Structure, synthesis of hierridin C and discovery of prevalent alkylresorcinol biosynthesis in picocyanobacteria

Margarida Costa,[a,☆] Ivo E. Sampaio-Dias,[b,☆] Raquel Castelo-Branco,[a,☆] Hugo Scharfenstein,[a] Roberta Rezende de Castro,[c] Artur Silva,[c] Maria Paula C. Schneider,[c] Maria João Araújo,[b] Rosário Martins,[a,d] Valentina F. Domingues,[e] Fátima Nogueira,[f] Vera Camões,[f] Vitor M. Vasconcelos[a,g] and Pedro N. Leão*[,a]

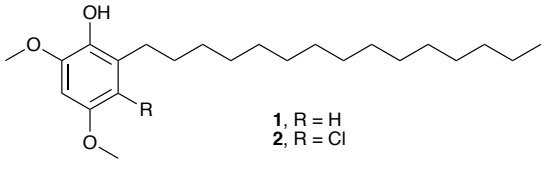
[a]*Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Avenida General Norton de Matos, s/n, 4450-208 Matosinhos, Portugal*
[b]*LAQV-REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*
[c]*Institute of Biological Sciences, Center of Genomic and System Biology, Federal University of Pará(UFPA), Belem, PA66075-110, Brazil*
[d]*Health and Environment Research Centre (CISA), School of Health, Polytechnic Institute of Porto, Rua Dr. António Bernardino de Almeida, 400, 4200-072, Porto, Portugal*
[e]*REQUIMTE/LAQV, Instituto Superior de Engenharia, Instituto Politécnico do Porto, Rua Drº António Bernardino de Almeida, 431, 4200-072 Porto, Portugal*
[f]*Global Health and Tropical Medicine, GHTM, Unidade de Ensino e Investigação de Parasitologia Médica, Instituto de Higiene e Medicina Tropical, IHMT, Universidade Nova de Lisboa, UNL, Rua da Junqueira nº100, 1349-008 Lisboa, Portugal*
[g]*Department of Biology, Sciences Faculty, University of Porto, Rua do Campo Alegre, Porto 4169-007, Portugal*
[☆]*These authors contributed equally to this work.*
†Electronic Supplementary Information (ESI) available with this preprint.
**Author to whom correspondence should be addressed; E-Mail: pleao@ciimar.up.pt (P.N.L.).*

Abstract

Small, single-celled planktonic cyanobacteria are ubiquitous in the world's oceans yet tend not to to be perceived as biosynthetically-prolific organisms. Here we report the isolation and structure elucidation of hierridin C, an antiplasmodial halogenated alkylresorcinol produced in very low amounts by the picocyanobacterium *Cyanobium* sp. LEGE 06113. We describe a simple, straightforward synthetic route to the scarcely-produced hierridins that relies on a key, highly regioselective halogenation step. In addition, we show that these compounds originate from a type III PKS pathway and that similar biosynthetic gene clusters are found in a variety of bacterial genomes, most notably those of the globally-distributed picocyanobacteria genera *Prochlorococcus*, *Cyanobium* and *Synechococcus*.

Introduction

Cyanobacteria are biosynthetically rich and produce a wide array of natural products with varied structural features.[1] Many of these compounds, in particular those of polyketide and/or non-ribosomal peptide nature show strong potency against pharmacologically-relevant targets, such as the toxins microcystins[2] or the promising anticancer compounds largazole.[3] The majority of cyanobacterial secondary metabolites have been isolated from field collections or from laboratory cultures of so-called complex cyanobacteria.[1] Such cyanobacteria display elaborate morphological features (e.g. differentiated cells), and have large genomes that harbour multiple biosynthetic gene clusters (BGCs).[4] Picoplanktonic cyanobacteria, such as those from the genera *Synechococcus/Cyanobium* and *Prochlorococcus*, are widespread throughout coastal and open waters around the globe and contribute strongly to primary productivity in the oceans.[5] These so-called picocyanobacteria have smaller genomes and do not appear to be as biosynthetically prolific, particularly regarding non-ribosomal peptides and polyketides.[4] We have recently reported that the marine picocyanobacterium *Cyanobium* sp. LEGE 06113 produces the alkylresorcinol hierridin B (**1**),[6] a methoxylated monoalkylresorcinol that features a long (C_{15}) aliphatic chain. This metabolite had been previously reported from the filamentous cyanobacterium *Leptolyngbya ectocarpii* SAG



60.90 and it was active as a mixture (with its C_{17} alkyl analog hierridin) against the malaria parasite *Plasmodium falciparum*.[7]
As part of an effort to reisolate **1**, we identified a related chlorinated alkylresorcinol – hierridin C (**2**), the isolation, structural elucidation, antiplasmodial activity and a straightforward synthesis of which are reported herein. To understand the biosynthesis of the hierridins in cyanobacteria, we sequenced and mined the genome of *Cyanobium* sp. LEGE 06113 to find the hierridin BGC, *hid*, of a type III PKS nature. We show that this gene cluster is present in a large number of cyanobacterial genomes, most notably those of picocyanobacterial genera.

Results and discussion

Isolation and structural elucidation of hierridin C (2).

With an initial purpose of re-isolating hierridin B (**1**) for biological activity studies, early-stationary-phase cultures of *Cyanobium* sp. LEGE 06113 were harvested and the biomass extracted with CH_2Cl_2:MeOH (2:1) as previously described.[6] The resulting crude extract was fractionated by a modified version of the normal-phase vacuum liquid chromatography (VLC) protocol employed to isolate hierridin B (**1**).[6] The modification consisted of an initial gentle gradient elution step with the purpose of obtaining less-complex hierridin B-containing fractions and facilitate the isolation of this metabolite. This was effectively achieved, as revealed by spectroscopic inspection of the fractions obtained by VLC. The ^{1}H NMR spectrum ($CDCl_3$) of fraction A1, eluting with an 87:13 mixture of *n*-hexane and EtOAc, exhibited the distinctive methoxy singlets of **1**, resonating at δ3.85 and δ3.76 (Fig. S1). This fraction was selected for further purification, which was carried out using consecutive semi-preparative and analytical-scale RP-HPLC separations, resulting in the isolation of **1**, as expected, but also of the novel metabolite **2**, both in sub-milligram amounts (0.5 and 0.3 mg, 0.003% and 0.002% d.w., respectively).

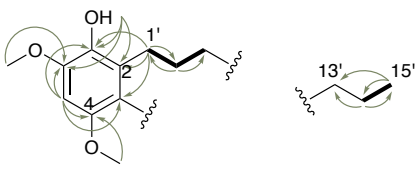


Figure 1 - HMBC (arrows) and COSY (thick bonds) correlations establishing the partial structures of the aromatic moiety and aliphatic chain terminus of compound **2**.

