

Biosynthesis of nocuolin A and related metabolites

Teresa Patrícia Pedrosa Martins

D
2023

Teresa Patrícia Pedrosa Martins. Biosynthesis of nocuolin A and
related metabolites



D.ICBAS 2023

**Biosynthesis of nocuolin A and related metabolites. Teresa Patrícia Pedrosa
Martins**



BIOSYNTHESIS OF NOCUOLIN A AND RELATED METABOLITES

Tese de Candidatura ao grau de Doutor em Ciências
Biomédicas submetida ao Instituto de Ciências
Biomédicas Abel Salazar da Universidade do Porto

Orientador

Doutor Pedro Nuno da Costa Leão

Investigador Principal no Centro Interdisciplinar de
Investigação Marinha e Ambiental

Co-orientador

Doutora Emily Patricia Balskus

Investigadora Principal do Instituto Médico Howard
Hughes e Professora Catedrática da Universidade de
Harvard

Co-orientador

Doutor Paulo Miguel Nunes Oliveira

Investigador no Centro Interdisciplinar de
Investigação Marinha e Ambiental e Professor Auxiliar
Convidado na Faculdade de Ciências da Universidade
do Porto

Declaração de Honra

Declaro que a presente tese é de minha autoria e não foi utilizada previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referência. Tenho consciência de que a prática de plágio e autoplágio constitui um ilícito académico.

Teresa Patrícia Pedrosa Martins

A rectangular box containing a handwritten signature in black ink. The signature is written in a cursive style and reads "Teresa Patrícia Pedrosa Martins".

Acknowledgments

First of all, I want to thank my supervisor, Dr. Pedro Leão for all the help, shared knowledge and most importantly, the extraordinary support and availability. Either in the laboratory bench, the results' analyses, or the writing of this thesis, his contribute and supervision were crucial to conclude this PhD.

Secondly, to my co-supervisor, Dr. Emily Balskus who was extremely supportive along these 4 years. Particularly, I would like to thank Emily for the opportunity to be part of her lab for 9 months which were incredibly productive and rewarding, despite the outbreak of the Covid pandemic. This experience was extremely positive to me, and I would gladly repeat it. I would also like to thank my co-supervisor Dr. Paulo Oliveira for being so understanding, positive and attentive to all small details. It is always a pleasure and an enriching experience to work with Paulo.

Furthermore, this work would not be possible without the outstanding contributions of Nathaniel Glasser which spent many hours teaching and helping me, since the very beginning of my stay at Harvard University.

I would like to thank all my colleagues from the Balskus lab, but specially my colleagues from the Leão lab who have become great friends throughout the years and make every day in the lab joyful and fun. Adriana, Raquel, Sara, Kat thank you for the friendship, long conversations and for all the support, João thank you for making me laugh, Sandra, Ana, Eunice, Anne, Amaranta thank you for the kind words and good humor, Lígia thank you for always being there and taking care of me, Bárbara and Diana for the kindness and finally Mariana and Mariane for the fun moments. Thank you for all the help, particularly from Marco, João and Kat, and all the good extra-work moments. I would also like to thank my old companions from BBE, Micaela, João, Tiago, Joana Almeida, Joana Azevedo and Flávio for always making me feel at home in BBE. I would also like to thank the Balskus lab members for receiving me with open arms, smiling faces and excellent professionalism.

I am incredibly grateful for all the support, sympathy and great work atmosphere shared in all labs, providing me with an amazing and unforgettable work experience.

Very special thanks to my parents for the support and to Edu for all the patience, for always being there, in the good and in the bad moments, for making me smile and simply make my life easier and happier. At last, to my old and incredible friends, Inês, João, Rúben, Joana, António, Catarina, Pedro, Luís, Sara C., Sara A., Cátia, Vitor and Flora, for making my life full of exciting experiences, good laughs, and simply amazing moments!

I would also like to thank the funding agencies that enabled this work, particularly Fundação para a Ciência e a Tecnologia for the PhD scholarship (SFRH/BD/138308/2018) and the institutions involved in this work: Interdisciplinary Centre of Marine and

Environmental Research (CIIMAR), Department of Chemistry and Chemical Biology from Harvard University and School of Medicine and Biomedical Sciences (ICBAS) from University of Porto. This work was developed under the European Research Council through a Starting Grant (759840), European Union's Horizon 2020 programme (Grant Agreement 952374) and Fundação para a Ciência e a Tecnologia through strategic funding grant UIDB/04423/2020 and UIDP/04423/2020.



Abstract

Cyanobacteria are a rich source of natural products (NPs) with intriguing structures and diverse bioactivities. Among these compounds is the cytotoxic nocuolin A, a molecule containing a 1,2,3-oxadiazine moiety, a structural feature unique among NPs, suggesting that novel enzymatic chemistry might be involved in its biosynthesis. So far, 1,2,3-oxadiazines have been inaccessible through organic synthesis. Nocuolin A is proposed to be produced by the *noc* biosynthetic gene cluster (BGC), but no biochemical evidence has been gathered to prove this connection. Furthermore, the *noc* BGC is co-localized with the *cly* BGC, involved in the biosynthesis of the chlorosphaerolactylates, which it turn can be esterified with nocuolin A, forming the hybrid nocuolactylates. In this work, we aim to study the biosynthesis of nocuolin A and related metabolites, in particular the unsaturated chlorosphaerolactylates and, consequently, the nocuolactylates.

In the first part of this work, we determine that the central C₁₃ chain of nocuolin A 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 an acyl carrier protein domain and undergo a non-decarboxylative Claisen condensation catalyzed by the ketosynthase (KS) NocG. This enzyme is part of a phylogenetically well-defined clade within similar genomic contexts, typifying a new class of KSs. Further supplementation experiments indicate that the 3-hydroxypropanoyl moiety of nocuolin A, originates from L-methionine, through an as-yet-uncharacterized mechanism. Here, we elucidate the origin of the carbon skeleton of nocuolin A and provide ample biochemical evidence connecting the putative *noc* biosynthetic gene cluster to nocuolin A.

In the second part, we report our efforts to elucidate the following steps on nocuolin A biosynthesis, including the incorporation of the first nitrogen and second oxygen atoms. Supplementation experiments and biochemical assays revealed that NocE, a flavin-dependent monooxygenase, hydroxylates 6-tridecanone, previously formed by NocG, through direct C-H bond activation. This hydroxylated substrate undergoes an unusual transamination reaction by a hybrid aminotransferase/kinase, NocF, which likely results in the formation of an imine intermediate. Both NocE and NocF seem to exhibit unique reactivities and cluster separately from the remaining protein family homologues. In this chapter, we further suggest that NocD is a methyltransferase involved in the methylation of nocuolin A to generate nocuolin B, the first reported nocuolin analogue.

In the last part of this work, we focus on the origin of the unusual β,γ unsaturation present in nocuolactylates and chlorosphaerolactylates E and F. Using *in vitro* enzymatic assays and *in vivo* supplementation experiments, we show that this β,γ unsaturation results

from the activity of a β,γ acyl-ACP dehydrogenase, NocC. The activity of this enzyme likely involves a multiple step reaction, which initiates by removal of the α proton to generate an α,β unsaturated intermediate and is followed by a shift of the unsaturation to the β,γ position. NocC shows promiscuity towards substrate acyl chain length (C_8 to C_{16}), branching and halogenation. This enzyme combines structural features of both medium-chain and very-long chain acyl-CoA dehydrogenases. With these findings, we have characterized the first acyl-ACP dehydrogenase known to generate a β,γ unsaturated product, which might be particularly interesting for future biocatalytic applications.

Key-words: natural products, cyanobacteria, nocuolin A, nocuolin B, chlorosphaerolactylates, nocuolactylates, biosynthesis, ketosynthase, fatty acyl-AMP ligase, aminotransferase, flavin-dependent monooxygenase, methyltransferase, acyl carrier protein, acyl-CoA dehydrogenase, β,γ acyl-ACP dehydrogenase, enzymatic assays, stable isotope-labelled supplementation, DiPaC

Resumo

As cianobactérias são fontes ricas de produtos naturais (PNs) com estruturas únicas e bioatividades diversas. Entre esses compostos está a nocuolina A, uma molécula citotóxica contendo um motivo 1,2,3-oxadiazina, único entre os PNs, sugerindo que nova química enzimática poderá estar envolvida na sua biossíntese. Até agora, as 1,2,3-oxadiazinas são ainda inacessíveis de obter por síntese orgânica. Foi proposto que a nocuolina A seja produzida pelo cluster de genes biossintéticos *noc* (CGB), mas evidências bioquímicas não foram ainda reunidas para provar essa ligação. Além disso, o CGB *noc* está co-localizado com o CGB *cly*, envolvido na biossíntese dos clorosferolactilatos, os quais, por sua vez, podem ser esterificados com a nocuolina A, formando os híbridos nocuolactilatos. Os principais objetivos deste trabalho foram o estudo da biossíntese da nocuolina A e compostos relacionados, particularmente os clorosferolactilatos insaturados e, consequentemente, os nocuolactilatos.

Na primeira parte deste trabalho, determinou-se que a cadeia C₁₃ central da nocuolina A é formada a partir de dois ácidos gordos de cadeia média, os ácidos hexanóico e octanóico. Usando ensaios bioquímicos, mostra-se que uma ácido gordo-AMP ligase, NocH, ativa ambos os ácidos gordos como acil adenilatos, e carrega-os num domínio de proteína transportadora de acil, sofrendo uma condensação de Claisen não descarboxilativa catalisada pela ketosintase (KS) NocG. Esta enzima faz parte de um clado filogeneticamente bem definido com contextos genómicos semelhantes, estabelecendo uma nova classe de KSs. Outras experiências de suplementação indicaram que o motivo 3-hidroxipropanoil da nocuolina A é proveniente da L-metionina, através de um mecanismo ainda não caracterizado. Aqui, elucida-se a origem do esqueleto de carbono da nocuolina A e é fornecida ampla evidência bioquímica para estabelecer a ligação entre o cluster de genes biossintéticos *noc* e a nocuolina A.

Na segunda parte, relatam-se os esforços para elucidar as etapas seguintes da biossíntese da nocuolina A, incluindo a incorporação do primeiro azoto e do segundo átomo de oxigénio. Experiências de suplementação e ensaios bioquímicos revelaram que a enzima NocE, uma monoxigenase dependente de flavina, hidroxila a 6-tridecanona, previamente formada pela NocG, através da ativação direta da ligação C-H. Posteriormente, este substrato hidroxilado sofre uma reação de transaminação incomum pela aminotransferase/cinase híbrida, NocF, que provavelmente resulta na formação de um intermediário imina. Ambas as enzimas NocE e NocF exibem reatividades únicas e agrupam-se separadamente dos restantes homólogos da família de proteínas. Ainda neste capítulo, sugere-se que a enzima NocD seja uma metiltransferase envolvida na metilação da nocuolina A, gerando nocuolina B, o primeiro análogo reportado da nocuolina.

A última parte deste trabalho foca-se na origem da insaturação incomum β,γ presente nos nocuolactilatos e clorosferolactilatos E e F. Usando ensaios enzimáticos *in vitro* e experiências de suplementação *in vivo*, mostra-se que esta insaturação β,γ resulta da atividade de uma β,γ -acil-ACP desidrogenase, NocC. A atividade desta enzima provavelmente envolve uma reação de múltiplos passos que se inicia pela remoção do próton α para gerar um intermediário α,β insaturado, sendo seguido pela deslocação da insaturação para as posição β,γ . NocC mostra promiscuidade em relação ao comprimento da cadeia acil (C_8 a C_{16}), à ramificação e halogenação dos substratos. Esta enzima combina características estruturais de acil-CoA desidrogenases de cadeia média e cadeia muito longa. Com estas descobertas, caracterizamos a primeira acil-ACP desidrogenase capaz de gerar um produto insaturado β,γ , o que pode ser particularmente interessante para futuras aplicações biocatalíticas.

Palavras-chave: produtos naturais, cianobactérias, nocuolina A, nocuolina B, clorosferolactilatos, nocuolactilatos, biossíntese, ketosintase, ácido gordo-AMP ligase, aminotransferase, monoxigenase dependente de flavina, metiltransferase, proteína transportadora de acil, acil-CoA desidrogenase, acil -ACP desidrogenase β,γ , ensaios enzimáticos, suplementação marcada com isótopos estáveis, DiPaC

Aims

In this work we aimed to study the biosynthesis of nocuolin A and related metabolites. Considering their unique structures, we hypothesized that highly novel biochemistry could be involved in their biosynthesis, and we aimed to characterize new biosynthetic steps that would expand the boundaries of known enzymatic reactivity.

To do so, this work was divided into 3 chapters, namely:

Chapter I: Biosynthesis of the unusual carbon skeleton of nocuolin A.

Chapter II: Diversification of the 6-tridecanone scaffold during nocuolin A biosynthesis

Chapter III: NocC is a β,γ acyl-ACP dehydrogenase involved in nocuolactylate biosynthesis

In the first two chapters we aimed to investigate the biosynthesis of nocuolin A, from the formation of its C₁₃ alkyl chain to its initial diversification, characterizing substrates, intermediates and enzymes involved in these steps. In the 3rd chapter, we aimed to study the biosynthetic origin of the β,γ unsaturation found in unsaturated chlorosphaerolactylates and nocuolactylates.

