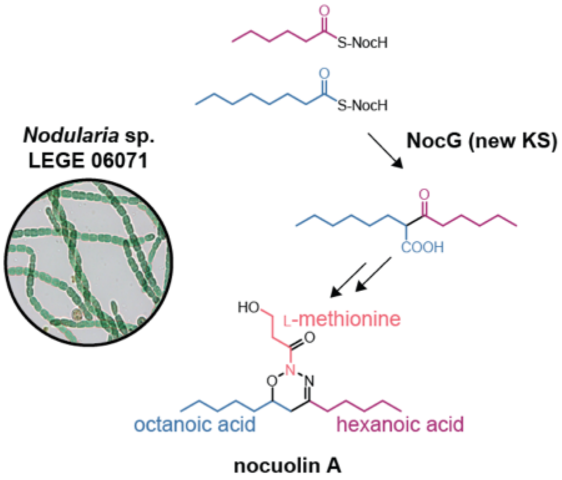
Biosynthesis of the unusual carbon skeleton of nocuolin A.

Teresa P. Martins[1,2], Nathaniel R. Glasser[3], Duncan J. Kountz[3], Paulo Oliveira[4,5,6], Emily P. Balskus[3]*, Pedro N. Leão[1,]*

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

Abstract

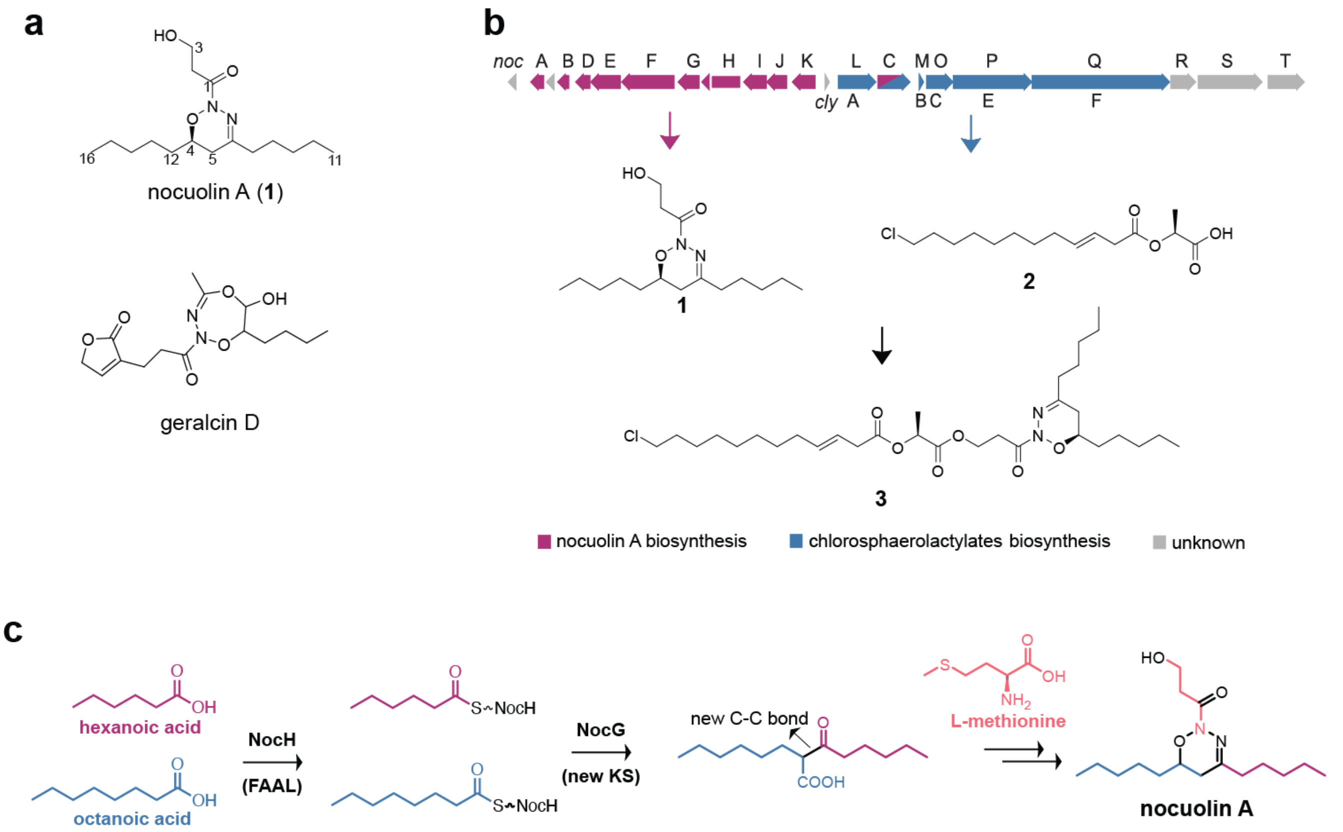
Nocuolin A is a cytotoxic cyanobacterial metabolite that is proposed to be produced by enzymes of the noc biosynthetic gene cluster. Nocuolin A features a 1,2,3-oxadiazine moiety, a structural feature unique among natural products and, so far, inaccessible through organic synthesis, suggesting that novel 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 C_{13} 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 polyketides and nonribosomal peptides.[1, 2] Cyanobacterial secondary metabolites also exhibit a wide range of pharmacologically-relevant bioactivities, mostly anticancer-related.[1] 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