The ^{1}H NMR spectrum of **2** in $CDCl_3$ resembled that of **1**[6] (Fig. S2). Still, a single aromatic proton resonance (δ6.44) was observed for **2**, suggesting a penta-substituted aryl moiety. A combination of 1D (^{1}H, ^{13}C) and 2D (HSQC, HMBC, COSY) NMR data for **2** in DMSO-d_6 (Table 1, Figs. S3-S7) revealed the architecture of the aromatic moiety as a 2,3,4,6-substituted phenol with an aliphatic chain and two methoxy groups as three of the substituents, confirming its relatedness to metabolite **1**. Establishing the full structure of **2** required nevertheless extensive correlational analysis, mostly owing to the determination of substituent arrangement in the aromatic moiety (Fig. 1). This was facilitated by HMBC correlations from the aromatic proton (δ6.66, H-5) to four downfield carbons at δ147.6, δ146.3, δ138.3 and δ112.8 (C-4, C-6, C-1 and C-3, respectively), suggesting that the carbon resonating at δ128.0 (C-2) should be *para* to H-5. The benzylic protons (δ2.66, H_2-1') resonance showed correlations to three aromatic carbons at δ138.3, δ128.0 and δ112.8 (C-1, C-2 and C-3, respectively). This allowed assigning the carbons resonating at δ147.6 (C-6) and δ146.3 (C-4) as *ortho* to H-5. In addition, each of these carbons was found to be methoxylated, on the basis of HMBC correlations from the methoxy protons at δ3.81 (H_3-8) and δ3.77 (H_3-7) to C-6 and C-4, respectively. The final substituent arrangement was established from HMBC correlations between the exchangeable phenol proton (OH-1) and three carbons at δ146.3, δ138.3 and δ128.0 (C-6, C-1 and C-2, respectively), leading to the assignment of the hydroxyl group to the carbon at δ138.3 (C-1) and of its position as adjacent to both C-6 and C-2. This and the previously assumed constraints resulted in the allocation of the alkyl substituent to the carbon at δ128.0 (C-2) and subsequent assignment of the carbons resonating at δ112.8 (C-3) and δ147.6 (C-4) as *ortho* and *meta* to C-2, respectively. The partial structure was finalized by the straightforward assignment, on the basis of COSY and HMBC correlations, of the methylene protons resonating at δ1.42 (H_2-2') as being in an intermediate position between the benzylic methylene and the CH_2 envelope. A second partial structure (Fig. 1) was extracted from the NMR data and corresponded to the terminus of an aliphatic chain and was COSY- and HMBC- correlated to the methylene envelope.

Table 1 - NMR Spectroscopic Data (600 MHz, DMSO-d_6) for hierridin C (**2**).

No.	δ_C[a]		δ_H, (*J* in Hz)[b]	HMBC[c]	COSY
1	138.3	ArOH	8.30 s	1, 2, 6	
2	128.0	C	-		
3	112.8	C	-		
4	147.6	C	-		
5	96.6	CH	6.66 s	1, 3, 4, 6	
6	146.3	C	-		
7	56.4	OCH_3	3.77 s	4	
8	55.9	OCH_3	3.81 s	6	
1′	27.0	CH_2	2.66 t (7.6)	1, 2, 3, 2′	2′
2′	28.2	CH_2	1.42 m	1′, 3′-12′	1′, 3′
3′-12′	28.7-29.0	CH_2	1.23-1.29 m	3′-12′	2′
13′	31.3	CH_2	1.23 m	14′	
14′	22.1	CH_2	1.24 m	13′, 15′	15′
15′	13.9	CH_3	0.85 t (7.0)	13′, 14′	14′

[a]*Recorded at 150 MHz;* [b]*Recorded at 600 MHz;* [c]*From proton to indicated carbon.*

From the NMR data, the substituent at C-3 was unknown and the precise number of CH_2 groups in the aliphatic chain difficult to estimate through the integration of the ^{1}H NMR data of **2**. These uncertainties were clarified by HRESIMS analysis. Despite the low ionization under ESI conditions, using direct infusion we were able to observe an $[M-H]^-$ anion with an *m/z* value of 397.25298 that exhibited a characteristic isotopic pattern for one chlorination (Fig. S8), corresponding to a molecular formula for the parent compound of $C_{23}H_{39}ClO_3$ and thereby establishing its proposed structure of **2**. Because fragmentation of this compound under ESI conditions was limited and uninformative, we resorted to gas chromatography-mass spectrometry (GC-MS/MS) analysis of **2** for structural

corroboration. A single-chlorination isotope pattern for the aromatic "head" following benzylic fragmentation was observed (Fig. S9), supporting the proposed structure.

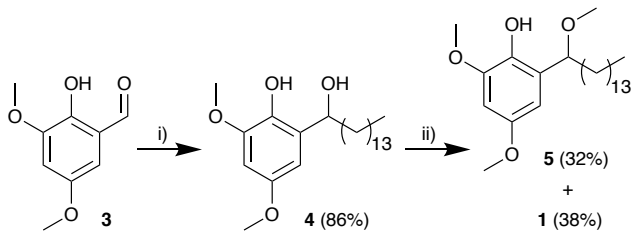
Metabolite **2** adds to a small number of previously reported chlorinated alkylresorcinols from the eukaryotic realm. The halogenated 5-methylresorcinol derivative 4-chloroorcinol was isolated from the phytopathogenic fungus *Colletotrichium higgisianum* and induced chlorosis in plant leaves, inhibited seed germination and was also toxic to the crustacean *Artemia salina*.[8] The monochasiols are a series of long chain alkylresorcinols, with a single chlorinated carbon in between the resorcinol hydroxy groups, that have been isolated from the slime mold *Dictyostelium* and have mild immunosuppressive activity.[9] Other examples include a series of dimerized chlorinated short-chain alkylresorcinols – the spiromastols – that have been reported recently from a deep-sea fungus and have antibacterial activity.[10]

Biological activity of hierridin C (2).

Taking into account the bioactivity profile of **1**, with negligible toxicity to human cell lines (showing only a mild selective activity towards HT-29 cancer cells)[6] and low micromolar antiplasmodial activity,[7] we decided to evaluate the cancer cell line cytotoxicity and antimalarial activity of **2**. Upon exposure of human cancer cell lines to up to 30 μg mL^{-1} (75.2 μM) of **2**, no cytotoxicity was observed (Fig. S10). Both **1** and **2** were tested for *in vitro* activity against *Plasmodium falciparum* (3D7 strain and the chloroquine-resistant Dd2 strain). While all IC_{50} values obtained were in the low-micromolar range (Fig. S11), compound **2** was slightly more potent than **1** against the 3D7 strain (IC_{50} values of 1.5 and 2.1 μM, respectively).[§] The two metabolites showed equal potency towards the Dd2 strain (IC_{50} = 2.3 μM). Thus, the presence of the chlorine atom seems to have a marginal influence on the antiplasmodial activity.

Total synthesis of hierridin C (2).

