The Sandarazols are Cryptic and Structurally Unique Plasmid-Encoded Toxins from a Rare Myxobacterium**

Fabian Panter, Chantal D. Bader, and Rolf Müller

Abstract: *Herein, we describe a new plasmid found in Sandaracinus sp. MSr10575 named pSa001 spanning 209.7 kbp that harbors a cryptic secondary metabolite biosynthesis gene cluster (BGC). Activation of this BGC by homologous-recombination-mediated exchange of the native promoter sequence against a vanillate inducible system led to the production and subsequent isolation and structure elucidation of novel secondary metabolites, the sandarazols A–G. The sandarazols contain intriguing structural features and very reactive functional groups such as an α-chlorinated ketone, an epoxyketone, and a (2R)-2-amino-3-(N,N-dimethylamino)-propionic acid building block. In-depth investigation of the underlying biosynthetic machinery led to a concise biosynthetic model for the new compound family, including several uncommon biosynthetic steps. The chlorinated congener sandarazol C shows an IC_{50} value of 0.5 μ$_M$ against HCT 116 cells and a MIC of 14 μ$_M$ against Mycobacterium smegmatis, which points at the sandarazols' potential function as defensive secondary metabolites or toxins.*

Introduction

Bacterial antibiotic resistance genes cause antimicrobial resistance (AMR) in pathogens thereby leading to enormous challenges in the treatment of infectious diseases in the clinic.[1] AMR genes are often encoded on plasmids, horizontally transferrable elements employed by bacteria to survive

[*] F. Panter, C. D. Bader, R. Müller
Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University, Campus E8 1
66123 Saarbrücken (Germany)
and
German Centre for Infection Research (DZIF)
Partner Site Hannover-Braunschweig (Germany)
E-mail: rolf.mueller@helmholtz-hzi.de

F. Panter, R. Müller
Helmholtz International Lab for Anti-Infectives, Campus E8 1
66123 Saarbrücken (Germany)

[**] A previous version of this manuscript has been deposited on a preprint server (https://doi.org/10.1101/2020.10.06.323741).

📄 Supporting information and the ORCID identification number(s) for
ⓘD the author(s) of this article can be found under:
https://doi.org/10.1002/anie.202014671.

antibiotics treatment and effectively spread AMR.[2] In addition to that, AMR genes co-localized with antibiotic biosynthesis gene clusters (BGCs) are important for self-resistance during antibiotic production in bacteria capable of respective biosyntheses. Similarly to plasmid-mediated exchange of antibiotic resistance, bacteria are able to gain an advantage over competing microorganisms by exchanging secondary metabolite BGCs in horizontal gene transfer events.[3] The mechanism for plasmid-mediated exchange of secondary metabolite BGCs is known for plasmid-encoded toxins, with the anthrax-causing *Bacillus anthracis* as the most prominent example.[4] The *Bacillus* plasmid is shared between different bacilli and thereby transmits the anthrax toxin virulence factor. Exchange mechanisms of such plasmids that encode toxins enable the respective bacteria to acquire the ability for chemical warfare. As bacteria are competing with other microbes for nutrients in their ecological niches, the encoded chemical warfare molecules optimized by evolution are attractive targets in the search for novel bioactive natural products. For actinobacteria it has already been shown that defensive secondary metabolites can be encoded as multi-modular BGCs on autonomously replicating plasmids. A BGC responsible for production of the highly cytotoxic mycolactones, for example, is also encoded on a large plasmid.[5] *Mycobacterium ulcerans* produces this cytotoxin and releases it during the infection process of human skin, while the bacterium feeds on skin cells. Streptomycetes also host plasmid-borne antibiotics BGCs as exemplified by the esmeraldins, a series of phenazine antibiotics.[6] With respect to the total output of natural products and natural product diversity, bacteria have emerged as major players as they are prolific producers of biologically active secondary metabolites.[7] Besides the well-described bacterial phyla such as actinobacteria, firmicutes, and cyanobacteria that are responsible for the majority of identified biologically active natural products to date, Gram-negative proteobacteria such as myxobacteria have also shown great promise for the discovery of novel bioactive secondary metabolites.[8–10] Especially novel and phylogenetically distant bacterial genera among the myxobacteria, such as the recently identified genus *Sandaracinus* studied herein, hold promise for finding interesting novel natural products chemistry.[11] In recent years, availability of cheap and reliable DNA sequencing technologies along with the development of BGC prediction tools such as antiSMASH depict the theoretical genetically encoded bacterial secondary metabolome as to date largely underexploited.[12,13] However, many BGCs that can be identified in silico are "cryptic" under laboratory conditions. In this case, the production of the corresponding secondary metabolite is completely repressed or remains below the

detection limit of modern analytical tools such as liquid chromatography (LC)-coupled mass spectrometry (MS) instrumentation. Activating such cryptic clusters by genetic tools holds promise for discovering new chemistry, but currently remains a non-automatable, labor-intensive process.[14] Therefore, a strong focus on rational prioritization of BGCs to be activated by heterologous expression or gene cluster overexpression using heterologous promoters remains crucial to focus discovery efforts on BGCs encoding for bioactive natural products.[15] Methods employed for BGC prioritization so far include expected target-guided mining for potential self-resistance genes and exploring biosynthetic gene cluster complexity for modular BGC architectures as well as the presence of certain tailoring enzymes such as halogenases or epoxidases.[16] Still, the presence of a BGC on a plasmid, as an example for a BGC that is likely to be a mobile genetic element, has not yet been used as a means for BGC prioritization. As part of our ongoing efforts to isolate taxonomically diverse myxobacteria we isolated a novel myxobacterial strain belonging to the *Sandaracinus* clade. Surprisingly, we found that the strain's genome contains a circular 209.7 kbp plasmid called pSa001 that features a large polyketide synthase (PKS) non-ribosomal peptide synthetase (NRPS) hybrid BGC in its sequence. Herein, we show that activation of this BGC by genetic engineering of this strain leads to production of several derivatives of the sandarazols, a previously unknown family of natural products. These compounds turned out to be potent and chemically novel toxins biosynthesized by a series of intriguing biosynthetic steps.

