Biosynthesis of the unusual carbon skeleton of nocuolin A.

Teresa P. Martins1,2, Nathaniel R. Glasser3, Duncan J. Kountz2, Paulo Oliveira4,5,6, Emily P. Balskus3*, Pedro N. Leão1,*

1CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Matosinhos, Portugal
2ICBAS – Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal
3Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
4i3S – Institute for Research and Innovation in Health, University of Porto, Porto, Portugal
5IBMC – Institute of Molecular and Cell Biology, University of Porto, Porto, Portugal
6Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal

Abstract

Nocuolin A is a cytotoxic cyanobacterial metabolite that is proposed to be produced by enzymes of the noc biosynthetic gene cluster. Nocuolin A features a 1,2,3-oxadiazine moiety, a structural feature unique among natural products and, so far, inaccessible through organic synthesis, suggesting that novel enzymatic chemistry might be involved in its biosynthesis. This heterocycle is substituted with two alkyl chains and a 3-hydroxypropanoyl moiety. We report here our efforts to elucidate the origin of the carbon skeleton of nocuolin A. Supplementation of cyanobacterial cultures with stable isotope-labeled fatty acids revealed that the central C13 chain is assembled from two medium-chain fatty acids, hexanoic and octanoic acids. Using biochemical assays, we show that a fatty acyl-AMP ligase, NocH, activates both fatty-acids as acyl adenylates, which are loaded onto acyl carrier protein domains and undergo a non-decarboxylative Claisen condensation catalyzed by the ketosynthase NocG. This enzyme is part of a phylogenetically well-defined clade within similar genomic contexts. NocG presents a unique combination of characteristics found in other ketosynthases, namely in terms of substrate specificity and reactivity. Further supplementation experiments indicate that the 3-hydroxypropanoyl moiety of 1 originates from methionine, through an as-yet-uncharacterized mechanism. This work provides ample biochemical evidence connecting the putative noc biosynthetic gene cluster to nocuolin A and identifies the origin of all its carbon atoms, setting the stage for elucidation of its unusual biosynthetic chemistry.

Introduction

Cyanobacteria are well-known producers of natural products with intriguing structures ranging from terpenes and alkaloids to polypeptides and nonribosomal peptides.1,2 Cyanobacterial secondary metabolites also exhibit a wide range of pharmacologically-relevant bioactivities, mostly anticancer-related.3 To synthesize such unique natural products, these organisms make use of complex secondary metabolic pathways.3 Enzymes in these pathways often catalyze challenging reactions currently unachievable by available synthetic methodologies.

Among the most structurally unique cyanobacterial secondary metabolites is nocuolin A (1, Fig. 1a), initially isolated by Hrouzek and co-authors from the cyanobacterium Nostoc sp. CCAP1453/38
on the basis of its cancer cell line cytotoxicity.\(^4\) Metabolite 1 was later isolated independently from *Nodularia* sp. LEGE 06071 and shown to impair mitochondrial oxidative phosphorylation.\(^5,6\) IC\(_{50}\) values for 1 in cancer cell line cytotoxicity assays ranged from the high nanomolar to low micromolar.\(^4,5\) The structure of 1 features a substituted 1,2,3-oxadiazine moiety, unprecedented among reported natural products. In fact, natural hydrazones per se are a very restricted class of compounds.\(^7\) A NP with close structural resemblance to the nocuolin A scaffold is geralcin D (Fig. 1a), isolated from *Streptomyces* sp. LMA-545.\(^8\) It is also noteworthy that while 1,3,4- and 1,2,4-oxadiazines have been prepared by chemical synthesis, 1,2,3-oxadiazines are yet to be accessed synthetically.\(^4,9\)

The unique structure of 1 is likely generated by previously unrecognized enzymatic chemistry or by a unique combination of known enzymes.
biochemistry. For instance, it is unclear whether the central carbon scaffold of I derives from a single fatty acid (FA) precursor or if it is the product of carbon-carbon (C-C) bond formation between two shorter-chain substrates. The biosynthetic origin of the 3-hydroxypropanoyl moiety and the oxadiazine ring are also intriguing, and so it the basis for N-N bond formation, as there are few examples of characterized N-N bond-forming enzymes and no homologs of such enzymes are present in the noc BGC.