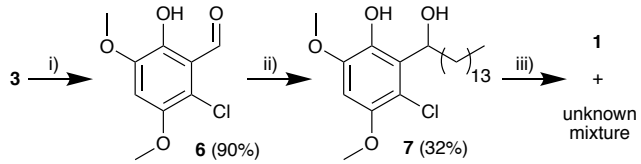
The low yields of **1** and **2** obtained from *Cyanobium* sp. LEGE 06113 cultures and their relatively simple and planar structures motivated our efforts towards synthesizing these compounds in a expeditious and convenient fashion, as a means to enable future biosynthetic and biological activity studies. We envisioned that the most straightforward synthetic route for the preparation of **2** would be from aromatic electrophilic chlorination of **1**. A previous synthetic procedure for obtaining a longer-chain analogue of **1** was available in the literature[11] but did not work in our hands. Therefore, we designed a convenient synthetic route for its preparation (Scheme 1) using the commercially available suitable aromatic ring moiety, 2-hydroxy-3,5-dimethoxybenzaldehyde (**3**) which can be easily prepared from 3,5-dimethoxybenzaldehyde.[12,13]



Scheme 1 - Synthesis of hierridin B (**1**). i) tetradecylmagnesium chloride, THF, 2h, rt; ii) H_2SO_4, MeOH, Pd/C 10% (w/w), H_2 (1 atm), 80 °C.

Grignard addition to **3** using commercial tetradecylmagnesium chloride afforded the corresponding carbinol **4** smoothly in high yields (86%). Reduction of the α-hydroxyl group is already described in literature for the monomethoxylated counterpart (*o*-vanillin derivative).[11] Applying the same protocol for derivative **4** using methanol as solvent (due to solubility issues), afforded **1** in moderate yields in contrast with the results described in literature for the monomethoxylated parent.[11] We found that this protocol led to the formation of an ether at the α-benzylic position as a side product (**5**), which can be explained by competition with methanolysis under acidic medium. An additional methoxyl group seems to influence the outcome of the reaction, since no formation of ether side products were reported by the authors.

We predicted that monochlorination at carbon 3 of the aromatic ring in **1** would be a bottleneck step for the preparation of **2**, since it is highly activated by the presence of electron donating groups which might end up competing with 3,5-dichlorinated and/or 3 and 5-monochlorinated products. Our attempts at this reaction revealed unfruitful using standard procedures commonly used for monohalogenation of phenol derivatives. Using NaClO/NaOH[14] no reaction took place and NCS-assisted chlorination[15] yielded an unknown polychlorinated adduct. Taking into consideration these results, we considered the monochlorination of the salicylaldehyde derivative **3** prior to the Grignard addition protocol (Scheme 2). An analogous strategy was used, for example, in the syntheses of certain monochasiols.[9]



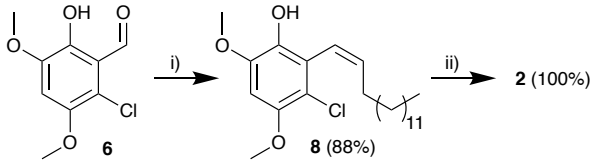
Scheme 2 - Preparation of monochlorinated **6** from **3** and synthesis of α-benzylic derivative of hierridin C (**5**). Reagents and conditions: i) NCS, HCl 37%, 15 min, RT; ii) tetradecylmagnesium chloride, THF, 2h, rt; iii) H_2SO_4, MeOH, Pd/C 10% (w/w), H_2 (1 atm).

Following a standard chlorination protocol using 1 eq. of NCS in presence of 0.1 eq of *p*-toluenesulfonic acid (TsOH)[15] no traces of **6** were detected after 6h. Surprisingly, replacement of TsOH by conc. HCl afforded **6** in very good yield (> 90%) in a short reaction time (15 min). Despite the high aromatic ring activation in compound **3**, the electrophilic chlorination revealed total regioselectivity (see NOESY experiment, Fig. S21).

With the 2-monochlorinated salicylaldehyde derivative **6** in hand, the alkyl chain was introduced by following the same protocol used before to synthesize **4**, affording the corresponding carbinol **7** in good yield (79%) (Scheme 2). Hydrogenation of **7** led to a mixture of products, the major being the dechlorinated parent, hierridin B (**1**). This can be related to the harsh reaction conditions employed for α-hydroxyl elimination.

To overcome this hurdle, we considered the Wittig olefination for the introduction of the alkyl chain and subsequent saturation of the olefin by catalytic hydrogenation (Scheme 3). The appropriate Wittig reagent was prepared by refluxing 1-bromotetradecane with triphenylphosphine, affording the corresponding phosphonium salt that was re-crystallized in Et_2O. The Wittig reagent was then treated with *n*-BuLi in THF followed by addition of salicylaldehyde **6**, affording the (*Z*)-alkene **8** in good yield (88%). Hydrogenation of **8** using catalytic amounts of Pd/C 10% in EtOAc at 0°C proceeded smoothly yielding hierridin C (**2**) quantitatively in 1h. Compound **2**

can thus be efficiently obtained in just three steps (**3** → **6** → **8** → **2**) in high global yield (80%).



Scheme 3 - Synthesis of Hierridin C (**2**) by Wittig approach. i) Wittig reaction: myristyltriphenylphosphonium bromide, *n*-BuLi, THF, rt; ii) Pd/C 10% (w/w), H_2 (1 atm), EtOAc, 0 °C.

Discovery of the hierridin biosynthetic gene cluster (*hid*)

Despite a number of reports on the biological activity of extracts from members of the picocyanobacteria (*Synechococcus*/*Cyanobium*/*Procholorococcus*),[16,17] this phylogenetically congruent group is notoriously poor in PKS or NRPS biosynthetic genes.[4] We had previously proposed that a type III PKS could give origin to the alkylresorcinol moiety,[6] based on the biosynthesis of cyanobacterial cylindrocyclophane monomers,[18] or of other prokaryotic alkylresorcinols.[19] In a previous effort using degenerate primers designed to target then-available type III PKS genes from cyanobacteria, we failed to PCR-amplify a homolog from *Cyanobium* sp. LEGE 06113.[6]

We envisioned that a genome-mining strategy could yield better results and sequenced the genome of this cyanobacterium, having obtained 2.6 Mbp distributed over 30 contigs. The genome data was searched for homologs of the type III PKS *cylI* from *Cylindrospermum licheniforme* UTEX B 2014 (involved in cylindrocyclophane biosynthesis)[18] using BlastP. A single homolog (27% identity, 42% similarity) was found, located in a ~387 kb contig. Annotation of the genomic context of this type III PKS homolog (Table S1), revealed genes with predicted functions consistent with hierridin structural features (one methyltransferase and one hydroxylase). Based on the analysis of predicted open reading frames, this putative hierridin BGC (*hid*) is proposed to comprise at least *hidA*, encoding a SAM-dependent methyltransferase; *hidB*, coding for a FixC superfamily FAD-dependent oxidoreductase and the type III PKS-encoding *hidC* (Fig. 2A). A PAP-fibrillin is encoded downstream of this proposed minimal *hid* gene cluster – but its relation to the *hid* genes is unclear at this stage.