Results and Discussion

Biosynthetic Gene Cluster Identification and Activation

Myxobacteria are Gram-negative δ-proteobacteria with a rich and diverse secondary metabolism.[9, 10] A survey of known and unknown natural products in 2300 myxobacteria showed structural novelty to be clearly correlated to phylogenetic distance and thus heavily reliant on in-depth investigation of novel myxobacterial genera.[11] We thus chose to dive into the secondary metabolome of a myxobacterial strain called MSr10575 that was recently isolated in-house. The strain belongs to the *Sandaracinus* genus, a rare myxobacterial genus little studied for secondary metabolism with the type strain *Sandaracinus amylolyticus* NOSO-4T.[17] This strain was the only member in the Sorangineae clade until MSr10575 was isolated and provided the indiacens A and B, two prenylated indole secondary metabolites.[18] We chose to sequence the genome of strain MSr10575 using single-molecule real-time sequencing technology.[19] We subsequently annotated the BGCs in its sequence by antiSMASH to obtain an overview about the strain's secondary metabolite production potential.[12] Sequencing coverage analysis and genome assembly revealed that the genome of MSr10575 consists not only of a bacterial chromosome of 10.75 Mbp but also of an autonomously replicating plasmid of 209.7 kbp, which we named pSa001 (see SI). As sequencing coverage per

base on the pSa001 plasmid is twice as large as observed for the bacterial chromosome, we assume a median value of two pSa001 copies per MSr10575 cell (see SI). The overall GC content of the plasmid is 70% and the codon usage bias favors GC-rich codons over AT-rich codons for the same amino acid at a rate of 3.6 to 1. As these values do not differ significantly from the corresponding parameters for the MSr10575 chromosome, the plasmid seems well adapted to its host strain (see SI). The presence of this plasmid is distinctive for this myxobacterial strain as autonomously replicating plasmids are rarely observed in myxobacteria. The only other example of a characterized myxobacterial plasmid is pMF1 from *Myxococcus fulvus*.[20] In comparison to pMF1, pSa001 is not only significantly larger, but also encodes a large type 1 in-*trans* acyl transferase (*trans*-AT) PKS–NRPS hybrid BGC, which was named sandarazol (*szo*) BGC (see SI). Besides the *szo* BGC, the plasmid contains five ORFs encoding putative transposases and one ORF encoding a putative integrase indicating its ability to integrate into (or transfer parts of its sequence into) foreign bacterial genomes after conjugation. It might thus serve as a BGC shuttle vector for horizontal gene transfer of the BGC. Cultivation of wild-type MSr10575 followed by extraction and liquid chromatography–tandem mass spectrometry (LC–MS²) analysis and GNPS based spectral networking of the bacterial metabolome did not reveal a family of secondary metabolites matching the BGC architecture on the plasmid in expected secondary metabolite size and fragmentation pattern (see SI).[21] We therefore assumed the plasmid borne *szo* BGC to be "cryptic" as seen in many other cases such as the pyxidicyclines or taromycin.[22,23] All the coding regions in the *szo* cluster spanning from *szoA* to *szoO* are encoded on the same DNA strand and the intergenic regions seemed too small to contain important elements other than ribosome binding sites. We therefore reasoned that the *szo* cluster is encoded as a single transcriptional unit even though it spans 44.5 kbp. Activating the *szo* BGC with a single promoter exchange in line with the experiments described in promoter exchange BGC activation in myxobacteria seemed therefore a reasonable strategy to unlock the corresponding natural products.[22,24] As we set out to create a plasmid for *szo* cluster overexpression by single crossover promoter exchange, we chose a plasmid with a pBelobac borne replication machinery, as this would allow the entire 209.7 kbp pSa001 to be extracted and transferred into *E. coli* as replicative plasmid for further investigations into the plasmid's replication mechanism.[25] While extracting the pBeloBacSa001 into *E. coli* is possible, retransformation of this plasmid into a plasmid-cured version of MSr10575 needs further optimization. To investigate the secondary metabolite products of the BGC, we chose BGC activation in the native host MSr10575 first, as it was assumed MSr10575 would be best suited to produce the corresponding small-molecule product. We activated the *szo* cluster on the plasmid by promoter exchange against the vanillate promoter and repressor cassette, since this tool achieved overexpression of BGCs in several other myxobacteria including *Myxococcus xanthus* and *Pyxidicoccus fallax* (see SI).[22,26] Compared to other promoters, the vanillate promoter and repressor system shows tight control of BGC expression as well as strong
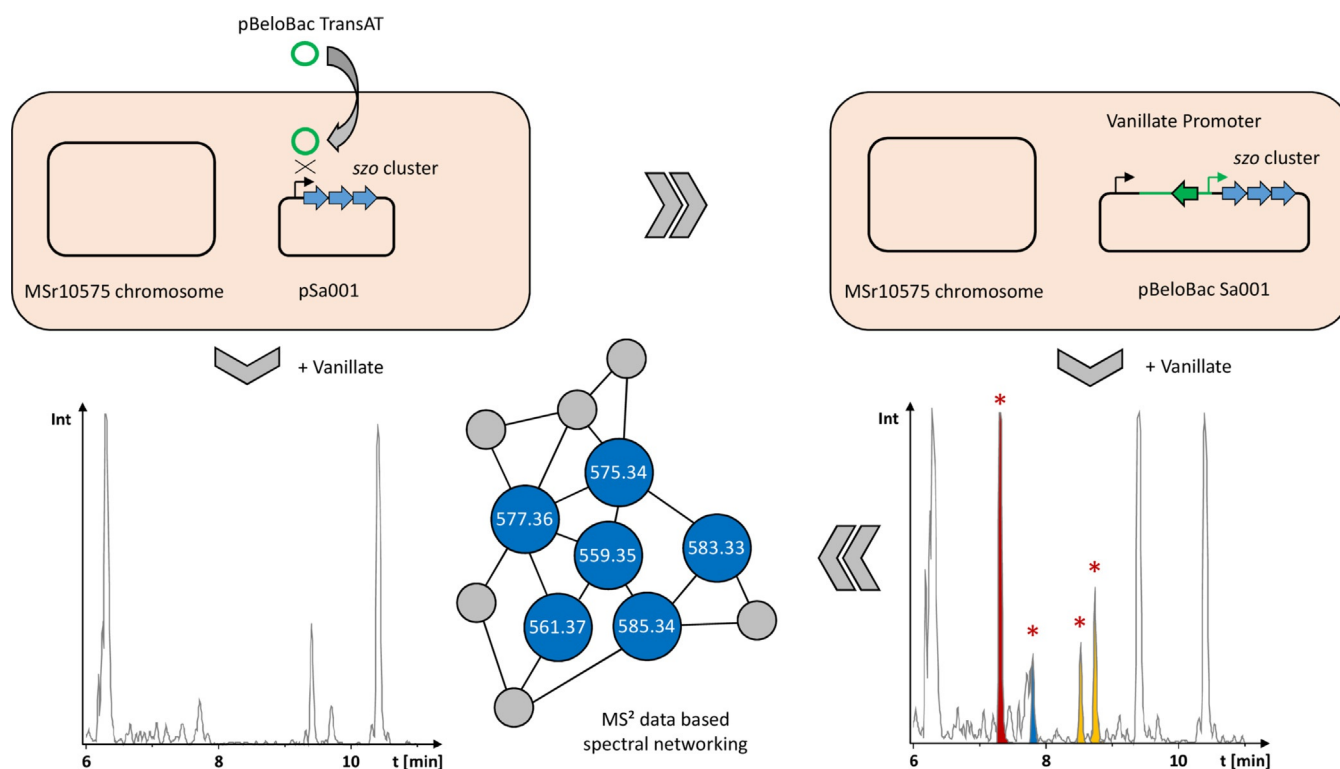
Figure 1. Schematic overview of the activation of the plasmid-borne *szo* BGC in wild-type *Sandaracinus* sp. MSr10575 by homologous recombination, the corresponding changes in the extracts' LC–MS chromatograms, and LC–MS²-based spectral networking for identification of the produced sandarazols.