The initial report of the discovery of I by Hrouzek et al. suggested that a 50 kb locus (noc, Fig. 1a) shared by the genomes of cyanobacteria producing I, was the likely nocoerin biosynthetic gene cluster (BGC). This proposal was based on the presence of genes encoding fatty acyl-AMP ligases (FAALs) and different nitrogen processing/incorporating enzymes in this locus (Fig. 1b). However, several of the putative noc genes have recently been implicated in the biosynthesis of chlorosphaerolactylates (e.g. 2, Fig. 1b). To acknowledge these findings, those genes were renamed as cly-F, constituting the cly BGC. Our group later showed that Nodularia sp. LEGE 06071 produces not only I and chlorosphaerolactylates, but also hybrids of these two molecules named nocoerolactylates (e.g. 3, Fig. 1a). This discovery led us to propose that the entire noc locus would be involved in the production of these larger metabolites. However, a direct experimental connection between noc genes and I has not been established.

In this study, we experimentally interrogate the formation of the carbon skeleton of I. Using stable-isotope labeled precursor supplementation experiments and in vitro enzymatic assays, we show that the carbon atoms of I originate from three different building blocks: hexanoic acid, octanoic acid and t-methionine. Both fatty-acids are activated by NocH (FAAL) and condensed by the ketosynthase (KS) NocG with high specificity. We show that NocG is part of phylogenetically well-supported clade with no other characterized members and that it enzyme combines, in a unique way, the specificity and reactivity of other stand-alone KSs. t-methionine is converted into the 3-hydroxypropanoyl moiety through an as-yet-uncharacterized mechanism. By identifying the origins of all carbon atoms in I and the enzymatic steps leading to the formation of its C10 alkyl moiety (Fig. 1c), our study provides a first glimpse at the biosynthetic events that underlie the unique molecular scaffold of I and firmly establishes its connection to the noc pathway.

**Results and Discussion**

**Annotation of the noc gene products and acyl-ACP dehydrogenase activity of Nocl and NocC**

To identify the Nocl enzymes potentially involved in the biosynthesis of the carbon skeleton of I, the previous noc gene product annotation from Hrouzek et al. was reviewed using the remote homology detection bioinformatics tools HHPredict and Swiss Model. The new data suggested that NocG could act as a ketosynthase (KS) and Nocl as an acyl-[acyl carrier protein (ACP)] dehydrogenase (Fig. 1d, Table S1). The noc locus also encodes a second predicted acyl-ACP dehydrogenase, NocC, as well as two putative FAALs, NocH and ClyA (NocL), and a stand-alone ACP, ClyB (NocM). Following heterologous expression and purification of NocC and Nocl as well as the ACP ClyB (Fig. S1), both enzymes were found to desaturate several fatty acyl-ClyB substrates (from hexanoyl- to dodecanoyl-thioesters), generated via incubation of different chain length acyl-CoAs with ClyB and the promiscuous 4-phosphopantetheinyl transferase, Sfp. This result confirmed the annotation of NocC and Nocl as acyl-ACP dehydrogenases (Fig. S2 and S3). We considered that these enzymes could be involved in the formation of the carbon skeleton of I by generating unsaturated intermediates suitable for C-C bond formation by the putative KS NocG. However, we detected unsaturated versions of the chlorosphaerolactylates in Nodularia sp. LEGE 06071 (Fig. S4), in coherence with the structures of the nocoerolactylates produced by this cyanobacterium, which are also unsaturated. We also observed a higher activity of NocC (compared to Nocl) towards ClyB-bound fatty acyl substrates (Fig. S2). Taken together these data are suggestive of an involvement of NocC in generating the β, γ-unsaturation of the C12 chain in the chlorosphaerolactylates.