To experimentally confirm the involvement of the *hid* gene cluster in the biosynthesis of the hierridins, we initially attempted to clone and express the entire cluster in E. *coli* BL21 cells, but were unable to observe any protein overexpression by gel electrophoresis or any differential metabolite profile by HPLC analysis. Because the *hidA* gene encodes a methyltransferase that does not contain an ICMT (isoprenylcysteine carboxyl methyltransferase) domain, we envisaged that HidA was not required for decarboxylation of the aromatized product of HidC[20] and that its natural substrate could be the alkylresorcinol **9** (Fig. 2B). The latter compound is commercially

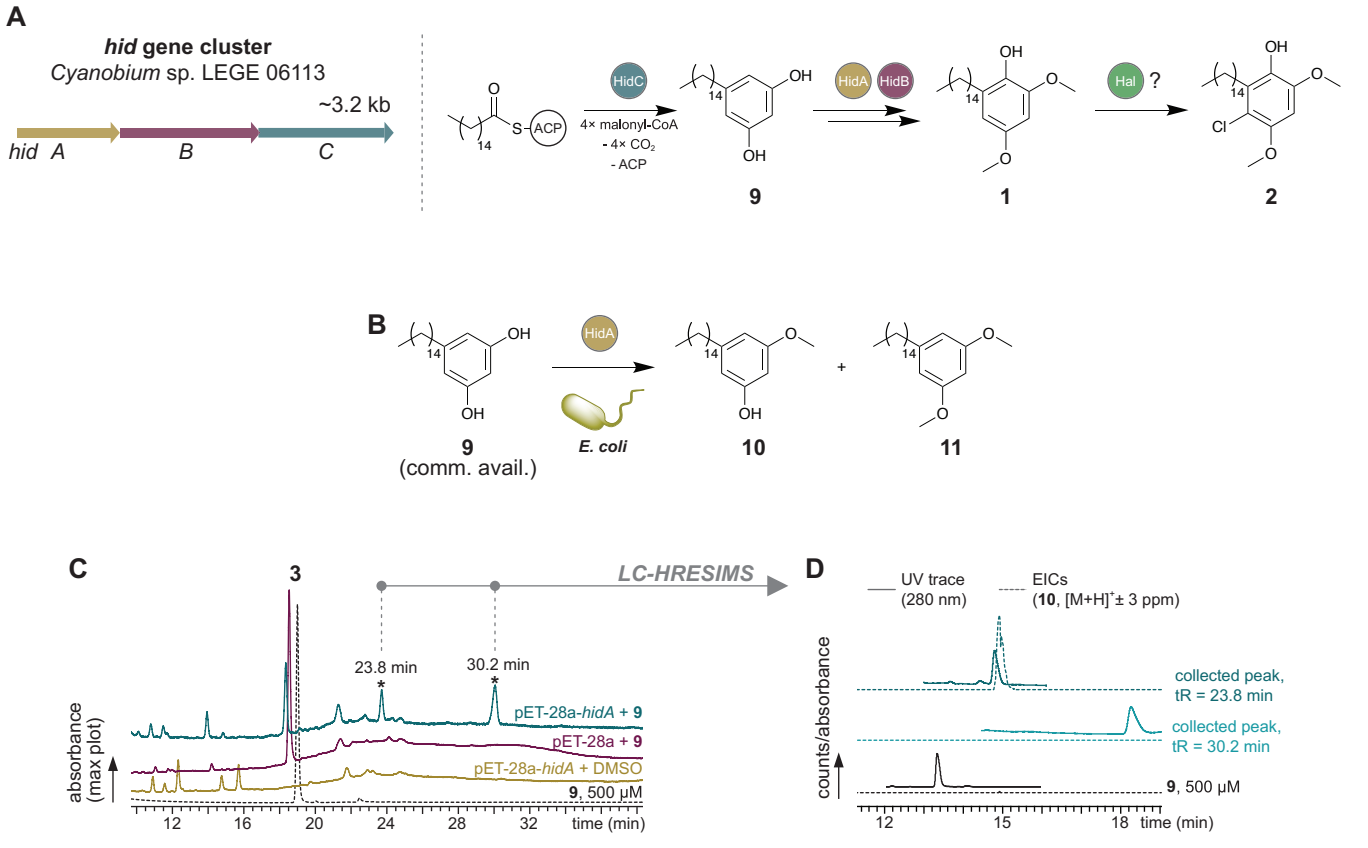


Figure 2 – Hierridin biosynthesis in Cyanobium sp. LEGE 06113. **A** – Overview of the hierridin biosynthetic gene cluster (hid) organization and proposed biosynthesis of hierridins (Hal – putative halogenase). **B** – Schematic representation of the experimental approach to test whether HidA methylates an alkylresorcinol substrate in hierridin biosynthesis. **C** – HPLC-PDA analysis of E. coli expressing the hidA gene, with two new peaks eluting at 23.8 and 30.2 minutes. **D** – LC-HRESIMS analysis of the two chromatographic peaks identified from HPLC-PDA analysis, indicating that an m/z value consistent with compound **10** is observed for one of the peaks, while for the other peak no ionization was observed.

