# Biosynthesis of chlorinated lactylates in *Sphaerospermopsis* sp. LEGE 00249

Kathleen Abt,[1] Raquel Castelo-Branco[1] and Pedro N. Leão[1,]*

[1]*Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Avenida General Norton de Matos, s/n, 4450-208 Matosinhos, Portugal*
*author to whom correspondence should be addressed; E-Mail: pleao@ciimar.up.pt.

Abstract
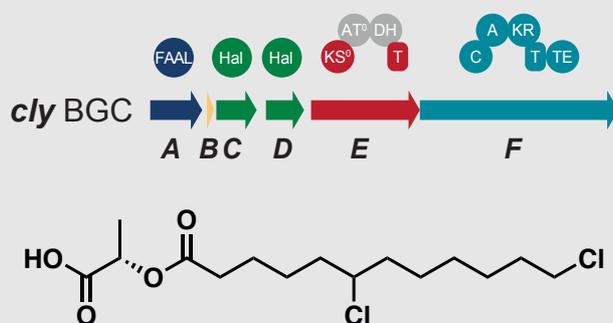
Lactylates are an important group of molecules in the food and cosmetic industries. A series of natural halogenated 1-lactylates – chlorosphaerolactyaltes (**1-4**) – were recently reported from *Sphaerospermopsis* sp. LEGE 00249. Here, we identify the *cly* biosynthetic gene cluster, containing all the necessary functionalities to generate and release the natural lactylates. Using a combination of stable isotope-labeled precursor feeding and bioinformatic analysis, we propose that dodecanoic acid and pyruvate are the key building blocks in the biosynthesis of **1-4**. We additionally report minor analogues of these molecules with varying alkyl chains. The discovery of the *cly* gene cluster paves the way to accessing industrially-relevant lactylates through pathway engineering.

Introduction

Humans have been functionalizing different organisms for the desirable effects of their secondary metabolites for thousands of years[1]. Of special interest nowadays are natural products (NPs) with pharmacological activities or biotechnological applications, for example anti-pathogenic[2,3], anticancer[4] activities or biofuels[5]. Repurposed natural products are derived from all kingdoms of life and in the last decades cyanobacteria have gained recognition as a plentiful source of NPs[6]. In these organisms, the genes for secondary metabolite production are typically organized in biosynthetic gene clusters (BGCs). Two of the major BGC classes are associated with polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) enzymes. BGCs that combine elements of these two pathways are also common[7]. Beyond the basic assembly logic of PKS/NRPS pathways based on a set of few essential protein domains[8], structural variety is greatly enhanced by additional specialized domains and tailoring enzymes such as methyltransferases[9], glycolsyltransferases[10] or halogenases[11]. NRPSs can further directly incorporate non-proteinogenic substrates including different amino acids, hydroxy acids and keto acids[7], overall providing a huge amount of combinatorial possibilities for natural product formation. Such non-proteinogenic substrates are for example used by depsipeptide synthetases, specialized NRPSs with the ability to form ester bonds[12].

Table 1. Annotation of the *cly* gene cluster products.

Protein	Length [aa]	Predicted function	Closest homolog and closest Noc homolog	Identity/ Similarity [%]	Accession No.
-1	397	transferase	DUF3419 family protein [*Moorea* sp. SIO2B7]	84/82	NES86017.1
ClyA	626	FAAL	fatty acyl-AMP ligase [*Anabaena* sp. PCC 7108]	82/92	WP_016949104.1
			NocL [*Nodularia* sp. HBU26]	79/87	AQX77690.1
ClyB	92	ACP	acyl carrier protein [*Moorea* sp. SIO2B7]	71/85	NES81554.1
			NocM [*Nodularia* sp. HBU26]	76/92	AQX77692.1
ClyC	471	halogenase	hypothetical protein [*Anabaena* sp. PCC 7108]	87/93	WP_016949101.1
			NocN [*Nostoc* sp. CCAP 1453/38]	82/90	AKL71647.1
ClyD	452	halogenase	hypothetical protein [*Trichormus variabilis*]	84/91	WP_127052821.1
			NocO [*Nodularia* sp. HBU26]	84/91	AQX77693.1
ClyE	1,286	PKS (KS^0 [1-496], AT^0 [543-769], DH [839-1128], T [1170-1254]	acyltransferase domain-containing protein [*Moorea* sp. SIO2B7]	68/81	NES81557.1
			NocP [*Nodularia* sp. HBU26]	77/86	AQX77694.1
ClyF	2,325	NRPS (C [58-517], A [538-1347], KR [1399-1849], T [1934-2009], TE [2042-2313]	NocQ [*Nodularia* sp. HBU26]	79/88	AQX77695.1
+1	426	lipase	NocR [*Nostoc* sp. CCAP 1453/38]	82/91	AKL71651.1

With recent advances in next-generation sequencing technologies, genomic data has become widely accessible. This has led to the accumulation of many so-called "orphan" BGCs, i.e. those without any known secondary metabolites assigned. Still, many known compounds do not have a cognate BGC[13]. Knowledge of the underlying biosynthetic machinery of NPs can uncover unprecedented enzymes which often find application as new biocatalysts in synthetic reactions[14]. It also enables the transfer of entire BGCs into a suitable host for heterologous expression[15] and pathway engineering, leading to increased yields or to the generation of unnatural analogues of economically relevant NPs[16].

A class of industrially important compounds are lactylates. They are mainly used as emulsifiers in the food and cosmetic industries[17]. Apart from the most common sodium or calcium stearoyl-2-lactylates, several analogues are used in different products[18]. Currently, commercial lactylates are produced by esterification of lactic acid and fatty acids and neutralization at elevated temperature[18]. Limitations include product impurity[18] and dependence on substrate supply chains that feed into other industries[19]. Direct microbial production of lactylates could therefore improve the current process.

Recently, lactylates of halogenated fatty acids have been isolated from the freshwater cyanobacterium *Sphaerospermopsis* sp. LEGE 00249[20]. These compounds, termed chlorosphaerolactylates A-D (**1-4**, Fig. 1a), are esters of (poly)chlorinated dodecanoic acid and L-lactic acid. They were discovered in an antibiofilm activity screening and displayed weak antibacterial, antifungal and antibiofilm properties[20]. Their structures bear some resemblance to columbamides (e.g. **5**), which, instead of esters, are polychlorinated acyl amides with cannabinomimetic properties[21]. Here, we propose the steps involved in chlorosphaerolactylate biosynthesis, notably the recruitment of dodecanoic acid and pyruvate to build the lactylate carbon skeleton, by uncovering its BGC – *cly* – using genome mining. The *cly* genes had previously been assigned, on the basis of comparative genomics, to the larger nocuolin A (**6**) gene cluster (*noc*)[22]. However, sequence alignments, homology modeling and stable-isotope precursor feeding experiments supported our proposal for a *cly*-encoded biosynthesis of **1-4**. In addition, we detected analogues of **1-4** with varying acyl chain length. Overall, these biosynthetic insights open up the possibility for pathway engineering and direct microbial production of different widely used lactylates.