List of publications, presentations and awards within the scope of the PhD thesis

List of Publications within the scope of the PhD thesis

Martins, T. P.; Glasser N. R.; Kountz, D. J.; Oliveira, P. N.; Balskus, E. P.; Leão, P. N. (2022). Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A. ACS Chemical Biology 17(9): 2528-2537 <https://doi.org/10.1021/acscchembio.2c00464>.

Unpublished articles to be submitted

Martins, T. P.; Glasser, N. R.; Balskus, E. P.; Leão, P. N. " NocC is a β,γ acyl-ACP dehydrogenase involved in nocuolactylate biosynthesis". *In preparation*

Martins, T.P.; Braffman, N. B.; Preto, M.; Glasser N. R.; Oliveira, P. N.; Balskus, E. P.; Leão, P. N. "Diversification of the 6-tridecanone scaffold during nocuolin A biosynthesis". *In preparation*

List of Publications during PhD period, not included on the present thesis

Leão, P. N.; **Martins, T. P.;** et al. (2023). "Incorporation and modification of fatty acids in cyanobacterial natural products biosynthesis." Chem Commun (Camb) **59**(30): 4436-4446. <https://doi.org/10.1039/D3CC00136A>

Martins, T. P.; Arsin, S.; Fewer, D. P.; Leão, P. N. (2022). Chapter 5 - UV-protective secondary metabolites from cyanobacteria. The Pharmacological Potential of Cyanobacteria. G. Lopes, M. Silva and V. Vasconcelos, Academic Press: 107-144. <https://doi.org/10.1016/B978-0-12-821491-6.00005-3>.

Figueiredo, S. A. C.; Preto, M.; Moreira, G.; **Martins, T. P.;** Abt, K.; Melo, A.; Vasconcelos, V.M.; Leão, P.N. (2021). "Discovery of Cyanobacterial Natural Products Containing Fatty Acid Residues." Angewandte Chemie. International Ed. In English 60(18): 10064-10072. <https://doi.org/10.26434/chemrxiv.13100564.v2>

Martins, T. P. et al. (2019). "Chemistry, bioactivity and biosynthesis of cyanobacterial alkylresorcinols." Natural Product Reports 36(10): 1437-1461. <https://doi.org/10.1039/C8NP00080H>

Martins, T. P.; Ramos, V.; Hentschke, G. S.; Castelo-Branco, R.; Rego, A.; Monteiro, M.; Brito, A. et al. (2019). "The Extremophile *Endolithella mcmurdoensis* gen. et sp. nov. (Trebouxiophyceae, Chlorellaceae), A New Chlorella -like Endolithic Alga From Antarctica". *Journal of Phycology* 56 1: 208-216. <http://dx.doi.org/10.1111/jpy.12940>

Rego, A.; Raio, F.; **Martins, T. P.;** Ribeiro, H.; Sousa, A. G.G.; Séneca, J.; Baptista, M. S.; et al. (2019). "Actinobacteria and cyanobacteria diversity in terrestrial Antarctic microenvironments evaluated by culture-dependent and independent methods": <https://doi.org/10.3389/fmicb.2019.01018>

List of Presentations within the scope of the PhD thesis

Martins, T. P.; Glasser, N. R.; Oliveira, P. N.; Balskus, E. P.; Leão, P. N. (2023). "NocC: a promiscuous two step β,γ Acyl-ACP dehydrogenase". BIOPROSP_23, Tromsø, Norway. *Poster presentation*

Martins, T. P.; Glasser, N. R.; Kountz, D.; Oliveira, P.; Balskus, E.; Leão, P. N. (2022). Biosynthesis of the unusual carbon skeleton of nocuolin A. Directing Biosynthesis VI, Edinburgh, United Kingdom. *Oral presentation*

Martins, T. P.; Glasser, N. R.; Kountz, D.; Oliveira, P.; Balskus, E.; Leão, P. N. (2022). Biosynthesis of the unusual carbon skeleton of nocuolin A. XII 2022 Applied Molecular Microbiology Summer School, Dubrovnik, Croatia. *Poster presentation*

Martins, T. P.; Glasser, N. R.; Kountz, D.; Oliveira, P.; Balskus, E.; Leão, P. N. (2021). Biosynthesis of the unusual carbon skeleton of nocuolin A. XII European Conference on Marine Natural Products, Galway, Ireland (hold online). *Oral presentation*

Martins, T. P.; Glasser, N. R.; Kountz, D.; Oliveira, P.; Balskus, E.; Leão, P. N. (2021). Biosynthesis of the unusual carbon skeleton of nocuolin A. II Blue Think Conference, Matosinhos, Portugal (hold online). *Oral presentation*

List of awards within the scope of the PhD thesis

Meeting Attendance Grant (2022) by FEMS – Federation of European Microbiological Societies

Best oral communication award (2021) by XII European Conference on Marine Natural Products

Best oral communication award (2021) by II Blue Think Conference

Fulbright research fellowship, with the support of FCT, 2019/2020 (2019) by Fulbright Commission

Table of contents

Abstract	V
Resumo	vii
Aims	ix
List of publications, presentations and awards within the scope of the PhD thesis	x
List of Publications within the scope of the PhD thesis	x
Unpublished articles to be submitted	x
List of Publications during PhD period, not included on the present thesis	x
List of Presentations within the scope of the PhD thesis	xi
List of awards within the scope of the PhD thesis	xi
Table of contents	xiii
List of Figures	xxiii
List of Abbreviations	xxx
 General Introduction	 35
Introduction	36
1. Natural Products: Chemistry, Bioactivity and Biosynthesis	36
1.1. Natural Products: Chemistry and Bioactivity	36
1.2. Natural Products Biosynthesis	38
1.2.1. PKs/NRPs biosynthesis	39
1.2.2. Alkaloid biosynthesis	41
1.2.3. Terpene biosynthesis	42
1.2.4. RiPP biosynthesis	42
1.2.5. Pathways lacking a signature enzyme	44
1.3. Biosynthetic enzymes and biocatalysis	44
1.3.1. Biosynthetic enzymes	44
1.3.2. Biocatalysis	46
2. Cyanobacteria Natural Products: Chemistry, Bioactivity and Biosynthesis	48
2.1. Cyanobacteria	48
2.2. Cyanobacterial natural products	49

2.3. Cyanobacterial Biosynthetic enzymes	52
3. Nocuolin A and related metabolites: Chemistry, Bioactivity and Biosynthesis.....	53
3.1. Nocuolin A	53
3.1.1. Enzyme classes encoded in the <i>noc</i> BGC.....	56
3.1.1.1. Rieske oxygenases	56
3.1.1.2. Acyl-CoA dehydrogenases	57
3.1.1.3. Fatty acyl-AMP ligases	58
3.1.1.6. Flavin containing monooxygenases	60
3.1.1.8. Methylthioribulose-1-phosphate dehydratase, acireductone dioxygenase and the methionine salvage pathway enzyme	62
3.2. Nocuolin A related metabolites	63
3.2.1. Chlorosphaerolactylates	63
3.2.2. Nocuolactylates.....	65
4. Methodologies for biosynthetic studies in cyanobacteria	65
4.1. Bioinformatic tools	67
4.2. Precursor supplementation experiments.....	67
4.3. Gene/BGC inactivation (knockout).....	68
4.4. Heterologous expression of full and partial BGCs	68
4.5. <i>In vitro</i> enzymatic reconstitution	69
Chapter I	71
Biosynthesis of the unusual carbon skeleton of nocuolin A	72
Abstract	72
Introduction.....	73
Results and Discussion	75
Annotation of the <i>noc</i> gene products and Acyl-ACP Dehydrogenase activity of NocI and NocC.....	75
The C ₁₃ alkyl chain in 1 is derived from hexanoic and octanoic acids.....	76
NocH activates hexanoic and octanoic acids.....	78
NocG is a ketosynthase that generates the C ₁₃ alkyl moiety in 1	79
The 3-hydroxypropanoyl moiety of 1 is derived from L-methionine.....	84
Conclusions	86

Acknowledgments	86
Author contribution	87
References	87
Chapter I – Supplementary Information	90
Methodology	91
1. General experimental procedures.....	91
2. Plasmids, Strains and culture conditions.....	92
3. Supplementation experiments.....	92
4. LC-HRMS and LC-HRMS/MS analysis - Orbitrap.....	93
5. LC-MS and LC-MS/MS analysis - qTOF.....	93
6. Metabolomics analysis (nocuolin intermediates search).....	94
7. Cloning, expression and purification of C-His ₆ -NocH.....	94
8. Expression and purification of C-His ₆ -Sfp.....	96
9. Cloning, expression and purification of C-His ₆ -NocML.....	96
10. Cloning, expression and purification of C-His ₆ -NocI.....	98
11. Cloning, expression and purification of N-His ₆ -NocG.....	99
12. Cloning, expression and purification of N-His ₆ -NocG E100A.....	100
13. Cloning, expression and purification of C-His ₆ -NocA.....	101
14. Cloning and expression of N-His ₆ -MtnA and N-His ₆ -MtnK.....	102
15. Small-scale expression tests and western blots.....	103
16. Enzymatic assays.....	103
16.1. NocC enzymatic assay.....	103
16.2. NocI enzymatic assay	104
16.3. NocL and NocH competition assay	104
16.4. NocA enzymatic assay	104
16.5. NocG enzymatic assay.....	105
16.6. NocG E100A enzymatic assay	105
17. Determination of kinetic parameters for N-His ₆ -NocG.....	105
18. Phylogenetic analysis.....	106
19. Sequence similarity networks and Genomic Neighborhood Diagrams.....	106
20. NocG structural model.....	106

21. Chemical synthesis procedures and characterization data.....	107
21.1. Synthesis of hexanoyl-SNAC	107
21.2. Synthesis of octanoyl-SNAC	107
21.3. Synthesis of 2-octenoyl-SNAC	108
21.4. Synthesis of 5-methylthioribose.....	108
Chapter II	140
Diversification of the 6-tridecanone scaffold during nocuolin A biosynthesis	141
Abstract	141
Introduction.....	142
Results and Discussion	143
Nocuolin A biosynthesis requires CH activation by NocE.....	143
NocF introduces N-2 in the biosynthesis of nocuolin A.....	147
Further intermediates in the biosynthesis of 1	150
NocD is a putative methyltransferase involved in the biosynthesis of nocuolin B.....	153
Conclusion.....	155
Acknowledgments	156
Author contribution	156
References	157
 Chapter II – Supplementary Information	 160
Methodology	161
1. General experimental procedures.....	161
1.1. Chemistry	161
1.2. Molecular Biology	161
2. Plasmids, Strains and culture conditions.....	162
3. Supplementation experiments in <i>Nodularia</i> sp. LEGE 06071.....	163
4. Organic extraction of cyanobacteria.....	163
5. LC-HRMS and LC-HRMS/MS analysis – Orbitrap.....	163
6. LC-MS and LC-MS/MS analysis – qTOF.....	164
7. Metabolomic analysis (nocuolin intermediates search).....	164
8. Small-scale expression tests and western blots.....	164

9. Cloning, expression and purification of NocE.....	165
9.1. Cloning of <i>nocE</i> in pET-28a/pET-29b.....	165
9.2. Cloning of <i>nocE</i> in pET-His ₆ -SUMO-TEV and pET-51b.....	166
9.3. Purification of N-SUMO-NocE-His ₆	166
10. Cloning, expression and purification of C-His ₆ -NocJ.....	167
10.1. Cloning of <i>nocJ</i> in pET-28a/pET-29b.....	167
10.2. Cloning of <i>nocJ</i> in pET-51b-SUMO.....	168
10.3. Purification of N-Strep-SUMO-NocJ-His ₆	168
11. Cloning, expression and purification of NocF.....	168
11.1. Cloning of <i>nocF</i> in pET-28a/pET-29b.....	168
11.2. Cloning of <i>nocF</i> in pET-22b.....	169
11.3. Cloning of <i>nocF/nocG</i> in pET-29b.....	169
11.4. Cloning of <i>pelB-nocF</i> in pET-41a.....	169
11.5. Cloning of <i>pelB-nocF</i> in pET-51b, pMAL-2cx, pET-His ₆ -SUMO-TEV-LIC, pET-32a and pET-43.1a.....	170
11.6. Cloning of N-SUMO-pelB-NocF165-His ₆ and N-SUMO-pelB-NocF1-His ₆ in pET-51b.....	170
11.7. Expression tests.....	170
11.8. Protein Purification.....	171
11.8.1. N-Strep-SUMO-pelB-NocF165-His ₆	171
11.8.2. N-His ₆ -SUMO-pelB-NocF.....	171
12. Cloning, expression and purification of C-His ₆ -NocD(LEGE) and N-His ₆ -NocD(PCC)...	172
13. Cloning, expression and purification of Nock.....	173
13.1. Cloning of <i>nocK</i> in pET-28a/pET-29b.....	173
13.2. Cloning of <i>nocK</i> in pET-51b-SUMO.....	173
13.3. Purification of N-Strep-SUMO-Nock-His ₆	173
14. Enzymatic assays.....	173
14.1. NocF enzymatic assays.....	173
14.2. NocE enzymatic assay.....	174
14.3. NocJ enzymatic assay.....	174
14.4. Nock enzymatic assay.....	174