d

***noc* locus (*Nodularia* sp. LEGE 06071)**

Protein	Predicted function	Protein	Predicted function
NocA	Methylthioribulose-1-phosphate dehydratase	**NocI**	Acyl-ACP dehydrogenase
NocB	Acireductone dioxygenase	**NocJ**	Rieske oxygenase
NocD	Methyltransferase	**NocK**	SAM methyltransferase
NocE	N-monooxygenase	**NocL** (ClyA)	Fatty acyl-AMP ligase (FAAL)
NocF	Aminotransferase	**NocC**	Acyl-ACP dehydrogenase
NocG	Ketosynthase	**NocM** (ClyB)	Acyl carrier protein (ACP)
NocH	Fatty acyl-AMP ligase (FAAL) ACP domain		

Figure 1. *Schematic highlighting the structure of Nocuolin A, the noc BGC and the proposed biosynthetic steps. (a) Chemical structures of nocuolin A (**1**) and geralcin D, the natural product with closest structural resemblance to the nocuolin A scaffold. (b) The noc locus and its proposed relationship with nocuolin A (**1**), the chlorosphaerolactylates (e.g. **2**) and nocuolactylates (e.g. **3**). (c) Representation of the proposed biosynthesis of **1**. (d) Detailed annotation of noc-encoded biosynthetic enzymes (some have previously been included in the cly BGC) with a potential role in the biosynthesis of **1**.*

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 biochemistry. For instance, it is unclear whether the central carbon scaffold of **1** 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[10, 11] and no homologs of such enzymes are present in the *noc* BGC.

The initial report of the discovery of **1** by Hrouzek *et al.* suggested that a 50 kb locus (*noc*, Fig. 1a) shared by the genomes of cyanobacteria producing **1**, was the likely nocuolin 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).[4, 12] However, several of the putative *noc* genes have recently been implicated in the biosynthesis of chlorosphaerolactylates (e.g. **2**, Fig. 1b).[12] To acknowledge these findings, those genes were renamed as *clyA-F*, constituting the *cly* BGC. Our group later showed that *Nodularia* sp. LEGE 06071 produces not only **1** and chlorosphaerolactylates,[6] but also hybrids of these two molecules named nocuolactylates (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.[6] However, a direct experimental connection between *noc* genes and **1** has not been established.

In this study, we experimentally interrogate the formation of the carbon skeleton of **1**. Using stable-isotope labeled precursor supplementation experiments and *in vitro* enzymatic assays, we show that the carbon atoms of **1** originate from three different building blocks: hexanoic acid, octanoic acid and L-methionine. Both fatty-acids are activated by NocH (FAAL) and condensed by the ketosynthtase (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. L-methionine is converted into the 3-hydroxypropanoyl moiety through an as-yet-uncharacterized mechanism. By identifying the origins of all carbon atoms in **1** and the enzymatic steps leading to the formation of its C_{13} alkyl moiety (Fig. 1c), our study provides a first glimpse at the biosynthetic events that underlie the unique molecular scaffold of **1** and firmly establishes its connection to the *noc* pathway.

Results and Discussion

Annotation of the noc gene products and acyl-ACP dehydrogenase activity of NocI and NocC.

To identify the Noc enzymes potentially involved in the biosynthesis of the carbon skeleton of **1**, the previous *noc* gene product annotation from Hrouzek *et al.* was reviewed using the remote homolog detection bioinformatics tools HHPred and Swiss Model.[13, 14] The new data suggested that NocG could act as a ketosynthase (KS) and NocI 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 NocI 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[15]. This result confirmed the annotation of NocC and NocI as acyl-ACP dehydrogenases (Fig. S2 and S3). We considered that these enzymes could be involved in the formation of the carbon skeleton of **1** 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 nocuolactylates produced by this cyanobacterium, which are also unsaturated.[6] We also observed a higher activity of NocC (compared to NocI) 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 C_{12} chain in the chlorosphaerolactylates.

The C13 alkyl chain in 1 is derived from hexanoic and octanoic acids.