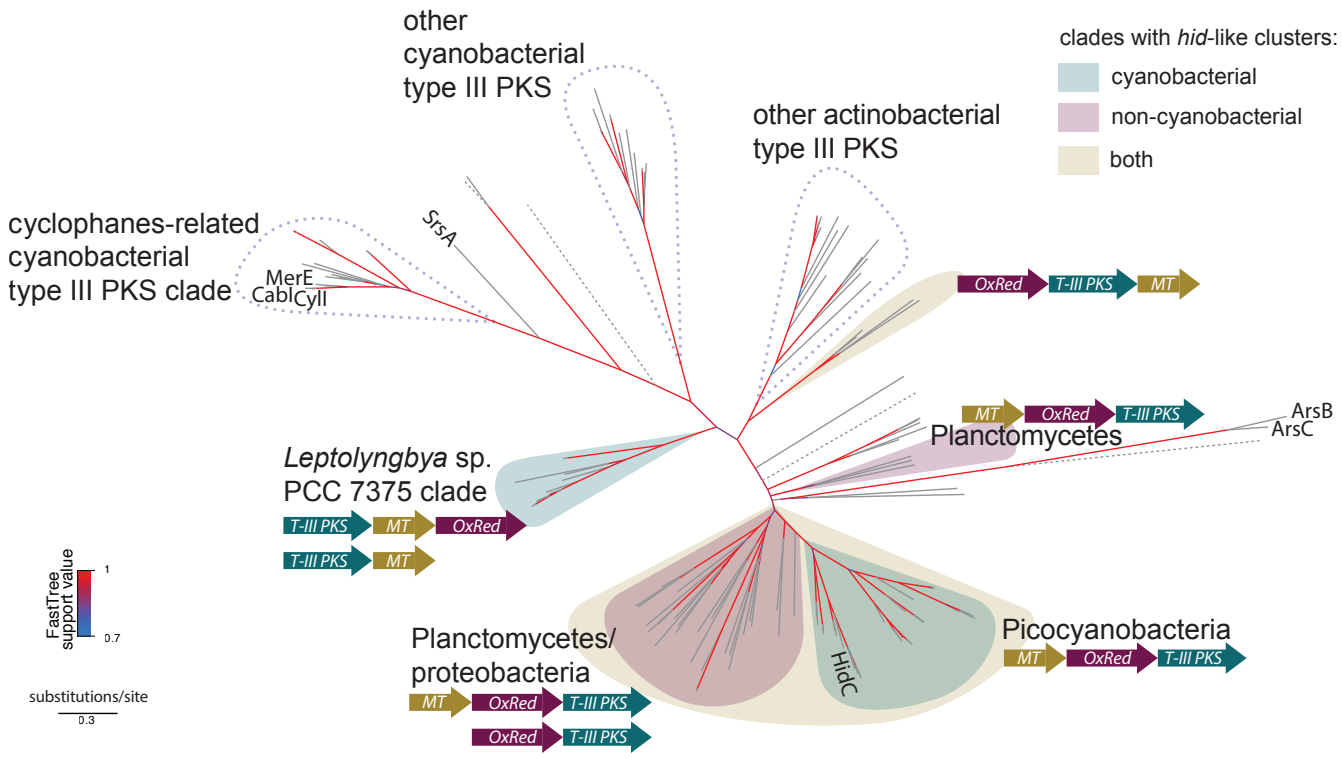


Figure 3 – Distribution of hid-like gene clusters among bacteria. FastTree 2 (approximately maximum-likelihood) phylogeny of bacterial homologs of type III PKS enzymes associated with alkylresorcinol biosynthesis, highlighting clades dominated by type III PKS enzymes that are encoded in clusters with the same general architecture of the hid gene cluster – i.e., those putatively coding for an additional SAM-dependent methyltransferase (MT) and a FAD-dependent oxidoreductase (OxRed). In the highlighted clades, the organization of clusters lacking either the MT or the OxRed are also shown. Dotted branches correspond to cyanobacterial type III PKSs that do not fall within the highlighted clades.

available, which presented an opportunity to test the ability of HidA to methylate this substrate and confirm the involvement of the *hid* BGC in hierridin production. In fact, *E. coli* BL21 cells carrying an expression vector with a N-His_6-HidA construct and grown in the presence of **9** produced at least two new compounds compared to an empty vector control. One of these compounds presented a HRESIMS-derived *m/z* value consistent with the mono-methylated alkylresorcinol **10**. The other compound, eluting later in reversed-phase conditions, showed no ionization in LC-HRESIMS analysis and remains unidentified (Fig. 2C). Nevertheless, the observed later elution (compared to **10**) and poor ionization would be expected for a dimethylated alkylresorcinol such as **11**.

Curiously, the *hid* gene cluster does not contain a halogenase, nor could one be found among the genome data of this cyanobacterium through blast searches using the sequence of flavin-dependent ChlA (that acts on phenolic rings)[21] or those of members of other halogenase classes[22] as queries. Exposure of hierridin B to a lysate of *Cyanobium* sp. LEGE 06113 cells did not lead to any observable conversion to hierridin C (not shown), therefore the nature of the halogenation event remains unclear. Interestingly, no halogenase was found in the genome of *Synechocystis salina* LEGE 06155 that could explain an analogous halogenation in the biosynthesis of bartoloside D.[23]

Distribution of the *hid* gene cluster among cyanobacterial genomes

The occurrence of hierridin B in phylogenetically distant cyanobacteria from the *Cyanobium* and *Leptolyngbya* genera[6] suggests that the *hid* gene cluster could be widely distributed in the cyanobacterial tree of life. BlastP searches, using the amino acid sequence of HidC as query against the NCBI database, revealed an elevated number of close homologs, found mostly among Cyanobacteria, but also in Planctomycetes and Proteobacteria. Many of these homologs were part of *hid*-like clusters, i.e. containing a FAD-dependent oxidoreductase and a SAM-dependent methyltransferase. To obtain a clearer picture of the distribution of these clusters in the Cyanobacteria phylum we retrieved and aligned the sequences of the top HidC homologs identified from BlastP searches (we retrieved a total of 149 sequences of top Blast hits from searches limited and not-limited to cyanobacterial sequences), together with the amino acid sequences of the alkylresorcinol-generating SrsA, ArsB and ArsC type III PKSs. A phylogenetic analysis of the resulting alignment revealed that several well-supported groups encoded homologs of HidC and that these were encompassed in a major phylogenetic clade comprised of picocyanobacterial as well as other bacterial type III PKSs that were part of *hid*-like gene clusters (Fig. 3). Within this larger clade, picocyanobacterial *hidC* homologs (and *hidC*) formed a well-supported independent sub-clade. In fact, the picocyanobacterial *hidC* homologs were found within *hidABC* gene clusters, while non-cyanobacterial *hidC* homologs were found in gene clusters with either *hidABC* or *hidBC* architectures. The genomes of filamentous cyanobacteria closely-related to the hierridin-producing *Leptolyngbya ectocarpi* SAG 60.90 encoded HidC homologs that constituted a separate clade in the resulting phylogenetic tree – these were mostly found within *hidCAB* or *hidCA* gene cluster architectures. Overall, *hid*-like gene clusters seem to be prevalent within certain cyanobacterial phylogenetic groups, most notably picocyanobacteria, which contrasts with the small number of alkylresorcinols reported from these organisms.

Experimental

General Experimental Procedures

UV spectra were acquired on a Spectramax spectrophotometer (Molecular Devices). NMR spectra were recorded on a 400 MHz spectrometer (Avance III, Bruker) and on a 600 MHz spectrometer equipped with a 5 mm cyroprobe (Avance III HD, Bruker). $CDCl_3$ or DMSO-d_6 were used as solvents, and the chemical shifts are reported relative to residual $CHCl_3$ or DMSO, respectively. Both semi-preparative and analytical-scale RP-HPLC for the isolation of hierridin C were performed using a system composed of a 1525 pump module and a 2485 UV-Vis detector (Waters) and using HPLC-grade solvents. All other solvents used were ACS grade. GC-MS analysis was performed as reported previously for **1**,[6] using a TRACE GC Ultra (Thermo Scientific) gas chromatograph coupled with a Polaris Q ion trap mass spectrometer. The system included an AS-3000 autosampler. A ZB-XLB capillary column (30 m × 0.25 mm × 0.25 µm) from Phenomenex (Torrance, CA, USA) was selected for chromatographic separation. Ultra-pure grade Helium was used as a carrier gas (flow 1.0 mL min^{-1}). HRESIMS data of pure **2** was acquired on an LTQ Orbitrap XL spectrometer, controlled by LTQ Tune Plus 2.5.5 and Xcalibur 2.1 (Thermo Scientific). The capillary voltage of the ESI was set to −3000 V. The capillary temperature was 275 °C. The sheath gas flow rate (nitrogen) was set to 40 units. The capillary voltage was −35 V and the tube lens voltage −200 V. Samples were injected at a concentration of approx. 50 µg/mL. The same instrument and analysis was used for synthetic compound characterization, but positive mode was used (ESI capillary voltage was 3.1 kV and temperature was set at 275 °C, sheath gas was set at 6 units and the capillary voltage was set at 35 V with a tube lens voltage set at 110 V). For the synthesis of hierridins, all chemicals were of reagent grade and were used without further purification: tetradecylmagnesium chloride solution 1M in THF (CAS: 110220-87-6, Aldrich) and 2-hydroxy-3,5-dimethoxybenzaldehyde (97% purity, CAS: 65162-29-0, Fluorochem). All reactions were carried out under argon atmosphere. Analytical TLC was carried out on pre-coated silica gel plates (Merck 60 F254, 0.25 mm) using UV light and ethanolic solution of phosphomolybdic acid (followed by gentle heating) for visualization. Melting points were determined using a Stuart Scientific Bibby apparatus and are uncorrected. Full experimental procedures and compound characterization regarding the synthesis of compounds **1** and **2** are provided as ESI.