**The C13 alkyl chain in I is derived from hexanoic and octanoic acids**

In our previous study, nocoerolactylates were found to incorporate three alkyl moieties derived from either hexanoic acid or longer fatty acids. The lactylate portion of these molecules accounts for one such incorporation, indicating that the two remaining moieties are used for the biosynthesis of I. We considered that these building blocks could be used to generate the C13 alkyl moiety of I. To explore this possibility, stable isotope-labeled FA supplementation experiments were carried out. In a first assay, Nodularia sp. LEGE 06071 was pulse fed with perdeuterated hexanoic acid (d1, hexanoic acid) and compared to a non-supplemented control. Following LC-HRMS analysis, we detected versions of I with shifts of m/z 11.069 and m/z 22.138, which are in agreement with incorporation of one and two d1-hexanoic acids, respectively, into I (Fig. 2a). A shift of m/z 21.132 was also evident (Fig. 2a) and could result from a loss of deuterium due to keto-enol tautomerism of a putative biosynthetic intermediate. MS/MS analysis confirmed that the two FAs are incorporated into the C13 chain (Fig. S5, S6). These results indicate that the C13 skeleton of I is formed from two FAs, implying C-C bond formation must occur during its assembly.

We next sought to determine the precise length of the FAs involved in the biosynthesis of I by performing supplementation experiments with a range of additional even-chain perdeuterated fatty acids (octanoic to tetradecanoic acids). The data revealed incorporation of a single d10-octanoic acid moiety with the loss of two deuterium atoms (d10-I) (Fig. 2b, Fig S6). No incorporation from FAs with longer chains was observed (Fig S7). We also supplemented Nodularia sp. LEGE 06071 with perdeuterated heptanoic acid (d11-heptanoic acid) and found no deuterium incorporation from this substrate (Fig. S7). Overall, these experimental results indicate that one hexanoic and one octanoic acid unit are used to generate the C10 moiety of I.

To pinpoint the positions of the hexanoic and octanoic acid derived atoms in the C13 chain of I, the [M+H]+ ion of d10-I, resulting from supplementation of Nodularia sp. LEGE 06071 with d10-octanoic acid (m/z 312.3145), was subjected to MS/MS analysis. Two diagnostic peaks that include position C-5 (Fig. S5) were prominent (m/z 110.096 and 127.123) and showed a mass shift of 1.006 amu
relative to the corresponding MS/MS peak for 1, indicating the incorporation of a single deuterium atom at position 5 (Fig. 2c). With these data, we concluded that hexanoic acid is incorporated into positions C-6 to C-11 and octanoic acid into the remaining portion of the C_{13} alkyl moiety, with a new C-C bond being formed between C-5 and C-6.

Based on the deuterium incorporation pattern from these previous experiments, we reasoned that C-C bond formation must occur with loss of C1 from octanoic acid, perhaps through decarboxylation. To test this hypothesis, Nodularia sp. LEGE06071 was supplemented with hexanoic acid-1-^{13}C (L-^{13}C-hexanoic acid) or octanoic acid-1-^{13}C (L-^{13}C-octanoic acid). Incorporation of one or two L-^{13}C-hexanoic acid units into 1 was evident, and MS/MS analysis of the labeled species was consistent with the predicted labeling at C-4 and C-6 (Fig. S8, S9). On the other hand, L-^{13}C-octanoic acid supplementation showed a low-abundance and unspecific (from MS/MS analysis) incorporation into 1 and other non-FA-derived molecules such as chlorophyll a and phaeophytin a, indicating that the observed labeling pattern is caused by \textsuperscript{13}C scrambling (likely as a result of \textsuperscript{13}CO\textsubscript{2} recycling through the Calvin cycle) (Fig. S10). Overall, these supplementation experiments with stable isotope-labeled fatty acids showed that construction of the C_{13} alkyl chain of 1 involves C-C bond formation between C-1 of hexanoic acid and C-2 of octanoic acid, with loss of the C-1 carboxylate from the latter substrate.