In our previous study,[6] nocuolactylates 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,[12] indicating that the two remaining moieties are used for the biosynthesis of **1**. We considered that these building blocks could be used to generate the C_{13} alkyl moiety of **1**. 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 (d_{11}-hexanoic acid) and compared to a non-supplemented control. Following LC-HRMS analysis, we detected versions of **1** with shifts of *m/z* 11.069 and *m/z* 22.138, which are in agreement with incorporation of one and two d_{11}-hexanoic acids, respectively, into **1** (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 C_{13} chain (Fig. S5, S6). These results indicate that the C_{13} skeleton of **1** 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 **1** 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 d_{15}-octanoic acid moiety with the loss of two deuterium atoms (d_{13}-**1**) (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 (d_{13}-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 C_{13} moiety of **1**.

To pinpoint the positions of the hexanoic and octanoic acid derived atoms in the C_{13} chain of **1**, the $[M+H]^+$ ion of d_{13}-**1** , resulting from supplementation of *Nodularia* sp. LEGE 06071 with d_{15}-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

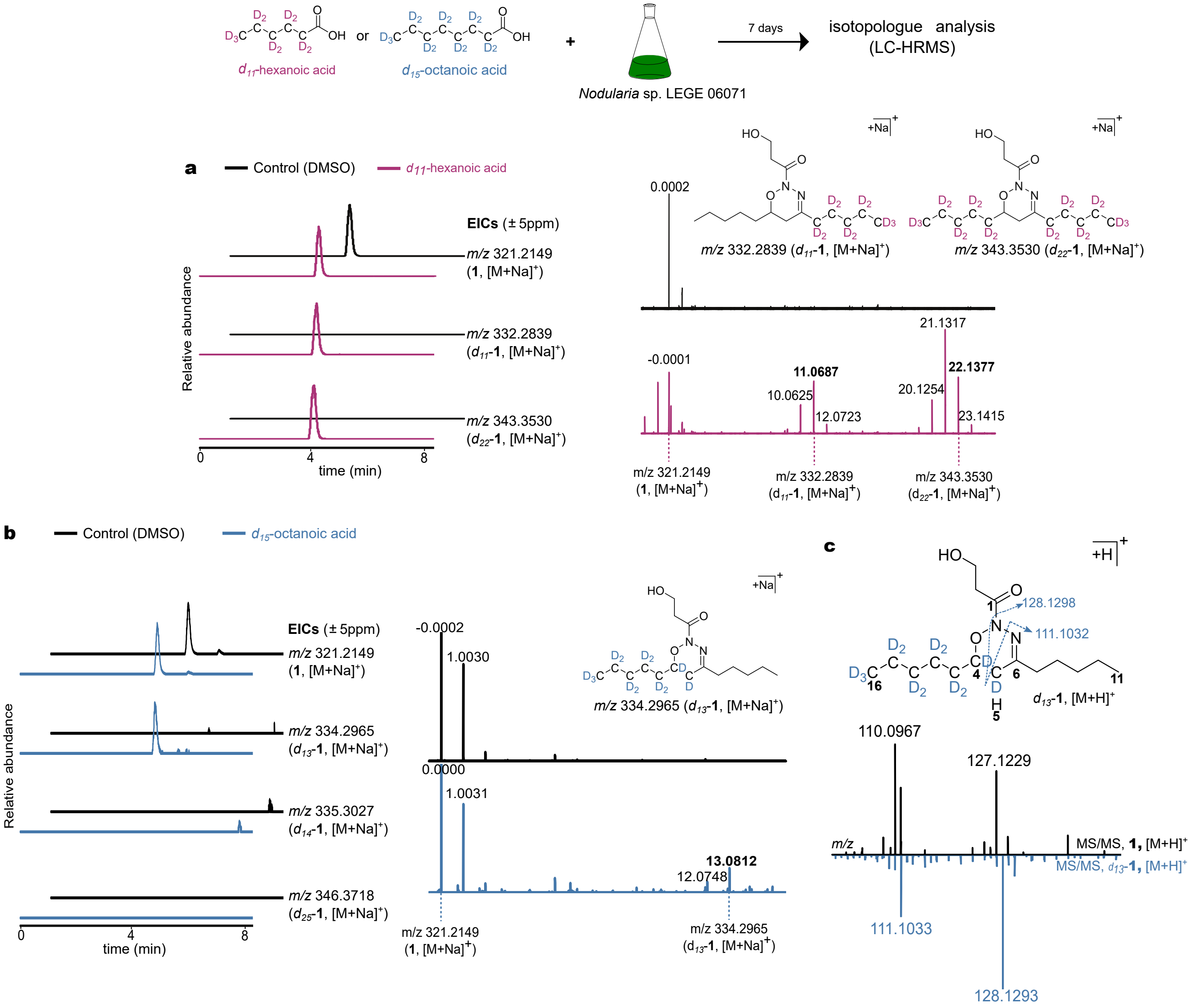


*Figure 2. Supplementation of Nodularia sp. LEGE06071 with d_{11}-hexanoic acid and d_{15}-octanoic acid leads to d_{11}-**1**/d_{22}-**1** and d_{13}-**1**, respectively. LC-HRMS-derived Extracted Ion Chromatograms (EICs) of **1** and labeled version with (a) d_{11}-hexanoic acid and (b) d_{15}-octanoic in Nodularia sp. LEGE 06071. (c) HRMS/MS spectrum of d_{13}-**1** and **1** ($[M+H]^+$) highlighting the diagnostic fragmentations supporting the structural proposal.*

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 (*1*-^{13}C-hexanoic acid) or octanoic acid-1-^{13}C (*1*-^{13}C-octanoic acid). Incorporation of one or two 1-^{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, *1*-^{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 ^{13}C scrambling (likely as a result of $^{13}CO_2$ 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-*cly*