Biological Material

The cyanobacterium *Cyanobium* sp. LEGE 06113 was obtained from the LEGEcc.[24] The strain was cultured in 6 L flasks each with 4 L of Z8 medium[25] supplemented with 20 g L^{-1} NaCl, under constant sterile aeration and a light (photon irradiance ~30 µmol m^{-2} s^{-1})/dark cycle of 14:10 h. Temperature was maintained at 26 °C. When cultures reached late exponential phase (six to eight weeks post inoculation), cyanobacterial cells were harvested by centrifugation and rinsed with deionized water. The biomass was then freeze-dried and kept at -20 °C until extraction.

Isolation of hierridin C (2)

Freeze-dried cells from *Cyanobium* sp. LEGE 06113 (16.0 g, d.w.) were extracted repeatedly with a warm (< 40 °C) mixture of CH_2Cl_2:MeOH (2:1). The resulting slurry was filtered through Whatman grade 1 qualitative filter paper (GE Healthcare Life Sciences, Little Chalfont, UK) and concentrated in a rotary evaporator to yield 2.58 g of crude extract. This was fractionated by Vacuum Liquid Chromatography (VLC) using silica gel 60 (0.015-0.040 mm, Merck KGaA, Damstadt, Germany) as stationary phase. Four fractions (A, A1, A2 and A3) were initially collected using 9:1, 87:13, 17:3 and 41:9 mixtures (*n*-hexane:EtOAc), respectively, as mobile phases. A steeper mobile-phase gradient was then applied to the column, from 4:1 (*n*-hexane:EtOAc) to 100% EtOAc and then to 100% MeOH, resulting in eight additional fractions. Fraction A1 (20.4 mg) was further fractionated by semi-preparative RP-HPLC with a Luna C18 column (10 µm, 250×10 mm, Phenomenex) under isocratic conditions (97% aqueous MeCN, 3 mL min^{-1}). A chromatographic peak eluting at RT = 27.1 min was collected and found to contain **1** as a major constituent. The sample was re-injected into the HPLC system fitted with the above-mentioned column, and further separated isocratically, using 90% MeCN (aq) as mobile phase. Pure **1** (0.4 mg) eluted at 41.5 min and an additional peak eluting at 50.7 min was collected. The ^{1}H NMR ($CDCl_3$, 400 MHz) of this latter peak showed a major constituent with a set of resonances that indicated relatedness to compound **1**. An additional round of analytical-scale HPLC with a Synergi Fusion-RP column (4 µm, 250×4.6 mm, Phenomenex) and using isocratic elution with 85% MeCN (aq), afforded **2** (tR = 38.9 min) as a white amorphous solid (0.3 mg).

Hierridin C (2). White, amorphous solid, UV/Vis (MeOH) λ_{max} (log ε) 211 (3.0), 249 (2.3), 271 (2.2), 281 (2.2) nm; ^{1}H and ^{13}C NMR, Table 1; HRMS *m/z* 397.25298 $[M-H]^-$ (calcd for $C_{23}H_{38}ClO_3^-$, 397.25150).

Biological assays

Human cell line cytotoxicity assays. The following cell lines were exposed to metabolite **2**: hepatocellular carcinoma HepG2, colon adenocarcinoma HT-29, neuroblastoma SH-SY5Y, breast carcinoma T47D and normal prostate cell line PNT2 (Sigma-Aldrich), breast adenocarcinoma and RKO colon carcinoma (ATCC), and MG-63 osteosarcoma (ATCC). Cells were cultured as previously described.[6] Cells were seeded in 96-well culture plates at a concentration of 10^4 cells cm^{-2}. After 24 h of adhesion, cells were exposed for 24 or 48 h to 99 µL of fresh medium supplemented with 1 µL of **2** in DMSO to final concentration of 3 and 30 µg mL^{-1}. After incubation, cells were exposed to 10 µL of 0.5 mg mL^{-1} MTT.[26] Following exposure, purple-colored formazan salts were dissolved in 100 µL DMSO and the absorbance measured at 550 nm in a microplate reader (Synergy HT, Biotek). All tests were run in triplicate.

Antiplasmodial assays. Laboratory-adapted *P. falciparum* 3D7 and Dd2 strains were continuously cultured as previously described,[27] with minor modifications. Briefly, parasites were cultivated on human erythrocytes suspended in RPMI 1640 medium supplemented with 25 mM HEPES, 6.8 mM hypoxanthine and 10% AlbuMAX II (Life Technologies), at pH 7.2. Cultures were maintained at 37 °C under an atmosphere of 5% O_2, 3–5% CO_2 and N_2, and synchronized by sorbitol treatment prior to the assays.[28] Staging and parasitaemia were determined by light microscopy of Giemsa-stained thin blood smears. The antimalarial activities of **1** and **2** were determined using the SYBR Green I (Life Technologies) assay as previously described[29] with modifications. Briefly, early ring stage parasites were tested in triplicate in a 96-well plate and incubated with drugs for 48h (37°C, 5% CO_2), parasite growth was assessed with SYBR Green I. Each compound was tested in a concentration ranging from 10-0.0097µM (0.2% DMSO).

Fluorescence intensity was measured with a multi-mode microplate reader (Triad, Dynex) with excitation and emission wavelengths of 485 nm and 535 nm, respectively, and analyzed by nonlinear regression using GraphPad Prism 5 (GraphPad Software).