**NocH activates hexanoic and octanoic acids.**

Having established the origin of the C_{13} alkyl moiety in 1, we tried to identify the biosynthetic machinery involved in its generation. We considered whether the *noc* locus, in particular the non-acyl
genes, could encode enzymes capable of performing hexanoic and octanoic acid activation. The noc locus encodes two putative FAALs, NocH and ClyA (NocL). The latter has been implicated in the biosynthesis of the chlorosphaerolactylates and was predicted to activate dodecanoic acid, but also decanoic and tetradecanoic acids to a minor extent. We heterologously expressed and purified each FAAL as well as with their cognate ACPS in *E. coli*. NocH contains both FAAL and ACPS domains and was expressed as a single polypeptide, while ClyA and its associated ACPS ClyB were expressed and purified separately. Competition assays were performed for each FAAL using a range of even chained FAs from hexanoic to dodecanoic acids. As expected, ClyA mainly activated dodecanoic acid (Fig. 3). NocH showed a clear preference for hexanoic and octanoic acids (Fig. 3, Fig. S11, S12), and can therefore activate and load both FA substrates involved in the generation of the C13 alkyl chain of 1. Additionally, we tested and observed loading of tetradecanoic acid by ClyA (Fig. S12), further supporting the role of this enzyme in chlorosphaerolactylate biosynthesis. These experiments provided the first biochemical evidence supporting the connection between the noc locus and the synthesis of nuculactylates, chlorosphaerolactylates and 1.

*NocG is a ketosynthe that generates the C13 alkyl moiety in 1.* We then considered candidate enzymes for carrying out C-C bond formation between hexanoyl- and octanoyl-NocH thioesters encoded within the noc BGC. Based on the bioinformatic analysis detailed above (Fig 1b, Table S1), NocG was annotated as a putative KS. Recombinant NocG was obtained after codon optimization, expression in the toxic-protein resistant *E. coli* C43 strain and cobalt resin purification (Fig. S1, S13 Table S2). We then carried out coupled enzymatic assays with NocG, the FAAL-ACP NocH and hexanoic and octanoic acid substrates. In assay mixtures, we detected a low-abundance feature with m/z 241.1809 (Fig. 4a), compatible with the formation of a 13-carbon alkyl-β-ketoacid (4). The corresponding decarboxylated molecule 6-tridecanone (5) was also detected in this assay, while neither 4 nor 5 were detected in control assays lacking NocG (Fig. 4a). Compound 4 (but not 5) was also detected in assays with NocG and hexanoyl- and octanoyl-S-N-acetylcysteamine (SNAC) thioester substrate surrogates, but in a lower amount (Fig. 4a). MS/MS analysis of 4 corroborated the proposed structure (Fig. S14). In addition, we repeated the NocG enzymatic assays with different combinations of 1H- and 13C-labeled FAs and observed the formation of products with the predicted m/z values for the corresponding labeled versions of 4 (Fig. S15). With these results, we confirmed the role of NocG as the KS responsible for condensation of hexanoyl- and octanoyl-thioesters in the C13 chain of 1.