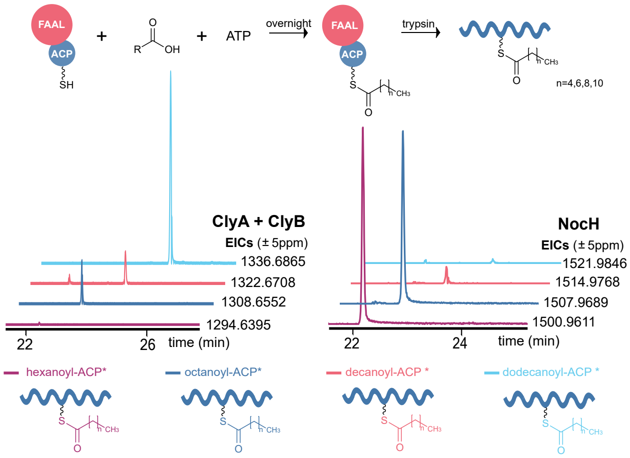


Figure 3. *The FAALs NocH and ClyA (NocM) activate hexanoic/octanoic acids and dodecanoic acid as acyl-ACP thioesters, respectively, in competition assays. Shown are LC-HRMS-derived EICs of a trypsinated ACP fragment loaded with each fatty-acid.*

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.[12] We heterologously expressed and purified each FAAL as well as with their cognate ACPs in *E. coli*. NocH contains both FAAL and ACP domains and was expressed as a single polypeptide, while ClyA and its associated ACP 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 C_{13} 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 nocuolactylates, chloroshpaerolactylates and **1**.

NocG is a ketosynthase that generates the C_{13} 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 ^{2}H- and ^{13}C-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 C_{13} chain of **1**.

To clarify whether the C-C bond forming reaction catalyzed by NocG is decarboxylative, we quenched assay mixtures with sodium borohydride ($NaBH_4$). 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 C_{13} 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-octenoyl-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.[16] 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 d_{11}-hexanoic acid supplementation of *Nodularia* sp. LEGE 06071 cultures (Fig. 2a).

Interestingly, despite the typical promiscuity of KSs regarding the alkyl chain length[17-19] and the fact that both hexanoic and octanoic acids are activated by NocH, we were unable to detect longer (C_{15}, from two octanoic acid substrates) or shorter (C_{11}, 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 C_8 to C_{16} acyl-CoA substrates[18] or PpyS, which can condense C_8 to C_{14} (with different branching) thioesters to generate photopyrones A to H.[17, 20] On the other hand, LstAB generates a C_{22} aliphatic skeleton with no longer or shorter homologs having been described.[16] However, to our knowledge, the promiscuity of the LstAB heterodimer towards fatty-acyl substrates of different chain length was not tested.

