

Fischerazoles A-C, cyanobacterial polychlorinated lipids featuring fatty acyl chain rearrangement.

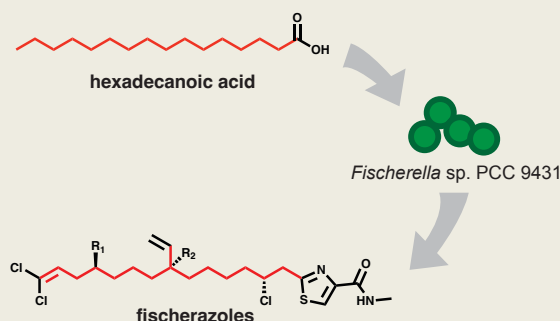
Sandra A. C. Figueiredo,^{1,§} Kathleen Abt,^{1,2,§} Teresa P. Martins,^{1,2} Iñaki Lacomba,¹ Abel M. Forero,³ Carlos Jiménez,³ Jaime Rodríguez,³ Pedro N. Leão^{1,*}

¹Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Avenida General Norton de Matos, s/n, 4450-208 Matosinhos, Portugal

²Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-313 Porto, Portugal

³Departamento de Química, Faculdade de Ciências, Centro de Investigações Científicas Avanzadas (CICA), Universidade da Coruña, 15071, A Coruña, Spain.

Abstract



Fatty acyl moieties are a very common feature of cyanobacterial secondary metabolites. Using a recently developed method to discover fatty acyl-containing natural products, we detected and isolated three new lipopeptides – fischerazoles A-C (**1-3**) – from the cyanobacterium *Fischerella* sp. PCC 9431. These metabolites present extensive halogenation in their fatty acyl moiety as well as an unusual pendant allyl alcohol. We analyzed the genome sequence of the producing cyanobacterium and identified a putative biosynthetic gene cluster associated with the fischerazoles, based on bioinformatic analysis. Using stable-isotope precursor feeding experiments, we identified the key substrates for the biosynthesis of **1-3**. Surprisingly, we found that hexadecanoic acid was fully incorporated into fischerazoles, despite the fact that these compounds have a linear 15-carbon chain. Additional feeding experiments with stable-isotope labeled fatty acids established a one-carbon contraction of the palmitic acid-derived alkyl chain, and that the remaining carbon branches out as part of the pendant allyl alcohol group. This rearrangement, which leads to the functionalization of a mid-chain aliphatic carbon, might be of interest for biocatalysis. Our findings reinforce the utility of metabolomics-based methods to uncover natural product scaffolds with a high degree of novelty.

Introduction

The discovery of a new natural product is often the starting point to reveal a pharmacologically relevant biological activity,¹ to uncover chemically mediated biological interactions,² or to discover uncharted biochemical transformations and associated enzymes.^{3,4}

The large number of secondary metabolite biosynthetic gene clusters (BGCs) found in publicly available genome and metagenome data greatly outnumbers currently known natural products.^{4,5} As a consequence, natural products discovery has been gradually shifting towards genome-guided approaches, as opposed to the more traditional bioassay-guided explorations.⁶ However, our ability to computationally recognize BGCs in genome data is greatly dependent on detecting sets of genes encoding well-established biochemistry.⁴ Even if a BGC can be pinpointed, it is often the case that reliable structural predictions for their associated natural products cannot be made, due to a large number of hypothetical proteins, low homology with known enzymes or non-canonical gene/domain organization.⁶

In contrast, BGC-agnostic discovery techniques, such as those based on activity screenings⁷ or metabolomics,⁸ have the potential to reveal highly novel natural products structures that would not be predictable based on contemporary biosynthetic knowledge. For the same reasons, such strategies are also likely to unearth structures generated by novel enzymatic chemistry.

Cyanobacteria have been extensively studied for their bioactive natural products, which include notable toxins, such as microcystins and drug leads, such as dolastatin 10.⁹ The chemical space occupied by cyanobacterial secondary metabolites is quite unique as illustrated by the intricate structures of the hapalindole-class metabolites¹⁰ or the lipidic curacins¹¹ and cylindrocyclophanes.¹² The study of the biosynthesis of cyanobacterial natural products has led to remarkable discoveries. Examples include unique polyketide biochemistry^{13,14}, the Friedel-Crafts alkylating enzyme CyLK,¹⁵ or the WelO5-like halogenases, which act on free-standing substrates.¹⁶ The phylum Cyanobacteria is still poorly covered in terms of genomic data, and even for currently available genomes, BGCs for which no natural product can be assigned (orphan BGCs) abound.¹⁷ Despite this richness of orphan BGCs, we currently lack genetic tools to manipulate most BGC-rich cyanobacteria.¹⁸ Therefore, BGC-guided natural product discovery in these organisms is typically restricted to leveraging structural predictions from BGC analysis^{19,20,20} or to the technically-challenging heterologous expression of the BGC.¹⁸

We have recently developed a BGC-agnostic metabolomics method aimed at the discovery of cyanobacterial natural products containing fatty-acid derived moieties.²¹ The method employs supplementation of cyanobacterial cultures with perdeuterated hexanoic acid. Cyanobacteria do not appreciably degrade exogenously supplemented fatty acids and instead elongate them and incorporate them into a wide range of metabolites.^{21,22} This method exploits this peculiarity to detect mass features that are shifted in mass spectra from cultures receiving perdeuterated fatty acid supplementa-

tion. For example, we used this method to discover several hapalolin analogues from *Fischerella* sp. PCC 9431.²¹ Here, we report the detection, in this same cyanobacterium, of a series of additional metabolites showing mass shifts for perdeuterated fatty acid incorporation. The compounds were isolated and their structures elucidated as extensively modified linear lipopeptides, which we named fischerazoles A-C (**1-3**). The biosynthesis of the fischerazoles was studied using bioinformatics and feeding experiments, which revealed an unusual alkyl chain rearrangement in **1-3**.

Results and Discussion

Discovery of the fischerazoles.

In our previous study,²¹ we supplemented *Fischerella* sp. PCC 9431 with pulses of *d*₁₁-hexanoic acid (*d*₁₁-C₆) and carried out LC-HRESIMS comparative analysis with a non-supplemented culture. After filtering out mass features presumably corresponding to primary metabolites, this analysis led to the identification of 124 labeled positive-mode MS features resulting from incorporation of deuterium atoms.²¹ Among these, a few exhibited *d*₇ and *d*₆ incorporation (Fig. 1A) and isotopic patterns consistent with multiple halogenations. A more detailed inspection clarified that these corresponded to four major compounds, for which in general [M+Na]⁺ ions were more abundant and showed *m/z* values of 537.067, 503.106, 521.073 and 487.111. Those that were labeled with *d*₇ (*m/z* 503 and 487) had an isotope pattern suggesting trichlorination and those labeled with *d*₆ (*m/z* 537 and 521) appeared to be tetrachlorinated. Dereplication of the corresponding high-resolution masses in Natural Products Atlas²³ indicated that these were likely new compounds and prompted their isolation. To this end, we performed large scale (80 L) culturing of *Fischerella* sp. PCC 9431. The resulting cyanobacterial cells, harvested in late exponential/stationary phase, were extracted with CH₂Cl₂/MeOH (2:1, v/v). LC-HRESIMS analysis of the resulting organic extract confirmed the presence of the target compounds. A series of normal- and reversed-phase chromatographic steps, guided by LC-HRESIMS analysis of the resulting fractions, led to the isolation of spectroscopically pure **1** (3.6 mg), **2** (2.4 mg) and **3** (1.0 mg) (Fig. 1B). A small amount (0.2 mg) of a partially purified metabolite with *m/z* 487.111, [M+Na]⁺ was also obtained, but additional purification was not pursued.

Compound **1**, which was isolated in higher amounts, was selected for initial structure elucidation based on combined analysis of HRESIMS (Fig. S1) and NMR (1D and 2D) data in DMSO-*d*₆ (Table 1, Fig. S2-S7). This tetrachlorinated metabolite showed a monoisotopic [M+H]⁺ ion of *m/z* 515.0852 (calcd 515.0855), corresponding to a molecular formula C₂₁H₃₀O₂N₂Cl₄S. The ¹H NMR data showed several conspicuous resonances, namely two protons (one quartet, one singlet) in the 8.5 to 8.0 ppm range, four olefinic protons (δ_H 6.2-5.0), two triplets of triplets (δ_H 4.45 and 4.16) that could correspond to aliphatic chlorinated or oxygenated methines, a sharp singlet resonating at δ_H 4.27 and a methyl doublet (δ_H 2.77)