To clarify whether the C-C bond forming reaction catalyzed by NocG is decarboxylative, we quenched assay mixtures with sodium borohydride (NaBH4). Under these conditions, the β-hydroxy acid (6) was obtained and no 6-tridecanol was obtained (Fig. 4b), indicating that NocG catalyzes a non-decarboxylative Claisen condensation between hexanoyl- and octanoyl-ACP thioesters, giving rise to a C13 β-ketoacid. Because the two acyl-ACP dehydrogenases encoded in the noc locus (NocC and NocI) were able to accept octanoyl-ACP substrates (Fig. S1, S2), we considered that NocG might accept 2-octanoyl-ACP as a substrate (since the hexanoic acid-derived alkyl portion remains unchanged in 1). However, NocG was unable to generate 4 in vitro when substituting the octanoyl substrate with 2-octenoyl (Fig. S16). C-C bond formation also did not occur with a single thioester and one free fatty acid (Fig. S17). Therefore, saturated acyl-ACP thioesters are the substrates for this enzyme and the C-C bond formation reaction it catalyzes does not require the action of a dehydrogenase. Likewise, we considered that a pre-hydroxylated (C-3) version of octanoic acid could be a substrate of NocG, as observed for the heterodimer KS LstAB, during lipstatin biosynthesis. Such a possibility would be consistent with the C-4 oxymethine in 1. However, no activity was detected when NocG and NocH were incubated with hexanoic acid and 3-hydroxyoctanoic acid (Fig. S18a) and no NocH loading was observed for the hydroxylated fatty acid (Fig. S18b). Considering the lack of C-5 functionalization in 1, we hypothesize that 4 spontaneously decarboxylates to give 5. Formation of such an enolizable ketone intermediate is consistent with the presence of an abundant +21 amu peak in addition to the expected but less prominent +22 amu peak observed upon d17-hexanoic acid supplementation of *Nodulalaria* sp. LEGE 06071 cultures (Fig. 2a).

Interestingly, despite the typical promiscuity of KSs regarding the alkyl chain length and the fact that both hexanoic and octanoic acids are activated by NocH, we were unable to detect longer (C15, from two octanoic acid substrates) or shorter (C11, from two hexanoic acid substrates) versions of the β-ketoacid in NocG assay mixtures (Fig. S19a). Such reactivity is contrasting with closely-related stand-alone KSs like OleA, which is able to accept C8 to C16 acyl-CoA substrates or PpyS, which can condense C6 to C16 (with different branching) thioesters to generate photozymones A to H. On the other hand, LstAB generates a C22 alkaphatic skeleton with no longer or shorter homologs having been described. However, to our knowledge, the promiscuity of the LstAB heterodimer towards fatty-acyl substrates of different chain length was not tested.
Figure 4. NocG belongs to a new class of KSs. (a) LC-HRMS-derived EICs of the β-ketoacid 4 and ketone 5 produced by NocG upon enzymatic assays with hexanoyl- and octanoyl-thioesters. (b) LC-HRMS-derived EICs of 6 upon reduction of 4 with sodium borohydride proving the non-decarboxylative nature of NocG reaction. (c) Phylogenetic tree composed of NocG (*), its homologues and other known ketosynthases (the scale bar indicates the degree of divergence as substitutions per site).
Consequently, the selectivity of NocG when generating 4 from its structurally very similar substrates (hexanoyl-NocH and octanoyl-NocH thioesters) is striking and unique. This selectivity possibly arises from the need to control the size of downstream products, as longer or shorter versions of 1 in extracts of Nodularia sp. LEGE 06071 could be found (Fig.S19b). Another interesting feature of the biosynthesis of the C13 chain of 1 arises from the fact that NocH loads and activates the two substrates later used by NocG in vitro. Typically, FAALs load one free fatty-acid substrate, which can be further elongated by KSs, usually as part of polyketide synthases assembly lines.21, 22 To our knowledge, no previous studies have described the use of the same FAAL/ACP pair in loading two different substrates to be used simultaneously by the same downstream KS.