Results and discussion

Identification of a putative chlorosphaerolactylate BGC (*cly*). We sought to identify the biosynthetic gene cluster responsible for the production of the chlorinated lactylates **1-4**. Recognizing the similarity of their halogenated fatty acyl moieties to those of the columbamides (e.g. **5**, Fig. 1a), we envisioned that similar enzymes might be involved in the biosynthesis of these natural products. After sequencing the genome of *Sphaerospermopsis* sp. LEGE 00249 (NCBI: PRJNA655889), we searched the resulting nucleotide data for genes encoding halogenases of the CylC-type[23]. This recently described dimetal-carboxylate halogenase class has been implicated in the chlorination of fatty acyl derived moieties of different cyanobacterial natural products, including the columbamides[21,23–25]. We found two adjacent homologs of *cylC* (*clyC* and *clyD*) in a ~225 kb contig. No additional *cylC* homologs (or genes homologous to non-heme iron halogenases, which may also act on unactivated carbon centers)[26,27] were found in the genome data. Annotation of the genomic context of the *clyC* and *clyD* halogenases (Table 1) revealed that these were part of a roughly 50 kb region containing multiple biosynthetic genes. This locus has high homology to the previously reported *noc* clusters (Fig. 1b), proposed to be involved
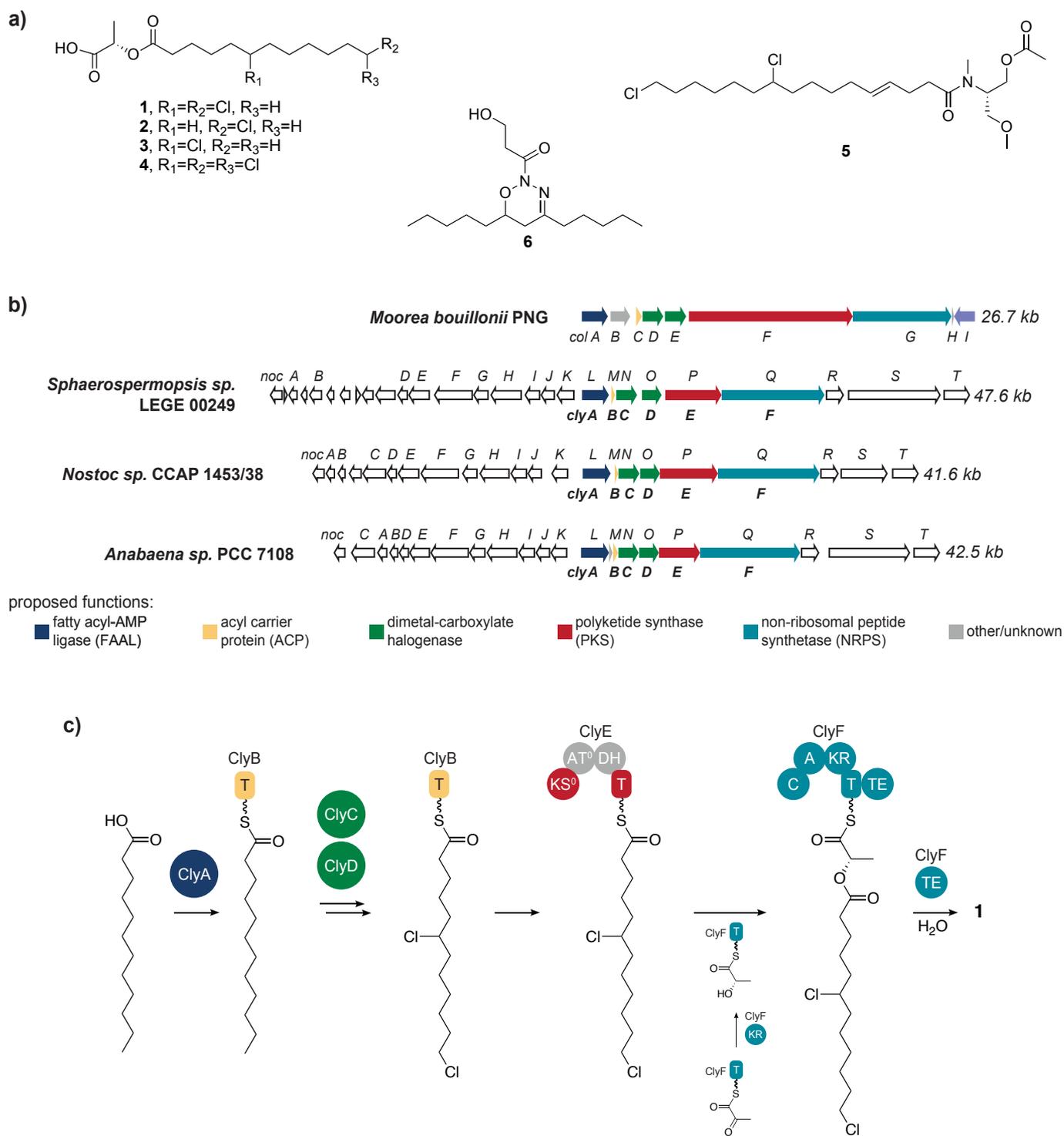
Figure 1. Structure and biosynthesis of chlorosphaerolactylates. a) structures of chlorosphaerolactylates (**1-4**), of the biosynthetically related columbamide A (**5**), and of nocuolin A (**6**), a metabolite that has been putatively associated with the *noc* cluster. b) Schematic representations of the proposed BGCs for columbamides (*col*), chlorosphaerolactylates (*cly*) and nocuolin A (*noc*). c) Proposed biosynthesis of the chlorosphaerolactylates (exemplified for compound **1**).

in the biosynthesis of nocuolin A (**6**, Fig. 1a). CylC homologs are known to carry out cryptic halogenations[23,28] and therefore their involvement in the biosynthesis of the non-halogenated **6** could not be ruled out. Hrouzek and co-workers[22] assigned this putative function for the *noc* cluster based on comparative genomics (strains that contained this locus were found to produce **6**).

Analysis of the *cly* gene cluster. Despite the fact that the *cly* genes were found embedded in the putative BGC encoding compound **6**,

the two dimetal-carboxylate halogenases were strongly indicative of an involvement in the biosynthesis of **1-4**, structurally unrelated to **6**. As such, we thoroughly inspected the genes neighboring the halogenases (Fig. 1b and Table 1). The upstream region of the two halogenases comprises a FAAL (*clyA*) and an ACP (*clyB*), an arrangement that is also observed in the columbamides (*col*), microginin (*mic*) or cylindrocyclophanes (*cyl*) BGCs[21,23,29]. Downstream of the halogenases, a polyketide synthase (*clyE*) is

found before a depsipeptide synthetase NRPS (*clyF*). Further downstream a putative lipase and a lectin-like protein are encoded, just upstream of a kinase. The ClyE PKS is unusual in containing a DH domain while lacking a KR domain. Furthermore, its KS domain lacks an active site histidine (detected by antiSMASH[30] and confirmed through sequence alignments, Fig. S1), and is expected to be a KS[0] domain[31–33]. In agreement with these observations, ClyE features an AT[0] domain, i.e. missing an active site serine residue (Fig. S1)[33,34]. ClyF – an NRPS – has a typical depsipeptide synthetase[35,36,12] domain architecture (C-A-KR-T) and contains also a thioesterase (TE) domain. We rationalized that the *clyA-F* (*nocL-Q*) genes would be sufficient to encode the production and release of **1-4**. We propose (Fig. 1c) that the biosynthesis of these natural lactylates begins with the activation of dodecanoic acid and transfer to ClyB, catalyzed by the FAAL ClyA. Next, the two halogenases, ClyC and D, would chlorinate the unactivated terminal and/or mid-chain carbon centers in the fatty acyl-ACP (ClyB) thioester (a similar substrate is halogenated by CylC in cylindrocyclophane biosynthesis)[23]. The ClyE KS[0] domain would then transfer the halogenated acyl moiety to the ClyE ACP (T) domain. KS[0] domains