secondary metabolite production upon BGC induction in myxobacteria.[27]

To exchange the promoter of the *szo* cluster, the first 2 kbp of the *szo* BGC were PCR-amplified, fused to a vanillate promoter and repressor cassette, and ligated into the commercial pBeloBac11. This plasmid featuring the pBeloBac replication machinery and a kanamycin resistance in its backbone was named pBeloBac-TransAT (see SI). Promoter exchange of the *szo* cluster's native promoter against the vanillate cassette is achieved by homologous recombination of pBeloBac-TransAT with the pSa001 plasmid after electroporation (Figure 1). MSr10575 clones harboring the recombined fused plasmid called pBeloBacSa001 were selected on 25 µg mL⁻¹ kanamycin and genotypically verified by PCR (see SI). To obtain a comprehensive overview about the metabolic differences between MSr10575 wild type and MSr10575:pBeloBacSa001, analytical-scale fermentation cultures were prepared in triplicates, extracted, and subjected to UHPLC–qTOF analysis. All detected LC–MS features in the triplicate analyses were grouped into 2D retention time–exact mass windows (so-called buckets), and all MSr10575 wild-type-derived liquid chromatography–mass spectrometry (LC–MS) features were subtracted from the detected LC–MS features using principal component analysis (PCA).[28,29] These features were then used for selective acquisition of LC–MS² data covering exclusively the mutant-derived LC–MS features that are subsequently used for spectral networking in GNPS.[21] We thus obtained spectral networks comprising all LC–MS features linked to activation of the sandarazol BGC (see Figure 1 and SI).

Isolation and Structure Elucidation

The novel secondary metabolites observed in this PCA-based analysis of LC–MS features were named the sandarazols. Besides sandarazol A (**1**; 575.344 Da [*M*+H]⁺; $C_{30}H_{47}N_4O_7$, $\Delta = 0.64$ ppm), we were able to identify and characterize seven structural variants from appearance of their characteristic MS signals, approximately half of which showed the characteristic isotope pattern for chlorine-containing compounds such as sandarazol C (**3**; 583.3256 Da [*M*+H]⁺; $C_{29}H_{48}ClN_4O_6$, $\Delta = 0.07$ ppm).[30] The sandarazols A, B, C, and F (**1**, **2**, **3**, and **6**) were isolated from vanillate-induced large-scale cultures of MSr10575:pBeloBac Sa001 using a sequence of different techniques such as liquid/liquid extraction, countercurrent partition chromatography (CPC), and HPLC under N₂ (see SI, Figure 2). It is worth noting that not even trace amounts of the sandarazols A–G (**1**–**7**) can be detected in the MSr10575 wild-type extracts indicating the *szo* cluster to be most likely fully cryptic in the wild-type strain under the laboratory conditions applied in our study. The structures of sandarazol D (**4**), E (**5**), and G (**7**) were assigned by comparison of their MS² spectra to the other derivatives (see SI). As the compounds are sensitive to oxygen and/or strongly acidic or basic pH values, compound isolation and purification was performed in ammonium-formate-buffered eluents and under constant N₂ stream wherever possible (Figure 2).

¹H NMR, ¹³C NMR and HSQC-dept. spectra of **1** reveal two α-protons based on their characteristic chemical shift at $\delta^1H = 4.78$ and 4.45 ppm (Figure 3). COSY and HMBC
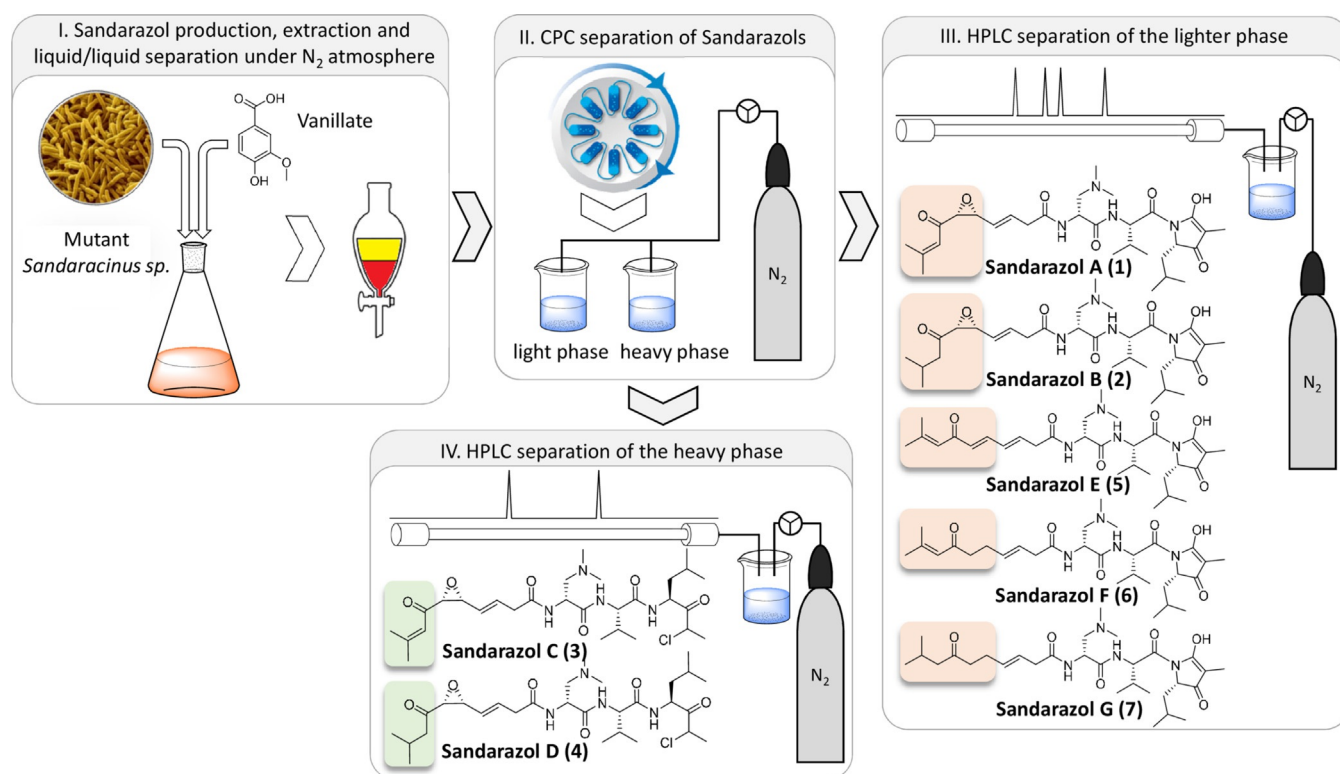
Figure 2. Workflow and isolation scheme for preparation of the sandarazols from *Sandaracinus* sp. MSr10575:pBeloBacSa001 including structure formulas for all structurally elucidated sandarazol derivatives.