Genome sequencing and mining

Genomic DNA was isolated from a fresh pellet of a *Cyanobium* sp. LEGE 06113 culture, using a CTAB-chloroform/isoamyl alcohol-based protocol.[30] The gDNA was fragmented using the AB Library Builder System (Applied Biosystems) and subsequently submitted for sequencing on the Ion Torrent PGM (ThermoFisher Scientific) platform. After sequencing, the reads were evaluated for quality using FastQC25[31] followed by assembling with SPAdes version 3.7.0.26.[32] Using the Geneious 8.0 software package (Biomatters), the generated contigs were used to setup a local database. The cyanobacterial type III PKS CylI (GenBank: AFV96143) amino acid sequence was used to perform a tblastn search against the local database of *Cyanobium* sp. LEGE 06113 genome data. A single hit was found (e-value 2.4e-29), within a 387 kb contig. The genomic context of these genes was annotated manually, by using the products of the predicted ORFs as queries in BlastP searches against the NCBI protein database (Table S1). Annotated nucleotide data have been deposited in GenBank under accession number MH824152.

Heterologous expression of *hidA* in *E. coli* and metabolite profiling

Cloning of hidA. Forward (NdeI-hidA-v2, 5' GAG ATC TCA TAT GGG TGG GCG CTC GGC 3') and reverse (HindIII-hidA-NHis, 5' GAT CCA AGC TTT CAT CCA GCC GGA GCC 3') primers containing recognition sequences for NdeI and HidIII, respectively, were used to amplify the *hidA* gene from the gDNA of *Cyanobium* sp. LEGE 06113. The *hid* gene cluster has a high GC content and successful amplification by PCR required an optimized two-step protocol that used the Phusion High Fidelity DNA Polymerase (New England Biolabs) with GC buffer and DMSO. The PCR reactions were prepared in a volume of 20 µL containing 1X GC buffer, 200 µM dNTPs, 0.75 µM of each primer, 3% DMSO and 0.4U Phusion polymerase and 1 µL of template gDNA. The thermal cycling protocol consisted of an initial denaturation step at 98 °C for 45 s, followed by 35 cycles of a 10 s denaturation step (98 °C) and a 30 s annealing/extension step at 72 °C, before a final extension at 72 °C for 7 min. The resulting PCR mixture was loaded onto a 1 % agarose gel stained with Sybr Safe and visualized under a blue light transilluminator, before excision of a slice containing the single band (~800 bp). The DNA was purified from the agarose using the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare) and ligated to a pET-28a vector (Novagen), using T4 DNA ligase (Takara) after double digestions with NdeI and HindIII (NZYTech). The resulting pET-28a-hidA construct encoded a $NHis_6$-tagged version of HidA and was cloned into *E. coli* BL21 DE3 chemically competent cells (NZYTech).

In vivo heterologous overexpression of hidA. The pET-28a-hidA construct led to overexpression of a ~35 kDa protein in *E. coli* BL21 DE3 cells in LB medium (expected MW for N-His_6-HidA is 31 kDa) that was most evident at 37 °C without IPTG addition (Fig. S12), therefore such conditions were used in the *in vivo* assays. These were carried out using *E. coli* DE3 cells carrying either the pET-28a-hidA or the empty pET-28a plasmid that were cultured at 37 °C in a shaker (200 rpm) until reaching an OD_{600} of 0.3. At that point, 30 µL of a 10 mM solution of 5-pentadecyl resorcinol (**9**) in DMSO (or 30 µL of DMSO, as a control) were added to each culture. Following overnight growth, cell pellets were extracted with EtOAc and MeOH (pooled and repeated three times for each solvent). After solvent removal in a rotary evaporator, the resulting extracts were dried and kept at -20 °C until being used for HPLC separation or LC-HRESIMS analysis.

HPLC analysis of E. coli extracts, peak collection and LC-HRESIMS analysis. The *E. coli* extracts resulting from the feeding experiments with substrate **9** were resuspended in MeOH (1 mg mL^{-1}) and analysed by HPLC-PDA. Separation was carried out in a Synergi 4u Fusion-RP column (250 × 10 mm, Phenomenex) using a linear gradient from 20% MeCN (aq) to 100% MeCN over 20 minutes, held for 25 additional minutes before returning to the initial conditions. The flow rate was 3 mL min^{-1} and the injection volume was 40 µL. Two peaks that were only observed for the extract of *E. coli* cells carrying the pET-28a-hidA plasmid and fed with compound **9** were collected individually (tR = 23.8 and 30.2 min) and pooled from multiple injections (scaled-up to 250 µL). After evaporation of the eluents in a rotary evaporator, the two collected peak samples were resuspended in a 1:10:4 solution of $CHCl_3$/MeOH/MeCN to a concentration of 1 mg mL^{-1} and filtered through 0.2 µm syringe filters. The samples were then used for LC-HRESIMS analysis on a system composed of a Dionex Ultimate 3000 HPLC system coupled to a qExactive Focus Orbitrap mass spectrometer, controlled by XCalibur 4.1 software (Thermo Fisher Scientific). Separation was carried out in a Luna C18 column (3 µm, 100 × 3 mm, Phenomenex) under a flow rate of 0.4 mL min^{-1}, with a gradient from 60% solution A (H_2O + 0.1% formic acid) 40% solution B (MeOH + 0.1% formic acid), to 100% solution B for 7 min and then held at 100% solution B for 8 additional minutes before returning to the initial conditions. The UV absorbance of the eluate was monitored at 280 nm and a Full MS acquisition was carried out in positive mode (spray voltage +2600 V, capillary temperature 320 °C, sheat gas parameter set to 35).

Phylogenetic analysis of the *hid* gene cluster

The sequences of HidC homologs were retrieved from the GenBank following a BlastP search and were aligned using the MUSCLE package in the Geneious software. The resulting alignment was used for phylogenetic analysis. An approximately-maximum-likelihood tree was inferred using FastTree 2[33] (from within the Geneious software package). In addition, the genomic context of the corresponding *hidC* genes was explored and annotated (using BlastP for the predicted amino acid sequences of the surrounding open reading frames).

Conclusions

Our investigations of hierridin-like alkylresorcinol biosynthesis in cyanobacteria suggest that these metabolites are prevalent in the world's oceans. Because many of the strains that harbour *hid* gene clusters are picoplanktonic unicellular cyanobacteria with a major primary production impact in the world's oceans, the ecological role of these compounds and associated ecological implications deserve further study. Phenolic lipids structurally related to hierridins interact with biological membranes[34] with known implications in encystment[19] and antibacterial resistance,[35] and it is likely that hierridins play similar role(s) in cyanobacteria, despite their low abundance in *Cyanobium* sp. LEGE 06113. On the other hand, also noteworthy is the similarity of hierridins to the chlorinated

hexaphenones DIF-1, DIF-2, DIF-3 (chlorinated hexaphenones)[36] and to MPBD (a methylated alkylresorcinol)[37] associated with developmental signaling in the slime mold *Dictyostelum*. Therefore, it is also possible that hierridins have evolved to carry out a signaling or interference role in the aquatic medium. In any case, our straightforward synthesis of the hierridins will facilitate future experimental explorations of these scarce metabolites, in particular of their biological function and antimalarial properties.