To gain insight into how NocG compares to previously characterized KSs, we performed a phylogenetic analysis of 303 different KSs sequences from the main phylogenetic groups described so far. In addition, the top 43 hits to NocG from BLASTp searches were also included in the analysis (Table S4). The resulting phylogenetic tree shows that NocG clades separately from all other characterized KSs (Fig. 4c). The NocG-containing clade comprises mostly proteins from Cyanobacteria and Actinobacteria, along with some Proteobacteria and Acidobacteria sequences. The “NocG clade” is most closely related to the clade that features PpyS enzymes.20 PpyS (identity/similarity to NocG: 36/57%, respectively) and its closest homologs are involved in the biosynthesis of photopyrones. Analogously to the NocG C-C bond formation, this class of KSs performs a non-decarboxylative Claisen condensation between a β-ketoacyl-ACP thioester and an acyl-ACP thioester partner. However, and unlike what is observed for NocG, this intermediate is not released and instead undergoes an intramolecular cyclization to generate the final pyrone product (Fig. 4c).20 Previous phylogenetic studies have shown that PpyS homologs clustered in two different clades, one including the two characterized enzymes involved in pyrone biosynthesis (PpyS and PyrS) and a second clade containing Nocardia, Microcystis and other genera with no clear biosynthetic role.20 It is within this second clade that NocG and its closest homologs cluster, hence this is a functionally separate clade from pyrone-forming KSs. Sequence similarity network (SSN) analysis of the NocG supported these findings, with NocG and PpyS forming distinct clusters (Fig. S20). Additionally, the Genomic Neighborhood Network (GNN) analysis of the SSN data indicated that the NocG and PpyS SSN clusters had entirely different genomic contexts (Fig. S21). Members of the SSN cluster that encompasses NocG and its closest homologs are most often associated with fatty acyl activation/tethering (NocH homologs) and amino acid transference (NocF homologs) (Fig. S21). Some of these clusters harbor additional homologs of the Noc enzymes, especially in Nocardia spp. (Fig. S22). On the other hand, PpyS homologs show much more diverse genomic contexts, mainly associated with sugar metabolism, and possess no obvious connection Noc-related enzymes (Fig. S21, Table 1). NocG also clades separately from other reported stand-alone KSs that use β-ketoacyl-CoA as substrates, like MxH (myxopyronin biosynthesis)23 and CorB (corallopyronin biosynthesis).24

Notably, our phylogenetic analysis also shows that despite showing similar reactivity, NocG is phylogenetically distant from OleA, a thiolase that performs a non-decarboxylative Claisen condensation between two long-chain fatty acyl-CoAs in the olefin biosynthetic pathway.19 The reaction catalyzed by NocG differs in its use of protein-tethered thioester substrates. Key catalytic residues are shared between NocG, PpyS and OleA homologs — in NocG these are E100 (deprotonation of the acyl intermediate), C125, H275 and N304 (covalent binding of the precursor) (Fig. S23, S24). Based on this, and on the similar position of the predicted catalytic residues (Fig. S25), we hypothesize a similar reaction mechanism that starts with the deprotonation of the α-carbon of octanoyl-NocH thioester by Glutamate 100, creating a nucleophile species that subsequently attacks the carbonyl carbon of hexanoyl-NocH thioester to form the new C-C bond.20 Thus, NocG is the first characterized member of a new class of KS and expands the diversity of this C-C bond forming enzyme family.

| Table 1 – Comparison between NocG and its closest homologs, PpyS and OleA. |
|--------------------------|--------------------------|--------------------------|
| **Substrates** | **Substrate Promiscuity** | **Reaction** |
| NocG | C6-ACP + C8-ACP | Non-decarboxylative Claisen condensation |
| PpyS | No | Claisen condensation + Lactonization |
| OleA | CoA | Non-decarboxylative Claisen condensation |
| **Predicted catalytic residues** | | |
| NocG | E100, C125, H275, N304 | sugar metabolism |
| PpyS | E105, C129, H281, N310 | - a |
| OleA | E117, C143, H285, N315 | - a |
| **Genomic context** | | |
| NocG | C13 alkyl-β-ketoacid | Photopyrone |
| PpyS | A-H (different chain lengths and branching) | - a |
| OleA | C19 to C12 alkyl-β-ketoacid | - a |