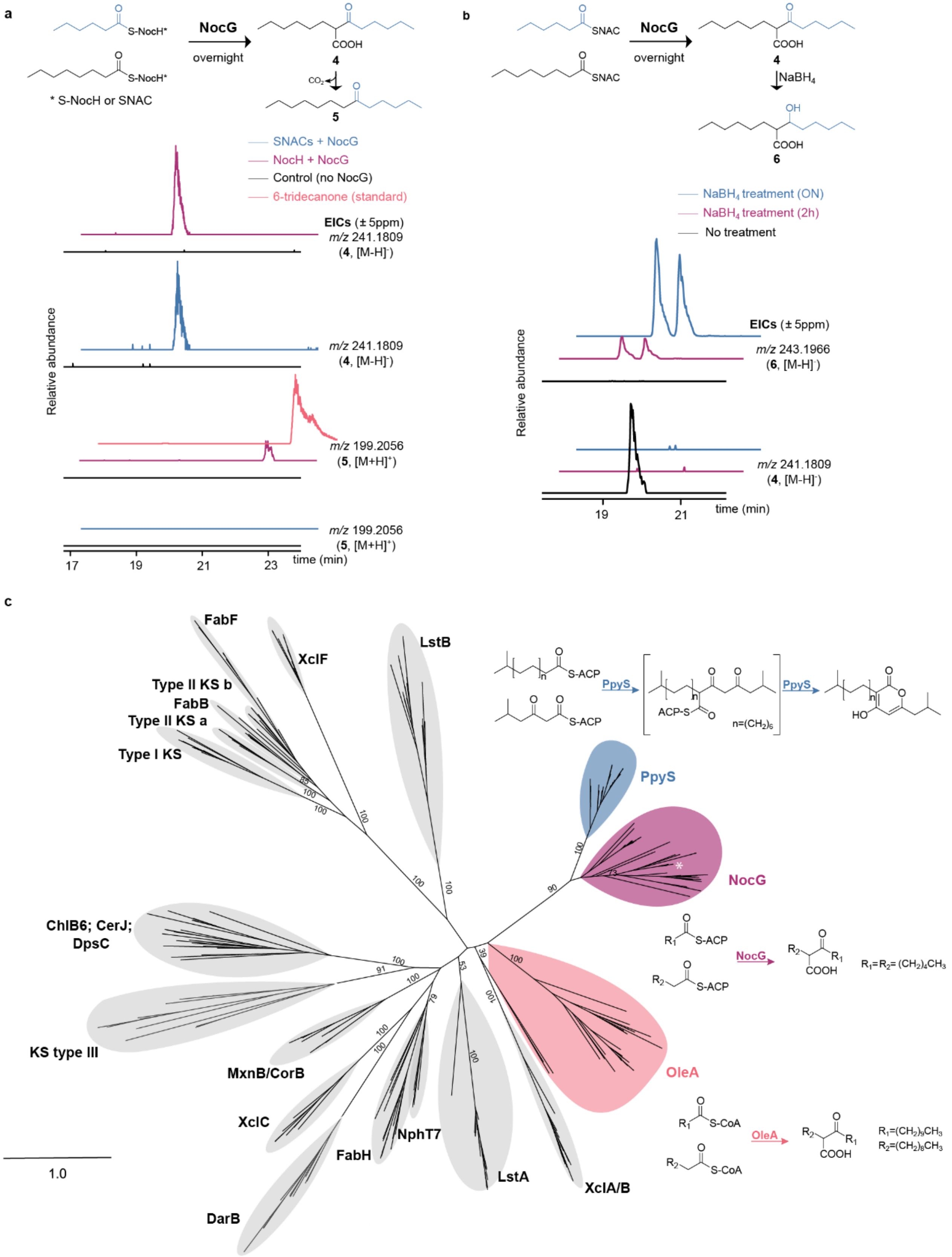


Figure 4. NocG belongs to a new class of KSs. (a) LC-HRMS-derived EICs of the β-ketoacid **4** and ketone **5** produced by NocG upon enzymatic assays with hexanoyl- and octanoyl-thioesters. **(b)** LC-HRMS-derived EICs of **6** upon reduction of **4** with sodium borohydride proving the non-decarboxylative 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 C_{13} 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.