correlations of the first methine group at $\delta^1 H = 4.78$ ppm to one methylene group, two methyl groups with a downfield $\delta^{13}C$ chemical shift of 45.2 ppm and a quaternary carbon with a characteristic amide shift, show that this α-proton is part of 2-amino-3-(N,N-dimethyl amino)-propionic acid (Me₂Dap). The second α-proton at $\delta^1 H = 4.45$ ppm exhibits COSY and HMBC correlations to one methine group, two methyl

Figure 3. A) NMR correlations important for structure elucidation. B) Reaction sequence to open the epoxide and form a methanol adduct. C) Mosher's esterification of **1** used to elucidate the epoxide's configuration.

groups, and another amide function. The second amino acid of the molecule was determined to be a valine. HMBC correlations of the valine α-proton to the Me₂Dap carboxylic acid function suggests their connection via the valine N-terminus. Further HMBC correlations of the Me₂Dap α-proton to another amide function reveal N-terminal elongation by the polyketide part of the molecule. Characteristic chemical shifts of two protons at $\delta^1 H = 3.47$ ppm and 3.46 ppm with correlations to two vinylic double bond protons at one side and one epoxide on the other side, suggest **1** to contain an epoxyketone, as well as unsaturation in β,γ position relative to the amide bond. Typical J_{H-H} coupling values of 15.4 Hz for the vinylic protons at the double bond indicate *E* configuration.[31] 1D and 2D NMR spectra reveal a trisubstituted vinyl group is conjugated to the ketone side of the epoxyketone. The C-terminal end of valine shows correlations to another methine group with characteristic chemical shifts close to an α-proton shift, suggesting further elongation of the molecule by another amino acid. This amino acid is determined to be leucine coupled to a propionate unit from 1D and 2D NMR data. The combination of a low-field proton shift and high-field carbon shift of an additional methyl group at $\delta^1 H = 1.60$ and $\delta^{13}C = 5.7$ ppm in this part of the molecule reveals heterocyclization of a propionate-extended valine cyclized to a tetramic acid, which also serves as connection between the valine and leucine building blocks (Figures 2 and 3).
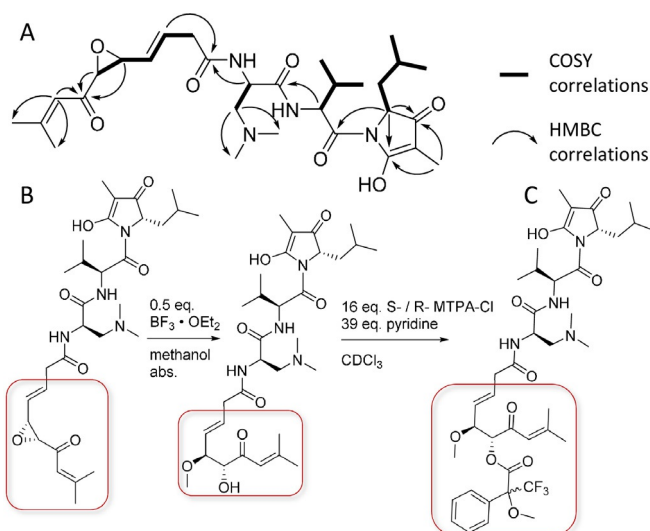
In contrast to **1**, HSQC spectra of **2** reveal a missing double bond in the terminal moiety of the polyketide part of the molecule (Figure 2). Sandarazol F differs in the poly-

ketide part of the molecule, as HSQC spectra reveal the two methines in the epoxide motif to be replaced by two double bond protons. In **3**, however, the polyketide part equates the one of sandarazol A, but 1D and HSQC spectra show a missing tetramic acid moiety. The methyl group at $\delta^{13}C = 6.0$ in **1** is shifted to $\delta^{13}C = 20.7$ ppm in **3**. In line with the *hr*MS spectra of **3**, chlorination of the molecule is observed at the α-position of the elongated leucine, confirmed by the characteristic chemical shifts and COSY as well as HMBC correlations of the surrounding methyl and methine groups.