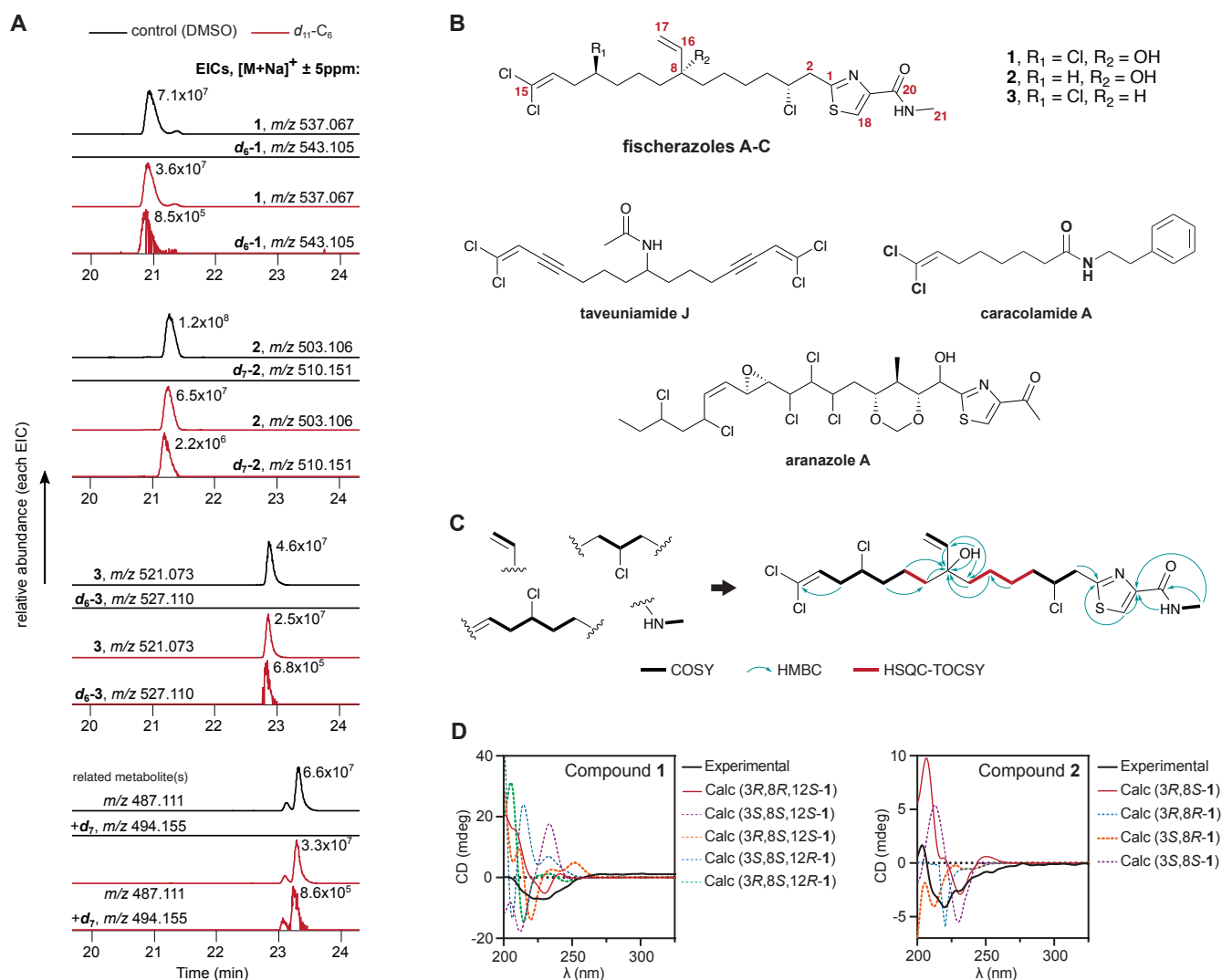


Figure 1. Discovery of fischerazoles A-C (**1-3**). A – LC-(+)-HERSIMS analysis of *Fischerella* sp. PCC 9431 cultures with or without supplementation of perdeuterated hexanoic acid ($d_{11}\text{-C}_6$) reveals a series of polyhalogenated mass features that indicate incorporation of six or seven deuterons. B – Structures of **1-3** and of cyanobacterial compounds sharing some structural features with these molecules. C – 1D and 2D NMR-based structure elucidation of **1**, showing the four spin systems obtained from COSY data, their extension using HSQC-TOCSY data and key HMBC correlations that connected these substructures. D – Assignment of absolute configuration to the stereogenic centers of **1** and **2**, based on the comparison of experimental and DFT-calculated ECD spectra.

coupled to the quartet at δ_{H} 8.28 ($J = 4.8$ Hz). Several other resonances were present in the spectrum, but an aliphatic methylene envelope, characteristic of unsubstituted fatty acyl moieties, could not be observed. ^{13}C NMR (APT) data revealed two signals in the carboxylic range (δ_{C} 166.5 and 160.9), six aromatic/olefinic carbons (δ_{C} 150-110, three of which protonated), one non-protonated carbon (δ_{C} 73.7), two carbons consistent with chlorinated or oxygenated methines (δ_{C} 62.5 and 61.9), several methylenes (δ_{C} 41-20) and a single methyl carbon (δ_{C} 25.8). Analysis of the combined 1D and 2D NMR data for **1**, in particular those of HSQC and COSY experiments, allowed to establish four spin systems, namely a terminal olefin, an *N*-methyl group, and two chlorinated aliphatic moieties (Fig. 1C). The terminal olefin protons (H-16, δ_{H} 5.75; H-17a, δ_{H} 5.13 and H-17b, δ_{H} 5.00) were HMBC-correlated to a non-protonated carbon (C-8, δ_{C} 73.7). A hydroxy group exchangeable proton singlet (δ_{H} 4.27) also exhibited an HMBC correlation to C-8.

Hence, this moiety was consistent with an allyl alcohol. The *N*-methyl spin system, on the other hand, could be connected to two deshielded carbons (C-19, δ_{C} 149.6 and C-20, δ_{C} 160.9) through HMBC correlations. These were particularly strong from H-21 (δ_{H} 2.77) to C-20 and from the NH proton (δ_{H} 8.28) to C-19; we considered that these were likely 3J correlations and in support of a substructure composed of an *N*-methylated amide connected to the fully substituted C-19. The singlet aromatic proton (H-18, δ_{H} 8.14) showed HMBC correlations to both C-19 and the fully substituted olefinic C-15 (δ_{C} 120.3). The resonances of C-19, CH-18 and C-1 were characteristic of a substituted thiazole, which would be consistent with the presence of the additional N and of the S atom in the molecular formula of **1**. HMBC-correlations indicated that one of the chlorinated aliphatic spin systems was connected to C-15, while the other was connected to C-1 (δ_{C} 166.5). In both spin systems, the other extremity showed COSY correlations to positions

with similar δ_{H} and δ_{C} values. HSQC-TOCSY was used to unambiguously extend these two chlorinated aliphatic spin systems (Fig. 1C, Fig S7-S8); these assignments were supported by extensive HMBC correlational data (Table 1). Finally, several HMBC correlations between the aliphatic chains and the allyl alcohol moiety enabled us to establish the connectivity between the different spin systems (Fig. 1C). The planar structure of **1** was established by assigning two Cl atoms as the substituents of the terminal olefinic C-15, which would be necessary to satisfy the molecular formula. (+)-HRESIMS/MS analysis supported the NMR-based structure elucidation, although the spectra interpretation was not straightforward due to HCl losses dominating low-energy fragmentations (Fig. S9).

The NMR-based structure elucidation of **2-3** used the proposed structure for **1** as reference. Compound **2** showed a monoisotopic $[\text{M}+\text{H}]^+$ ion of m/z 481.1242 (Fig. S10). This corresponded to a molecular formula $\text{C}_{21}\text{H}_{31}\text{O}_2\text{N}_2\text{Cl}_3\text{S}$ (calcd m/z 481.1245, $[\text{M}+\text{H}]^+$) and therefore **2** differs from **1** by one less Cl atom and one more proton. Comparison of the ^1H and ^{13}C NMR data for **2** (Figs. S11-12) to those of **1** allowed us to unambiguously determine that the difference lied at position 12. The NMR spectra of **2** display the

lack of the resonances for the CH-12 at δ_{C} 61.9/ δ_{H} 4.16 present in those of **1**, showing instead a larger methylene envelope. Additionally, the identical carbon and proton chemical shift values of the spin system containing CH-3 (C-2 to C-7) for both compounds (see NMR data for both compounds and Figs. S13-S14) is in agreement with the assignment. Compound **3** showed a characteristic $[\text{M}+\text{H}]^+$ ion cluster in its (+)-HRESIMS spectrum indicating the presence of four chlorine atoms. The complete formula of **3** as $\text{C}_{21}\text{H}_{30}\text{ON}_2\text{Cl}_4\text{S}$ was established from the monoisotopic $[\text{M}+\text{H}]^+$ ion peak at m/z 499.0908 (calcd 499.0906, Fig. S15). Therefore, the composition of **1** and **3** differed by a single O atom. We considered that in **3**, the hydroxyl group (OH-8) was not present. We acquired 1D and 2D NMR data for this compound in DMSO- d_6 (Table S1, Figs. S16-S20). Analysis of the combined NMR data indicated that the difference was in fact in position 8. The non-protonated C-8 and the exchangeable hydroxyl proton in **1** are not found in the NMR data for **3**, which instead features an additional methine (CH-8, δ_{C} 43.2, δ_{H} 1.94). ^1H and ^{13}C NMR resonances in the vicinity of this moiety are also shifted in **3**, relative to those in **1** (Table 1, Table S1). The NMR-derived structures for **2** and **3** were consistent with the respective (+)-HRESIMS/MS data (Fig. S9).

Table 1 – NMR Spectroscopic Data (^1H 600 MHz, ^{13}C 150 MHz, DMSO- d_6) for fischerazole A (**1**).