have been shown to transfer acyl intermediates between ACPs or between an ACP and a PCP[31,32,37]. Activation of pyruvate and stereospecific reduction of its alpha-keto group by the depsipeptide synthetase ClyF A and KR domains, respectively, would prompt the condensation of the lactyl and acyl moieties by the C domain of ClyF, yielding a halogenated dodecanoyl-lactyl-PCP (T) thioester. Finally, thioester hydrolysis mediated by the TE domain in ClyF would release the final lactylate product (Fig. 1c). To obtain further support towards this hypothesis, we turned our attention to the cyanobacterium *Anabaena* sp. PCC 7108. This strain had been previously reported to contain the *noc* gene cluster and produce **6**[22]. It has a *clyA-F* locus (Fig. 1b) with high homology (74%, nucleotide level) to that of *Sphaerospermopsis* sp. LEGE 00249 and the same structure and PKS/NRPS domain organization, missing only the region corresponding to the DH domain in ClyE (Fig. S2). LC-HRESIMS analysis of an organic extract of *Anabaena* sp. PCC 7108 revealed the presence of **1-4**, but these compounds could not be detected in extracts of other cyanobacterial strains whose genomes do not have a *cly* locus (Fig. S2). Overall, these observations support a role for the *cly* cluster in the biosynthesis of **1-4**.
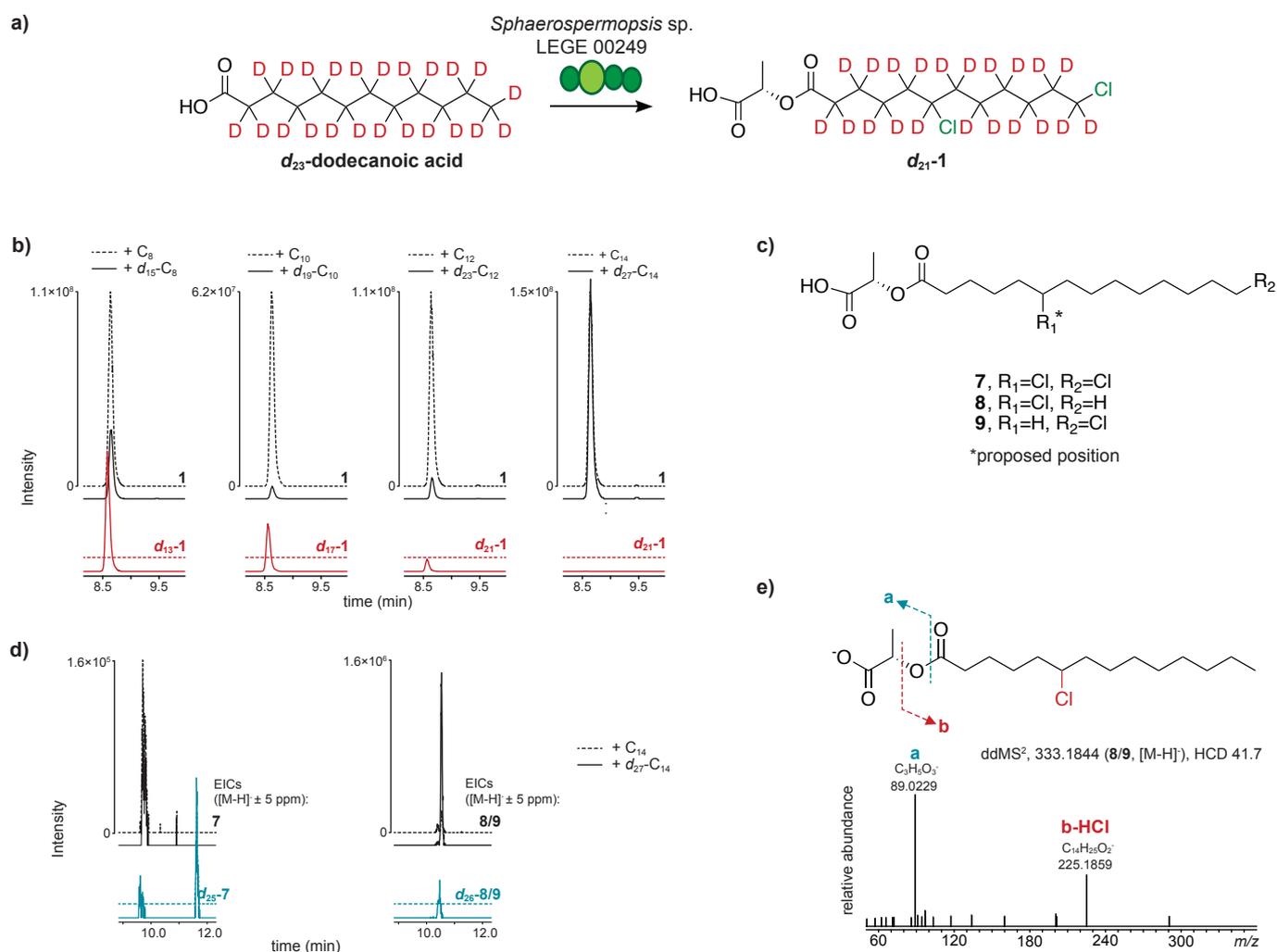

Figure 2. Supplementation of *Sphaerospermopsis* sp. LEGE 00249 with deuterated fatty acids reveals the origin of the acyl group in **1-4** and additional lactylate diversity. a) Schematic representation of the incorporation of a fully deuterated dodecanoic acid-derived moiety into compound **1**. b) LC-HRESIMS analysis of organic extracts of *Sphaerospermopsis* sp. LEGE 00249 following supplementation with different fatty acids: shown are Extracted Ion Chromatograms (EICs) of fully deuterium-labeled (red lines) and non-labeled (black lines) isotopologues of **1**. c) Proposed structures for tetradecanoic acid-derived chlorosphaerolactylates **7-9**, based on: d) LC-HRESIMS detection of dichlorinated (**7**) and monochlorinated (**8/9**) chlorosphaerolactylate isotopologues and (e) LC-HRESIMS/MS analysis of **8/9** (the source of the major observed fragments is exemplified for compound **8**).