14.5. NocF enzymatic assay with lysate and clarified lysate.....	175
14.6. NocD enzymatic assay with lysate and clarified lysate	175
15. Sequence Similarity networks.....	175
16. Attempt to generate knockout mutants in <i>Nodularia</i> sp. LEGE 06071.....	176
16.1. Cloning.....	176
16.2. Sucrose sensibility test	176
16.3. Transformation in <i>Nodularia</i> sp. LEGE 06071	177
16.3.1. Triparental mating.....	177
16.3.2. Electroporation.....	177
17. Direct Pathway Cloning -DiPaC.....	178
17.1. Cloning.....	178
17.2. Heterologous expression and LC-HRMS analysis	179
17.3. Supplementation experiments in <i>E. coli</i>	179
17.4. Metabolomics analysis (nocuolin A intermediate search in <i>E. coli</i>).....	180
17.5. Heterologous expression of <i>noc</i> in large scale	180
17.6. Isolation of 5 and 6	180
17.6.1. Organic extraction	180
17.6.2. Flash chromatography.....	180
17.2.1. High Performance Liquid Chromatography (HPLC) – Compound 5	181
17.2.2. High Performance Liquid Chromatography (HPLC) – Compound 6	181
17.2.3. Structural elucidation - NMR analysis	181
17.2.4. Construction of pET-28:: <i>noc</i> mutants.....	181
18. Structural models.....	182
19. Organic synthesis.....	182
19.1. Synthesis of 2	182
19.2. Synthesis of 3	182
19.3. Synthesis of <i>d</i> ₁₂ - 2	183
19.4. Synthesis of <i>d</i> ₁₂ - 3	183
References	220
Chapter III	222
NocC is a β,γ acyl-ACP dehydrogenase involved in nocuolactylate biosynthesis	223

Abstract	223
Introduction.....	223
Results and Discussion	226
Identification of NocC as a putative β - γ acyl-ACP dehydrogenase.....	226
NocC is a promiscuous β , γ acyl-ACP dehydrogenase.....	227
NocC represents a new class of ACADs.....	230
Conclusion.....	232
Acknowledgments	232
Author contribution	233
References	233
Chapter III – Supplementary Information	236
Methodology.....	237
1. General experimental procedures.....	237
2. Plasmids, Strains and culture conditions.....	238
3. Supplementation experiments.....	238
4. LC-HRMS and LC-HRMS/MS analysis.....	238
5. Cloning, expression and purification.....	239
5.1. Cloning, expression and purification of C-His ₆ -ClyAB and C-His ₆ -NocI	239
5.2. Cloning, expression and purification of C-His ₆ -NocC	239
5.3. Cloning, expression and purification of C-His ₆ -NocC E373A	240
6. Enzymatic assays.....	241
7. Phylogenetic analysis.....	242
8. Sequence similarity networks and Genomic Neighborhood Diagrams.....	242
9. Chemical synthesis procedures and characterization data.....	243
9.1. Synthesis of dodecanoyl-SNAC	243
9.2. Synthesis of d ₂₃ -dodecanoyl-SNAC.....	243
10.NocC structural model.....	244
References	262
General Discussion and Conclusions.....	263
General Discussion and Conclusions.....	264

General References	269
Appendix - Chapter I	289
Copyrights ACS ChemBio	290

List of Tables

General Introduction

Table A.1 - Relevant examples of FDA-approved NP or NP derivatives drugs	36
Table A.2 - Relevant examples of cyanobacterial bioactive NPs.....	50
Table A.3 - Relevant examples of fatty-acid containing cyanobacterial NPs.....	50

Chapter I - Biosynthesis of the unusual carbon skeleton of nocuolin A

Table I.1 – Comparison between NocG and its closest homologues, PpyS and OleA.	83
--	----

Chapter I - Supplementary Information

Table SI.1 - Detailed annotation of the <i>noc</i> gene cluster in <i>Nodularia</i> sp. LEGE 06071.....	110
Table SI.2 - Expression conditions used for all proteins purified in this work	111
Table SI.3 – Primers used in this work	111
Table SI.4 – Accession numbers of all sequences used in the KSs phylogentic tree	112

Chapter II - Supplementary Information

Table SII.1 - Detailed annotation of the <i>noc</i> gene cluster in <i>Nodularia</i> sp. LEGE 06071. Annotation based on HHPred and Swiss-Model.....	184
Table SII.2 - NMR data of compound 6 in CD ₃ Cl (600 MHz)	185
Table SII.3 – Primers used to clone NocE in the several expression vectors, using restriction cloning or Gibson assembly	185
Table SII.4 – Primers used to clone NocJ in the several expression vectors, using restriction cloning or Gibson assembly	186
Table SII.5 – Primers used to clone NocF in the several expression vectors, using restriction cloning or Gibson assembly	187
Table SII.6 – Expression plasmids used for NocF cloning, their antibiotic resistance and expected recombinant protein and size	188
Table SII.7 – Primers used to clone NocD from LEGE 06071 and PCC 7108 in the several expression vectors, using restriction cloning	188
Table SII.8 – Primers used to clone NocK in the several expression vectors, using restriction cloning or Gibson assembly	189
Table SII.9 – Parameters used for the construction of NocE, NocF, NocJ and NocK SSNs	189
Table SII.10 - Primers used to clone NocE and NocF flanking regions for knockout construction	189

Table SII.11 - List of Plasmids used for the knockout creation and its antibiotic resistances	190
Table SII.12 - Primers used for the DiPaC strategy	190

Chapter III - Supplementary Information

Table SIII.1 - <i>noc</i> and <i>cly</i> annotation from LEGE 06071 and LEGE 00249	244
Table SIII.2 – Molecular formula and calculated mass of nocuolin A, chlorosphaerolactylates and nocuolactylates ions mentioned in this work	245
Table SIII.3 - Expression conditions used for all proteins purified in this work	245
Table SIII.4 – Primers used in this work.	246
Table SIII.5 - BlastP results of NocC against nr database of NCBI (20/12/2022)	247
Table SIII.6 - Accession numbers of all sequences used in the ACADs phylogenetic tree.....	248

List of Figures

General Introduction

Figure A.1 - Chemical structures of relevant FDA-approved NP drugs.....	37
Figure A.2 - Chemical structures of limonene, ramoplanin and maitotoxin.	38
Figure A.3 – Schematic representation of PKS, NRPS and alkaloid biosynthetic pathways.....	40
Figure A.4 – Schematic representation of terpenes, RiPPS and no signature enzyme pathways.	43
Figure A.5 – Example of remarkable biosynthetic enzymes and their unusual C-H activation, C-C formation and N-N formation reactions	45
Figure A.6 – Biocatalysts application in industrial context.....	47
Figure A.7 – Cyanobacteria morphologic diversity in LEGE culture collection (LEGEcc).	48
Figure A.8 - Chemical structures of cyanobacterial NPs including terpenes, PKS-NRPs, alkaloid, PKs, NRPs and fatty-acid rich compounds.....	51
Figure A.9 – Examples of remarkable cyanobacterial biosynthetic enzymes and their unusual C-H activation, C-C formation.	52
Figure A.10 - Nocuolin A is a 1,2,3-oxadiazine cyanobacterial NP encoded by the putative <i>noc</i> BGC.....	55
Figure A.11 – Rieske oxygenases with diverse reactivities: AmbP, VioQ, MarG and PrnD perform C-C desaturation, C-hydroxylation, carbocyclization and N-oxidation, respectively.	56
Figure A.12 – Acyl-CoA dehydrogenase enzymatic activity.....	57
Figure A.13 – Fatty acyl-AMP ligases enzymatic activity.	58
Figure A.14 – Ketosynthases enzymatic activity.....	59
Figure A.15 – Aminotransferase enzymatic activity.	60
Figure A.16 – Flavin containing monooxygenases enzymatic activity.....	61
Figure A.17 – Methyltransferase enzymatic activity.....	62
Figure A.18 – Schematic representation of the "universal" MTA salvage pathway	63
Figure A.20 – Nocuolactylates A-C are hybrid molecules formed by nocuolin A and chlorosphaerolactylates in <i>Nodularia</i> sp. LEGE 06071.....	66
Figure A.21 - Overview of the most common techniques used in cyanobacterial biosynthetic studies	70

Chapter I - Biosynthesis of the unusual carbon skeleton of nocuolin A

Figure I.1 - Schematic highlighting the structure of nocuolin A, the <i>noc</i> BGC and the proposed biosynthetic steps.	74
Figure I.2 - Supplementation of <i>Nodularia</i> sp. LEGE06071 with <i>d</i> ₁₁ -hexanoic acid and <i>d</i> ₁₅ -octanoic acid leads to <i>d</i> ₁₁ -1/ <i>d</i> ₂₂ -1 and <i>d</i> ₁₃ -1, respectively.	77
Figure I.3 - FAALs Noch and ClyA (NocM) activate hexanoic/octanoic acids and dodecanoic acid as acyl-ACP thioesters, respectively, in competition assays.	78

Figure I.4 - NocG belongs to a new class of KSs.	82
Figure I.5 - 3-Hydroxypropanoyl moiety of 1 is derived from C-2 to C-4 and N-1 of L-methionine.....	85

Chapter I - Supplementary Information

Figure SI.1 – SDS-Page of NocA, NocB, NocC, NocG, NocI, NocML, MtnA and MtnK purifications.	113
Figure SI.2 - Extracted ion chromatograms of the expected enoyl-ACP fragments predicted for NocM previously loaded with hexanoyl-, octanoyl-, decanoyl- and dodecanoyl-CoA.....	114
Figure SI.3 - Extracted ion chromatograms of the expected enoyl-ACP fragments predicted for NocM previously loaded with hexanoyl-, octanoyl-, decanoyl- and dodecanoyl-CoA.....	115
Figure SI.4 - Extracted ion chromatograms of saturated and unsaturated versions of chlorosphaerolactylates A and B in <i>Nodularia</i> LEGE sp. 06071.	116
Figure SI.5 – MS/MS fragmentation of nocuolin A, highlighting key fragments that include position C-5 .	117
Figure SI.6 – HRMS/MS spectrum of <i>d</i> ₁₁ - 1 (<i>m/z</i> 332.2839) highlighting the diagnostic fragmentations that support the incorporation of <i>d</i> ₁₁ -hexanoic acid in the C ₁₃ carbon skeleton of nocuolin A.	118
Figure SI.7 – LC-EICs of 1 and labeled version upon supplementation of <i>Nodularia</i> sp. LEGE06071 with <i>d</i> ₁₃ -heptanoic acid, <i>d</i> ₁₉ -decanoic acid <i>d</i> ₂₃ -dodecanoic acid and <i>d</i> ₂₇ -tetradecanoic acids.....	118
Figure SI.8 - HRMS/MS spectrum of 1 ([<i>M</i> + <i>H</i>] ⁺ and [<i>M</i> + <i>Na</i>] ⁺ isotope clusters are shown) in supplementation experiments with 1- ¹³ C-FAs.	119
Figure SI.9 – HRMS/MS analysis of the <i>M</i> , <i>M</i> +1 and <i>M</i> +2 isotope peaks for the [<i>M</i> + <i>H</i>] ⁺ species of 1 after supplementation with 1- ¹³ C-labeled hexanoic and octanoic acids.	120
Figure SI.10 – HRMS analysis of chlorophyll a and pheophytin a in the crude extracts of <i>Nodularia</i> sp. LEGE 06071	121
Figure SI.11 - LC-HRMS ² spectra of the NocH fragment loaded with hexanoic acid highlighting the diagnostic C6:0-Ppant fragment.....	121
Figure SI.12 – LC-HRMS analysis of the fatty-acyl AMP ligases – NocH and NocL – competition assay with hexanoic, octanoic, decanoic, dodecanoic and tetradecanoic acids.....	122
Figure SI.13 – <i>E. coli</i> codon-optimized sequence (5' to 3') of the gene encoding NocG (NocG synthetic gene).....	123
Figure SI.14 - HRMS/MS spectrum of 4 [<i>M</i> - <i>H</i>] ⁻ (collision energy 30 eV), supporting its identification as 2-hexyl-3-oxooctanoic acid.	123
Figure SI.15 – LC-HRMS analysis of NocG and NocH coupled enzymatic assays with different combinations of ² H- and ¹³ C-labeled FAs	124
Figure SI.16 – LC-HRMS analysis of NocG enzymatic assays to test the potential role of NocI in the C-C bond forming reaction.....	125
Figure SI.17 – LC-HRMS-derived analysis of NocG enzymatic assays with free fatty-acids – hexanoic and octanoic acids – and/or activated thioesters – hexanoyl-SNAC, octanoyl-SNAC, hexanoyl-NocH, octanoyl-NocH, hexanoyl-CoA and octanoyl-CoA.....	125
Figure SI.18 – 3-hydroxyoctanoic acid is not a NocG and NocH substrate.	126

