Carretero-Molina, M. Sanchez-Hidalgo, J. Martin, I. Gonzalez, F. Roman-Hurtado, M. de la Cruz, S. Garcia-Fernandez, F. Reyes, J. P. Deisinger, A. Muller, T. Schneider, O. Genilloud, *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 12654-12658.

[14] a) S. Ma, Q. Zhang, *Nat. Prod. Rep.* **2020**, *37*, 1152-1163; b) W. Ding, N. Yuan, D. Mandalapu, T. Mo, S. Dong, Q. Zhang, *Org. Lett.* **2018**, *20*, 7670-7673.

[15] J. Vara, M. Lewandowska-Skarbek, Y. G. Wang, S. Donadio, C. R. Hutchinson, *J. Bacteriol.* **1989**, *171*, 5872-5881.

[16] D. R. Goodlett, P. A. Abuaf, P. A. Savage, K. A. Kowalski, T. K. Mukherjee, J. W. Tolan, N. Corkum, G. Goldstein, J. B. Crowther, *J. Chromatogr. A* **1995**, *707*, 233-244.

[17] S. Vijayasarathy, P. Prasad, L. J. Fremlin, R. Ratnayake, A. A. Salim, Z. Khalil, R. J. Capon, *J. Nat. Prod.* **2016**, *79*, 421-427.