Identification of dodecanoic acid as a building block for chlorosphaerolactylate biosynthesis. To experimentally test our biosynthetic hypothesis, we carried out feeding experiments with stable isotope-labeled precursors. We focused first on the fatty acid building block incorporated into **1-4**. If the KS^0 domain is – as hypothesized – non-elongating, then the entire acyl chain should derive from dodecanoic acid (Fig. 1c). We supplemented cultures of *Sphaerospermopsis* sp. LEGE 00249 with a range of fully deuterated, saturated fatty acids (d_{15}-octanoic - d_{15}-C$_8$, d_{19}-decanoic - d_{19}-C$_{10}$, d_{23}-dodecanoic – d_{23}-C$_{12}$ and d_{27}-tetradecanoic – d_{27}-C$_{14}$ acids) and used LC-HRESIMS to detect incorporation of the deuterium labels into **1-4**. As expected, for deuterated C$_8$-C$_{12}$ fatty acids, we observed incorporation of all deuterons in the supplemented substrates into the final products, with the exception of those that were removed as a consequence of chlorination (Fig. 2a,b, Fig. S3). Surprisingly, we also detected *m/z* values consistent with tetradecanoic-acid derived mono- and dichlorinated chlorosphaerolactylates (**7-9**, Fig. 2c). In these cases, supplementation with d_{27}-C$_{14}$ resulted in the expected d_{25} or d_{26} incorporation (Fig. 2d). LC-HRESIMS/MS analysis of the monochlorinated analogue(s) **8/9** confirmed their relatedness to **1-4** (Fig. 2e). However, we could not determine the positioning of the Cl atom in their structures. Considering the structures of columbamides A-E[21,24], the mid-chain halogenated position relative to the fatty acyl-thioester substrates seems to be conserved, which might be observed for the chlorosphaerolactylates as well. Still, this requires experimental validation and the structures of **7-9** presented herein are mere proposals. The discovery of these additional analogues prompted us to revisit the LC-HRESIMS data for the organic extracts of *Sphaerospermopsis* sp. LEGE 00249 in search of other chlorosphaerolactylates with varying acyl chains. As a result, we found traces of metabolites with *m/z* values consistent with decanoic acid-derived chlorosphaerolactylates (Fig. S4). Taken together, these data were consistent with our proposal of ClyA activating and loading dodecanoic acid to generate **1-4** and suggest that this enzyme also activates decanoic and tetradecanoic acids to generate additional chlorosphaerolactylate diversity. Varying degrees of relaxed substrate specificity has been observed for other FAALs[e.g. 38].

Identification of pyruvate as a precursor of the lactate moiety in 1-4. We sought to clarify whether pyruvate would be incorporated directly into the lactate portion of **1-4**, as per our biosynthetic proposal. We supplemented *Sphaerospermopsis* sp. LEGE 00249 cultures with ^{13}C$_3$-pyruvate and analyzed the incorporation of ^{13}C into **1-4** using LC-HRESIMS and LC-HRESIMS/MS analyses in the resulting organic extracts. Due to the central metabolic role of pyruvate, in particular its decarboxylative conversion to acetyl-CoA, we expected scrambling of the label to occur and ^{13}C incorporation to be observed potentially in all carbon positions of **1-4**, even if pyruvate is not a substrate for ClyF. This was, in fact, observed in ^{13}C$_3$-supplemented cultures (Fig. 3a,b), with a notable enrichment in ^{13}C$_2$-(**1-4**) and, to a lesser extent, ^{13}C$_1$-(**1-4**) isotopologues (~95% and ~58% of the monoisotopic base peak). Enrichment was clearly observable up to the M+12 peak, indicating multiple incorporation of ^{13}C atoms. A simulated mixture of isotopologues of **1** that matched the M, M+1 and M+2 fine structure indicated that the heavier M+3 peak only had a minor contribution from ^{13}C$_1$-**1** and ^{13}C$_2$-**1** isotope patterns and, therefore, was generated mostly from ^{13}C$_3$ isotopomers (Fig. 3b). To clarify if an intact ^{13}C$_3$-pyruvate-derived unit would be incorporated directly into the ^{13}C$_3$-**1**
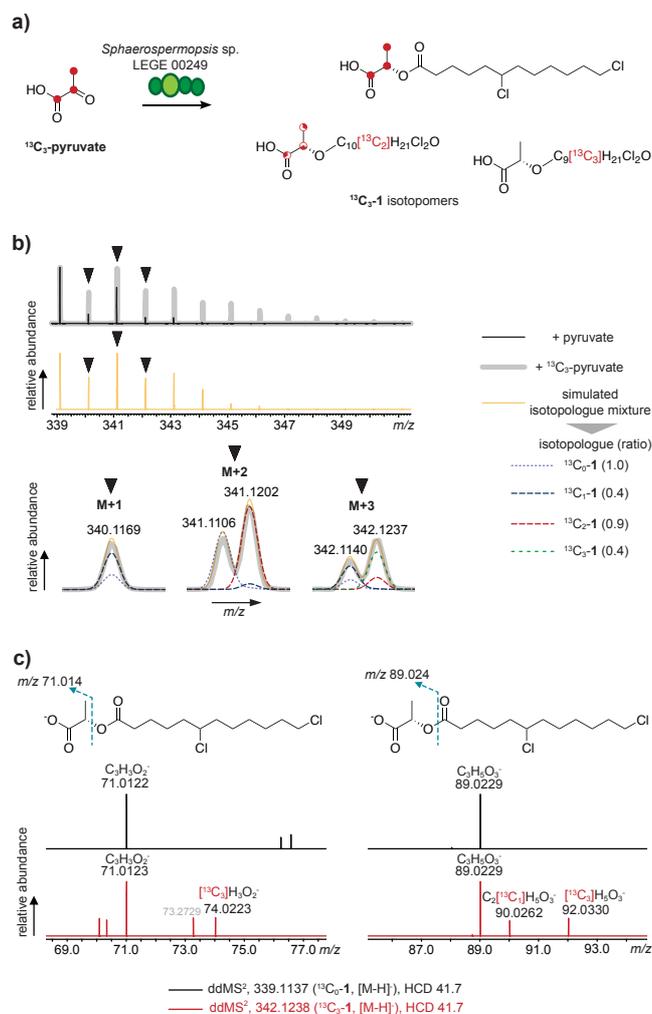
Figure 3. Supplementation of *Sphaerospermopsis* sp. LEGE 00249 with ^{13}C$_3$-pyruvate. a) Schematic representation of observed ^{13}C$_3$-**1** isotopomers following supplementation (full red circles represent incorporation of ^{13}C in that position, partially filled red circles represent three positions where ^{13}C incorporation might have occurred). b) LC-HRESIMS isotope cluster for **1** ([M-H]$^-$), following supplementation of *Sphaerospermopsis* sp. LEGE 00249 with non-labeled pyruvate and ^{13}C$_3$-pyruvate, and for a simulation of a mixture of isotopologues up to ^{13}C$_3$. Expanded regions for the M+1, M+2 and M+3 isotopic peaks (black arrowheads) are shown. c) LC-HRESIMS/MS analysis of **1** and ^{13}C$_3$-**1**, depicting the two spectral regions where fragments corresponding to the lactate portion of the molecule were observed.

isotopomer pool, we resorted to LC-HRESIMS/MS. The MS/MS spectra for both analysis of both ^{13}C$_3$-**1** and ^{13}C$_0$-**1** isotopologues (Fig. 3c) showed a major fragment at *m/z* 89.023 (calcd. for C$_3$H$_5$O$_3$, 89.024), corresponding to the loss of the fatty acyl moiety and confirming that pyruvate-derived carbons were incorporated into the fatty acyl moiety of the chlorosphaerolactylates under the supplementation conditions used. In addition, the MS/MS spectrum for the ^{13}C$_3$-**1** isotopologue showed a ^{13}C$_3$ lactate-derived fragment at *m/z* 92.033 (calcd. *m/z* 92.034). A corresponding ^{13}C$_2$ fragment could not be detected, but a ^{13}C$_1$-derived fragment at *m/z* 90.026 (calcd. *m/z* 90.028) was also present. Furthermore, loss of a dichlorododecanoic acid equivalent resulted in a less prominent fragment at *m/z* 71.012 (calcd. *m/z* 71.014) for ^{13}C$_0$-**1** and ^{13}C$_3$-**1** isotopologues; in this case, only the corresponding ^{13}C$_3$-derived fragment was observed in the fragmentation of ^{13}C$_3$-**1** (Fig. 3c).