Figure SI.19 - Michaelis-Menten kinetic data for N-His ₆ -NocG.....	126
Figure SI.20 – NocG performs a tight control of the β-ketoacid chain length.	127
Figure SI.21 – NocG and PpyS form distinct clusters in sequence similarity network (SSN) analysis	128
Figure SI.22 - Genome Neighborhood Network (GNN) hub-and-spoke clusters depicting the Pfam domains associated with NocG and PpyS clusters from the initial SSN.	129
Figure SI.23 - Diversity and genomic context of NocG-like enzymes BGCs. Genomic Neighborhood Diagrams (GNDs) showing examples of the different BGC architectures found among the SSN cluster encoding NocG homologues.....	130
Figure SI.24 - Protein alignment (Geneious alignment - Global alignment with free ends and gaps) of NocG, OleA (CAE17216) and OleA homologues.....	131
Figure SI.25 – Protein alignment (Geneious alignment - Global alignment with free ends and gaps) of NocG, PpyS (AGO97060) and PyrS (WP_007967127).	131
Figure SI.26 – Structural model of NocG (LEGE06071) and comparison with PpyS model and OleA (crystal structure).	132
Figure SI.27 – Proposed mechanism for NocG activity.....	133
Figure SI.28 – NocG E100A mutation results in loss of detectable product formation.	133
Figure SI.29 – Protein alignment (Geneious alignment – Global alignment with free ends and gaps) of NocA and characterized MtnB enzymes.	133
Figure SI.30 – Protein alignment (Geneious alignment – Global alignment with free ends and gaps) of NocB and characterized MtnD enzymes.	134
Figure SI.31 – LC-HRMS-derived analysis of NocA coupled enzymatic assay with MtnK and MtnA from <i>B. subtilis</i>	134
Figure SI.32 – HRMS/MS data of peak 303.2459 (C ₁₂ ¹³ C ₄ H ₃₁ O ₃ N ¹⁵ N) in the ¹⁵ N, ¹³ C ₅ -L-methionine feeding experiment from <i>Nodularia</i> sp. LEGE 06071.....	134
Figure SI.33 - The 3-hydroxypropanoyl moiety of 1 is derived from carbons C-2, C-3 and C-4 from L-methionine.	135
Figure SI.34 - HRMS spectrum of synthetic compounds hexanoyl-SNAC (a) and octanoyl-SNAC (b)	135
Figure SI.35 - HRMS spectrum of synthetic compounds of 2-octenoyl-SNAC (a) and MTR (b).....	136
Figure SI.36 - ¹ H NMR (400 MHz) of hexanoyl-SNAC in CD ₃ Cl.....	136
Figure SI.37 - ¹ H NMR (400 MHz) of octanoyl-SNAC in CD ₃ Cl.....	137
Figure SI.38 - ¹ H NMR (400 MHz) of 2-octenoyl-SNAC in CD ₃ Cl.	137
Figure SI.39 - ¹ H NMR (400 MHz) of MTR in DMSO.	138

Chapter II - Diversification of the 6-tridecanone scaffold during nocuolin A biosynthesis

Figure II.1 - Schematic highlighting the Nocuolin A structure, the <i>noc</i> BGC and the biosynthetic steps previously characterized to form nocuolin A C ₁₃ alkyl chain.....	143
--	-----

Figure II.2 –Supplementation of <i>Nodularia</i> sp. LEGE 06071 with <i>d</i> ₁₂ - 2 , <i>d</i> ₁₂ - 3 leads to the formation of <i>d</i> ₁₁ - 1 and <i>d</i> ₁₂ - 1 .	145
Figure II.3 – NocF is predicted to be an unusual aminotransferase	148
Figure II.4 – Expression of the <i>noc</i> BGC in <i>E. coli</i> yield the putative intermediates 3-hydroxy-N-(tridecan-6-yl)propenamide (6) and putative 6-nitrosotridecane (5).	151
Figure II.5 – Nocuolin B is the methylated analogue of 1 , likely resulting from NocD activity	154
Figure II.6 – Proposed biosynthesis of 1 and 7	156

Chapter II - Supplementary Information

Figure SII.1 - Putative intermediates found in <i>Nodularia</i> sp. LEGE06071 supplementation experiments.	191
Figure SII.2 - SDS-Page of N-His ₆ -SUMO-NocE (a), N-Strep-SUMO-NocJ-His ₆ (b), N-Strep-SUMO-pelB-NocF165-His ₆ (c), N-His ₆ -SUMO-pelB-NocF (d) and N-Strep-SUMO-NocK-His ₆ (e) purifications.	192
Figure SII.3 - Western blots of NocE expression tests using pET-29b and pET-28a, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation	193
Figure SII.4 - Western blots of NocE expression tests using pET-29b, pET-51b, pET-His ₆ -SUMO-TEV, terrific broth (TB), 0.5 mM IPTG, 25 °C overnight	194
Figure SII.5 - Western blots of NocJ expression tests using pET-29b and pET-28a, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation	195
Figure SII.6 - Western blots of NocJ expression tests using pET-51b-SUMO, terrific broth (TB), 0.5 mM IPTG, 25 °C or 15 °C and 1 and 4 days incubation	196
Figure SII.7 - HR-MS/MS spectrum of 2 (in blue) from NocE enzymatic assay with 6-tridecanone.	196
Figure SII.8 - LC-HRMS-derived EICs of 2 (a) and 3 (b) in NocJ enzymatic assays with 6-tridecanone. NocJ was not able to generate 2 or 3 .	197
Figure SII.9 - LC-HRMS-derived EICs of 3 in NocE enzymatic assays with 6-tridecanone. NocE was not able to generate 3 .	197
Figure SII.10 – Predicted domains of NocF and cell location.	197
Figure SII.11 – Swiss model (a) and InterPro (b) prediction of NocF location in the cell.	198
Figure SII.12 – NocF sequence after codon-optimization for <i>E. coli</i> .	198
Figure SII.13 – Western blots of NocF (codon optimized) expression tests using pET-29b, pET-28a and pET-22b, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation	199
Figure SII.14 – Western blots of NocF (codon optimized) expression tests using pET-41a, pET-29b, pET-51b or pET-SUMO, terrific broth (TB), 0.5 mM IPTG at 25 °C and overnight incubation	200
Figure SII.15 – Western blots of NocF (codon optimized) expression tests using pET-His ₆ -SUMO-TEV, terrific broth (TB), 0.5 mM IPTG at 12 °C and several days of incubation	201

Figure SII.16 - Western blots of NocF expression tests using pET-32a, pET-43.1a, and N-His6-SUMO-pelB-NocF using pET-51b in terrific broth (TB), 0.5 mM IPTG at 25 °C for 1 day or 12 °C or 5 days at 12 °C	202
Figure SII.17 - NocF purification attempts.	202
Figure SII.18 - LC-HRMS-derived EICs of the putative amine and imine products generated by NocF (except for 8-iminotrican-6-ol – Fig 3 of the main manuscript) in enzymatic assays with 2	203
Figure SII.19 – Spectrum of m/z feature 226.2919 (<i>d</i> ₁₂ - 4) detected upon NocF assay with <i>d</i> ₁₂ - 2	203
Figure SII.20 - LC-HRMS-derived EICs of the putative amine and imine product generated by NocF in enzymatic assays with 6-tridecanone (a) and 3 (b).	204
Figure SII.21 - LC-HRMS-derived EICs of 4 (a) and <i>d</i> ₁₂ - 4 (b) generated by enzymatic assays of N-His6-SUMO-pelB-NocF at room temperature and 37°C overnight, with 2 and 3 , respectively.	205
Figure SII.22 - LC-HRMS-derived EICs of 4 generated by <i>in vivo</i> assays of C43 harboring pET-SUMO::pelB-NocF supplemented with 2	205
Figure SII.23 - Sequence similarity network (SSN) of NocF and the 1000 closest homologues	206
Figure SII.24 – BG11 neomycin plates after 1 month of conjugation with pRL278::nocE showed no cyanobacterial colonies.	206
Figure SII.25 - LC-HRMS-derived EICs of 1 in several growth conditions of BAP1 pET-28::noc transformants.	207
Figure SII.26 – Putative intermediates of the <i>noc</i> biosynthetic pathway identified in BAP1 <i>E. coli</i> harboring pET28::noc.	208
Figure SII.27 - HR-MS/MS spectrum of 5 (a) and 6 (b) under upon <i>d</i> ₁₁ -hexanoic acid and <i>d</i> ₈ -L-methionine supplementation.	208
Figure SII.28 - LC-HRMS-derived EICs of m/z 214.2165 (4 and 5) in enzymatic assays of NocF with 2 and in the crude extracts of BAP1 pET-28::noc.	209
Figure SII.29 - ¹ H NMR spectrum (600 MHz) of 6 in CD ₃ Cl.	209
Figure SII.30 – ¹³ C APT NMR spectrum (600 MHz) of 6 in CD ₃ Cl.	210
Figure SII.31 - HSQC spectrum (600 MHz) of 6 in CD ₃ Cl.	210
Figure SII.32 - COSY NMR spectrum (600 MHz) of 6 in CD ₃ Cl.	211
Figure SII.33 - HSQC spectrum (600 MHz) of 6 in CD ₃ Cl.	211
Figure SII.34 - LC-HRMS-derived EICs of 6 in NocF enzymatic assays incubated with 2 and 3-hydroxypropanamide as putative substrate	212
Figure SII.35 – Nocuolin A and B are produced by <i>Anabaena aphanizomenioides</i> LEGE 00250 and <i>Nodularia</i> sp. LEGE 04288.	212
Figure SII.36 – Coomassie blue gels of the purification attempts of N-His6-NocD(PCC) (a) and C-His6-NocD(LEGE) (b) from BL21(DE3) cells.	213
Figure SII.37 - LC-HRMS-derived EICs of 7 in <i>in vitro</i> assays of 1 with NocD lysates and clarified lysates.	213
Figure SII.38 - LC-HRMS-derived EICs of putative products of NocJ oxygenation of 4	214
Figure SII.39 - Sequence similarity network (SSN) of NocJ and the 1000 closest homologues.	214

Figure SII.40 - Sequence similarity network (SSN) of Nock, the 1000 closest homologues and the IPR021829 InterPro family	215
Figure SII.41 - Western blots of Nock expression tests using pET-28a and pET-51b-SUMO, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation	216
Figure SII.42 - LC-HRMS-derived EICs of putative products of Nock products generated from 4 and 6-tridecanone with 3-hydroxypropanamide.	217
Figure SII.43 - HRMS-derived EICs of the synthetic compounds 2 (a) and 3 (b), confirming the expected masses.	217
Figure SII.44 - HR-MS spectrum of synthetic compounds of <i>d</i> ₁₂ - 2 (a) and <i>d</i> ₁₂ - 3 (b) confirming the expected masses.	218
Figure SII.45 - ¹ H NMR (400 MHz) of 2 in CD ₃ Cl.	218
Figure SII.46 - ¹ H NMR (400 MHz) of 3 in CD ₃ Cl.	219
Figure SII.47 - ¹ H NMR (400 MHz) of <i>d</i> ₁₂ - 2 in CD ₃ Cl.....	219
Figure SII.48 - ¹ H NMR (400 MHz) of <i>d</i> ₁₂ - 3 in CD ₃ Cl	220

Chapter III - NocC is a β,γ acyl-ACP dehydrogenase involved in nocuolactylate biosynthesis

Figure III.1 – Acyl-CoA dehydrogenases and their unsaturated natural products.	225
Figure III.2 – <i>nocC</i> is present in cyanobacteria producing unsaturated chlorosphaerolactylates like <i>Nodularia</i> and absent in cyanobacteria producing fully saturated chlorosphaerolactylates	226
Figure III.3 – NocC is a β,γ acyl-ACP dehydrogenase	228
Figure III.4 – NocC represents the first characterized member of a unique group of ACADs.	231

Chapter III - Supplementary Information

Figure SIII.1 - EICs of nocuolin A, nocuolactylates A, B, C and putative saturated versions in the crude extract of <i>Sphaerospermopsis</i> LEGE 00249.	249
Figure SIII.2 - Schematic representation of the <i>noc</i> and <i>cly</i> BGCs from five cyanobacterial genomes. NocC (in orange) is the only mobile gene of the cluster, being positioned either on the <i>noc</i> or <i>cly</i>	249
Figure SIII.3 – SDS-Page of C-His ₆ -NocI (A), C-His ₆ -NocC (B), C-His ₆ -NocC E373A (C), C-His ₆ -ClyAB (D) purifications. All proteins were purified by Ni-IDA packed columns (Macherey Nagel).	250
Figure SIII.4 – FAD binding of NocC causes the yellow coloration of the enzyme	250
Figure SIII.5 –NocC enzymatic assay peaks compared with <i>trans</i> -2-octenoyl-ClyB and <i>trans</i> 3-octenoyl-ClyB generated by incubation of the <i>trans</i> -2-octenoic and <i>trans</i> -3-octenoic acids with ClyAB (zoom in)..	251
Figure SIII.6 – HR-LCMS ² of NocC generated C8-ClyB thioester (<i>m/z</i> 1307.6474) and comparison with the <i>trans</i> -2-octenoyl-ClyB and <i>trans</i> -3-octenoyl-ClyB.	251