All sandarazol derivatives contain five stereocenters, two of which are linked by the epoxide. The chlorinated derivatives **3** and **4** feature an additional stereocenter at the chlorinated carbon atom, which epimerizes fast at room temperature (see SI). Its original stereoconfiguration can therefore not be determined as epimerization occurs already under fermentation conditions. The stereocenters contained in the L-valine (L-Val), L-leucine (L-Leu), and (2R)-2-amino-3-(N,N-dimethyl amino)-propionic acid (D-Me$_2$Dap) building blocks were confirmed by Marfey's analysis using commercially available standards (see SI).[32]

The stereocenters in the polyketide part of the molecule are determined by the orientation of the epoxide. Direct configurative assignment was not possible, as established derivatization methods like the Mosher's method rely on reaction of functional groups like alcohols.[33] To tackle this issue, the epoxide was transformed into an alcohol by Lewis-acid-catalyzed addition of methanol to the epoxide (Figure 3). The resulting secondary alcohol inherits its stereochemistry from the epoxide due to retention of the epoxide's stereochemistry in the S$_N$2-based ring opening reaction (Figure 3). Subsequent Mosher's esterification of the secondary alcohol revealed the alcohol to be in *R*-configuration and thus the epoxide in sandarazols to be *R,R*-configured. With full structure elucidation, the sandarazols' absolute configuration, as well as the *szo* biosynthetic gene cluster at hand we were able to devise a biosynthetic model for the sandarazols.

Biosynthesis of the Sandarazols

The sandarazol biosynthesis pathway is based on a type 1-*trans*-AT PKS–NRPS hybrid BGC spanning 44.5 kbp and 15 open reading frames (ORFs, *szoA* to *szoO*). The megasynthase genes are encoded on the ORFs *szoE* to *szoH*. The in-*trans* acting AT and the very unusual in-*trans* acting DH domains are encoded on *szoC* and *szoD*. Tailoring enzymes such as the epoxidase (SzoP), the halogenase (SzoI), the short-chain reductase (SzoA), the β-branching cassette (SzoK to SzoO), and the biosynthetic machinery to supply the (2S)-2,3-diamino-propionic acid precursor (SzoB and SzoJ) are encoded on the remaining tailoring enzymes in the BGC (Figure 4).

Sandarazol biosynthesis starts with a *trans*-AT PKS starter module on SzoE that loads an acetate unit onto the first acyl carrier protein (ACP). The peculiarity of this first module is the presence of a β-branching ACP, which acts at the site where the β-branching cassette SzoK to SzoO adds a β-methyl branch to the first PKS extension (Figure 4B).[34] In contrast

to standard β-branching cassettes like PyxK to PyxO in the pyxipyrrolone BGC consisting of an ACP, a ketosynthase (KS), a hydroxymethylglutaryl-CoA synthase (HMG), and two enoyl-CoA dehydratases (ECH), the latter ECH domain is replaced with SzoO, a thioesterase (TE).[35] This enzyme is able to replace the decarboxylation function of the second ECH as described in the curacin biosynthesis.[36] While most branched-chain-end moieties in polyketide synthase are biosynthesized by incorporation of an isovaleryl or an isoamyl starter unit as, for example, in fulvuthiacene biosynthesis, bongkrekic acid biosynthesis contains a similar branched-chain starter unit biosynthesis based on a β-branching cassette.[29,37] As the *szo* cluster does not encode enoyl reductase (ER) domains, the standalone short-chain reductase protein *szoA* is likely responsible for the reduction steps leading to the production of the saturated tail group seen in the sandarazol variants **2**, **4**, and **7**. The enzyme is also likely to be responsible for the other α,β enoyl reduction reaction next to the ketone in **6** and **7** (Figures 2 and 4). Module 2 is most likely either skipped or its ACP solely takes part in a trans-acylation reaction to forward the polyketide chain to module 3, as we see no polyketide extension in this module even though both the KS and the ACP in this module seem functional based on sequence alignment analysis. The following modules three to five incorporate three units of malonyl-CoA by successive decarboxylative Claisen condensation to attach six carbon atoms to the growing polyketide chain. In this process, we observe consecutive formation of two double bonds through β-ketoreduction by a ketoreductase (KR) and subsequent dehydration most likely performed by the in-*trans*-acting dehydratase SzoD in modules four and five. In-*trans*-acting dehydratases have already been described in *trans*-AT PKS pathways such as the CurE and CorN proteins from curacin and corallopyronin, respectively.[38,39] The KR domain located in module 3 is most likely responsible for β-ketoreduction in module 4, as module 4 does not contain a KR domain and no β-ketoreduction is observed in module 3. Both double bonds formed in modules 4 and 5 are isomerized from an α,β-double bond to a β,γ-double bond by the double-bond-shifting DH$_2$ domain encoded on *szoF*.[40] Similarly to the double-bond-shifting domains in corallopyronins, the double-bond-shifting DH$_2$ on SzoF contains the first DH domain consensus motif but lacks the second DxxxQ consensus motif, a characteristic that is conserved in double-bond-shifting DH domains (see SI).[39,41] The subsequent NRPS module on *szoF* contains an adenylation (A) domain with a specificity code close to serine specificity according to NRPS predictor 2. This domain is assumed to accept (2S)-2,3-diamino-propionic acid (L-Dap) that is then incorporated into the nascent chain (Figure 4).[42] Biosynthesis of L-Dap is presumed in analogy to the described pathway in *Staphylococcus aureus*.[43] The proteins carrying out the corresponding biosynthetic steps in MSr10575 are the SbnA homolog SzoB and the SbnB homolog SzoJ. SbnA is a pyridoxal phosphate (PLP)-dependent aminotransferase-type enzyme that connects the amino group of glutamate to replace the phosphate moiety of an *O*-phosphoserine molecule from the bacterium's primary metabolism (Figure 4D).[44] SbnB is an NAD-dependent oxidase that releases α-ketoglutarate from
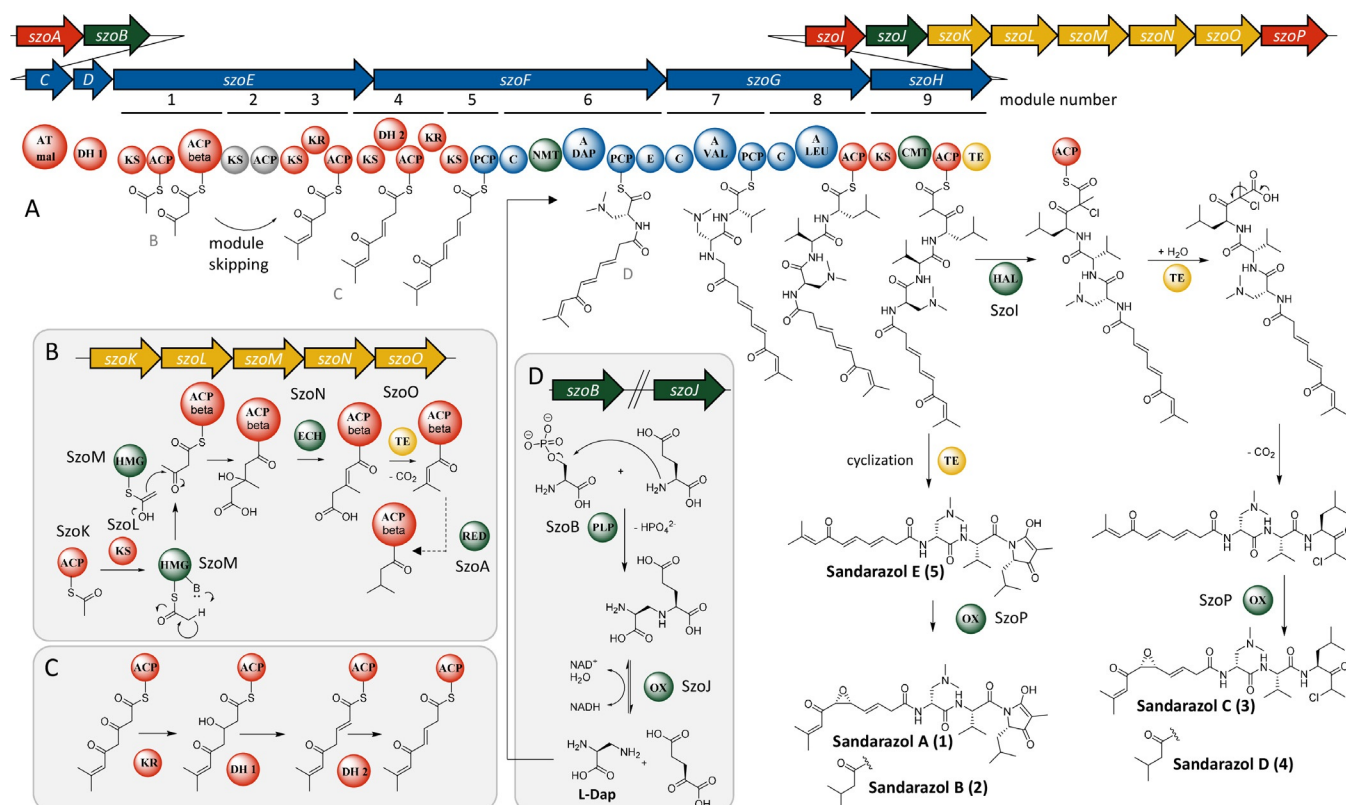
Figure 4. A) Scheme of the proposed sandarazol BGC and sandarazol biosynthesis by its megasynthase including the corresponding tailoring reactions. B) β-branching reaction cascade to form the branched-chain tail group. C) Reaction cascade that leads to the formation of two isomerized double bonds. D) Reaction sequence that supplies the amino acid L-Dap to module 6 of the assembly line (arrows: genes; circles: domains; A: adenylation; ACP/PCP: acyl/peptidyl carrier protein; AT: acyl transferase; C: condensation; CMT/NMT: *C/N* methyl transferase; DH: dehydratase; E: epimerization; ECH: enoyl-CoA hydratase; HAL: halogenase; HMG: hydroxymethylglutaryl-CoA synthase; KR: ketoreductase; KS: ketosynthase; PLP: pyridoxal-phosphate-dependent enzyme; RED: reductase; TE: thioesterase; OX: oxidation).