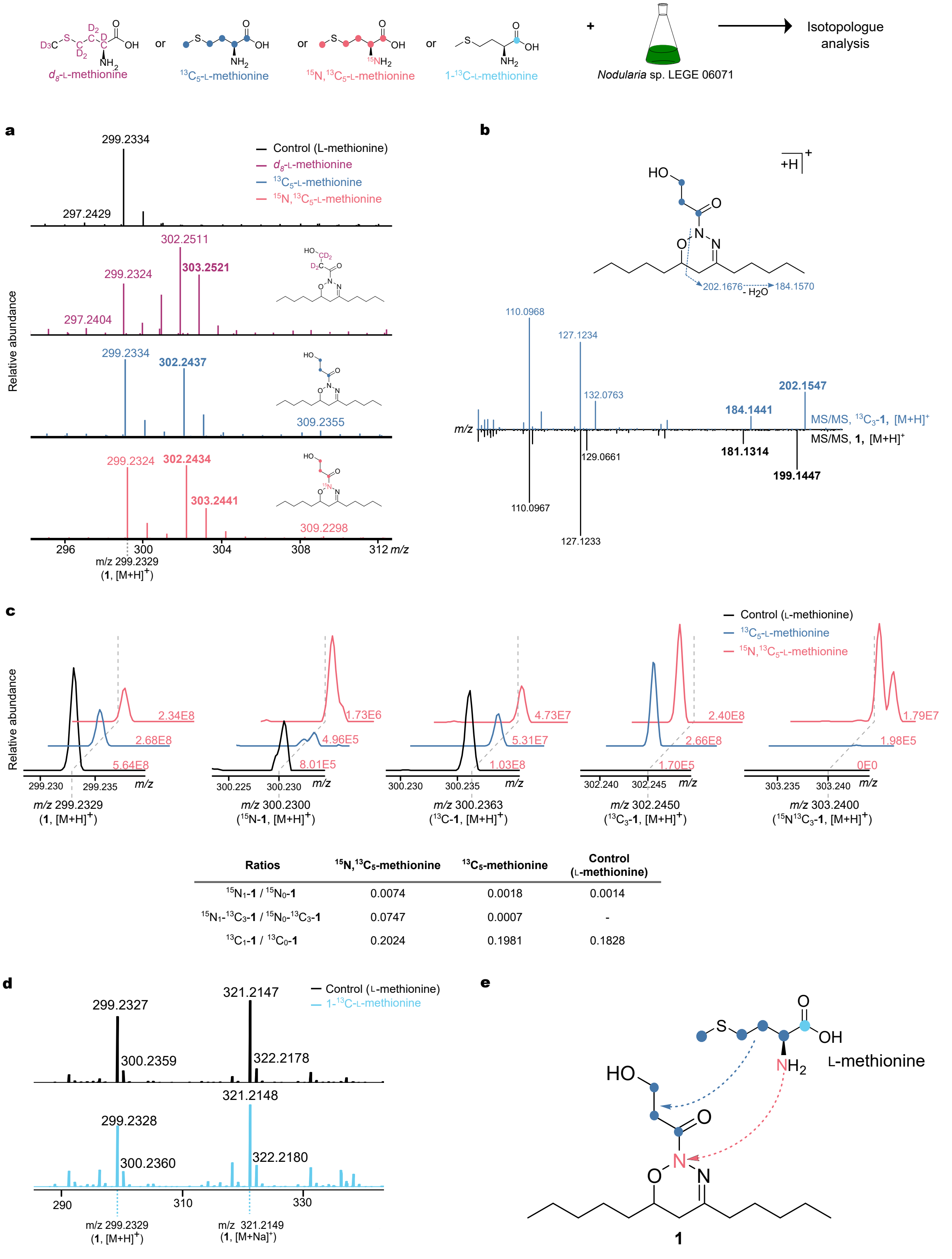
To gain insight into how NocG compares to previously characterized KSs, we performed a phylogenetic analysis of 303 different KSs sequences from the main phylogenetic groups described so far. In addition, the top 43 hits to NocG from BLASTp searches were also included in the analysis (Table S4). The 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-CoAs as substrates, like MxnB (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.[18] 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.*

	NocG	**PpyS**	**OleA**
Substrates	C6-ACP + C8-ACP	C_8 to C_{14} (different branching) + 3-oxo-C6-ACP	C_8 to C_{16} acyl-CoA
Substrate Promiscuity	No	Yes	Yes
ACP/CoA	ACP (a single FAAL/ACP for both substrates)	ACP or CoA	CoA
Reaction	Non-decarboxylative Claisen condensation	Non-decarboxylative Claisen condensation + Lactonization	Non-decarboxylative Claisen condensation
Predicted catalytic residues	E100, C125, H275, N304	E105, C129, H281, N310	E117, C143, H285, N315
Genomic context	fatty acyl activation/tethering and amino acid transference	sugar metabolism	- [a]
Final Product	C_{13} alkyl-β-ketoacid	Photopyrone A-H (different chain lengths and branching)	C_{19} to C_{32} alkyl-β-ketoacid

[a] *Genomic context of OleA was not investigated.*



Ratios	$^{15}N,^{13}C_5$-methionine	$^{13}C_5$-methionine	Control (L-methionine)
$^{15}N_1$-**1** / $^{15}N_0$-**1**	0.0074	0.0018	0.0014
$^{15}N_1$-$^{13}C_3$-**1** / $^{15}N_0$-$^{13}C_3$-**1**	0.0747	0.0007	-
$^{13}C_1$-**1** / $^{13}C_0$-**1**	0.2024	0.1981	0.1828