Figure SIII.7 – NocC and NocI enzymatic assays with C12, 2,2- <i>d</i> ₂ -C12 and <i>d</i> ₂₃ -C12. EICs and spectra for C12 assay (A, D), <i>d</i> ₂₃ -C12 (B, E) and 2,2- <i>d</i> ₂ -C12 (C, F).....	252
Figure SIII.8 – NocC substrate scope. EICs of the NocC assay with C6 (A), <i>d</i> ₁₁ -C6 (B), <i>d</i> ₁₃ -C7 (C), C10 (D), <i>d</i> ₁₉ -C10 (E), C12 (F), <i>d</i> ₂₃ -C12 (G), 2,2- <i>d</i> ₂ -C12 (H), C14 (I), <i>d</i> ₂₇ -C14 (J), C16 (K), <i>d</i> ₃₁ -C16 (L), C18 (M), <i>d</i> ₃₅ -C18 (N), 3-methyl-C12 (O), 2-methyl-C12 (P), 6-heptynoic acid (Q), 5-hexynoic acid (R), 6-bromo-C6 (S), 6-bromo-C7 (T), 2-hydroxy-C8 (U).	253
Figure SIII.9 – NocC substrate scope. MS Spectra of NocC assay with C6 (A), <i>d</i> ₁₁ -C6 (B), <i>d</i> ₁₃ -C7 (C), C10 (D), <i>d</i> ₁₉ -C10 (E), C12 (F), <i>d</i> ₂₃ -C12 (G), 2,2- <i>d</i> ₂ -C12 (H), C14 (I), <i>d</i> ₂₇ -C14 (J), C16 (K), <i>d</i> ₃₁ -C16 (L), C18 (M), <i>d</i> ₃₅ -C18 (N), 3-methyl-C12 (O), 2-methyl-C12 (P), 6-heptynoic acid (Q), 5-hexynoic acid (R), 6-bromo-C6 (S), 6-bromo-C7 (T), 2-hydroxy-C8 (U).	255
Figure SIII.10 – NocC enzymatic assays with C12-SNAC (A) and <i>d</i> ₂₃ -C12-SNACs (B).....	256
Figure SIII.11 – <i>Nodularia</i> sp. LEGE 06071 supplementation experiments with <i>d</i> ₂₃ -C12, highlighting the labeled peaks of nocuolactylate A with 19 deuterons (A), nocuolactylate B with 18 deuterons (B) and nocuolactylate C with 20 deuterons (C).....	256
Figure SIII.12 – Sequence similarity network of NocC with the top 5000 BLAST hits	257
Figure SIII.13 - Diversity and genomic context of NocC-like enzymes (in red) BGCs.	258
Figure SIII.14 - ClustalΩ alignment of NocC from <i>Nodularia</i> and HBU26 with characterized SCADs, MCADs, FadE, FadE5, VLCAD and the γ,σ ACADs TcsD and HliR.....	259
Figure SIII.15 – EICs of NocC and NocC E373A enzymatic assays with octanoic acid. The presence of loaded ClyAB-C8 in the assay lacking C8 is caused by <i>in vivo</i> loading of the substrate since ClyAB was expressed in BAP1 <i>E. coli</i> which introduce the PPant arm and consequently can load FA directly <i>in vivo</i>	260
Figure SIII.16 – Structure comparison of TcsD (magenta), VLCAD 2uwx (light pink) and NocC (blue) (A) and detailed active site highlighting the catalytic glutamate (B).	260
Figure SIII.17 – MS spectrum of the synthesized C12-SNAC (A) and <i>d</i> ₂₃ -C12-SNAC (B).	261
Figure SIII.18 - 400 MHz ¹ H spectrum of C12-SNAC (CDCl ₃).	261
Figure SIII.19 - 400 MHz ¹ H spectrum of <i>d</i> ₂₃ -C12-SNAC (CDCl ₃).	262

General Discussion and Conclusions

Figure B.1. - Partial biosynthesis of nocuolin A and nocuolin B based on the work of chapters 1 and 2 .	266
Figure B.2 - Proposed biosynthetic pathway of the β,γ unsaturated chlorosphaerolactylate F in <i>Nodularia</i> sp. LEGE 06071, highlighting NocC and the β,γ unsaturation introduced by its activity	268

List of Abbreviations

A	Adenylation domain
ACAD	Acyl-CoA dehydrogenase
ACN	Acetonitrile
ACP	Acyl carrier protein
Amp	Ampicillin
AMP	Adenosine monophosphate
AmT	Aminotransferase
AO	Aspartate oxidase
ARG	Arginase
AT	Acyltransferase
ATA	Amine aminotransferase
ATP	Adenosine triphosphate
ATT	Amino acid aminotransferase
BBL	Berberine-bridge enzyme
BGC	Biosynthetic gene cluster
BSA	Bovin Serum Albumine
BVMO	Baeyer-Villiger Monooxygenases
C	Condensation domain
C10	Decanoic acid
C12	Dodecanoic acid
C14	Tetradecanoic acid
C16	Hexadecanoic acid
C18	Octadecanoic acid
C6	Hexanoic acid
C7	Heptanoic acid
C8	Octanoic acid
CCAP	Culture Collection of Algae and Protozoa
CEMUP	Centro de Estudo de Materiais da Universidade do Porto
CIIMAR	Interdisciplinary Centre of Marine and Environmental Research
Cm	Chloramphenicol
CN/DAB	Chloronaphthol/diaminobenzidine
CoA	Coenzyme A
COSY	^1H - ^1H Correlation Spectroscopy
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats

Cy	Cyclization domain
DAO	Diamine oxidase
DCM	Dichloromethane
DH	Dehydratase domain
DHK-MTPene	1,2-dihydroxy-3-keto-5-methylthiopent(1)ene
DiPaC	Direct Pathway Cloning
DMAPP	Dimethylallyl diphosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
DXMT	3,7-dimethylxanthine 1- <i>N</i> -methyltransferase
E	Epimerization domain
EC	Energy of collision
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EFI	Enzyme Function Initiative
EIC	Extracted ion chromatogram
ER	Enoyl reductase
EtOAc	Ethyl acetate
FA	Fatty-acid
FAAL	Fatty acyl-AMP ligase
FAD	Flavin adenine dinucleotide
FDA	Food and Drug Administration
FMO	Flavin-containing monooxygenases
GCAD	Glutaryl-CoA dehydrogenase
gDNA	genomic deoxyribonucleic acid
GNN	Genome neighborhood network
GPP	Geranyl pyrophosphate
GPPTS	Geranyl pyrophosphate synthase
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human Immunodeficiency Virus
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry

HSQC	Heteronuclear Single Quantum Coherence
ICBAS	School of Medicine and Biomedical Sciences
IBD	Isobutyryl-CoA dehydrogenase
IDA	Iminodiacetic acid
IPA	Isopropanol
iPD	Trans-isopiperitenol dehydrogenase
iPI	Cis-isopulegone isomerase
IPP	Isopentenyl pyrophosphate
iPR	Trans-isopiperitenone reductase
IPTG	Isopropyl- β -D-thiogalactopyranoside
IVD	Isovaleryl-CoA dehydrogenase
Km	Kanamycin
KR	Ketoreductase
KS	Ketosynthase
L3OH	limonene 3-hydroxylase
LB	Luria-Bertani Lennox medium
LCAD	Long-chain acyl-CoA dehydrogenase
LC-HRMS	Liquid chromatography-high-resolution mass spectrometry
LDA	Lithium diisopropylamide
LEGEcc	Blue Biotechnology and Ecotoxicology Culture Collection
MBP	Maltose-binding protein
MCAD	Medium-chain acyl-CoA dehydrogenase
MCADD	Medium-chain ACAD disease
MeOH	Methanol
MR	Menthone reductase
MS	Mass spectrometry
MT	Methyltransferase
MTA	Methylthioadenosine
MTR	Methylthioribose
MTR-1-P	5'-methylthioribose-1-phosphate
MTRu-1-P	5'-methylthioribulose 1-phosphate
MXMT	7- <i>N</i> -methylxanthine 3- <i>N</i> -methyltransferase
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
Nm	Neomycin
NMO	N-hydroxylating monooxygenase

NMR	Nuclear magnetic resonance
NP	Natural Product
NRP	Nonribosomal peptide
NRPS	Non-ribosomal peptide synthetases
NSF	National Science Foundation
NTA	Nitriloacetic acid
OD	Optical density
ODC	Ornithine decarboxylase
Ox	Oxidation domain
PCCcc	Pasteur Culture Collection
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PK	Polyketide
PKS	polyketide synthase
PLP	Pyridoxal 5'-phosphate
PR	pulegone reductase
PTM	putrescine N-methyltransferase
PTMs	post-translational modifications
PVDF	Polyvinylidene fluoride
QPT	Quinolate phosphor-ribosyltransferase
QS	Quinolate synthase
RiPP	Ribossomally-produced and post-translationally modified peptide
SAH	S-adenosyl-L-homocysteine
SAM	S'adenosyl-I-methionine
SBCAD	Short/branched chain acyl-CoA dehydrogenase
SCAD	Short-chain acyl-CoA dehydrogenase
SDS	Sodium Dodecyl Sulfate
Sfp	4'-phosphopantetheinyl transferase
SLiC	Sequence- and ligation-independent cloning
SNAC	N-acetylcysteamine
SSN	Sequence similarity network
SUMO	Small Ubiquitin-like Modifier
TAE	Tris-acetate-EDTA
TAR	Transformation-associated recombination
TB	Terrific broth
TBST	Tris-buffered saline

TC/TS	terpene cyclases/terpene synthases
TE	Thioesterase
TGS	Tris-Glycine-SDS
TIC	Total ion chromatogram
TLC	Thin layer chromatography
Trx	Thioredoxin
UV	Ultraviolet
VLC	Vacuum liquid chromatography
VLCAD	Very-long acyl-CoA dehydrogenase
XMT	Xanthosine 7- <i>N</i> -methyltransferase

General Introduction

Introduction

1. Natural Products: Chemistry, Bioactivity and Biosynthesis

1.1. Natural Products: Chemistry and Bioactivity

Natural products (NPs) or specialized metabolites are often defined as small molecules that are not directly involved in cell growth or development, but represent an evolutive/adaptative advantage for the producing organism.² In nature, they can serve diverse ecological roles such as defense, symbiosis, metal transport, competition, nutrient capture from the environment, intra- and inter-species communication and others.³ On the other hand, specialized metabolites are also known for their technological applications which includes not only pharmaceuticals, but also flavors, fragrances, fuels, biopolymers, pesticides and others.⁴

Humankind has been taking advantage of the properties of NPs since early times.⁵ For instance, cypress (*Cupressus sempervirens*) oils were used to treat coughs, fever and inflammation since 2600 B.C..⁶ Furthermore, the discovery of famous NPs such as morphine (1803) and penicillin (1929) revolutionized the History of Modern Medicine⁷ and helped to double the human life expectancy during the 20th century. In the past two decades, NPs, as well as their derivatives and mimics, represented over half of all new Food and Drug Administration (FDA) approved drugs,⁸ including anticancer, analgesic, immunosuppressant, antibacterial, antiparasitic, antifungal drugs among others (Table A.1, Fig. A.1).^{8,9}

Table A.1 - Relevant examples of FDA-approved NP or NP derivatives drugs

Bioactivity	Compound name	Type of compound	Source organism	FDA-approval date	Reference
Antibacterial	Penicillin	nonribosomal peptide	<i>Penicillium rubens</i>	1945	10
	Streptomycin	aminoglycoside	<i>Streptomyces griseus</i>	1946	11
Antimalaric	Quinine	alkaloid	<i>Cinchona succirubra</i>	2004	12
Anticancer	Dolastatin 10	peptide	<i>Symploca</i> sp.	2011	13
	Taxol	terpene	<i>Taxus brevifolia</i>	1992	14
Analgesic	Morphine	alkaloid	<i>Papaver somniferum</i>	1827	15
Antifungal	Amphotericin B	polyketide	<i>Streptomyces nodosus</i>	1997	16
Immunosuppressant	Rapamycin	polyketide	<i>Streptomyces hygroscopicus</i>	1999	17

The majority of all known NPs or secondary metabolites are produced by plants (~80%), followed by bacteria, fungi, and other marine organisms.¹⁸ Additionally, genome data indicates that 90% of microbial NPs are still cryptic¹⁹, i.e. they have not been isolated nor their structures elucidated,²⁰ which underlines a huge potential for the discovery of new and unique NPs and their corresponding biosynthetic enzymes. This genomic dark matter is particularly significant for microbial NPs, which are the source of an impressive number of highly effective chemotherapeutic agents despite their outstanding unexplored biodiversity.⁶