Position	δ_{C}	Type	δ_{H}^a	mult., J (Hz)	HMBC ^b	COSY
1	166.5	C=N	-	-	-	-
2a	41.2	CH ₂	3.53	dd, 15.4, 4.4	1, 3, 4, 18, 19	2b, 3
2b			3.38	dd, 15.4, 8.8	1, 3, 4	2a, 3
3	62.5	CH	4.45	tt, 8.7, 4.3	1, 2, 4, 5	2a, 2b, 4a, 4b
4a	37.3	CH ₂	1.84	m	2, 3, 5, 6	3, 4b, 5a, 5b
4b			1.71	m	2, 3, 5, 6	3, 4a, 5a, 5b
5a	26.3	CH ₂	1.46	m	3, 4, 6, 7	4a, 4b, 5b, 6b
5b			1.39	m	3, 4, 6, 7	4a, 4b, 5a, 6b
6a	22.6	CH ₂	1.33	m	5, 7, 8	6b
6b			1.20	m	5, 7, 8	5a, 5b, 6a, 7
7	40.4	CH ₂	1.40	m	4, 5, 6, 8, 9, 16	6b
8	73.7	C	-	-	-	-
		OH	4.27	s	8, 9, 16	-
9	39.8 ^c	CH ₂	1.40	m	8, 10, 16	-
10a	20.2	CH ₂	1.42	m	11, 12	10b, 11a, 11b
10b			1.37	m	11, 12	10a, 11a, 11b
11a	37.8	CH ₂	1.71	m	9, 10, 13	10a, 10b, 11b, 14
11b			1.62	m	9, 10, 12, 13	10a, 10b, 11a, 14
12	61.9	CH	4.16	tt, 8.8, 4.5	10, 11, 14	11a, 11b, 13a, 13b
13a	38.1	CH ₂	2.64	ddd, 15.3, 6.6, 4.6	11, 12, 14, 15	12, 13b, 14
13b			2.55	ddd, 15.4, 7.8, 7.8	11, 12, 14, 15	12, 13a, 14
14	127.4	CH	6.18	dd, 7.7, 6.6	12, 13, 15	13a, 13b
15	120.1	C	-	-	-	-
16	144.6	CH	5.75	dd, 17.3, 10.7	7, 8, 9	17a, 17b
17a	112.0	CH ₂	5.13	dd, 17.3, 2.2	7, 8, 9	16, 17b
17b			5.00	dd, 10.7, 2.2	7, 8, 9	16, 17a
18	123.6	CH	8.14	s	1, 19, 20	-
19	149.6	C=C	-	-	-	-
20	160.9	C=O	-	-	-	-
		NH	8.28	q, 4.8	18, 19, 20, 12	21
21	25.8	CH ₃	2.77	d, 4.8	19, 20	20-NH

^afrom HSQC; ^bfrom proton to indicated carbon; ^cextracted from HMBC.

To determine the absolute configuration of the chiral centers in the fischerazoles, we resorted to ECD measurements for compounds **2** (two stereogenic centers) and **1** (three stereogenic centers), which were compared with simulated ECD spectra obtained by time-dependent density functional theory (TDDFT)^{24,25} at the CAM-B3LYP/6-311++G(2d,p) in MeOH. The overall calculated ECD spectra were generated by Boltzmann weighting of their lower energy conformers by relative free energies. The absolute configurations were then extrapolated by comparison of the calculated and experimental ECD spectra of **1** and **2**. The experimental ECD spectrum of **1** was nearly identical to the calculated ECD curve of 3*R*,8*S*,12*S* (Fig 1D), suggesting the 3*R*,8*S*,12*S* absolute configuration. Comparison of the experimental and calculated ECD spectra of **3** displayed excellent agreement with 3*R*,8*R* diastereoisomer (Fig 1D). Similar configurations are assumed for the stereogenic centers in **3**. Compounds **1-3** bear distinct structural features from previously described secondary metabolites. In particular, we were unable to find another natural product with a pendant allyl alcohol moiety. The *gem*-dichlorovinylidene aliphatic chain termini are rare in nature²⁶ and found in two other cyanobacterial natural products – caracolamide A²⁶ and the taveuniamides²⁷ (Fig. 1B). The peptidic portion of **1-3** has some resemblance to the aranazoles²⁸ (Fig. 1B), which are also extensively chlorinated cyanobacterial lipopeptides. We could not find additional examples of the methyl carboxamide-

thiazole motif among natural products; this seems to be a unique feature of the fischerazoles.

Identification of the *fsh* biosynthetic gene cluster

Considering the tetrachlorination of fischerazoles A and C, we expected the presence of several halogenases in the corresponding BGC. We therefore mined the genome of *Fischerella* sp. PCC 9431 for different characterized halogenases known to act on alkyl moieties, namely dimetal-carboxylate halogenases²⁹ and iron(II)/ α -ketoglutarate-dependent halogenases.³⁰ We detected several BGCs that contained more than one halogenase, so we examined these BGCs further, particularly for genes that would explain the structural feature of a thiazole. This prompted us to consider a single candidate genomic region predicted to contain, among others, two CylC homologs, one iron(II)/ α -ketoglutarate-dependent halogenase and a non-ribosomal peptide synthetase (NRPS) that contained the necessary domains to create a thiazole moiety. As we established that this putative BGC encoded additional biosynthetic machinery that could enable the production of the fischerazoles, we named it the *fsh* BGC (Figure 2A, Table S2). FshA is a putative fatty acyl-AMP ligase (FAAL) responsible for activating and loading a fatty acid onto an acyl carrier protein (ACP). FshE and FshM are two free-standing ACPs which could be the acceptors of the activated fatty acid. This insight resulted from analysis of their N-terminal docking domains, which showed highest homology to

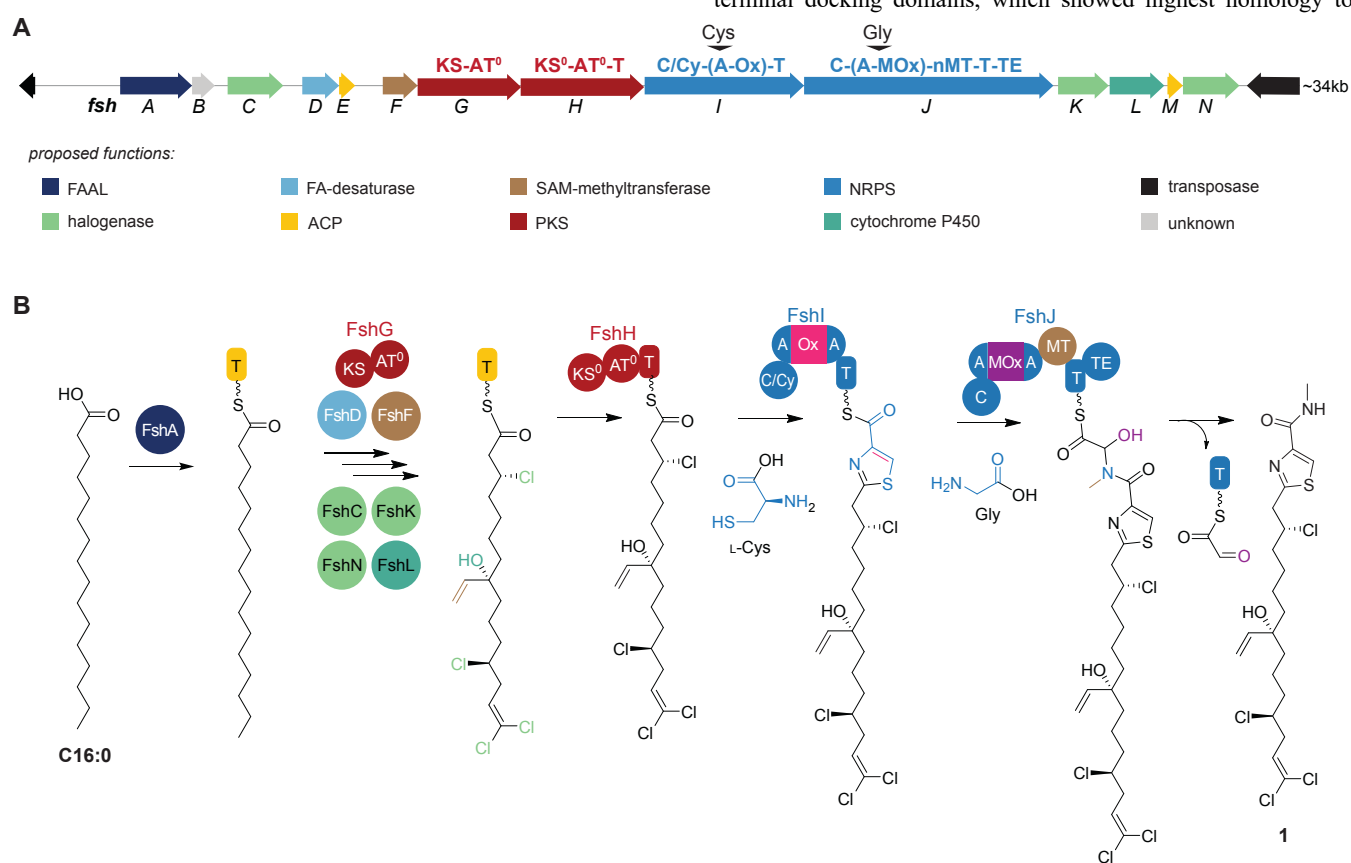


Figure 2. Proposed biosynthesis of the fischerazoles in *Fischerella* sp. PCC 9431, based on bioinformatics analysis and substrate feeding experiments. **A** – The putative fischerazole BGC (*fsh*) and bioinformatics-based annotation of its encoded gene products. **B** – Biosynthetic proposal for the generation of **1**, involving fatty acid activation, a series of acyl chain tailoring events, incorporation of Cys to generate the thiazole, followed by Gly incorporation/*N*-methylation and the unusual termination through the release of a glyoxylidene-T domain thioester.