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 d_8-l-methionine, $^{13}C_5$-l-methionine and $^{15}N,^{13}C_5$-l-methionine, confirming incorporation of this precursor into* ***1****. (b) HRMS/MS spectrum of $^{13}C_3$-**1** 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-^{13}C-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 methylthioribulose-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 methylthioribose-1-phosphate (MTR-1P) by methylthioribose-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 methylthioribulose-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 **1** 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.[25, 26] The assay yielded 2,3-DK-MTP-1-P in LC-HRMS analyses, confirming that NocA has methylthioribulose-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 **1**. There is precedence for SAM decarboxylation followed by transfer of its aminopropyl group to various metabolites.[27] 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_8-L-methionine. LC-HRMS data showed a prominent M+4 peak in the isotope cluster of **1** ($[M+H]^+$, Fig. 5a), indicating that methionine is in fact incorporated into **1**. To clarify which atoms in **1** derive from methionine, we supplemented *Nodularia* sp. LEGE 06071 with $^{13}C_5$-L-methionine and $^{13}C_5$-^{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 **1**. LC-HRMS/MS data from $^{13}C_5$-^{15}N-L-methionine supplementation experiments revealed a much higher ^{15}N incorporation into **1** 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 ($^{15}N_1{}^{13}C_3$-**1**/$^{15}N_0{}^{13}C_3$-**1**) versus scrambled single ^{15}N incorporation in **1** ($^{15}N_1$-**1**/$^{15}N_0$-**1**) 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-^{13}C-L-methionine also allowed us to conclude that the C-1 of L-methionine C-1 is not incorporated into **1**, 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 [27]. Recently, Van Lanen *et al.*[28] have reported the transfer of a C_3N group derived from SAM to the muraymycins 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 **1** (Fig. 5e). Overall, our findings strongly indicate that generation of the 3-hydroxypropanoyl scaffold in **1** involves unprecedented biochemical transformations.

Conclusions

This study shows that the carbon skeleton of the cytotoxic cyanobacterial metabolite **1** 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 **1** 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 **1** and reveals several instances of unusual biochemistry leading to a unique molecular scaffold. Further investigations on the biosynthesis of **1**, 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