Overall, these data are consistent with direct pyruvate incorporation into the lactate portion of **1-4**, with $^{13}C_3$ incorporation directly from the supplemented $^{13}C_3$-pyruvate. The observed $^{13}C_1$ incorporation can be explained by fixation of $^{13}CO_2$ (from decarboxylation of $^{13}C_3$-pyruvate) via the Calvin cycle into $^{13}C_1$-3-phosphoglycerate, eventually leading to $^{13}C_1$-pyruvate[39] (Fig. S5).

Sequence alignment and homology modeling of ClyF. An NRPS-like depsipeptide synthetase, (StsA, PDB ID:6ULW), which utilizes alpha-ketoisocaproic acid as a substrate, has recently been structurally characterized[12]. In that study, Alonzo et al.[12] pinpointed two key sequential residues (Gly414 and Met415 in StsA) as conferring selectivity to alpha-keto acids vs. amino acids, by promoting an antiparallel carbonyl-carbonyl interaction between the amide bond connecting the two residues and the alpha-keto group. Depsipeptide synthetases were also found to contain a hydrophobic residue replacing the Asp featured in canonical NRPSs that is involved in interaction with the amino group[40]. Alonzo and co-workers also show that depsipeptide synthetases contain a unique split motif, so-called pseudo Asub domain, composed of ~30 residues from the N-terminal region and ~70 residues located between the KR and T domains. This motif appears to be exclusive to keto acid-utilizing NRPSs[12]. In light of the results of our pyruvate supplementation experiments, we aimed to understand, by bioinformatic analysis, whether ClyF contained these structural features associated with depsipeptide synthetases. A BlastP search of the full-length ClyF sequence showed that the cyanobacterial depsipeptide synthetases HctE, HctF and CrpD were the closest characterized homologs (47.9, 45.8 and 40.4% identity, respectively). HctE and HctF, both involved in hectochlorin biosynthesis[41], contain C-A-KR-T modules; CrpD (part of the cryptophycins BGC[35]) contains a C-A-KR-T-TE module. The three enzymes are responsible for incorporating alpha-keto acids into depsipeptides. Alignment of ClyF with other depsipeptide or canonical NRPS enzymes showed that ClyF contained the Gly-Met motif (Gly1115 and Met1116) and the hydrophobic residue (Val1007) in lieu of the amino group-interacting Asp residue (Fig. 4a). A homology model of ClyF based on the structure of StsA (PDB ID:6ULW, Fig. S6) showed a similar arrangement of these key residues (Fig. 4b). A pseudo A_{sub} domain could be modeled from the N-terminus and the region before the thiolation domain (Fig. 4c), despite a lower quality of the model in these regions (Fig. S6). Further substantiating the involvement of ClyF in pyruvate incorporation and modification, the stereoselectivity of the KR domain predicted by antiSMASH analysis (Fig. S7) matches the experimentally obtained value for the lactate stereocenter (2*S*), although a single stereospecificity-conferring motif[42] is found in depsipeptide synthetase KR domains[12]. Overall, the results of the in silico analysis of ClyF were entirely consistent with our biosynthetic proposal, regarding its role in pyruvate loading, reduction and condensation of the resulting lactyl moiety.

To conclude, we disclose here a natural 1-lactylate biosynthetic pathway from a photoautotrophic bacterium. We show that the *cly* locus contains all the functions necessary for the biosynthesis of the chlorosphaerolactylates **1-4** from dodecanoic acid and pyruvate precursors. We have also detected additional congeners of these cyanobacterial metabolites in *Sphaerospermopsis* sp. LEGE 00249 cells. The *cly* locus is embedded in the putative nocuolin A (**6**) (*noc*) BGC, but **1-4** and **6** are structurally unrelated. The biosynthesis of **6** has not been clarified and the genes surrounding the *cly* BGC are not
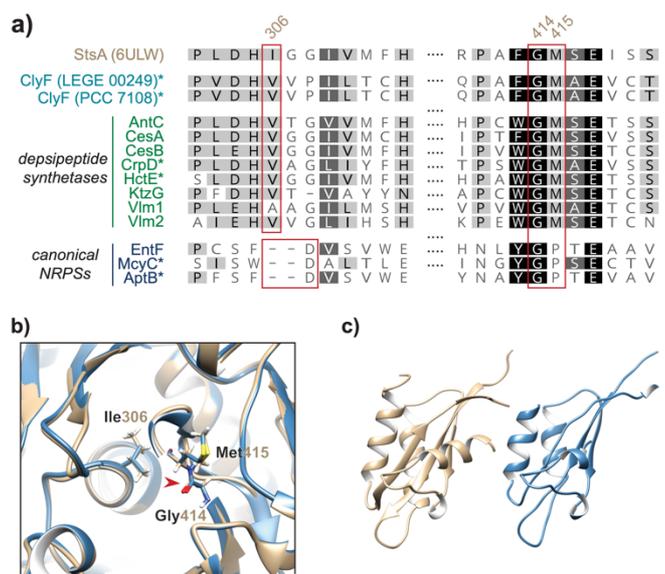

Figure 4. In silico analysis of ClyF. a) Sequence alignment of ClyF with StsA (the single depsipeptide synthetase with a currently available crystallographic structure), additional depsipeptide synthetases and canonical NRPSs. Shown are the regions of previously identified key residues in StsA (Ile306, Gly414 and Met415) that are implicated in the specificity of depsipeptide synthetases towards alpha-keto acids. Asterisks denote cyanobacterial enzymes. b) Superimposed view of the crystallographic model of StsA (beige) and the homology model obtained for ClyF (blue), highlighting the position of the key residues outlined above. Residue numbering corresponds to StsA. The red arrow depicts the amide carbonyl previously proposed to interact with the alpha-keto group of the substrates. c) Comparison of the pseudo Asub domain, characteristic of depsipeptide synthetases, in the crystallographic model of StsA (beige) and the homology model of ClyF (blue).

necessary for the biosynthesis of **1-4**. Hence, apart from the proposed potential role in the biosynthesis of **6**[22], the function of the large *noc* locus remains unclear. However, Gutiérrez-del-Río et al.[20] have reported minor components related to **1-4**, with *m/z* values consistent with 2-lactylates and therefore some of the genes neighboring the *cly* cluster could be associated with these larger metabolites. The chlorosphaerolactylates are assembled under photoautotrophic conditions in a small number of steps by a relatively small BGC with simple and easily accessible intermediates. For all these reasons we consider that the *cly* cluster is an attractive target for engineering the microbial production of industrially relevant lactylates.