FAAL-ACP interfaces (Figure S21). The C-terminal docking domains of the two ACPs aligned best with FAAL-ACP to polyketide synthase (PKS) interfaces, suggesting a downstream transfer of the substrate from these ACPs to the PKS FshH (Figure 2B). The *fsh*-encoded halogenases comprise a homolog of the dimetal-carboxylate halogenase CylC (FshC, 45% identity), a homolog of the Fe(II)/ α -ketoglutarate-dependent halogenase ArzI from the aranzoles BGC (FshK, 70% identity) and another CylC homolog (FshN, 47% identity), which together install up to four chlorine atoms in the herein reported fischerazoles. FshD is annotated as a membrane-associated fatty acid desaturase and could generate the unsaturation in the *gem*-dichlorovinylidene terminus of **1-3**. FshF is predicted to be a SAM-methyltransferase of the mycolic acid cyclopropane synthetase superfamily. These enzymes are known to generate cyclopropane moieties, *O*-Me and *C*-Me groups in fatty acid-derived substrates.³¹ However, if FshF could indeed create a cyclopropane during the biosynthesis of the fischerazoles is uncertain at this point. The PKS FshG is composed of a ketosynthase (KS) and acyltransferase (AT) domain. A thiolation (T) domain, which together with one KS and one AT domain configures a minimal PKS module, is absent in FshG. A second PKS, FshH, contains all three minimal PKS domains. The AT domains of both FshG and FshH and the KS domain of FshH are predicted by antiSMASH to be inactive. This prediction is supported by sequence alignments with known active PKS domains (Figure S22). Analysis of the FshH docking domains indicated an upstream connectivity to the FAAL-ACPs and downstream to an NRPS, in this case FshI whose N-terminal docking domain matches PKS connectivity (Figure S21). This suggests that although inactive in their canonical function, the PKSs are still involved in fischerazole biosynthesis. The NRPS FshI consists of a condensation/cyclization (C/Cy) domain, of an interrupted adenylation (A) domain containing an oxidase (Ox) domain between the conserved motifs a8 and a9³² and of a thiolation (T) domain. FshI is predicted (AntiSMASH) to carry out the key biosynthetic step of incorporating a cysteine and, according to its domain structure, forming the thiazole heterocycle as previously described for other natural products such as epothilone³³ (Figure 2B). The second NRPS FshJ consists of a condensation (C) domain, an adenylation (A) domain which is interrupted between the core motifs a4 and a5 by a monooxygenase (MOx) domain³² and followed by N-methyltransferase (nMT), thiolation (T) and thioesterase (TE) domains. The architecture of this module is unprecedented and similar only to myxothiazol and melithiazol biosynthesis.^{32,34,35} The predicted amino acid specificity for FshJ could not be determined by antiSMASH. However, after *in silico* excision of the MOx domain, antiSMASH recognized the A domain as glycine-specific (Fig. S23). Loading and N-methylation of glycine is in accordance with the fischerazole structures and suggests a biosynthetic step as for myxothiazol/melithiazol. In this step, after incorporation, the carbon backbone of glycine is excised, and the product is released with a terminal carboxamide. In the case of the fischerazoles, this would be a methylated carboxamide (Figure 2B). We speculate that the cytochrome P450 FshL performs the C-8 hydroxylation on fischerazole A and B and may be involved in additional chemistry leading to the pendant allyl alcohol. Finally, FshB has no significant homology to any characterized protein and

therefore is of unknown function. The *fsh* BGC is flanked by transposases on both sides, supporting the FAAL FshA and the halogenase FshN as its first and last gene, respectively.

Fischerazole biosynthesis involves carbon rearrangement of hexadecanoic acid

To learn more about the biosynthesis of the fischerazoles, in particular the unusual C₁₅ chain with a pendant allyl (alcohol), we performed stable isotope incorporation experiments. As mentioned before, in cyanobacteria, deuterated fatty acids do not lose their label to being catabolized, but are elongated and incorporated into various metabolites.^{21,36} We used a sequence of perdeuterated even-chain fatty acids from octanoic to octadecanoic acid to supplement *Fischerella* sp. PCC 9431 cultures. We detected label incorporation into the fischerazoles from all fatty acids up to hexadecanoic acid (*d*₃₁-C₁₆) (Figure 3A, Figure S24-25). The activation of a C₁₆ fatty acid for the biosynthesis of **1-3** was surprising, but – at the same time – consistent with non-elongating PKS modules, as predicted bioinformatically. With no more than 15 carbons in the linear chain of **1-3**, we considered that this label incorporation could be explained only through a rearrangement of the carbon skeleton into a branched chain, ultimately resulting in the pendant allyl alcohol. Carbon rearrangements can be mediated by cytochrome P450s,³⁷ an enzyme that is encoded in the *fsh* BGC. Still, a rearrangement of this kind, in which a fatty acyl linear chain is contracted to generate a side chain, has no precedent in biological chemistry, and we investigated this hypothesis further.

The deuterium incorporation into the fischerazoles (data for **1** and **3** are presented), from *d*₁₅-C₈ to *d*₃₁-C₁₆ feeding, was consistent with the structures of **1** and **3** derived from NMR (Fig. 3A, Fig. S24). Additional losses of deuterons occurred only for *d*₁₉-C₁₀ and *d*₃₁-C₁₆, corresponding to the positions of the allyl alcohol on C-8 and chlorination on C-3. Notably, the incorporation of only 11 deuterons from *d*₁₉-C₁₀ into **1** (Figure 3A) and 12 deuterons into **3** (Figure S24) further supports the fatty acyl-origin of the allyl group C-16 atom. The stand-alone SAM-methyltransferase FshF and NRPS nMT domain of FshJ suggested SAM-dependent methylation in two positions of **1-3**. In particular, we considered that one of the allyl group carbons should originate from SAM, since the other would originate from hexadecanoic acid. To validate this, we used methyl-¹³C,*d*₃-L-methionine for precursor feeding. Indeed, we observed a mass shift corresponding to the incorporation of two methyl groups from methionine into **1** and **3**, one of which had lost one deuteron (Figure 3B). This was in accordance with one methylation giving rise to C-17 position in the allylic double bond and the other being part of the methyl carboxamide moiety (C-21 position). Similarly, we sought to confirm hexadecanoic acid incorporation into **1** and **3** by supplementation with [U-¹³C]-hexadecanoic acid. We detected masses corresponding to ¹³C₁₆-labeled **1** and **3** (Figure 3C, Fig. S24). To assess the position of the labels from the different feeding experiments in the fischerazoles, we resorted to HRESIMS/MS analysis of **3** which allowed the best [M+H]⁺ detection. Fragmentation of **3** resulted in characteristic loss of HCl but also the N-methyl group, which for the ¹³C₂,*d*₅-**3** isotopologue indicated label incorporation in this position from ¹³C,*d*₃-L-methionine (Figure 3D). Additional diagnostic fragments supported the proposed SAM-derived methylation on both the allyl chain and the

terminal carboxamide. On the other hand, fragmentation of $^{13}\text{C}_{16}\text{-3}$ generated several diagnostic fragments indicating the labels were restricted to the alkyl-chain and allyl group (Figure 3D). We

thereby confirmed our hypothesis that the fischerazoles undergo a carbon rearrangement during their biosynthesis.