this intermediate and forms L-Dap. A SAM-dependent *N*-methyl transferase domain in module 6 subsequently transfers two methyl groups to the free amino group of the attached L-Dap building block to form L-Me₂Dap, which is subsequently epimerized to D-Me₂Dap by the following epimerase domain (Figure 4). In contrast to most *N*-methyl transferase domains built into NRPS systems, which transfer a methyl group onto the α-nitrogen atom that will later form the peptide bond, this methylation reaction consecutively transfers two methyl groups to the nitrogen atom in 3-position of L-Dap.[45] The following two modules incorporate L-Val and L-Leu according to NRPS textbook logic.[46] Incorporation of all three amino acids was proven by feeding of stable-isotope-labeled precursors (see SI). Module 9 is a PKS module that incorporates another malonyl-CoA unit by decarboxylative Claisen condensation and adds an α-methyl branch via a SAM-dependent *C*-methyl transferase domain to the molecule (Figure 4).

We did not find any non-chlorinated, non-cyclized sandarazols in the culture broth and cyclization of an amide nitrogen with a carboxylic acid function to form the tetramic acid heterocycle is very unlikely to occur spontaneously. We therefore suspect that chlorination of sandarazols by the halogenase SzoJ occurs on the assembly line. This reaction is likely to govern whether sandarazol is released from the

megasynthase as a chlorinated open-chain molecule prone to decarboxylation as seen in **3** or as a non-chlorinated cyclized product like **1** (Figure 4). Halogenation by SzoJ occurs via the accepted FAD-dependent halogenation mechanism similar to synthetic α-chlorination of β-keto acids.[47] This chlorinated ACP-bound intermediate is released from the assembly line by the TE and quickly loses its terminal carboxylic acid moiety by decarboxylation, a reaction occurring spontaneously in α-chlorinated β-keto acids. If chlorination does not occur, the molecule is subsequently cleaved off the assembly line by a TE domain with a much slower rate, thus cyclizing the product to form a 5-membered ring system as seen, for example, in **1** and **2** (Figure 4). To finalize sandarazol biosynthesis, epoxidation needs to occur next to the ketone moiety in the polyketide part of the molecule. Catalysis of such a reaction is likely performed by FAD-dependent monooxygenase SzoP that putatively performs a reaction similar to the one performed by the FAD-dependent epoxidizing styrene monooxygenase.[48] We assume that sandarazols are finally exported into the surrounding medium by an ABC exporter of broad range specificity encoded somewhere else in the MSr10575 genome, as no specific exporter system is encoded within or in close proximity to the sandarazol BGC or elsewhere on the pSa001 plasmid.

Biological Activity of the Sandarazols

Due to the low stability of the sandarazols in aerobic environments, antimicrobial and anti-proliferative activities were only determined for **1** and **3** as representatives for one chlorinated and one cyclized member of the sandarazol compound family, respectively. Under the specified assay conditions, the non-chlorinated **1** displays only very limited biological activity, while the chlorinated **3** shows prominent cytotoxicity as well as antibiotic activity against *Mycobacterium smegmatis* and a variety of other indicator bacteria for Gram-positive pathogens (Table 1).