Methods

General Experimental Procedures

LC-HRESIMS and LC-HRESIMS/MS data were acquired with an UltiMate 3000 UHPLC (Thermo Fisher Scientific) system composed of a LPG-3400SD pump, WPS-3000SL autosampler and coupled to a Q Exactive Focus hybrid quadrupole-Orbitrap mass spectrometer controlled by Q Exactive Focus Tune 2.9 and Xcalibur 4.1 (Thermo Fisher Scientific). The capillary voltage of HRESI in negative mode was set to -3.3 kV, the capillary temperature to 320 °C and the sheath gasflow rate to 5 units. For analysis in switching mode these parameters were -3.3 kV, 300 °C and 35 units, respectively. LC-MS-grade solvents were purchased from Thermo Fisher Scientific and Carlo Erba. Solvents used for extraction (Thermo Fisher Scientific, VWR) were ACS grade.

Cyanobacterial Strains

Sphaerospermopsis sp. LEGE 00249 was obtained from the LEGE Culture Collection. *Anabaena* sp. PCC 7108, *Anabaena cylindrica* PCC 7122 and *Synechocystis* sp. PCC 6803 were obtained from the Pasteur Culture Collection. All strains were cultured in Z8 medium[43] at 25 °C under a 16:8 hours light/dark cycle with a light intensity of 30 µmol photons m^{-2} s^{-1}. Biomass from stationary-phase batch cultures was harvested by centrifugation (5411 ×*g*, 12 min, 4 °C, Gyrozen 2236R), lyophilized (LyoQuest, Telstar) and stored at -20 ºC until extraction.

Genome Sequencing, Mining and BGC annotation

Total genomic DNA was isolated from a fresh pellet of 50 mL culture of *Sphaerospermopsis* sp. LEGE 00249 using the using the commercial PureLink Genomic DNA Mini Kit (Life Technologies), according to the manufacturer's instructions. The genome of *Sphaerospermopsis* sp. LEGE 00249 was sequenced elsewhere (MicrobesNG) using Illumina technology and 2 × 250 bp paired-end reads. Since the *Sphaerospermopsis* sp. LEGE 00249 culture was not axenic, the resulting genomic data was treated as a metagenome. Quality-filtered raw reads were assembled into contigs by the sequencing services provider. These were re-analyzed in our lab using the binning tool MaxBin 2.0[44] and checked manually in order to obtain only cyanobacterial contigs. This yielded a draft genome of *Sphaerospermopsis* sp. LEGE 00249 (NCBI: PRJNA655889) with an estimated size of 5.3 Mb assembled into 177 contigs. The genome data was mined for homologs of CylC (NCBI: ARU81117.1) and the non-heme iron halogenases SyrB2 (PDB ID: 2FCT_A) and WelO5 (NCBI: AHI58816.1) with the tblastn tool in Geneious 2019.2.1 (Biomatters). The candidate BGC and its translated proteins were annotated based on antiSMASH version 5.1.2[30], NCBI BLAST and InterProScan. Sequences for Cly proteins can be found in the NCBI under accession numbers MBC5793764-MBC5793765 and MBC5793737-MBC5793740.

Feeding Experiments

Sodium pyruvate (99%, Acros Organics) and $^{13}C_3$ sodium pyruvate (99%, Cambridge Isotope Laboratories) were diluted in ultrapure water and filtered through 0.2 µm sterile filters. Octanoic acid (98%, Alfa Aesar), decanoic acid (99%, Alfa Aesar), sodium dodecanoate (98%, Acros Organics) and sodium tetradecanoate (98%, Sigma) were diluted in DMSO (Fisher BioReagents) at a stock concentration of 500 mM. The corresponding perdeuterated fatty acids (d_{15}-C8, d_{19}-C10, d_{23}-C12-potassium salt, d_{27}-C14-potassium salt, 98%, CDN isotopes) were also diluted in DMSO. Fresh Z8 medium (100 mL for pyruvate and decanoic acid feeding, 50 mL for the remaining feeding conditions) was inoculated with *Sphaerospermopsis* sp. LEGE 00249 cells to a starting OD_{750} of 0.04. Cultures were supplemented with the different substrates in two equal pulses (right after and 3 days post inoculation) for a cumulative concentration of 450 µM pyruvate or 100 µM fatty acid. After one-week incubation on an orbital shaker (Mini-Shaker, VWR) at 190 rpm under otherwise standard culture conditions, the biomass was harvested by centrifugation (5411 ×*g*, 12 min, 4 °C, Gyrozen 2236R) and stored at -20 °C until extraction.

Biomass Extraction

Lyophilized biomass from batch cultures or fresh biomass from feeding experiments was fully immersed in CH_2Cl_2/MeOH (2:1), sonicated for 10 min at 30-35 °C and filtered through grade 1 filter paper (Whatman) where it was further extracted with CH_2Cl_2/MeOH (2:1) until no further colour could be extracted from the cells. Solvents were evaporated in a rotary evaporator, the resulting extracts were weighed and resuspended in MeOH at 2.0 mg mL^{-1}, filtered (0.2 µm) and used for LC-HRESIMS analyses.

High-Performance Liquid Chromatography and Mass Spectrometry

All LC-HRESIMS analyses were performed on an UltraCore 2.5 SuperC18 column (75 × 2.1 mm, ACE) at a flow rate of 0.4 mL min^{-1}. The injection volume for all samples was 5 µL. For lactylate detection in the different extracts, the HPLC gradient started with 80 % H_2O with 0.1 % formic acid (eluent A) and 20 % acetonitrile with 0.1 % formic acid (eluent B), continued in a linear gradient over 10 min to 100 % eluent B and held for 10 min before returning to the initial conditions. Spectra were recorded from the spectrometer running in switching mode ($^{13}C_3$-pyruvate, d_{19}-decanoic acid feeding and organic extract of *Sphaerospermopsis* sp. LEGE 00249 batch culture) or in negative mode (all other). The scan range was set to *m/z* 100-900. The resolution in full scan mode was 70,000. For LC-HRESIMS/MS analysis, the scan range was reduced to *m/z* 50-450 and the resolution was 35,000 with an isolation window of *m/z* 0.4, offset of *m/z* 0.2 and a stepped collision energy of 30/40/55 eV.

Bioinformatic Analysis of ClyE and ClyF

Sequences were aligned with the MUSCLE algorithm and Blosum62 matrix from within Geneious 2019.2.1 (Biomatters). ClyE was aligned with Bamb_5919 (NCBI accession number ABI91464, KS2-AT2), DEBS III (CAA44449, KS1-AT1), CurL (AEE88278, KS1-AT1) and Slna9 (AEZ53953, KS2-AT2). ClyF was aligned with AntC (AGG37764), CesA (ABK00751), CesB (ABD14712), CrpD (ABM21572), HctE (AAY42397), KtzG (ABV56587), Vlm1 (ABA59547), Vlm2 (ABA59548), EntF (EMX32470), McyC (AAL82384) and AptB (GU174493). Homology models were built with SWISS-MODEL[45] from ClyF residues 670-1941 with chain B and chain D of partial StsA (PDB ID: 6ULW) as templates. Models were visualized in Chimera 1.14[46].

Associated Content

A Supporting Information file is associated with this manuscript.

Acknowledgments

Author contributions

KA and PNL conceived the project, designed experiments and lead the writing of the manuscript. KA and RC-B performed experimentation. The final manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

References

(1) Newman, D. J.; Cragg, G. M.; Snader, K. M. The Influence of Natural Products upon Drug Discovery. *Nat. Prod. Rep.* **2000**, *17* (3), 215–234. https://doi.org/10.1039/a902202c.