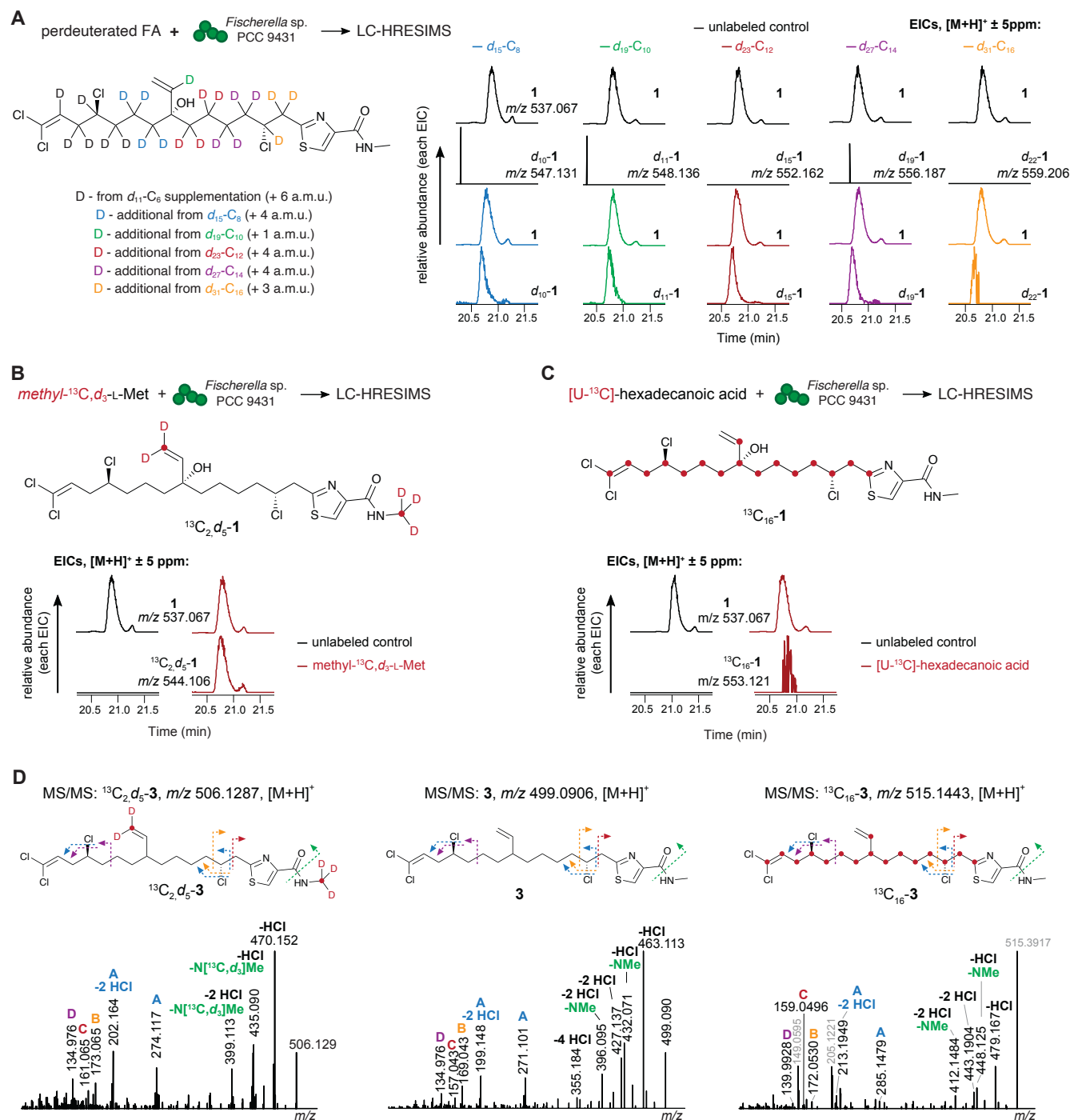


Figure 3. Substrate feeding experiments reveal an alkyl chain rearrangement in fischerazole biosynthesis. Proposed incorporation pattern into **1** and LC-HRESIMS analysis of organic extracts of *Fischerella* sp. PCC 9431 cell biomass from cultures supplemented with: A – a perdeuterated fatty acid (hexanoic, octanoic, decanoic, dodecanoic, tetradecanoic and hexadecanoic acids were used in the experiments). EICs for the $[\text{M}+\text{H}]^+$ ions corresponding to non-labeled and observed deuterium-labeled **1**. (EICs for perdeuterated hexanoic acid ($d_{11}\text{-C}_6$) feeding experiments are as show in Fig. 1); B – L-Met-(methyl- $^{13}\text{C}_2, d_3$); C – $[\text{U-}^{13}\text{C}]$ -hexadecanoic acid. D – LC-(+)-HRESIMS/MS of compound **3** (as a reference) and of heavier versions resulting from $^{13}\text{C}_2, d_3$ incorporation (L-Met-(methyl- $^{13}\text{C}_2, d_3$) feeding) and from $^{13}\text{C}_{16}$ incorporation ($[\text{U-}^{13}\text{C}]$ -hexadecanoic acid feeding), highlighting several diagnostic fragments that support the proposed heavy atom incorporation pattern.

Taking into account the bioinformatic analysis and substrate feeding studies, we propose that the fatty acyl-AMP ligase FshA activates and loads hexadecanoic acid onto one or both of the stand-alone ACPs FshE/FshM (Figure 2B). The tethered substrate is extensively modified by the halogenases FshC, FshK and FshN, the desaturase FshD, the SAM-methyltransferase FshF and the cytochrome P450 FshL. The highly modified fatty acyl chain would then be passed on to the adjacent thiolation domain. While several domains in the PKSs FshG and FshH are bioinformatically predicted to be inactive, we cannot exclude that these PKSs are involved in fischerazole formation beyond substrate shuttling. Such non-canonical activity could potentially be involved in the intriguing C₁₆ rearrangement. The last steps of fischerazole biosynthesis are more aligned with currently known biosynthetic logic. We propose that cysteine is incorporated and cyclized by the first NRPS FshI, thereby creating the thiazole. The subsequent NRPS FshJ would incorporate and N-methylate a glycine, oxygenate the glycine to excise the glyoxylyl-FshJ thioester and release the fischerazole with a terminal N-methylated carboxamide (Figure 2B). The TE domain of FshJ would release glyoxylate from the T domain to allow for further catalytic cycles.

Conclusions

We have used a combined perdeuterated fatty acid feeding and metabolomics approach²¹ to uncover several polychlorinated lipopeptides (**1-3**). These metabolites contain unusual structural features, such as a pendant allyl alcohol moiety, an *N*-methylcarboxamide moiety and a *gem*-dichlorovinylidene aliphatic chain terminus. By identifying a candidate fischerazole BGC (*fsh*), we clarified several aspects of their biosynthesis. The biosynthetic events leading to the rare *gem*-dichlorovinylidene moiety are unclear at this stage, but by identifying the *fsh* BGC, our study provides the groundwork for their dissection. Most strikingly, we found that the alkyl chain scaffold of **1-3** derives from hexadecanoic acid, which is rearranged to eventually generate the unique pendant allyl moiety. This is likely to occur as the result of the concerted action of multiple enzymes and deserves to be further investigated. Rearrangements of this kind, which happen with concomitant mid-chain functionalization of fatty acid-derived moieties, provide access to unactivated carbon centers and could in principle be explored for biocatalytic strategies.³ Our study illustrates the usefulness of BGC-agnostic methods to reveal both natural products with a high degree of structural novelty as well as unusual biosynthetic events.

Methods

General experimental procedures

Stable-isotope labeled fatty acids (*d*₁₁-hexanoic, *d*₁₅-octanoic, *d*₁₉-decanoic *d*₂₃-dodecanoic, *d*₂₇-tetradecanoic, *d*₃₁-hexadecanoic acids and potassium *d*₃₅-octadecanoate) were obtained from CDN Isotopes Inc. [U-¹³C]-hexadecanoic acid was obtained from Cambridge Isotope Laboratories. L-methionine-(methyl-¹³C, *d*₃) sodium salt was obtained from Sigma-Aldrich. All solvents, purchased from VWR Chemicals, Thermo Fisher Scientific, Carlo Erba and Chem Lab NV were ACS grade, except for HPLC solvents

(HPLC gradient grade) and LC-MS solvents (MS-grade). Deuterated solvents for NMR were purchased from Cambridge Isotope Laboratories and Alfa-Aesar. LC-HRESIMS and LC-HRESIMS/MS analyses were obtained on a Thermo Scientific UltiMate 3000 HPLC system, which consists of the following components: LPG-3400RS pump, WPS-3000RS autosampler, TCC-3x00RS column compartment, and MWD-3000RS UV/VIS detector coupled to a Q Exactive Focus Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Fisher Scientific), controlled by Q Exactive Focus (Exactive Series) 2.9 and Thermo Scientific Xcalibur 4.1.31.9 software. Optical rotations were obtained on a P-2000 polarimeter (Jasco) in a 1 mL sample cell. The electronic circular dichroism (ECD) spectrum of **1** and **2** was acquired on a J-815 CD Spectrometer (Jasco). UV spectra were acquired on a UV-1600PC spectrophotometer (VWR). Infrared spectra were determined on a Thermo Scientific Nicolet™ iS™ 5 Fourier transform IR spectrometer controlled by OMNIC 9.8.372 software. NMR spectra were acquired using a DMSO-matched Shigemi tube on a Bruker Avance III HD, 600 MHz, equipped with a 5 mm cryoprobe and controlled by TopSpin 3.6.1., in the Materials Center of the University of Porto (CEMUP) and the NMR data were analyzed in MestReNova 12.0.3 (MestrelabResearch). The chemical shifts values (δ) are presented in parts per million (ppm) and the coupling constants (*J*) in hertz (Hz). The purification of **1-3** was performed on a Thermo Scientific UltiMate 3000 HPLC system and fitted with an ACE 10 C18 column (250 mm × 10 mm, 10 μ m, 100 Å, ACE).

Strains and culture conditions

The cyanobacterium *Fischerella* sp. PCC 9431 was acquired from Pasteur Culture Collection of Cyanobacteria (PCC), Institute Pasteur (France). Cyanobacterial cultures were carried out in glass Erlenmeyer flasks (small-scale culturing) or in 20-L polycarbonate carboys (large-scale culturing), in Z8 medium with aeration, 25 °C, 30 μ mol photons m⁻² s⁻¹ (16 h light/ 8 h dark cycle). For large-scale cultures, the biomass was harvested in late-exponential to early-stationary phase by centrifugation (5000×g), freeze-dried and stored at -20 °C until further use.

LC-HRESIMS and MS/MS analysis

HRESIMS data were obtained in Full Scan positive mode with a capillary voltage of HESI set to 3.5 kV and the capillary temperature to 263 °C. Sheath gas flow rate was set to 50 units. For LC-HRESIMS analyses, separation was performed in an ACE UltraCore 2.5 SuperC18 column (75 x 2.1 mm, ACE).

For flow injection analysis (during HPLC isolation of **1-3**), the ESI parameters were 3.8 kV a capillary temperature of 300 °C and a sheath gas flow rate of 5 units.

Samples from stable isotope supplementation experiments (full MS) and pure compounds were separated at a flow rate of 0.3 mL min⁻¹ with a mobile phase of 0.1% formic acid in H₂O (eluent A), MeOH (eluent B) and iPrOH (eluent C). The gradient for the crude extracts started with 5% B increasing to 100% B over 20 min, held for 15 min, then B was decreased to 30% while C was increased to 70% over 5 min which was held for 7 min before returning to initial conditions over 2 min.