Table 1: Antimicrobial and cytotoxic activities of sandarazol A (**1**) and C (**3**) as minimum inhibitory concentrations (MIC) and inhibitory concentrations at 50% inhibition (IC_{50}).

Microbial strain	MIC **1** [μM]	MIC **3** [μM]
C. albicans	>110	110
P. anomala	>110	110
C. freundii	>110	>110
A. baumanii	>110	>110
S. aureus	110	55
B. subtilis	>110	55
E. coli	>110	>110
P. aeruginosa	>110	55
M. smegmatis	>110	14
Cell line	IC_{50} **1** [μM]	IC_{50} **3** [μM]
Human colon cancer HCT116	>64	0.5

We observe that the biological activity of the sandarazols mainly stems from the chlorinated sandarazols such as **3**, which shows both the best antimicrobial and anti-proliferative activity. Its cytotoxicity of 0.5 μM against HCT116 cells might make **3** a promising candidate for compound optimization and mode-of-action studies. Such approaches directed towards improvements of the sandarazols' stability against oxygen would probably increase the observable IC_{50} values, as the values shown here cannot be determined under anaerobic conditions that stabilize the sandarazols. The antimicrobial and anti-proliferative activities of **3** indicate that those secondary metabolites enhance the bacterium's capability to defend itself against competitors like other bacteria or eukaryotic microorganisms. The position of the *szo* BGC on the replicative plasmid pSa001 indicates that this bacterial self-defense strategy can be transferred between bacteria via conjugation.

Conclusion and Outlook

Herein, we identified pSa001 as the second myxobacterial autonomously replicating plasmid known to date. It was found in the newly isolated *Sandaracinus* sp. MSr10575, the second myxobacterial isolate belonging to the *Sandaracinus* clade. The 209.7 kbp pSa001 plasmid itself encodes the *szo* biosynthetic gene cluster as well as five transposase- and one integrase-type ORFs. Therefore, the plasmid's content in-cluding the *szo* BGC may be transferred between bacterial species, not only by conjugative transfer of the plasmid as an autonomously replicating unit, but also by transposition or integration into the acceptor strain's genome. We chose to investigate and characterize the pSa001-borne *szo* BGC that turned out to be cryptic under laboratory conditions. Previous examples of such plasmid-borne secondary metabolite pathways have shown to be responsible for esmeraldin and mycolactone production, both of which show significant biological activity.[5,6] The biosynthetic origin of these two compounds led to the theory that the BGC encoded on pSa001 also encodes the biosynthetic machinery for toxin production. After activation of the BGC located on the plasmid pSa001 by promoter exchange, we observed the production of the sandarazols, a series of type1-*trans*-AT PKS–NRPS hybrid secondary metabolites. The chemical structure of the sandarazols features a permanent positive charge at physiological conditions at the tertiary amine, a Michael acceptor system, an epoxyketone as well as an α-chlorinated ketone. These compounds are therefore prone to display a variety of different reactivities that might well translate into diverse biological activities. More precisely, the sandarazols show promising anti-bacterial and anti-proliferative activities. Following in-depth investigation of the *szo* BGC, we developed a concise biosynthetic model, which explains the sandarazol biosynthesis on its megasynthase protein. Furthermore, we shed light on a variety of uncommon biosynthetic steps such as the first description of incorporation of Me_2Dap into natural products, polyketide β-branching, isomerization of two consecutive double bonds by a single shifting DH-domain, and α-chlorination-governed release of the sandarazols from the assembly line, all of which warrant further investigations of biosynthetic details. The discovery of the sandarazols does not only highlight the capability of myxobacteria to biosynthesize diverse and biologically active secondary metabolites, it also emphasizes the promise of activating biosynthetic gene clusters by inducible heterologous promoters. Vanillate-inducible promoter systems have proven once more to be able to unlock a larger fraction of the cryptic secondary metabolome of myxobacteria in promoter exchange experiments.[8] The sandarazols combine biological activity with both their intriguing chemical structure and biosynthesis. Thus, natural products encoded on other bacterial plasmids might feature similar intriguing properties. The discovery of sandarazols therefore highlights plasmid-encoded BGCs as valuable targets in the search for novel biologically active natural products in the future. As for myxobacteria, encoding toxin BGCs on plasmids that are transferrable between themselves may provide myxobacteria with a competitive edge in their ecological niche. This transfer of toxin production machinery would make particular sense, as myxobacteria do not live individually but rather as part of a swarm which shows wolf-pack-like predatory behavior.[49] Thus, distribution of the genetic means for toxin production among its kind should be advantageous for these microorganisms in their collective predation and survival strategies.

[1] N. R. Naylor, R. Atun, N. Zhu, K. Kulasabanathan, S. Silva, A. Chatterjee, G. M. Knight, J. V. Robotham, *Antimicrob. Resist. Infect. Control* **2018**, *7*, 58.

[2] A. Brenciani, G. Morroni, S. Pollini, E. Tiberi, M. Mingoia, P. E. Varaldo, G. M. Rossolini, E. Giovanetti, *J. Antimicrob. Chemother.* **2016**, *71*, 307.

[3] S. M. Soucy, J. Huang, J. P. Gogarten, *Nat. Rev. Genet.* **2015**, *16*, 472.

[4] V. A. Luna, D. S. King, K. K. Peak, F. Reeves, L. Heberlein-Larson, W. Veguilla, L. Heller, K. E. Duncan, A. C. Cannons, P. Amuso, et al., *J. Clin. Microbiol.* **2006**, *44*, 2367.

[5] T. P. Stinear, G. A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, et al., *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1345.

[6] Z. Rui, M. Ye, S. Wang, K. Fujikawa, B. Akerele, M. Aung, H. G. Floss, W. Zhang, T.-W. Yu, *Chem. Biol.* **2012**, *19*, 1116.

[7] a) G. M. Cragg, D. J. Newman, *Biochim. Biophys. Acta Gen. Subj.* **2013**, *1830*, 3670; b) D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2016**, *79*, 629.

[8] S. C. Wenzel, R. Müller, *Mol. Biosyst.* **2009**, *5*, 567.

[9] C. D. Bader, F. Panter, R. Müller, *Biotechnol. Adv.* **2020**, *39*, 107480.

[10] J. Herrmann, A. A. Fayad, R. Müller, *Nat. Prod. Rep.* **2017**, *34*, 135.

[11] T. Hoffmann, D. Krug, N. Bozkurt, S. Duddela, R. Jansen, R. Garcia, K. Gerth, H. Steinmetz, R. Müller, *Nat. Commun.* **2018**, *9*, 803.