*Genomic context of OleA was not investigated.*
Figure 5. The 3-hydroxypropanoyl moiety of 1 is derived from C-2 to C-4 and N-1 of l-methionine. (a) HRMS spectra of 1 and its stable isotope labeled versions following supplementation with d3-l-methionine, 13C5-l-methionine and 15N,13C5-l-methionine, confirming incorporation of this precursor into 1. (b) HRMS/MS spectrum of 13C5-l-methionine confirming methionine incorporation into 3-hydroxypropanoyl moiety. (c) Fine structure of the M peak in the isotope clusters of [M+H]+ ions of key isotopologues of 1 (values next to each peak correspond to peak areas and their ratios are tabulated), supporting incorporation of a nitrogen atom from methionine into 1. (d) HRMS spectra of 1 and its stable isotope labeled versions following supplementation with 1,13C5-l-methionine, supporting a non-specific incorporation of C-1 from l-methionine. (e) schematic representation of incorporation of C and N atoms derived l-methionine into the 3-hydroxypropanoyl moiety of 1.
The 3-hydroxypropanoyl moiety of 1 is derived from L-methionine.

Having elucidated the biosynthesis of the C$_{13}$ alkyl chain of 1, we then focused on the origin of the remaining carbon atoms in this compound, namely the 3-hydroxypropanoyl moiety. We noticed that NocA and NocB are homologous to methylthioribulohe-1-phosphate dehydratase (MtnB) and acireductone dioxygenase (MtnD), respectively. These enzymes are part of the methionine salvage pathway (MTA salvage pathway) (Fig. 1b, Table S1), which regenerates L-methionine from its downstream metabolic products. In Bacillus subtilis, the MTA salvage pathway converts SAM (or a product of SAM metabolism) to methylthioribose (MTR), which is then phosphorylated by a methylthioribose kinase (MtnK) and converted to methylthioribulohe-1-phosphate (MTR-1P) by methylthioribulohe-1-phosphate isomerase (MtnA). MTR-1P is then converted to 1,2-diketo-5-methylthiopentyl-1-phosphate (2,3-DK-MTP-1-P) by MtnB. Alignment of NocA with characterized methylthioribulohe-1-phosphate dehydratases and alignment of NocB with characterized acireductone dioxygenases showed conservation of key residues (Fig. S26, S27), which could indicate the involvement of methionine in the biosynthesis of I and potential recycling of MTA by the salvage pathway. To confirm the suspected activity of NocA, this enzyme was heterologously expressed in E. coli, purified and tested in a coupled assay with MtnK and MtnA (from B. subtilis) with MTR-1P as substrate. The assay yielded 2,3-DK-MTP-1-P in LC-HRMS analyses, confirming that NocA has methylthioribulohe-1-phosphate dehydratase activity (Fig. S28).