Samples during the isolation of **1-3** (after VLC, flash chromatography and HPLC) were separated with a shorter gradient at a flow rate of 0.4 mL min⁻¹ and employed a mobile phase of 0.1% formic acid in MeOH/H₂O 1:1 (v/v) (eluent A) and in iPrOH (eluent B). The gradient started with 10% B for 1 min, a gradient from 10% to 65% B over 5 min, held at 65% B for 12 min,

then a gradient to 85% B over 2 min, and held again at 85% B for 9 min, before returning to the initial conditions.

Samples from stable isotope supplementation experiments for MS/MS analysis were separated over a similar shorter gradient of MeOH/H₂O 1:1 (v/v) (eluent A) and in iPrOH (eluent B), both containing 0.1% formic acid. Starting with 10% B for 1 min, the gradient increased to 23% B over 2 min, which was held for 10 min until another increase to 90% B over 2 min. This was held for 7 min, then returned to initial conditions in 2 min.

For LC-HRESIMS analysis, the parameters used were a resolution of 70,000 with an AGC target of 1×10^6 and a scan range of 150–2,000 *m/z*. Extracts from stable isotope labelling experiments were analyzed with the same parameters but a scan range from 100 to 1,500 *m/z*.

LC-HRESIMS/MS analysis was performed using a resolution of 35 000, with a 0.4 *m/z* isolation window, a loop count of 1, AGC target of 2×10^5 and collision energy of 25 (arbitrary units). For low abundance compounds, the full MS scan range was between *m/z* 400 and 650.

Identification of new mass features in LC-HRESIMS data from the initial *d*₁₁-hexanoic acid substrate incorporation experiments was performed with MZmine 2.53 as previously described.²¹

Extraction and MS-guided isolation of compounds 1–3

The freeze-dried biomass resulting from large-scale (80 L) culturing of *Fischerella* sp. PCC 9431 (24.5 g, d.w.) was repeatedly extracted by immersion on a 2:1 mixture of CH₂Cl₂/MeOH. After filtration of the resulting slurry and solvent evaporation, 3.2 g of crude extract were obtained. The extract was subjected to normal-phase (Si Gel 60, 0.040–0.063 mm, Merck) Vacuum Liquid Chromatography (VLC). A gradient of increasing solvent polarity, from *n*-hexane to EtOAc to MeOH, resulting in twelve fractions (A – L). All fractions were subjected to LC-HRESIMS analysis to determine those which contained the mass features corresponding to the poly-halogenated compounds of interest. Fractions F, G and H, eluting between 3:2 and 1:4 *n*-hexane/EtOAc (v/v), were enriched in target ([M+H]⁺) *m/z* 515.085 and 481.124, and were pooled (combined: 592.2 mg) for further separation. Likewise, fractions D and E, eluting with 4:1 and 7:3 *n*-hexane/EtOAc (v/v), respectively, were enriched in *m/z* 499.091 and 465.130 ([M+H]⁺) and were combined (520.3 mg) before being further fractionated. Pooled fractions F/G/H were loaded onto a 12 g normal phase silica cartridge (Silicycle), which was fitted to a flash chromatography system (Pure C-815 Flash, Büchi). A gradient from *n*-hexane to EtOAc to MeOH was used for the automated separation. A total of 247 collected fractions were pooled into 18 fractions on the basis of their ELSD and UV profiles. These were analyzed by LC-HRESIMS, which revealed that fractions 8 (16.1 mg) and 9 (15.5 mg) were highly enriched in the two target compounds (*m/z* 515.085 and 481.124, [M+H]⁺). In addition, fractions 3–5 (combined: 16.4 mg) were enriched in the other two target mass features (*m/z* 499.091 and 465.130). Semi-preparative HPLC separation of fraction 8 (gradient of 50% MeOH aq. (solvent A) and iPrOH (solvent B) starting with isocratic conditions of 33% B for 25 min, increasing to 90% B over 5 min, holding for 10 min and returning to initial conditions over 5 min), afforded pure **1** (t_R = 22.5 min) and **2** (t_R = 26.5 min). Similar HPLC conditions were used to further separate fraction 9 (the initial isocratic step was held for 30 min as the only change), which also afforded pure **1** and **2**. The two aliquots for each compound were pooled to give 3.6 mg of **1** (0.015% of biomass d.w.) and 2.0 mg of **2** (0.008% of biomass d.w.). Pooled fractions D/E and flash

chromatography fractions 3–5 were combined and further separated by automated flash chromatography using the same conditions as described above. This procedure afforded 13 fractions, which were analyzed by LC-HRESIMS. The target mass features were present in two fractions with a combined mass of 320.0 mg. These were pooled together and further fractionated by a new round of automated flash chromatography, this time using a 4g normal-phase silica column (Silicycle). Once again, the resulting fractions were pooled on the basis of ELSD and UV profiles (to give 14 fractions) and analyzed by LC-HRESIMS. This revealed that the mass features of interest were highly enriched in fraction 7 (36.0 mg). This fraction was further separated by semi-preparative HPLC (gradient of 50% MeOH aq. – solvent A – and iPrOH – solvent B – starting with isocratic conditions of 55% B for 23 min, increasing to 95% B over 3 min, holding for 7 min and returning to initial conditions over 3 min) to afford pure **3** (1.0 mg, 0.004% biomass d.w., t_R = 15.0 min) and a small amount (0.2 mg) of a partially purified analogue with an [M+H]⁺ ion of *m/z* 465.130 (t_R = 18.0 min).

Fischerazole A (1) (pale yellow amorphous solid): [α_D^{24}] +14.7 (c 0.1, MeOH); IR (thin film) ν_{\max} 3412, 2919, 2851, 1651, 1553, 1496, 1462, 1413, 1028 cm⁻¹; UV (MeOH) λ_{\max} (log ϵ) 234 (3.4), 281 (2.7); ¹H and ¹³C NMR data, see Table 1; HRESIMS *m/z* 515.0852 [M+H]⁺ (calcd. for C₂₁H₃₁O₂N₂³⁵Cl₄S 515.0855).

Fischerazole B (2) (pale yellow amorphous solid): [α_D^{24}] +17.6 (c 0.1, MeOH); IR (thin film) ν_{\max} 3417, 2936, 2858, 1658, 1651, 1556, 1495, 1463, 1412, 1311, 1252, 1108, 996, 918, 878, 506 cm⁻¹; UV (MeOH) λ_{\max} (log ϵ) 239 (3.0), 276 (2.0); ¹H NMR (600 MHz, DMSO-*d*₆) δ 8.28 (q, *J* = 4.9 Hz, 1H), 8.14 (s, 1H), 6.12 (t, *J* = 7.4 Hz, 1H), 5.74 (dd, *J* = 17.2, 10.8 Hz, 1H), 5.11 (dd, *J* = 17.3, 2.3 Hz, 1H), 4.98 (dd, *J* = 10.8, 2.3 Hz, 1H), 4.44 (tt, *J* = 8.7, 4.3 Hz, 1H), 4.20 (s, 1H), 3.53 (dd, *J* = 15.4, 4.4 Hz, 1H), 3.38 (dd, *J* = 15.4, 8.8 Hz, 1H), 2.77 (d, *J* = 4.8 Hz, 3H), 2.11 (q, *J* = 7.4 Hz, 2H), 1.88 – 1.80 (m, 1H), 1.71 (dtd, *J* = 14.1, 9.4, 4.6 Hz, 1H), 1.46 (m, 2H), 1.42 – 1.31 (m, 8H), 1.31 – 1.13 (m, 6H); ¹³C NMR (150 MHz, DMSO-*d*₆) δ 166.5, 160.9, 149.6, 144.8, 131.2, 123.6, 117.9, 111.8, 73.7, 62.5, 41.2, 40.4, 40.3, 37.3, 29.2, 29.1, 27.5, 26.3, 25.8, 22.8, 22.6; HRESIMS *m/z* 481.1242 [M+H]⁺ (calcd. for C₂₁H₃₂O₂N₂³⁵Cl₅S 481.1245).

Fischerazole C (3) (pale yellow amorphous solid): [α_D^{24}] +41.4 (c 0.1, MeOH); IR (thin film) ν_{\max} 3418, 2922, 2859, 1659, 1551, 1055, 1033 cm⁻¹; UV (MeOH) λ_{\max} (log ϵ) 238 (3.4), 281 (3.1); ¹H and ¹³C NMR data, see Table S1; HRESIMS *m/z* 499.0908 [M+H]⁺ (calcd. for C₂₁H₃₁O₂N₂³⁵Cl₄S 499.0906).

ECD measurements and computational simulations

Conformational search was conducted using Free Maestro 11.4 (Schrödinger) with force field OPLS_2005, RMSD cutoff of 1.2 Å and a dielectric constant of 32.70 (MeOH). DFT calculations were performed via the Gaussian 09 program package. Each stereoisomer was geometrically optimized using the DFT method at the B3LYP/6-31 G(d,p) PCM level in MeOH. Single conformers with no negative frequencies, were selected to ECD calculations. The ECD spectrum was calculated using the TDDFT method at the CAM-B3LYP/6-311++G(2d,p) level in methanol. The ECD spectra were generated by SpecDis1 software using a Gaussian band shape with a 0.3 eV exponential half-width from dipole-length dipolar and rotational strengths.