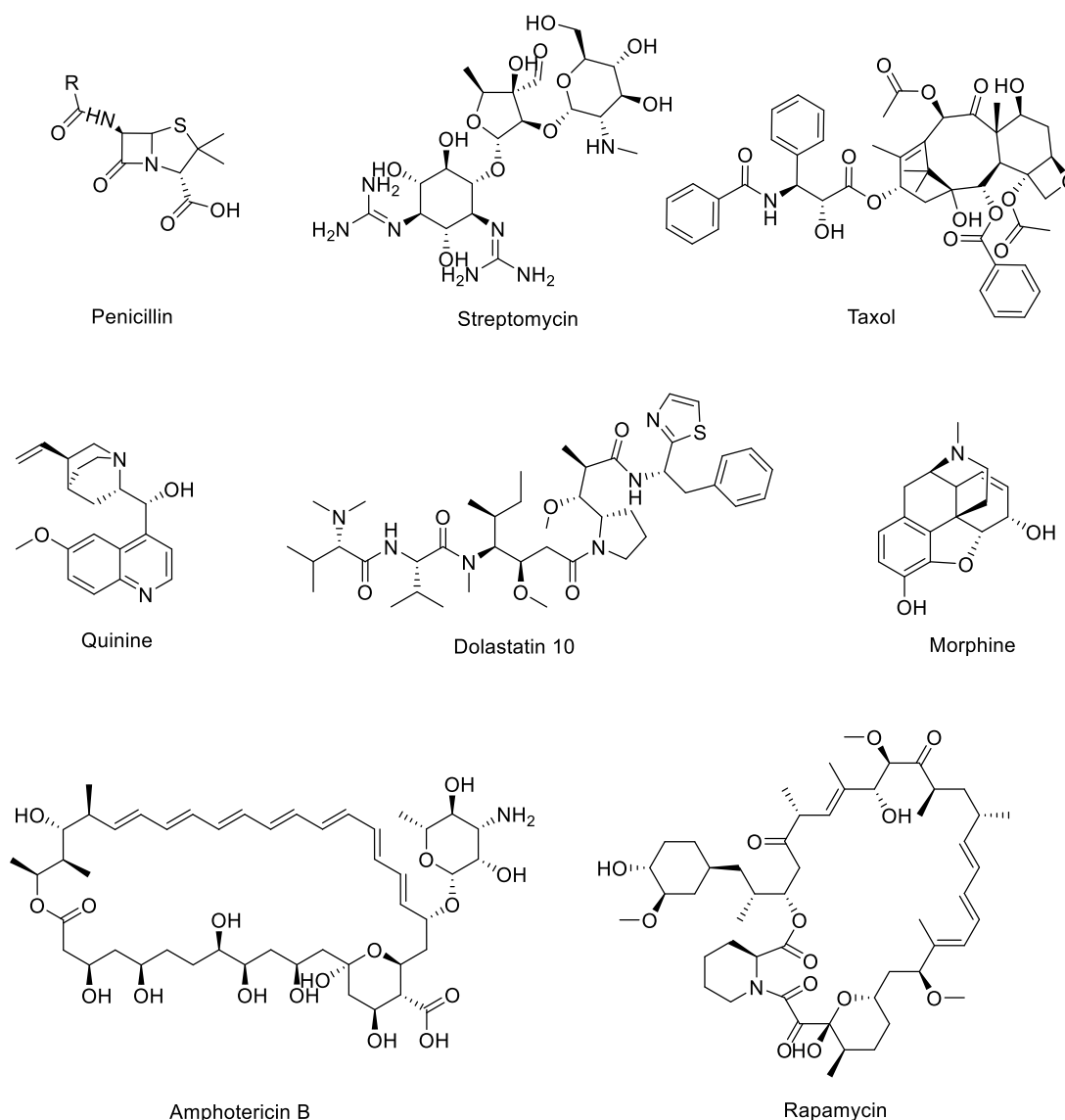


Figure A.1 - Chemical structures of relevant FDA-approved NP drugs.

This recognized impact of natural small molecules in pharmacology and chemical biology,⁸ is mainly derived from the broad chemical diversity of NPs,²¹ covering a unique chemical space from very small molecules, such as limonene, to bigger and more

complex structures, such as ramoplanin or maitotoxin (Fig. A.2).^{22,23} This structural diversity encompasses a large diversity of molecular properties like molecular weight, logP, ring systems, functionalizations, chiral centers and others.²⁴ Compared with synthetic compounds, NPs usually present a higher density of functional groups, more complex ring systems and more chiral centers.²⁵ Additionally, the complexity and scaffold diversity of NPs has been a continued source of inspiration for chemists to develop new total synthesis strategies and produce synthetic and semisynthetic NP derivatives, which has led to new synthetic methodology.¹⁹

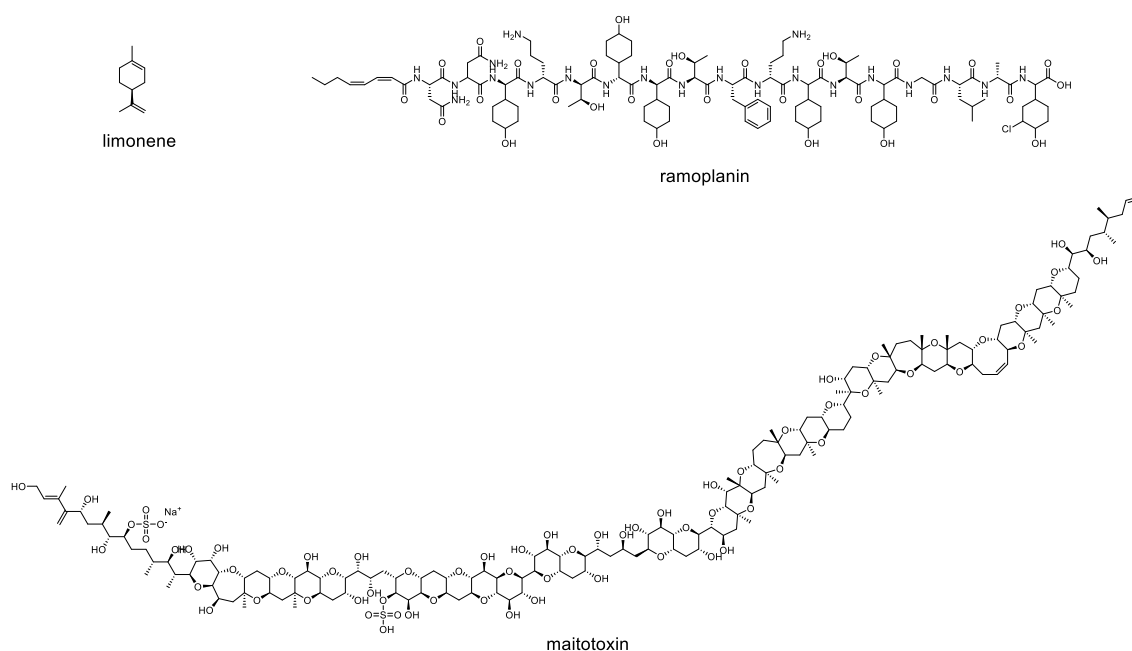


Figure A.2 - Chemical structures of limonene, ramoplanin and maitotoxin.

Overall, the complexity, diversity and bioactivity of NPs, when considered in the context of the large extent of cryptic biosynthetic information in genome data, lead to belief that the contributions of NPs to ecology, biotechnology and medicine are yet to be fulfilled.²⁶

1.2. Natural Products Biosynthesis

Despite the high structural diversity of NPs, the biosynthetic machineries involved in the formation of their core structures are often highly conserved,²⁷ presenting a high amino-acid-sequence similarity which allows the identification of biosynthetic genes from genome data. Based on their biosynthetic origin/machinery, NPs can be classified as terpenes, alkaloids, polyketides (PKs), nonribosomal peptides (NRPs), ribosomally synthesized and post-translationally modified peptides (RiPPs), among other classes.²⁷

However, some molecules can be produced by other biosynthetic logics for which we currently lack signature biosynthetic genes.

In prokaryotic organisms, NPs result from the assembly of building blocks usually from primary metabolism that are transformed by enzymes encoded in biosynthetic gene clusters (BGCs).²⁰ These are groups of two or more genes co-localized in the genome that encode a common biosynthetic pathway for the production of a specialized metabolite.²⁸ Despite the conserved core-machineries, NPs diversification is highly impacted by the enzymatic activity of many divergent and unusual tailoring enzymes.^{29,30}

1.2.1. PKs/NRPs biosynthesis

In PKs and NRPs, the biosynthesis is usually performed by large multifunctional modular enzymes – polyketide synthases (PKS) or nonribosomal peptides synthetases (NRPS), respectively - acting on covalently bound substrates,³¹ like in the biosynthesis of erythromycin and thermoactinoamide (Fig. A.3a, A.3b).^{32,33} A similar biosynthetic logic is involved in the generation of PKs and NRPs. In PKs biosynthesis by PKSs, the initial monomer, acetyl-coenzyme A (or analogues) is activated to a thioester formed with the phosphopantetheine (Ppant) arm of the acyl carrier protein (ACP) domain.³⁴ The Ppant arm, originated from coenzyme A, is attached to a serine residue of the *apo*-ACP by the 4'-phosphopantetheinyltransferase Sfp, forming a phosphodiester linkage and consequently generating *holo*-ACP.³⁵ Secondly, the ACP transfers the intermediate to the ketosynthase domain (KS) which performs an extension with two-carbon linear monomers, traditionally malonyl-CoA, using a decarboxylative Claisen condensation reaction.³⁶ The extender units are recognized by the acyltransferase (AT) domain and loaded onto the ACP. These three domains (ACP, KS and AT) constitute the minimum module in a traditional type I PKS.³⁷ Additional reductive domains are often present in the modules of the PKS assembly line, like ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains, consecutively reducing the initial β -keto group to a hydroxyl group, double bond or single bond, respectively.³⁴ At the end of the assembly line, a thioesterase domain (TE) performs thioester hydrolysis, releasing the final PK.³⁷ Most PKSs follow the collinearity rule, which dictates that the domains are present in the BGC in the same order as they act on the biosynthesis. However, several exceptions to these rules are found, including iterative type I PKSs where the same module can perform several rounds of elongation/modification and type I *trans*-AT PKSs which use stand-alone AT domains to load multiple extension units.³⁸ Additionally, PKS diversity comprises type II and type III PKS, which present different architectures and catalytic mechanisms from type I PKS.^{39,40} Several variations in terms of starter extension units

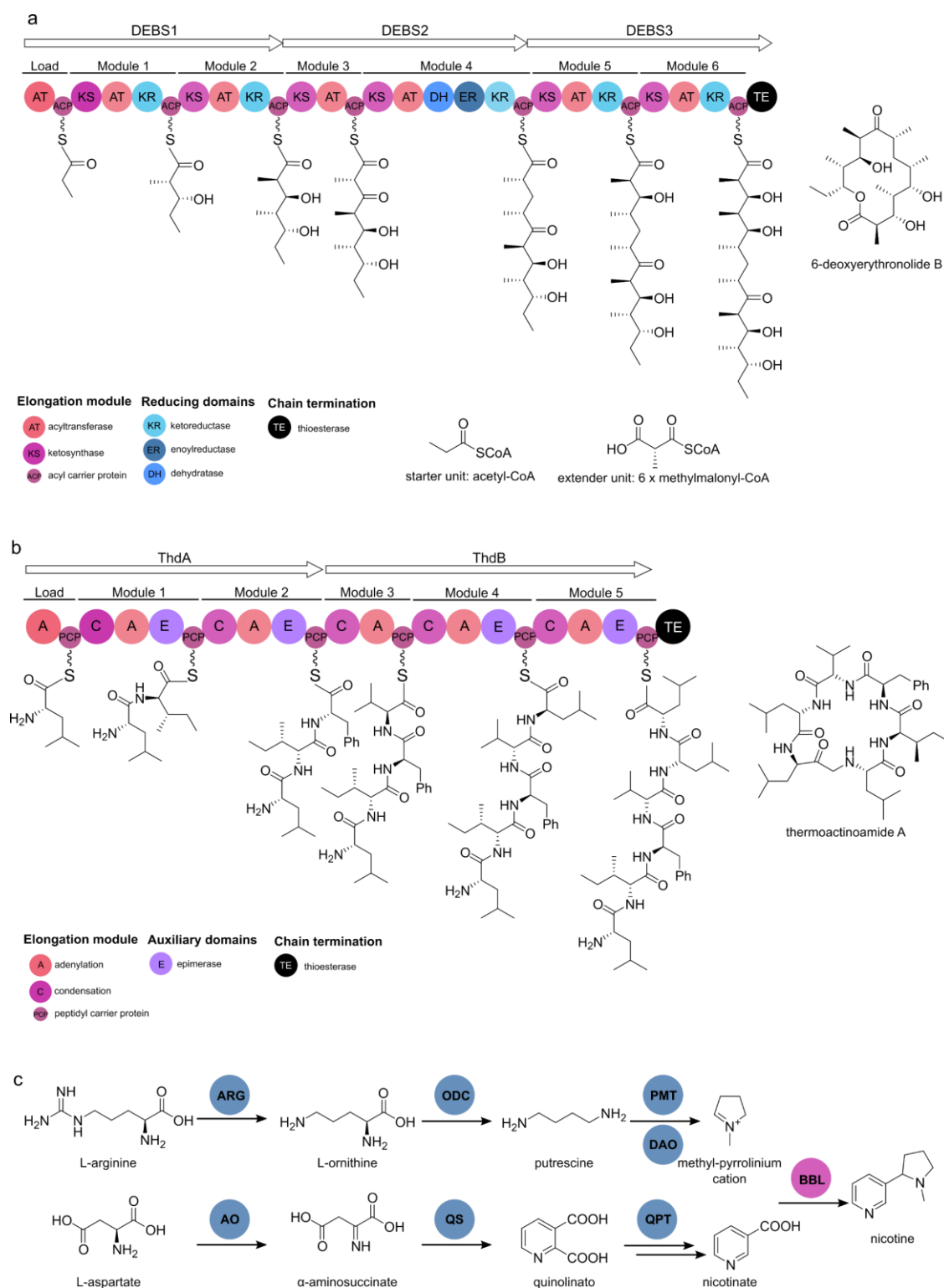


Figure A.3 – Schematic representation of PKS, NRPS and alkaloid biosynthetic pathways. (a) 6-deoxyerythronolide B, an erythromycin intermediate is formed by a 5 modules PKS assembly line that starts with acetyl-CoA as is followed by 6 extension rounds using methyl-malonyl-CoA and several reduction steps. (b) The biosynthesis of thermoactinoamide A, an antimicrobial NRP from *Thermoactinomyces vulgaris*, involves 5 amino acid residues (3 leucines, 1 isoleucine and 1 tyrosine) and several epimerization domains. (c) Nicotine biosynthesis requires arginine and aspartate which are transformed by several biosynthetic enzymes (ARG – arginase, ODC – ornithine decarboxylase, PMT – putrescine N-methyltransferase, DAO – diamine oxidase, AO – aspartate oxidase, QS – quinolinolate synthase, QPT – quinolinolate phosphoribosyltransferase) to produce the methyl-pyrrolinium cation and nicotinate, respectively, which are used by a berberine-bridge enzyme (BBL) to originate the final nicotine molecule.