Acknowledgements

We thank CEMUP for HRESIMS and NMR analyses. This research was supported by the Structured Program of R&D&I INNOVMAR - Innovation and Sustainability in the Management and Exploitation of Marine Resources (reference NORTE-01-0145-FEDER-000035, Research Line NOVELMAR), funded by the Northern Regional Operational Program (NORTE2020) through the European Regional Development Fund (ERDF). This work was also funded by Fundação para a Ciência e Tecnologia (FCT), under the framework of projects UID/Multi/04423/2013, GHTM - UID/Multi/04413/2013 and PTDC/MAR-BIO/2818/2013, IFCT contract IF/01358/2014 to PNL and PhD scholarship SFRH/BD/93632/2013 to IESD. In addition, this work was also partially supported by the Associate Laboratory for Green Chemistry- LAQV which is financed by national funds from FCT/MCTES (UID/QUI/50006/2013) and co-financed by the ERDF under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007265).

Notes and references

§ The antiplasmodial activity of hierridin C (2) has formed the basis for of a patent application (WO2016207869).

1 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.

2 S. Yoshizawa, R. Matsushima, M. F. Watanabe, K. Harada, A. Ichihara, W. W. Carmichael and H. Fujiki, J. Cancer Res. Clin. Oncol., 1990, 116, 609–614.

3 J. Y. Hong and H. Luesch, Nat. Prod. Rep., 2012, 29, 449–456.

4 P. M. Shih, D. Wu, A. Latifi, S. D. Axen, D. P. Fewer, E. Talla, A. Calteau, F. Cai, N. Tandeau de Marsac, R. Rippka, M. Herdman, K. Sivonen, T. Coursin, T. Laurent, L. Goodwin, M. Nolan, K. W. Davenport, C. S. Han, E. M. Rubin, J. A. Eisen, T. Woyke, M. Gugger and C. A. Kerfeld, Proc. Natl. Acad. Sci. U. S. A., 2013, 110, 1053–1058.

5 D. J. Scanlan, M. Ostrowski, S. Mazard, A. Dufresne, L. Garczarek, W. R. Hess, A. F. Post, M. Hagemann, I. Paulsen and F. Partensky, Microbiol. Mol. Biol. Rev., 2009, 73, 249–299.

6 P. N. Leão, M. Costa, V. Ramos, A. R. Pereira, V. C. Fernandes, V. F. Domingues, W. H. Gerwick, V. M. Vasconcelos and R. Martins, PloS ONE, 2013, 8, e69562.

7 O. Papendorf, G. M. König and A. D. Wright, Phytochemistry, 1998, 49, 2383–2386.

8 M. Masi, A. Cimmino, A. Boari, A. Tuzi, M. C. Zonno, R. Baroncelli, M. Vurro and A. Evidente, J. Agric. Food Chem., 2017, 65, 1124–1130.

9 H. Kikuchi, I. Ito, K. Takahashi, H. Ishigaki, K. Iizumi, Y. Kubohara and Y. Oshima, J. Nat. Prod., 2017, 80, 2716–2722.

10 S. Niu, D. Liu, P. Proksch, Z. Shao and W. Lin, Mar. Drugs, 2015, 13, 2526–2540.

11 A. G. Gonzalez, J. Bermejo Barrera and E. M Rodriguez Perez, Phytochemistry, 1992, 31, 1436–1439.

12 J. E. Ellis and S. R. Lenger, Synth. Commun., 1998, 28, 1517–1524.

13 Y.-R. Liao, P.-C. Kuo, S.-C. Huang, J.-W. Liang and T.-S. Wu, Tetrahedron Lett., 2012, 53, 6202–6204.

14 C. Y. Hopkins and M. J. Chisholm, Can. J. Res., 1946, 24b, 208–210.

15 T. Mahajan, L. Kumar, K. Dwivedi and D. D. Agarwal, Ind. Eng. Chem. Res., 2012, 51, 3881–3886.

16 M. S. Costa, M. Costa, V. Ramos, P. N. Leão, A. Barreiro, V. Vasconcelos and R. Martins, J. Toxicol. Environ. Health A, 2015, 78, 432–442.

17 S. Śliwińska-Wilczewska, J. Maculewicz, A. Barreiro Felpeto and A. Latała, Toxins, 2018, 10, 48.

18 H. Nakamura, H. A. Hamer, G. Sirasani and E. P. Balskus, J. Am. Chem. Soc., 2012, 134, 18518–18521.

19 N. Funa, H. Ozawa, A. Hirata and S. Horinouchi, Proc. Natl. Acad. Sci. U. S. A., 2006, 103, 6356–6361.

20 C. Nakano, N. Funa, Y. Ohnishi and S. Horinouchi, J. Bacteriol., 2012, 194, 1544–1551.

21 C. S. Neumann, C. T. Walsh and R. R. Kay, Proc. Natl. Acad. Sci., 2010, 107, 5798–5803.

22 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.

23 P. N. Leão, H. Nakamura, M. Costa, A. R. Pereira, R. Martins, V. Vasconcelos, W. H. Gerwick and E. P. Balskus, Angew. Chem. Int. Ed., 2015, 54, 11063–11067.

24 LEGE Culture Collection Database, http://lege.ciimar.up.pt/, (accessed May 24, 2017).

25 J. Kotai, Nor. Inst. Water Res. Blindern Oslo, 1972, 11/69, 5 pp.

26 M. C. Alley, D. A. Scudiero, A. Monks, M. L. Hursey, M. J. Czerwinski, D. L. Fine, B. J. Abbott, J. G. Mayo, R. H. Shoemaker and M. R. Boyd, Cancer Res., 1988, 48, 589–601.

27 W. Trager and J. B. Jensen, Science, 1976, 193, 673–675.

28 C. Lambros and J. P. Vanderberg, J. Parasitol., 1979, 65, 418–420.

29 M. Machado, F. Murtinheira, E. Lobo and F. Nogueira, Ann. Clin. Med. Microbiol., 2016, 2, 1010.

30 S. P. Singh, R. P. Rastogi, D.-P. Häder and R. P. Sinha, World J. Microbiol. Biotechnol., 2010, 27, 1225–1230.

31 S. Andrews, 2010, Available online, http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

32 A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev and P. A. Pevzner, J. Comput. Biol., 2012, 19, 455–477.

33 M. N. Price, P. S. Dehal and A. P. Arkin, PLOS ONE, 2010, 5, e9490.

34 M. Stasiuk and A. Kozubek, Cell. Mol. Life Sci., 2010, 67, 841–860.

35 M. Funabashi, N. Funa and S. Horinouchi, J. Biol. Chem., 2008, 283, 13983–13991.

36 R. R. Kay, M. Berks and D. Traynor, Development, 1989, 107, 81–90.

37 T. Saito, G. W. Taylor, J.-C. Yang, D. Neuhaus, D. Stetsenko, A. Kato and R. R. Kay, Biochim. Biophys. Acta BBA - Gen. Subj., 2006, 1760, 754–761.