Identification of the *fsh* BGC

The genome data of PCC 9431 (NCBI accession number: ALVX00000000) was used to populate a local BLAST database with the Geneious software package 11.1.4 (Biomatters). tBLASTn searches against this database, using the halogenases CylC (NCBI: ARU81117.1) and ArzI (NCBI: DAB41911.1) revealed the candidate *fsh* gene locus (accession numbers in Table S2). The genome was submitted to AntiSMASH (6.0.1 bacterial version) for BGC annotation. Manual annotation of the Fsh proteins was performed using BLASTp searches against the NCBI nr database. Domains were annotated based on NCBI BLAST and InterProScan from within Geneious. Docking domains for sequence alignments were downloaded from the DDAP database,³⁸ containing 382 entries at the time of accession (November 2020). Sequences of C-terminal heads were aligned using the 100 C-terminal residues of FshA, FshE, FshM, FshG, FshH and FshI; sequences of N-terminal tails were aligned using the 50 N-terminal residues of FshE, FshM, FshH, FshI and FshJ, by using MAFFT with a BLOSUM62 matrix in Geneious.

Stable-isotope labeled substrate incorporation experiments

The initial supplementation of *Fischerella* sp. PCC 9431 with *d*₁₁-hexanoic (*d*₁₁-C₆) was performed in three pulses with a final concentration of 0.5 mM in triplicates as previously described.²¹ Small-scale (25 mL) cultures in Z8 medium in Erlenmeyer flasks (starting chlorophyll *a* content of ~0.8–2.9 µg mL⁻¹), were inoculated with *d*₁₅-octanoic acid (*d*₁₅-C₈), *d*₁₉-decanoic (*d*₁₉-C₁₀), *d*₂₃-dodecanoic (*d*₂₃-C₁₂), *d*₂₇-tetradecanoic acid (*d*₂₇-C₁₄) and *d*₃₁-hexadecanoic (*d*₃₁-C₁₆) acids, to a final concentration of 0.5 mM from a 1000× concentrated solution of each acid in DMSO. L-methionine-(*methyl*-¹³C, *d*₃) was provided directly to the culture at 0.5 mM. Potassium *d*₃₅-octadecanoate (*d*₃₅-C₁₈) was provided at a final concentration of 0.25 mM from a 1000× concentrated solution in H₂O. For hexanoic acid, octanoic acid and methionine, DMSO was used as a control, for all others, the control was the respective non-labeled substrate. After a seven-day growth period with continuous orbital shaking (200 rpm, mini shaker, VWR), cells were harvested by centrifugation at 5000 ×g for 15 min, rinsed with deionized water, centrifuged again and the water drained as much as possible. The fresh biomass was extracted with ca. 7 mL of CH₂Cl₂/MeOH (2:1, v/v) and vigorous vortexing. All samples were incubated for 30 min on a nutating mixer in solvent-resistant tubes. The extracts were filtered through Whatman No. 1 filter paper, concentrated using a rotary evaporator (R-210, Buchi) and transferred to glass vials. The solvent was fully evaporated. Crude extracts were analyzed by LC- HRESIMS at a concentration of 2 mg mL⁻¹. For improved incorporation from [U¹³C]-hexadecanoic acid ([U¹³C]-C₁₆), small-scale (25 mL) cultures of PCC 9431 in Z8 medium in Erlenmeyer flasks (starting chlorophyll *a* content of ~4.0 µg mL⁻¹), were inoculated with 0.25 mM [U¹³C]-hexadecanoic acid and harvested after an incubation of 48 h with continuous orbital shaking (200 rpm, mini shaker, VWR). Centrifugation, extraction and LC- HRESIMS analysis were performed as for the other supplementation samples as described above. To achieve fragmentation of [U¹³C]-C₁₆-**3**, the extract was concentrated by solid phase extraction (Strata C18-E, 100 mg, 1 mL, 55 µm, 70 Å, Phenomenex) over a stepwise gradient of MeOH in H₂O, starting with 30% MeOH. [U¹³C]-C₁₆-**3** was more concentrated in the fraction eluting with 100% MeOH.

SUPPLEMENTARY INFORMATION

Supplementary Information document (PDF) containing Supplementary Figures and Tables.

DATA AVAILABILITY

The stable-isotope labeled substrate feeding datasets are available in MassIVE (MSV000088815). HRESIMS/MS spectra for **1-3** have been added to the GNPS Library under accessions CCMSLIB00009918437, CCMSLIB00009918439 and CCMSLIB00009918438, respectively. Raw NMR data for **1-3** are available at DOI:10.6084/m9.figshare.19146884.

AUTHOR INFORMATION

Corresponding Author

*to whom correspondence should be addressed: Pedro N. Leão, pleao@ciimar.up.pt.

Author Contributions

§SACF and KA contributed equally to this work.

SACF, KA, TPM and PNL designed research. SACF, KA, TPM and IL carried out experimentation. AM, CJ and JR carried out the planning, experiments and interpretation of ECD-based stereochemical assignments. KA and PNL drafted the manuscript, with contributions from all co-authors.

ACKNOWLEDGMENTS

We acknowledge funding from the European Research Council through a Starting Grant (grant agreement 759840) to PNL. Parts of the work were supported by Fundação para a Ciência e a Tecnologia through grants PTDC/BIA-BQM/29710/2017, UIDB/04423/2020 and UIDP/04423/2020 and through scholarships to KA (SFRH/BD/146003/2019) and TPM (SFRH/BD/138308/2018). This project has also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952374 and by BLUEBIOLAB (0474_BLUEBIOLAB_1_E), Programme INTERREG V A of Spain-Portugal (POCTEP).

REFERENCES

- 1 A. G. Atanasov, S. B. Zotchev, V. M. Dirsch and C. T. Supuran, *Nat. Rev. Drug Discov.*, 2021, **20**, 200–216.
- 2 J. Gershenzon and N. Dudareva, *Nat. Chem. Biol.*, 2007, **3**, 408–414.
- 3 E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner and S. L. Flitsch, *Nat. Rev. Methods Primer*, 2021, **1**, 46.
- 4 T. A. Scott and J. Piel, *Nat. Rev. Chem.*, 2019, **3**, 404–425.
- 5 S. Nayfach, S. Roux, R. Seshadri, D. Udway, N. Varghese, F. Schulz, D. Wu, D. Paez-Espino, I.-M. Chen, M. Huntemann, K. Palaniappan, J. Ladau, S. Mukherjee, T. B. K. Reddy, T. Nielsen, E. Kirton, J. P. Faria, J. N. Edirisinghe, C. S. Henry, S. P. Jungbluth, D. Chivian, P. Dehal, E. M. Wood-Charlson, A. P. Arkin, S. G. Tringe, A. Visel, IMG/M Data Consortium, H. Abreu, S. G. Acinas, E. Allen, M. A. Allen, L. V. Alteo, G. Andersen, A. M. Anesio, G. Attwood, V. Avila-Magaña, Y.