can also be found, as well as accessory domains like, methyltransferases (MT) and other tailoring enzymes acting *in trans* or after release of the compound from the assembly line, increasing the complexity and structural diversity of the final PK products.^{41,42}

Analogous to PKS assembly lines, in the biosynthesis of NRPs by NRPSs, a peptidyl carrier protein (PCP) harboring the Ppant arm is loaded with the initial amino acid which is transferred to the condensation (C) domain. This domain forms peptide- or ester-bonds with the extender unit, previously selected and activated in a ATP-dependent manner by the adenylation (A) domain.⁴³ Like in PKSs, several variations of the starter unit can be found, as well as additional domains like cyclization (Cy), oxidation (Ox) and epimerization (E), which introduce structural diversity into the final NRPs.⁴⁴

Due to their shared biosynthetic logic, there can be crosstalk between PKSs and NRPSs and these often form hybrid pathways. These can be mixed PKS/NRPS modules or discrete PKS and NRPS modules acting on a progressive line.⁴⁵

1.2.2. Alkaloid biosynthesis

Unlike PKs and NRPs, the biosynthesis of the remaining classes of NPs uses discrete enzymes acting on free-standing substrates. Among all NPs, alkaloids are likely the most heterogeneous group, being broadly defined as heterocyclic structures containing one or more nitrogen atoms, with alkaline character.⁴⁶ Although most of the NPs have a recognizable biochemical logic throughout the pathway like mentioned before, the biosynthetic pathways of alkaloids are as diverse as their structures.⁴⁷ However, the biosynthesis of most alkaloids makes use of amino acids like tryptophan or aspartate, and takes advantage of the formation and reactivity of the iminium cation, like for example in the case of nicotine biosynthesis (Fig. A.3c).⁴⁸⁻⁵⁰ Usually, the first biosynthetic steps entail the accumulation of an amine and an aldehyde precursor, followed by the formation of an iminium cation and a Mannich-like reaction, forming the basic alkaloid scaffold.⁴⁹ Furthermore, these reactions can occur intermolecularly like in benzyloquinolines biosynthesis, or intramolecularly like in polyamines.⁴⁹ Depending on the alkaloid class (pyrrolidines, pyridines, tropanes, pyrrolizidines, isoquinolines, indoles, quinolines, terpenoids and steroids), the following biosynthetic steps can diverge significantly.⁵¹ Because of this variability, it is difficult to predict the biosynthesis of an alkaloid only based on biosynthetic precursor and structural resemblance to previously characterized molecules, making alkaloid biosynthesis a challenging, but rewarding area of study.

1.2.3. Terpene biosynthesis

Terpenes are assembled from simple C5 isoprenoid units, usually isopentyl pyrophosphate (IPP) from the methylerythritol phosphate pathway⁵² and dimethylallyl diphosphate (DMAPP) from the mevalonate pathway,⁵³ which are joined into a polyene chain. Following this extension step, the chain is cyclized by terpene cyclases (TCs), also known as terpene synthases (TSs), which position the chain in a defined conformation that allows carbocation-driven cyclizations and rearrangements.⁵⁴ This creates the basic hydrocarbon skeleton of the terpene that might be further modified by downstream tailoring enzymes. Despite the simple biosynthetic logic, terpenes are the most chemically diverse class of NPs, being very common in plants, but also widely spread among bacteria.⁵⁵ This diversity is justified by the large diversity of TCs that can perform cyclization in very different patterns, even when using the same building blocks.⁵⁵ Additionally, several downstream modifications like acetylation, glycosylation, methylation and oxidation are common and able to generate several terpenes like in the case of limonene and menthol (Fig. A.4a).⁵⁶⁻⁵⁸ Particularly, most of these oxidation steps are performed by members of the cytochrome P450 family which is among the most diverse and versatile classes of enzymes.⁵⁹

1.2.4. RiPP biosynthesis

In the biosynthesis of RiPPs, a precursor gene is translated by the ribosome into a precursor peptide, which is typically 20-120 amino acids in length.⁶⁰ This precursor usually comprises a core and a leader region.⁶¹ The core region is where the post-translational modifications (PTMs) take place and is the region kept in the final peptide.⁶² The leader peptide provides binding affinity for many of the PTM enzymes and in some cases is also required to ensure that the PTMs are carried out in the correct order.⁶¹ The extensive and diverse post-translational modifications carried out in the core peptide provide a high degree of structural complexity and diversity like in the case of the antibacterial thiomuracins (Fig. A.4b).^{60,63-65} Once the core peptide has been fully modified, it is cleaved from the leader region, yielding the final natural product. In other cases, precursors carry multiple core peptides and multiple recognition sequences that allow different chemical transformation on the core peptide, and thus one precursor peptide results in more than one RiPP product.⁶¹

BIOSYNTHESIS OF NOCUOLIN A AND RELATED METABOLITES

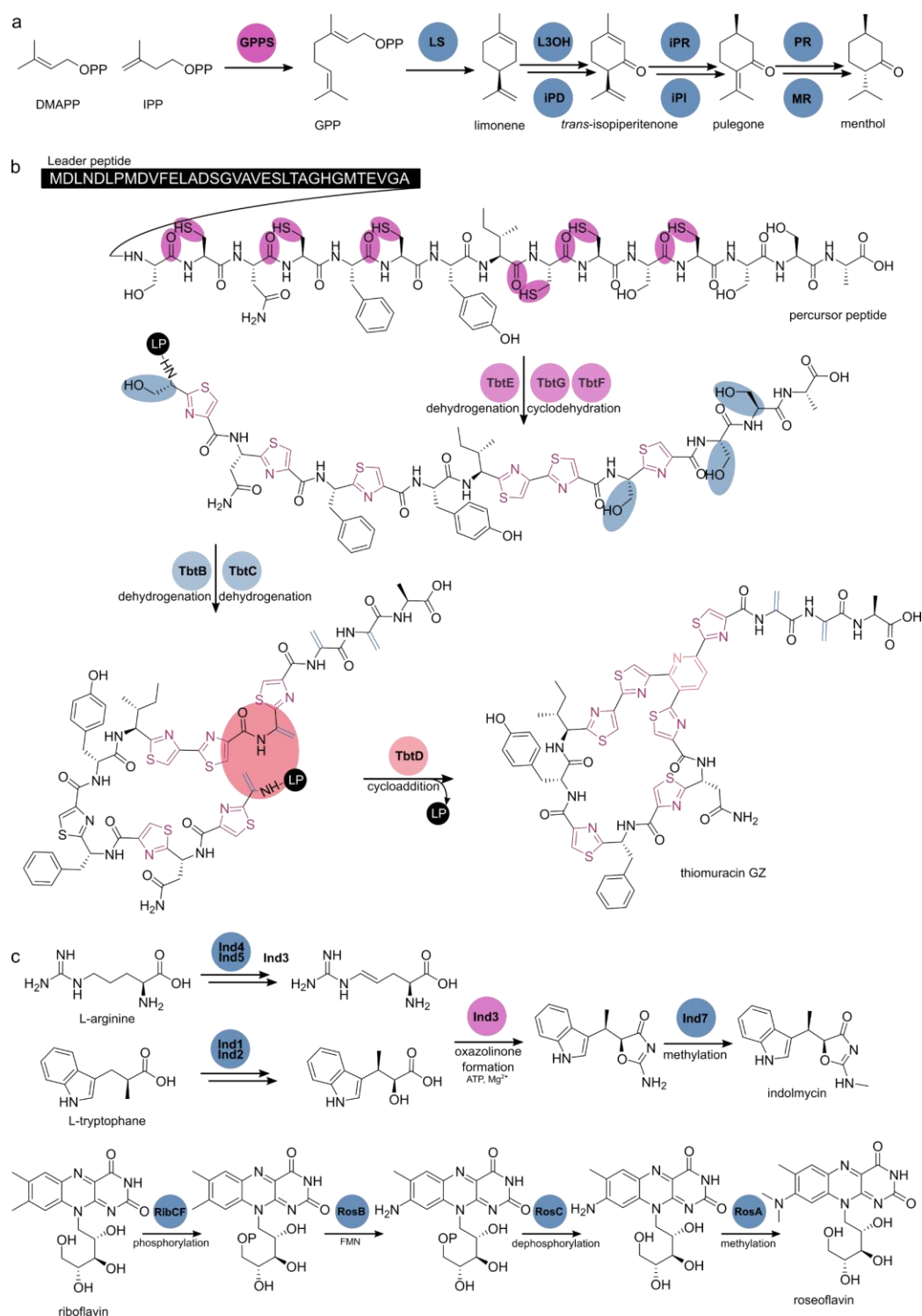


Figure A.4 – Schematic representation of terpenes, RiPPs and no signature enzyme pathways. (a) The biosynthesis of menthol and limonene is derived from the same biosynthetic pathway that initiates with the formation of geranyl pyrophosphate (GPP) by the GPP synthase (GPPS) from IPP and DMAPP. GPP is then converted to limonene by the limonene synthase (LS). Limonene is converted to menthol by the action of several biosynthetic enzymes (L3OH - limonene 3-hydroxylase, iPD – trans-isopiperitenol dehydrogenase, iPR – trans-isopiperitenone reductase, iPI – cis-isopulegone isomerase, PR – pulegone reductase, MR – menthone reductase). (b) The biosynthesis of the antibacterial thiomuracins is initiated by the ribosomal production of a precursor peptide composed by the leader and the core peptide. This initial peptide is subjected to cyclodehydration and dehydrogenation by TbtF/G (F-protein-dependent cyclodehydratase) and TbtE (thiazoline dehydrogenase), respectively, originating the 6 thiazole rings. TbtB and TbtC (LanB-type dehydratases) perform the following PTMs carrying out dehydrogenation that results in the formation of dehydroalanines. The biosynthesis is finalized by the formation of the pyridine ring as result of the cycloaddition of TbtD (pyridine synthase). (c) The biosynthesis of indolmycin and roseoflavin involves no signature enzyme pathways.

1.2.5. Pathways lacking a signature enzyme

Apart from the above-mentioned well-studied pathways, some BGCs lack the signature enzymes referred to previously, making them particularly enigmatic. The biosynthesis of these compounds makes use of alternative strategies. For instance, in indolmycin biosynthesis, an antibacterial compound, the peptide bond is formed not by the traditional NRPS pathway, but instead by a phenyl acetate CoA ligase, Ind3, which also catalyzes the oxazolinone ring formation of the molecule (Fig. A.4c).⁶⁶ Some of these BGCs can also use primary metabolites, lacking dedicated signature enzymes in the clusters. This is demonstrated in the biosynthesis of roseoflavin, which uses riboflavin as substrate, only requiring four enzymes to produce the final metabolite (Fig. A.4c).^{67,68} Because of the lack of signature genes, the identification of these BGCs, their connection to the synthesized NP and the elucidation of the biosynthetic pathway is particularly challenging. However, despite these difficulties, these pathways represent exciting possibilities to unveil unique enzymology that significantly diverges from the traditional pathways and that remains cryptic. In addition, elucidating the biosynthesis of such pathways contributes greatly to the body of knowledge by revealing previously unrecognized signature enzymes that can be used to bioinformatically classify BGCs and predict structural features.⁶⁹

1.3. Biosynthetic enzymes and biocatalysis

1.3.1. Biosynthetic enzymes

Biosynthetic enzymes are responsible for introducing complex motifs typically found in NPs (Fig. A.1), usually with high reaction rates, regio-, stereo- and enantioselectivity.⁷⁰ As such, they represent a source of inspiration for devising new synthetic methodology and to develop bioinspired catalysts⁷¹ which results in major energy, time and material savings.⁷² Additionally, they can catalyze unusual reactions that are currently inaccessible or limited using conventional synthetic methodologies, as for example C-H bond activation.^{73,74} In fact, perhaps the most state-of-the-art biocatalysts developed so far regard late-stage stereoselective C-H functionalization and C-C bond formation.^{75,76} This includes the P450 monooxygenase Tyll that performs the last hydroxylation step on the biosynthesis of the antibiotic juvenimicin B1 (Fig. A.5).⁷⁷ Other remarkable enzymes include C-C bond forming enzymes, such as CylK that performs an intermolecular Friedel–Crafts alkylation of two halogenated resorcinol derivatives to construct the cylindrocyclophanes⁷⁸ and KtnC that performs an oxidative coupling of coumarin derivatives in a stereoselective fashion to form P-orlandin (Fig. A.5).⁷⁹ Other NP biosynthetic enzymes play key roles in early-stage biosynthetic steps with great impact in synthetic applications. This includes enzymes like GriE, a non-heme

iron-dependent enzyme that hydroxylates a L-leucine derivative, forming an early intermediate in the biosynthesis of manzacidin C,⁸⁰ the α -oxoamine synthase SxtA AONS that deuterates an alanine methyl ester in the biosynthesis of the Parkinson's pharmaceutical safinamide⁷⁵ and BesD, a radical halogenase that chlorinates free amino acids to generate non-proteinogenic amino acids (Fig. A.5).⁸¹ Other relevant examples include N-N bond forming enzymes like the N-nitrosing SznF that performs N-N bond formation of *N*^ω-methyl-L-arginine, resulting in the formation a N-nitroso group that is an intermediate in the biosynthesis of the cancer chemotherapeutics streptozotocin^{82,83} and CreM (Fig. A.5), involved in the biosynthesis of the antibacterial cremeomycin⁸⁴.