The biochemical characterization of NocA and the bioinformatically predicted function of NocB as MTA salvage pathway enzymes support a role for methionine or S-adenosyl methionine (SAM) in the biosynthesis of I. There is precedence for SAM decarboxylation followed by transfer of its aminopropyl group to various metabolites. We hypothesized that a similar transformation could give rise to the 3-hydroxypropanoyl moiety in the biosynthesis of 1. To test this proposal, we supplemented Nodularia sp. LEGE 06071 with d-l-methionine. LC-HRMS data showed a prominent M+4 peak in the isotope cluster of I ([M+H]$^+$, Fig. 5a), indicating that methionine is in fact incorporated into I. To clarify which atoms in I derive from methionine, we supplemented Nodularia sp. LEGE 06071 with $^{13}$C$_{5}$-L-methionine and $^{13}$C$_{2}$-$^{15}$N-L-methionine (Fig. 5a). Three $^{13}$C$_{5}$-L-methionine carbon atoms were found to be incorporated into the 3-hydroxypropanoyl moiety, as revealed by MS/MS analysis (Fig. 5b), thereby establishing the origin of all carbon atoms in I. LC-HRMS/MS data from $^{15}$N-L-methionine supplementation experiments revealed a much higher $^{15}$N incorporation into I in MS/MS fragments containing N-1, suggesting a direct incorporation of one nitrogen atom from methionine. Additionally, the calculated ratios between direct $^{15}$N incorporation ([$^{15}$N-$^{13}$C$_{5}$-I/$^{15}$N-$^{13}$C$_{5}$-I]) versus scrambled single $^{15}$N incorporation in I ($^{15}$N-I/$^{15}$N-I) indicate a ten-fold increase as result of direct incorporation from $^{13}$C$_{5}$-$^{15}$N-L-methionine (Fig. 5c, Fig. S29). Further feeding with 1-3-13C-L-methionine also allowed us to conclude that the C-1 of L-methionine C-1 is not incorporated into I, and, therefore, that the 3-hydroxypropanoyl moiety is formed by L-methionine carbons C-2 to C-4 (Fig. 5d, S30).

Despite the role of SAM as a reactive one-carbon donor in the methylation of a wide range of substrates, the prosthetic group can also transfer amino groups, ribosyl groups and (as mentioned above) aminopropyl groups. Recently, Van Lanen et al. have reported the transfer of a C$_{3}$N group derived from SAM to the mu-raymymcins by a PLP-dependent enzyme, further highlighting the potential of SAM as a versatile biosynthetic precursor. Still, none of these reactions can explain by themselves the observed incorporation pattern for stable isotope-labeled methionine substrates into I (Fig. 5e). Overall, our findings strongly indicate that generation of the 3-hydroxypropanoyl scaffold in I involves unprecedented biochemical transformations.

Conclusions

This study shows that the carbon skeleton of the cytotoxic cyanobacterial metabolite I is formed from three different building blocks: hexanoic acid, octanoic acid and L-methionine (Fig. 1c). Both fatty acids are ultimately condensed by NocG with remarkable selectivity, forming a new C-C bond between carbons 4 and 5. A detailed study of NocG shows that it belongs to a new class of KSs, using saturated acyl-ACP thioesters as substrates and whose members are found mainly in cyanobacteria and actinobacteria. L-Methionine is incorporated into the 3-hydroxypropanoyl moiety of I through an as-yet-unclear mechanism that involves bond cleavage between C-1 and C-2. This work provides the first biochemical evidence connecting the putative noc pathway to I and reveals several instances of unusual biochemistry leading to a unique molecular scaffold. Further investigations on the biosynthesis of I, namely of the transformations leading to the 1,2,3-oxadiazine moiety are now facilitated by the identification of such key substrates and biosynthetic intermediates.

SUPPLEMENTARY INFORMATION

Supplementary information document (PDF) containing Materials and Methods, Supplementary Figures and Tables.

AUTHOR INFORMATION

*to whom correspondence should be addressed:
Emily P. Balskus, balskus@chemistry.harvard.edu;
Pedro N. Leão, pleao@ciimar.up.pt.

ACKNOWLEDGMENTS

PNL acknowledges funding from the European Research Council through a Starting Grant (759840), European Union’s Horizon 2020 programme (WIDESPREAD, Grant Agreement 952374) and Fundação para a Ciência e a Tecnologia (FCT) through strategic funding grant UIDB/04423/2020. TPM was supported by a scholarship SFRH/BD/138308/2018 from FCT and by the Fulbright Commission. EPB acknowledges support from Harvard University and by the National Science Foundation (NSF) through grants CHE-1454007 and CHE-2003436. NRG acknowledges the NSF Postdoctoral Research Fellowship in Biology (Grant No 1907240). We thank João Morais for assistance with cyanobacterial cultures.
and Matt Volpe for assistance with the synthesis of hexanoyl- and octanoyl-SNACs.

REFERENCES