- Badis, J. Bailey, B. Baker, P. Baldrian, H. A. Barton, D. A. C. Beck, E. D. Becraft, H. R. Beller, J. M. Beman, R. Bernier-Latmani, T. D. Berry, A. Bertagnolli, S. Bertilsson, J. M. Bhatnagar, J. T. Bird, J. L. Blanchard, S. E. Blumer-Schuetz, B. Bohannan, M. A. Borton, A. Brady, S. H. Brawley, J. Brodie, S. Brown, J. R. Brum, A. Brune, D. A. Bryant, A. Buchan, D. H. Buckley, J. Buongiorno, H. Cadillo-Quiroz, S. M. Caffrey, A. N. Campbell, B. Campbell, S. Carr, J. Carroll, S. C. Cary, A. M. Cates, R. A. Cattolico, R. Cavicchioli, L. Chistoserdova, M. L. Coleman, P. Constant, J. M. Conway, W. P. Mac Cormack, S. Crowe, B. Crump, C. Currie, R. Daly, K. M. DeAngelis, V. Deneff, S. E. Denman, A. Desta, H. Dionisi, J. Dodsworth, N. Dombrowski, T. Donohue, M. Dopson, T. Driscoll, P. Dunfield, C. L. Dupont, K. A. Dynarski, V. Edgcomb, E. A. Edwards, M. S. Elshahed, I. Figueroa, B. Flood, N. Fortney, C. S. Fortunato, C. Francis, C. M. M. Gachon, S. L. Garcia, M. C. Gazitua, T. Gentry, L. Gerwick, J. Gharechahi, P. Girguis, J. Gladden, M. Gradoville, S. E. Grasby, K. Gravuer, C. L. Grettenberger, R. J. Gruninger, J. Guo, M. Y. Habteselassie, S. J. Hallam, R. Hatzenpichler, B. Hausmann, T. C. Hazen, B. Hedlund, C. Henny, L. Herfort, M. Hernandez, O. S. Hershey, M. Hess, E. B. Hollister, L. A. Hug, D. Hunt, J. Jansson, J. Jarett, V. V. Kadnikov, C. Kelly, R. Kelly, W. Kelly, C. A. Kerfeld, J. Kimbrel, J. L. Klassen, K. T. Konstantinidis, L. L. Lee, W.-J. Li, A. J. Loder, A. Loy, M. Lozada, B. MacGregor, C. Magnabosco, A. Maria da Silva, R. M. McKay, K. McMahon, C. S. McSweeney, M. Medina, L. Meredith, J. Mizzi, R. Mock, L. Momper, M. A. Moran, C. Morgan-Lang, D. Moser, G. Muyzer, D. Myrold, M. Nash, C. L. Nesbø, A. P. Neumann, R. B. Neumann, D. Noguera, T. Northen, J. Norton, B. Nowinski, K. Nüsslein, M. A. O'Malley, R. S. Oliveira, V. Maia de Oliveira, T. Onstott, J. Osvatic, Y. Ouyang, M. Pachiadaki, J. Parnell, L. P. Partida-Martinez, K. G. Peay, D. Pelletier, X. Peng, M. Pester, J. Pett-Ridge, S. Peura, P. Pjevac, A. M. Plominsky, A. Poehlein, P. B. Pope, N. Ravin, M. C. Redmond, R. Reiss, V. Rich, C. Rinke, J. L. M. Rodrigues, W. Rodriguez-Reillo, K. Rossmassler, J. Sackett, G. H. Salekdeh, S. Saleska, M. Scarborough, D. Schachtman, C. W. Schadt, M. Schrenk, A. Sczyrba, A. Sengupta, J. C. Setubal, A. Shade, C. Sharp, D. H. Sherman, O. V. Shubenkova, I. N. Sierra-Garcia, R. Simister, H. Simon, S. Sjöling, J. Slonczewski, R. S. Correa de Souza, J. R. Spear, J. C. Stegen, R. Stepanauskas, F. Stewart, G. Suen, M. Sullivan, D. Sumner, B. K. Swan, W. Swingle, J. Tarn, G. T. Taylor, H. Teeling, M. Tekere, A. Teske, T. Thomas, C. Thrash, J. Tiedje, C. S. Ting, B. Tully, G. Tyson, O. Ulloa, D. L. Valentine, M. W. Van Goethem, J. VanderGheynst, T. J. Verbeke, J. Vollmers, A. Vuillemin, N. B. Waldo, D. A. Walsh, B. C. Weimer, T. Whitman, P. van der Wielen, M. Wilkins, T. J. Williams, B. Woodcroft, J. Woolet, K. Wrighton, J. Ye, E. B. Young, N. H. Youssef, F. B. Yu, T. I. Zemskaya, R. Ziels, T. Woyle, N. N. Mouncey, N. N. Ivanova, N. C. Kyrpidis and E. A. Eloe-Fadrosh, *Nat. Biotechnol.*, 2021, **39**, 499–509.
- 6 J. J. van der Hooft, H. Mohimani, A. Bauermeister, P. C. Dorrestein, K. R. Duncan and M. H. Medema, *Chem. Soc. Rev.*, 2020, **49**, 3297–3314.
- 7 C. C. Thornburg, J. R. Britt, J. R. Evans, R. K. Akee, J. A. Whitt, S. K. Trinh, M. J. Harris, J. R. Thompson, T. L. Ewing, S. M. Shipley, P. G. Grothaus, D. J. Newman, J. P. Schneider, T. Grkovic and B. R. O'Keefe, *ACS Chem. Biol.*, 2018, **13**, 2484–2497.
- 8 R. A. Quinn, L.-F. Nothias, O. Vining, M. Meehan, E. Esquenazi and P. C. Dorrestein, *Trends Pharmacol. Sci.*, 2017, **38**, 143–154.
- 9 K. Tidgewell, B. R. Clark and W. H. Gerwick, in *Comprehensive Natural Products II Chemistry and Biology*, eds. L. Mander and H.-W. Lui, Elsevier, Oxford, 2010, vol. 2, pp. 141–188.
- 10 R. M. Hohlman and D. H. Sherman, *Nat. Prod. Rep.*, 2021, **38**, 1567–1588.
- 11 W. H. Gerwick, P. J. Proteau, D. G. Nagle, E. Hamel, A. Blokhin and D. L. Slate, *J. Org. Chem.*, 1994, **59**, 1243–1245.
- 12 B. S. Moore, J. L. Chen, G. M. L. Patterson, R. E. Moore, L. S. Brinen, Y. Kato and J. Clardy, *J. Am. Chem. Soc.*, 1990, **112**, 4061–4063.
- 13 L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Håkansson, J. L. Smith and D. H. Sherman, *Science*, 2007, **318**, 970–974.
- 14 L. Gu, B. Wang, A. Kulkarni, T. W. Geders, R. V. Grindberg, L. Gerwick, K. Håkansson, P. Wipf, J. L. Smith, W. H. Gerwick and D. H. Sherman, *Nature*, 2009, **459**, 731–735.
- 15 H. Nakamura, E. E. Schultz and E. P. Balskus, *Nat. Chem. Biol.*, 2017, **13**, 916–921.
- 16 A. J. Mitchell, Q. Zhu, A. O. Maggiolo, N. R. Ananth, M. L. Hillwig, X. Liu and A. K. Boal, *Nat. Chem. Biol.*, 2016, **12**, 636–640.
- 17 R. V. Popin, D. O. Alvarenga, R. Castelo-Branco, D. P. Fewer and K. Sivonen, *Front. Microbiol.*, 2021, **12**, 684565.
- 18 D. Dhakal, M. Chen, H. Luesch and Y. Ding, *J. Ind. Microbiol. Biotechnol.*, 2021, **48**, kuab003.
- 19 J. Martins, N. Leikoski, M. Wahlsten, J. Azevedo, J. Antunes, J. Jokela, K. Sivonen, V. Vasconcelos, D. P. Fewer and P. N. Leão, *Sci. Rep.*, 2018, **8**, 14537.
- 20 M. Purushothaman, S. Sarkar, M. Morita, M. Gugger, E. W. Schmidt and B. I. Morinaka, *Angew. Chem. Int. Ed.*, 2021, **60**, 8460–8465.
- 21 S. A. C. Figueiredo, M. Preto, G. Moreira, T. P. Martins, K. Abt, A. Melo, V. M. Vasconcelos and P. N. Leão, *Angew. Chem. Int. Ed.*, 2021, **60**, 10064–10072.
- 22 J. Beld, R. Abbriano, K. Finzel, M. Hildebrand and M. D. Burkart, *Mol. Biosyst.*, 2016, **12**, 1299–1312.
- 23 J. A. van Santen, G. Jacob, A. L. Singh, V. Aniebok, M. J. Balunas, D. Bunsco, F. C. Neto, L. Castaño-Espriu, C. Chang, T. N. Clark, J. L. Cleary Little, D. A. Delgadillo, P. C. Dorrestein, K. R. Duncan, J. M. Egan, M. M. Gale, F. P. J. Haeckl, A. Hua, A. H. Hughes, D. Iskakova, A. Khadilkar, J.-H. Lee, S. Lee, N. LeGrow, D. Y. Liu, J. M. Macho, C. S. McCaughey, M. H. Medema, R. P. Neupane, T. J. O'Donnell, J. S. Paula, L. M. Sanchez, A. F. Shaikh, S. Soldatou, B. R. Terlou, T. A. Tran, M. Valentine, J. J. J. van der Hooft, D. A. Vo, M. Wang, D. Wilson, K. E. Zink and R. G. Lington, *ACS Cent. Sci.*, 2019, **5**, 1824–1833.
- 24 N. Berova, L. D. Bari and G. Pescitelli, *Chem. Soc. Rev.*, 2007, **36**, 914.
- 25 G. Bringmann, T. Bruhn, K. Maksimenka and Y. Hemberger, *Eur. J. Org. Chem.*, 2009, **2009**, 2717–2727.
- 26 C. B. Naman, J. Almaliti, L. Armstrong, E. J. Caro-Diaz, M. L. Pierce, E. Glukhov, A. Fenner, C. Spadafora, H. M. Deboni, P. C. Dorrestein, T. F. Murray and W. H. Gerwick, *J. Nat. Prod.*, 2017, **80**, 2328–2334.
- 27 R. T. Williamson, I. P. Singh and W. H. Gerwick, *Tetrahedron*, 2004, **60**, 7025–7033.
- 28 P. Moosmann, R. Ueoka, M. Gugger and J. Piel, *Org. Lett.*, 2018, **20**, 5238–5241.
- 29 N. Eusebio, A. Rego, N. R. Glasser, R. Castelo-Branco, E. P. Balskus and P. N. Leão, *BMC Genomics*, 2021, **22**, 633.
- 30 V. Agarwal, Z. D. Miles, J. M. Winter, A. S. Eustáquio, A. A. El Gamal and B. S. Moore, *Chem. Rev.*, 2017, **117**, 5619–5674.
- 31 H. Marrakchi, M.-A. Lanéelle and M. Daffé, *Chem. Biol.*, 2014, **21**, 67–85.
- 32 K. J. Labby, S. G. Watsula and S. Garneau-Tsodikova, *Nat. Prod. Rep.*, 2015, **32**, 641–653.
- 33 H. Chen, S. O'Connor, D. E. Cane and C. T. Walsh, *Chem. Biol.*, 2001, **8**, 899–912.
- 34 S. Weinig, H.-J. Hecht, T. Mahmud and R. Müller, *Chem. Biol.*, 2003, **10**, 939–952.
- 35 B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordstick, P. Brandt, H. Blöcker, G. Höfle, S. Beyer and R. Müller, *J. Biol. Chem.*, 1999, **274**, 37391–37399.
- 36 K. Abt, R. Castelo-Branco and P. N. Leão, *J. Nat. Prod.*, 2021, **84**, 278–286.
- 37 P. R. Ortiz de Montellano and S. D. Nelson, *Arch. Biochem. Biophys.*, 2011, **507**, 95–110.
- 38 T. Li, A. Tripathi, F. Yu, D. H. Sherman and A. Rao, *Bioinform. Oxf. Engl.*, 2020, **36**, 942–944.