[12] K. Blin, T. Wolf, M. G. Chevrette, X. Lu, C. J. Schwalen, S. A. Kautsar, H. G. Suarez Duran, E. L. C. de Los Santos, H. U. Kim, M. Nave, et al., *Nucleic Acids Res.* **2017**, *45*, W36–W41.

[13] P. Cimermancic, M. H. Medema, J. Claesen, K. Kurita, L. C. Wieland Brown, K. Mavrommatis, A. Pati, P. A. Godfrey, M. Koehrsen, J. Clardy, et al., *Cell* **2014**, *158*, 412.

[14] A. C. Ross, L. E. S. Gulland, P. C. Dorrestein, B. S. Moore, *ACS Synth. Biol.* **2015**, *4*, 414.

[15] D. Mao, B. K. Okada, Y. Wu, F. Xu, M. R. Seyedsayamdost, *Curr. Opin. Microbiol.* **2018**, *45*, 156.

[16] J. J. Hug, C. D. Bader, M. Remškar, K. Cirnski, R. Müller, *Antibiotics* **2018**, *7*, 44.

[17] K. I. Mohr, R. O. Garcia, K. Gerth, H. Irschik, R. Müller, *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 1191.

[18] H. Steinmetz, K. I. Mohr, W. Zander, R. Jansen, K. Gerth, R. Müller, *J. Nat. Prod.* **2012**, *75*, 1803.

[19] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, et al., *Science* **2009**, *323*, 133.

[20] J. Feng, X. J. Chen, X. Sun, N. Wang, Y. Z. Li, *Plasmid* **2012**, *68*, 105.

[21] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, et al., *Nat. Biotechnol.* **2016**, *34*, 828.

[22] F. Panter, D. Krug, S. Baumann, R. Müller, *Chem. Sci.* **2018**, *9*, 4898.

[23] K. Yamanaka, K. A. Reynolds, R. D. Kersten, K. S. Ryan, D. J. Gonzalez, V. Nizet, P. C. Dorrestein, B. S. Moore, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1957.

[24] N. S. Cortina, D. Krug, A. Plaza, O. Revermann, R. Müller, *Angew. Chem. Int. Ed.* **2012**, *51*, 811; *Angew. Chem.* **2012**, *124*, 836.

[25] H. Shizuya, B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, M. Simon, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8794.

[26] J. J. Hug, F. Panter, D. Krug, R. Müller, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319.

[27] A. A. Iniesta, F. García-Heras, J. Abellón-Ruiz, A. Gallego-García, M. Elías-Arnanz, *J. Bacteriol.* **2012**, *194*, 5875.

[28] T. Hoffmann, D. Krug, S. Hüttel, R. Müller, *Anal. Chem.* **2014**, *86*, 10780.

[29] F. Panter, D. Krug, R. Müller, *ACS Chem. Biol.* **2019**, *14*, 88.

[30] M. Meusel, F. Hufsky, F. Panter, D. Krug, R. Müller, S. Böcker, *Anal. Chem.* **2016**, *88*, 7556.

[31] H. Duddeck in *Houben-Weyl Methods of organic chemistry*, Vol. E 21 a, Thieme, Stuttgart, **1995**.

[32] a) K.-i. Harada, K. Fujii, K. Hayashi, M. Suzuki, Y. Ikai, H. Oka, *Tetrahedron Lett.* **1996**, *37*, 3001; b) P. Marfey, *Carlsberg Res. Commun.* **1984**, *49*, 591.

[33] T. R. Hoye, C. S. Jeffrey, F. Shao, *Nat. Protoc.* **2007**, *2*, 2451.

[34] F. P. Maloney, L. Gerwick, W. H. Gerwick, D. H. Sherman, J. L. Smith, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10316.

[35] L. Kjaerulff, R. Raju, F. Panter, U. Scheid, R. Garcia, J. Herrmann, R. Müller, *Angew. Chem. Int. Ed.* **2017**, *56*, 9614; *Angew. Chem.* **2017**, *129*, 9743.

[36] J. J. Gehret, L. C. Gu, W. H. Gerwick, P. Wipf, D. H. Sherman, J. L. Smith, *J. Biol. Chem.* **2011**, *286*, 14445.

[37] N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth, C. Hertweck, *Chem. Biol.* **2012**, *19*, 1164.

[38] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J. L. Smith, *Structure* **2010**, *18*, 94.

[39] D. Pogorevc, F. Panter, C. Schillinger, R. Jansen, S. C. Wenzel, R. Müller, *Metab. Eng.* **2019**, *55*, 201.

[40] D. C. Gay, P. J. Spear, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2014**, *9*, 2374.

[41] B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, *Angew. Chem. Int. Ed.* **2010**, *49*, 1460; *Angew. Chem.* **2010**, *122*, 1502.

[42] M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, *Nucleic Acids Res.* **2011**, *39*, W362–W367.

[43] F. C. Beasley, J. Cheung, D. E. Heinrichs, *BMC Microbiol.* **2011**, *11*, 199.

[44] M. J. Kobylarz, J. C. Grigg, S.-i. J. Takayama, D. K. Rai, D. E. Heinrichs, M. E. P. Murphy, *Chem. Biol.* **2014**, *21*, 379.

[45] T. Hornbogen, S.-P. Riechers, B. Prinz, J. Schultchen, S. Lang, S. Schmidt, C. Mügge, S. Turkanovic, R. D. Süssmuth, E. Taubergerer, et al., *ChemBioChem* **2007**, *8*, 1048.

[46] C. T. Walsh, *Nat. Prod. Rep.* **2016**, *33*, 127.

[47] a) E. Yeh, L. C. Blasiak, A. Koglin, C. L. Drennan, C. T. Walsh, *Biochemistry* **2007**, *46*, 1284; b) X. Guan, D. An, G. Liu, H. Zhang, J. Gao, T. Zhou, G. Zhang, S. Zhang, *Tetrahedron Lett.* **2018**, *59*, 2418.

[48] a) A. Kantz, G. T. Gassner, *Biochemistry* **2011**, *50*, 523; b) E. Morrison, A. Kantz, G. T. Gassner, M. H. Sazinsky, *Biochemistry* **2013**, *52*, 6063.

[49] J. Muñoz-Dorado, F. J. Marcos-Torres, E. García-Bravo, A. Moraleda-Muñoz, J. Pérez, *Front. Microbiol.* **2016**, *7*, 781.