(2) McCormick, M. H.; McGuire, J. M.; Pittenger, G. E.; Pittenger, R. C.; Stark, W. M. Vancomycin, a New Antibiotic. I. Chemical and Biologic Properties. *Antibiot. Annu.* **1955**, *3*, 606–611.

(3) Klayman, D. L. Qinghaosu (Artemisinin): An Antimalarial Drug from China. *Science (80-.).* **1985**, *228* (4703), 1049–1055. https://doi.org/10.1126/science.3887571.

(4) Luesch, H.; Moore, R. E.; Paul, V. J.; Mooberry, S. L.; Corbett, T. H. Isolation of Dolastatin 10 from the Marine Cyanobacterium Symploca Species VP642 and Total Stereochemistry and Biological Evaluation of Its Analogue Symplostatin 1. *J. Nat. Prod.* **2001**, *64* (7), 907–910. https://doi.org/10.1021/np010049y.

(5) Zhuang, X.; Kilian, O.; Monroe, E.; Ito, M.; Tran-Gymfi, M. B.; Liu, F.; Davis, R. W.; Mirsiaghi, M.; Sundstrom, E.; Pray, T.; Skerker, J. M.; George, A.; Gladden, J. M. Monoterpene Production by the Carotenogenic Yeast Rhodosporidium Toruloides. *Microb. Cell Fact.* **2019**, *18* (1). https://doi.org/10.1186/s12934-019-1099-8.

(6) Burja, A. M.; Banaigs, B.; Abou-Mansour, E.; Grant Burgess, J.; Wright, P. C. Marine Cyanobacteria - A Prolific Source of Natural Products. *Tetrahedron.* November 12, 2001, pp 9347–9377. https://doi.org/10.1016/S0040-4020(01)00931-0.

(7) Walsh, C. T.; O'Brien, R. V.; Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angew. Chemie - Int. Ed.* **2013**, *52* (28), 7098–7124. https://doi.org/10.1002/anie.201208344.

(8) Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. https://doi.org/10.1021/cr0503097.

(9) Skiba, M. A.; Sikkema, A. P.; Moss, N. A.; Tran, C. L.; Sturgis, R. M.; Gerwick, L.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L. A Mononuclear Iron-Dependent Methyltransferase Catalyzes Initial Steps in Assembly of the Apratoxin A Polyketide Starter Unit. *ACS Chem. Biol.* **2017**, *12* (12), 3039–3048. https://doi.org/10.1021/acschembio.7b00746.

(10) Nguyen, H. P.; Yokoyama, K. Characterization of Acyl Carrier Protein-Dependent Glycosyltransferase in Mitomycin C Biosynthesis. *Biochemistry* **2019**, *58* (25), 2804. https://doi.org/10.1021/acs.biochem.9b00379.

(11) Mori, S.; Pang, A. H.; Thamban Chandrika, N.; Garneau-Tsodikova, S.; Tsodikov, O. V. Unusual Substrate and Halide Versatility of Phenolic Halogenase PltM. *Nat. Commun.* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-09215-9.

(12) Alonzo, D. A.; Chiche-Lapierre, C.; Tarry, M. J.; Wang, J.; Schmeing, T. M. Structural Basis of Keto Acid Utilization in Nonribosomal Depsipeptide Synthesis. *Nat. Chem. Biol.* **2020**, *16* (5), 493–496. https://doi.org/10.1038/s41589-020-0481-5.

(13) Jensen, P. R. Natural Products and the Gene Cluster Revolution. *Trends Microbiol.* **2016**, *24* (12), 968–977. https://doi.org/10.1016/j.tim.2016.07.006.

(14) Wiltschi, B.; Cernava, T.; Dennig, A.; Galindo Casas, M.; Geier, M.; Gruber, S.; Haberbauer, M.; Heidinger, P.; Herrero Acero, E.; Kratzer, R.; Luley-Goedl, C.; Müller, C. A.; Pitzer, J.; Ribitsch, D.; Sauer, M.; Schmölzer, K.; Schnitzhofer, W.; Sensen, C. W.; Soh, J.; Steiner, K.; Winkler, C. K.; Winkler, M.; Wriessnegger, T. Enzymes Revolutionize the Bioproduction of Value-Added Compounds: From Enzyme Discovery to Special Applications. *Biotechnology Advances.* Elsevier Inc. May 1, 2020, p 107520. https://doi.org/10.1016/j.biotechadv.2020.107520.

(15) Li, Y.; Li, S.; Thodey, K.; Trenchard, I.; Cravens, A.; Smolke, C. D. Complete Biosynthesis of Noscapine and Halogenated Alkaloids in Yeast. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (17), E3922–E3931. https://doi.org/10.1073/pnas.1721469115.

(16) Baltz, R. H. Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology to Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synthetic Biology.* American Chemical Society October 17, 2014, pp 748–758. https://doi.org/10.1021/sb3000673.

(17) Wang, F. C.; Marangoni, A. G. Advances in the Application of Food Emulsifier α-Gel Phases: Saturated Monoglycerides, Polyglycerol Fatty Acid Esters, and Their Derivatives. *Journal of Colloid and Interface Science.* Academic Press Inc. December 1, 2016, pp 394–403. https://doi.org/10.1016/j.jcis.2016.08.012.

(18) Boutte, T.; Skogerson, L. Stearoyl-2-Lactylates and Oleoyl Lactylates. In *Emulsifiers in Food Technology: Second Edition*; Wiley Blackwell: Chichester, UK, 2015; Vol. 9780470670, pp 251–270. https://doi.org/10.1002/9781118921265.ch11.

(19) Alves de Oliveira, R.; Komesu, A.; Vaz Rossell, C. E.; Maciel Filho, R. Challenges and Opportunities in Lactic Acid Bioprocess Design—From Economic to Production Aspects. *Biochemical Engineering Journal.* Elsevier B.V. May 15, 2018, pp 219–239. https://doi.org/10.1016/j.bej.2018.03.003.

(20) Gutiérrez-del-Río, I.; Brugerolle de Fraissinette, N.; Castelo-Branco, R.; Oliveira, F.; Morais, J.; Redondo-Blanco, S.; Villar, C. J.; Iglesias, M. J.; Soengas, R.; Cepas, V.; Cubillos, Y. L.; Sampietro, G.; Rodolfi, L.; Lombó, F.; González, S. M. S.; López Ortiz, F.; Vasconcelos, V.; Reis, M. A. Chlorosphaerolactylates A–D: Natural Lactylates of Chlorinated Fatty Acids Isolated from the Cyanobacterium Sphaerospermopsis Sp. LEGE 00249 . *J. Nat. Prod.* **2020**. https://doi.org/10.1021/acs.jnatprod.0c00072.

(21) Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.; Duggan, B. M.; Di Marzo, V.; Sherman, D. H.; Dorrestein, P. C.; Gerwick, L.; Gerwick, W. H. Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. *J. Nat. Prod.* **2015**, *78* (7), 1671–1682. https://doi.org/10.1021/acs.jnatprod.5b00301.

(22) Voráčová, K.; Hájek, J.; Mareš, J.; Urajová, P.; Kuzma, M.; Cheel, J.; Villunger, A.; Kapuscik, A.; Bally, M.; Novák, P.; Kabeláč, M.; Krumschnabel, G.; Lukeš, M.; Voloshko, L.; Kopecký, J.; Hrouzek, P. The Cyanobacterial Metabolite Nocuolin a Is a Natural Oxadiazine That Triggers Apoptosis in Human Cancer Cells. *PLoS One* **2017**, *12* (3), e0172850. https://doi.org/10.1371/journal.pone.0172850.