1. Burja, A. M.; Banaigs, B.; Abou-Mansour, E.; Burgess, J. G.; Wright, P. C., Marine cyanobacteria - a prolific source of natural products. *Tetrahedron* **2001,** *57* (46), 9347-9377.
2. Tidgewell, K.; Clark, B. R.; Gerwick, W. H., The Natural Products Chemistry of Cyanobacteria. *Comprehensive Natural Products Ii: Chemistry and Biology, Vol 2: Natural Products Structural Diversity-Ii: Secondary Metabolites: Sources, Structures and Chemical Biology* **2010**, 141-188.
3. Scott, T. A.; Piel, J., The hidden enzymology of bacterial natural product biosynthesis. *Nat Rev Chem* **2019,** *3* (7), 404-425.
4. 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.; Kopecky, 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.
5. Sousa, M. L.; Preto, M.; Vasconcelos, V.; Linder, S.; Urbatzka, R., Antiproliferative Effects of the Natural Oxadiazine Nocuolin A Are Associated With Impairment of Mitochondrial Oxidative Phosphorylation. *Front. Oncol.* **2019,** *9*, 224.
6. Figueiredo, S. A. C.; Preto, M.; Moreira, G.; Martins, T. P.; Abt, K.; Melo, A.; Vasconcelos, V. M.; Leao, P. N., Discovery of Cyanobacterial Natural Products Containing Fatty Acid Residues. *Angew. Chem. Int. Ed. Engl.* **2021,** *60* (18), 10064-10072.
7. Le Goff, G.; Ouazzani, J., Natural hydrazine-containing compounds: Biosynthesis, isolation, biological activities and synthesis. *Bioorg Med Chem* **2014,** *22* (23), 6529-6544.
8. Le Goff, G.; Martin, M. T.; Iorga, B. I.; Adelin, E.; Servy, C.; Cortial, S.; Ouazzani, J., Isolation and characterization of unusual hydrazides from *Streptomyces* sp. impact of the cultivation support and extraction procedure. *J. Nat. Prod.* **2013,** *76* (2), 142-149.
9. M. O. Hunsen, 1,2,3-Oxadiazines and 1,2,3-Thiadiazines. In *Comprehensive Heterocyclic Chemistry III* Katritzky;, A. R.; Ramsden;, C. A.; Scriven;, E. F. V.; Taylor;, R. J. K., Eds. Elsevier: Oxford, 2008; pp 291-299.
10. Ng, T. L.; McCallum, M. E.; Zheng, C. R.; Wang, J. X.; Wu, K. J. Y.; Balskus, E. P., The l-Alanosine Gene Cluster Encodes a Pathway for Diazeniumdiolate Biosynthesis. *ChemBioChem* **2020,** *21* (8), 1155-1160.
11. Du, Y. L.; He, H. Y.; Higgins, M. A.; Ryan, K. S., A heme-dependent enzyme forms the nitrogen-nitrogen bond in piperazate. *Nat. Chem. Biol.* **2017,** *13* (8), 836-838.
12. Abt, K.; Castelo-Branco, R.; Leao, P. N., Biosynthesis of Chlorinated Lactylates in *Sphaerospermopsis* sp. LEGE 00249. *J. Nat. Prod.* **2021,** *84* (2), 278-286.
13. Soding, J.; Biegert, A.; Lupas, A. N., The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005,** *33* (Web Server issue), W244-8.
14. 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.
15. Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T., A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **1996,** *3* (11), 923-36.
16. Zhang, D. Z.; Zhang, F.; Liu, W., A KAS-III Heterodimer in Lipstatin Biosynthesis Nondecarboxylatively Condenses C-8 and C-14 Fatty Acyl-CoA Substrates by a Variable Mechanism during the Establishment of a C-22 Aliphatic Skeleton. *J. Am. Chem. Soc.* **2019,** *141* (9), 3993-4001.
17. Brachmann, A. O.; Brameyer, S.; Kresovic, D.; Hitkova, I.; Kopp, Y.; Manske, C.; Schubert, K.; Bode, H. B.; Heermann, R., Pyrones as bacterial signaling molecules. *Nat. Chem. Biol.* **2013,** *9* (9), 573-U73.
18. Frias, J. A.; Richman, J. E.; Erickson, J. S.; Wackett, L. P., Purification and Characterization of OleA from *Xanthomonas campestris* and Demonstration of a Non-decarboxylative Claisen Condensation Reaction. *J. Biol. Chem.* **2011,** *286* (13), 10930-10938.
19. Fuchs, S. W.; Bozhuyuk, K. A.; Kresovic, D.; Grundmann, F.; Dill, V.; Brachmann, A. O.; Waterfield, N. R.; Bode, H. B., Formation of 1,3-cyclohexanediones and resorcinols catalyzed by a widely occurring ketosynthase. *Angew. Chem. Int. Ed. Engl.* **2013,** *52* (15), 4108-12.
20. Kresovic, D.; Schempp, F.; Cheikh-Ali, Z.; Bode, H. B., A novel and widespread class of ketosynthase is responsible for the head-to-head condensation of two acyl moieties in bacterial pyrone biosynthesis. *Beilstein J. Org. Chem.* **2015,** *11*, 1412-1417.
21. Zhu, T.; Scalvenzi, T.; Sassoon, N.; Lu, X. F.; Gugger, M., Terminal Olefin Profiles and Phylogenetic Analyses of Olefin Synthases of Diverse Cyanobacterial Species. *Applied and Environmental Microbiology* **2018,** *84* (13).
22. Mareš, J.; Hájek, J.; Urajová, P.; Kopecký, J.; Hrouzek, P., A Hybrid Non-Ribosomal Peptide/Polyketide Synthetase Containing Fatty-Acyl Ligase (FAAL) Synthesizes the beta-Amino Fatty Acid Lipopeptides Puwainaphycins in the Cyanobacterium *Cylindrospermum alatosporum*. *PLoS One* **2014,** *9* (11), e111904.
23. Sucipto, H.; Sahner, J. H.; Prusov, E.; Wenzel, S. C.; Hartmann, R. W.; Koehnke, J.; Muller, R., In vitro reconstitution of alpha-pyrone ring formation in myxopyronin biosynthesis. *Chem. Sci.* **2015,** *6* (8), 5076-5085.
24. Zocher, G.; Vilstrup, J.; Heine, D.; Hallab, A.; Goralski, E.; Hertweck, C.; Stahl, M.; Schaberle, T. F.; Stehle, T., Structural basis of head to head polyketide fusion by CorB. *Chem. Sci.* **2015,** *6* (11), 6525-6536.
25. Ashida, H.; Saito, Y.; Kojinia, C.; Yokota, A., Enzymatic characterization of 5-methylthioribulose-1-phosphate dehydratase of the methionine salvage pathway in *Bacillus subtilis*. *Bioscience Biotechnology and Biochemistry* **2008,** *72* (4), 959-967.
26. Ashida, H.; Saito, Y.; Kojima, C.; Kobayashi, K.; Ogasawara, N.; Yokota, A., A functional link between RuBisCO-like protein of Bacillus and photosynthetic RuBisCO. *Science* **2003,** *302* (5643), 286-290.
27. Sauter, M.; Moffatt, B.; Saechao, M. C.; Hell, R.; Wirtz, M., Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochem. J.* **2013,** *451*, 145-154.
28. Cui, Z.; Overbay, J.; Wang, X. C.; Liu, X. D.; Zhang, Y. A.; Bhardwaj, M.; Lemke, A.; Wiegmann, D.; Niro, G.; Thorson, J. S.; Ducho, C.; Van Lanen, S. G., Pyridoxal-5 '-phosphate-dependent alkyl transfer in nucleoside antibiotic biosynthesis. *Nat. Chem. Biol.* **2020,** *16* (8), 904-911.