(23) Nakamura, H.; Schultz, E. E.; Balskus, E. P. A New Strategy for Aromatic Ring Alkylation in Cylindrocyclophane Biosynthesis. *Nat. Chem. Biol.* **2017**, *13* (8), 916–921. https://doi.org/10.1038/nchembio.2421.

(24) Lopez, J. A. V.; Petitbois, J. G.; Vairappan, C. S.; Umezawa, T.; Matsuda, F.; Okino, T. Columbamides D and E: Chlorinated Fatty Acid Amides from the Marine Cyanobacterium Moorea Bouillonii Collected in Malaysia. *Org. Lett.* **2017**, *19* (16), 4231–4234. https://doi.org/10.1021/acs.orglett.7b01869.

(25) Leão, P. N.; Nakamura, H.; Costa, M.; Pereira, A. R.; Martins, R.; Vasconcelos, V.; Gerwick, W. H.; Balskus, E. P. Biosynthesis-Assisted Structural Elucidation of the Bartolosides, Chlorinated Aromatic Glycolipids from Cyanobacteria. *Angew. Chemie Int. Ed.* **2015**, *54* (38), 11063–11067. https://doi.org/10.1002/anie.201503186.

(26) Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L. Crystal Structure of the Non-Haem Iron Halogenase SyrB2 in Syringomycin Biosynthesis. *Nature* **2006**, *440* (7082), 368–371. https://doi.org/10.1038/nature04544.

(27) Hillwig, M. L.; Fuhrman, H. A.; Ittiamornkul, K.; Sevco, T. J.; Kwak, D. H.; Liu, X. Identification and Characterization of a Welwitindolinone Alkaloid Biosynthetic Gene Cluster in the Stigonematalean Cyanobacterium *Hapalosiphon Welwitschii*. *ChemBioChem* **2014**, *15* (5), 665–669. https://doi.org/10.1002/cbic.201300794.

(28) Reis, J. P. A.; Figueiredo, S. A. C.; Sousa, M. L.; Leão, P. N. BrtB Is an O-Alkylating Enzyme That Generates Fatty Acid-Bartoloside Esters. *Nat. Commun.* **2020**, *11* (1), 1458. https://doi.org/10.1038/s41467-020-15302-z.

(29) Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E. P. Cylindrocyclophane Biosynthesis Involves Functionalization of an Unactivated Carbon Center. *J. Am. Chem. Soc.* **2012**, *134* (45), 18518–18521. https://doi.org/10.1021/ja308318p.

(30) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. AntiSMASH 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Res.* **2019**, *47* (W1), W81–W87. https://doi.org/10.1093/nar/gkz310.

(31) Huang, Y.; Tang, G. L.; Pan, G.; Chang, C. Y.; Shen, B. Characterization of the Ketosynthase and Acyl Carrier Protein Domains at the LnmI Nonribosomal Peptide Synthetase-Polyketide Synthase Interface for Leinamycin Biosynthesis. *Org. Lett.* **2016**, *18* (17), 4288–4291. https://doi.org/10.1021/acs.orglett.6b02033.

(32) Masschelein, J.; Sydor, P. K.; Hobson, C.; Howe, R.; Jones, C.; Roberts, D. M.; Ling Yap, Z.; Parkhill, J.; Mahenthiralingam, E.; Challis, G. L. A Dual Transacylation Mechanism for Polyketide Synthase Chain Release in Enacyloxin Antibiotic Biosynthesis. *Nat. Chem.* **2019**, *11* (10), 906–912. https://doi.org/10.1038/s41557-019-0309-7.

(33) Keatinge-Clay, A. T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. **2017**. https://doi.org/10.1021/acs.chemrev.6b00683.

(34) Jiang, C.; Wang, H.; Kang, Q.; Liu, J.; Bai, L. Cloning and Characterization of the Polyether Salinomycin Biosynthesis Gene Cluster of Streptomyces Albus XM211. *Appl. Environ. Microbiol.* **2012**, *78* (4), 994–1003. https://doi.org/10.1128/AEM.06701-11.

(35) Magarvey, N. A.; Beck, Z. Q.; Golakoti, T.; Ding, Y.; Huber, U.; Hemscheidt, T. K.; Abelson, D.; Moore, R. E.; Sherman, D. H. Biosynthetic Characterization and Chemoenzymatic Assembly of the Cryptophycins. Potent Anticancer Agents from *Nostoc* Cyanobionts. *ACS Chem. Biol.* **2006**, *1* (12), 766–779. https://doi.org/10.1021/cb6004307.

(36) Ehling-Schulz, M.; Vukov, N.; Schulz, A.; Shaheen, R.; Andersson, M.; Märtlbauer, E.; Scherer, S. Identification and Partial Characterization of the Nonribosomal Peptide Synthetase Gene Responsible for Cereulide Production in Emetic Bacillus Cereus. *Appl. Environ. Microbiol.* **2005**, *71* (1), 105–113. https://doi.org/10.1128/AEM.71.1.105-113.2005.

(37) Helfrich, E. J. N.; Piel, J. Biosynthesis of Polyketides by Trans-AT Polyketide Synthases. *Nat. Prod. Rep.* **2016**, *33* (2), 231–316. https://doi.org/10.1039/C5NP00125K.

(38) Hemmerling, F.; Lebe, K. E.; Wunderlich, J.; Hahn, F. An Unusual Fatty Acyl:Adenylate Ligase (FAAL)–Acyl Carrier Protein (ACP) Didomain in Ambruticin Biosynthesis. *ChemBioChem* **2018**, *19* (10), 1006–1011. https://doi.org/10.1002/cbic.201800084.

(39) Schwarz, D.; Orf, I.; Kopka, J.; Hagemann, M. Recent Applications of Metabolomics Toward Cyanobacteria. *Metabolites* **2013**, *3* (1), 72–100. https://doi.org/10.3390/metabo3010072.

(40) Fujimori, D. G.; Hrvatin, S.; Neumann, C. S.; Strieker, M.; Marahiel, M. A.; Walsh, C. T. Cloning and Characterization of the Biosynthetic Gene Cluster for Kutznerides. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (42), 16498–16503. https://doi.org/10.1073/pnas.0708242104.

(41) Ramaswamy, A. V.; Sorrels, C. M.; Gerwick, W. H. Cloning and Biochemical Characterization of the Hectochlorin Biosynthetic Gene Cluster from the Marine Cyanobacterium *Lyngbya Majuscula*. *J. Nat. Prod.* **2007**, *70* (12), 1977–1986. https://doi.org/10.1021/np0704250.

(42) Keatinge-Clay, A. T.; Stroud, R. M. The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* **2006**, *14* (4), 737–748. https://doi.org/10.1016/j.str.2006.01.009.

(43) Rippka, R. Isolation and Purification of Cyanobacteria. *Methods Enzymol.* **1988**, *167* (C), 3–27. https://doi.org/10.1016/0076-6879(88)67004-2.

(44) Wu, Y. W.; Simmons, B. A.; Singer, S. W. MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets. *Bioinformatics* **2016**, *32* (4), 605–607. https://doi.org/10.1093/bioinformatics/btv638.

(45) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303. https://doi.org/10.1093/nar/gky427.

(46) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.