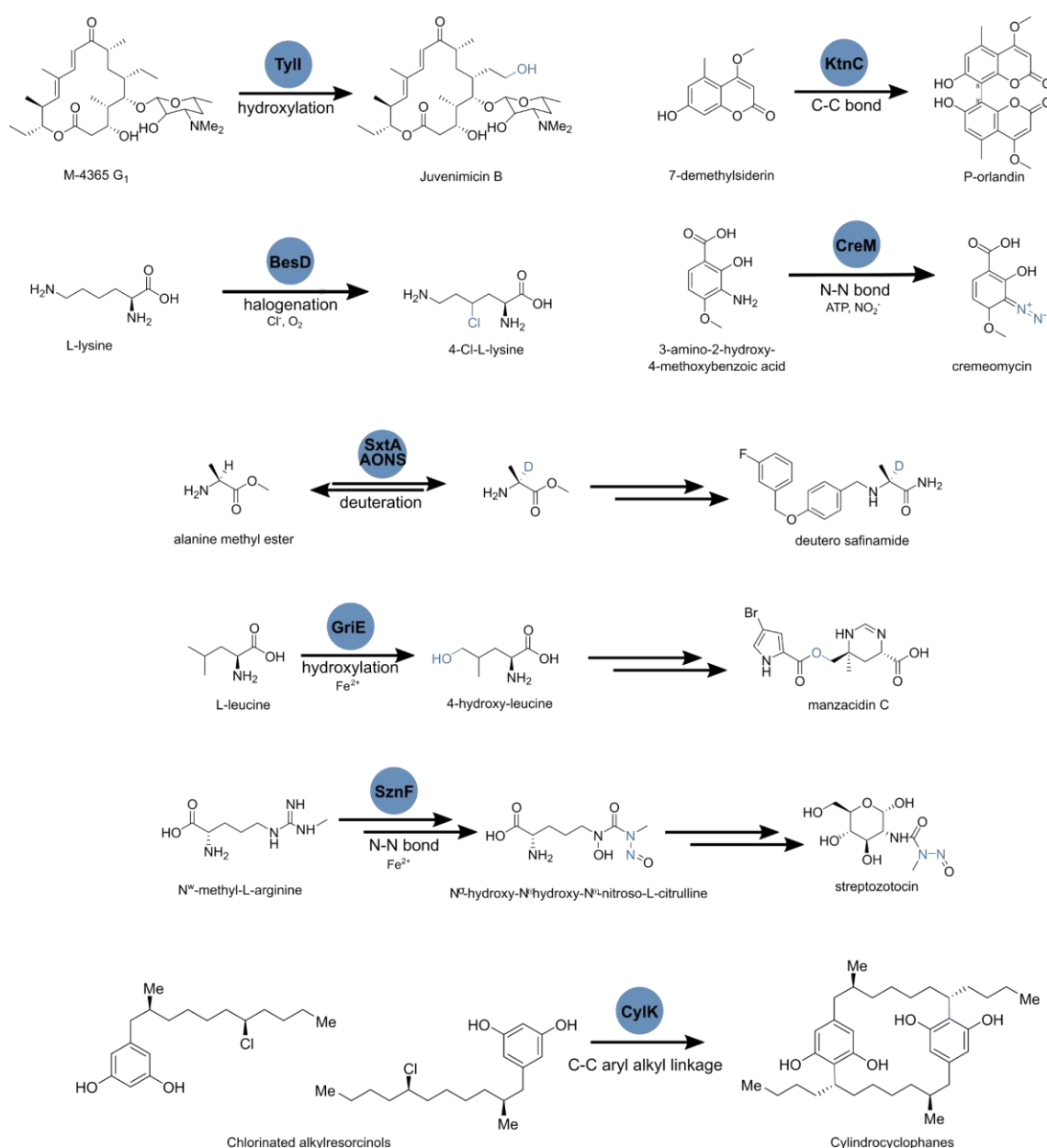


Figure A.5 – Example of remarkable biosynthetic enzymes and their unusual C-H activation, C-C formation and N-N formation reactions, including TyII, KtnC, BesD, CreM, SxtA, GriE, SznF and CylK.

The diversity of NP biosynthetic enzymes and their reactivities, reflected in the diversity of the resulting NPs, is a remarkable starting point to develop biocatalysts that perform unconventional reactions.^{73,74} However, enzymes involved in NP biosynthesis frequently present a narrow substrate scope and low turnovers, challenging its direct application in synthetic processes.⁸⁵ To overcome this issue, protein engineering and direct evolution can be applied to convert enzymes from primary metabolism with higher catalytic turnovers and broader substrate scope into specialized enzymes mimicking NP biosynthetic enzymes.⁸⁵ Additionally, NP biosynthetic enzymes have played a key role in drug discovery programs, by creating more efficient catalytic routes and by allowing access to building blocks that explore new areas of the chemical space and can be used in screening for lead compounds.⁸⁶

Furthermore, with the increase accessibility to genomic data, it is clear that the biosynthetic 'dark matter' represents invaluable potential to access cryptic pathways and discover new biosynthetic enzymes²⁵ whose biocatalysis interest has been key to rekindle the importance of NPs chemistry and biosynthesis in the genomics era.

1.3.2. Biocatalysis

Naturally occurring enzymes have been applied for human benefit for thousands of years, first through fermentation and food preservation processes⁸⁷ and more recently at the molecular level for numerous applications, particularly in organic synthesis.⁷⁵ Like mentioned above, biocatalysts allow remarkable reaction rates and selectivity, using mild reaction conditions, producing less waste, being more biodegradable than traditional chemical synthesis and being able to interact with other enzymes in a one-pot reaction.^{88,89} These features represent major advantages for the use of enzymes in biocatalysis processes, and are particularly important from an environmental point of view, with the increased need to use greener alternatives and develop a circular bioeconomy (Fig. A.6a).⁷³

Nowadays, biocatalysis is used in processes that range from simple chemical manipulations like reductive aminations, alcohol oxidations, chiral resolutions⁹⁰⁻⁹², to complex, multistep enzymatic cascades to produce high-value molecules on an industrial scale.⁹² Biocatalysts can be used in combined strategies with synthetic systems, but also in total enzymatic systems. In fact, total enzymatic systems have been developed in recent years to produce didanosine and the experimental anti-HIV islatravir.^{92,93}

Some of these enzymes, from the most traditional lipases, used for decades for the production of cheese and detergents⁹⁴ to the state-of-the-art NP biosynthetic enzymes have been at the core of several industrial biotechnological applications developed in recent years.^{75,95,96} Further examples include KabC, a α -ketoglutarate-

dependent enzyme used in biocatalysis of the neurochemical kainic acid⁹⁷ to perform oxidative cyclization and CYP71BE5, a cytochrome P450 that generates (-)-rotundone, a molecule used in the fragrance industry (Fig. A.6b).⁹⁸ Additionally, transaminases, also known as aminotransferases, have also been used to synthesize valuable chiral amines from a prochiral precursor in a single step,⁹⁹ like in the synthesis of the anti-epilepsy pregabalin^{100,101} or the herbicide L-glufosinate (Fig. A.6b).^{102,103} Very large industries like Cosmetic, Food, Materials, Bulk chemistry, Pharmaceutical, Water treatment and Biofuels¹⁰⁴⁻¹⁰⁸ have largely benefited from the innovations associated with using biocatalysts^{107,108} which is reflected in large economical outcomes. Therefore, functional characterization of new biosynthetic enzymes is a major research interest field that meets both the fundamental and applied science interests.¹⁰⁴

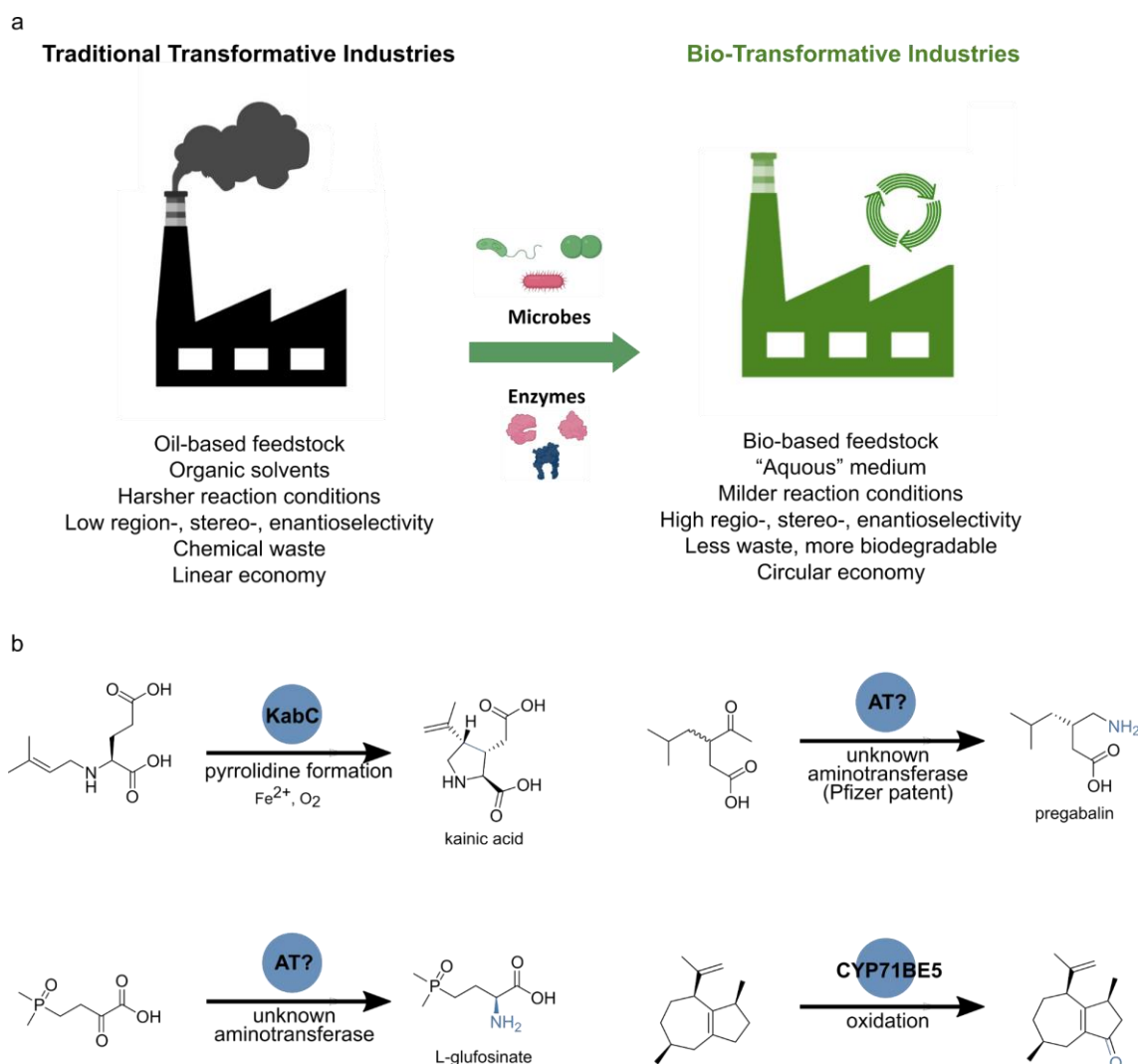


Figure A.6 – Biocatalysts application in industrial context. (a) The application of microbes or enzymes in bio-transformative industries leads to several advantage features with high environmental impact, contributing to a circular bioeconomy. (b) Examples of enzymes and their reactions currently used in industrial context for the production of kainic acid (KabC), pregabalin (aminotransferase), L-glufosinate (aminotransferase) and (-)-rotundone (CYP71BE5).

2. Cyanobacteria Natural Products: Chemistry, Bioactivity and Biosynthesis

2.1. Cyanobacteria

Cyanobacteria are a highly diverse prokaryotic group that includes all Gram-negative bacteria capable of oxygenic photosynthesis.^{109–110} They are ancient (2.7 to even 3.5 billion years old)^{111,112} organisms and were responsible for the oxygenic atmosphere formation¹¹³ in the Proterozoic era (also called “the age of cyanobacteria”), representing one of the most successful and oldest life forms on the planet. Additionally, they are proposed to be ancestors of today's chloroplasts of higher plants¹¹⁴ and are still

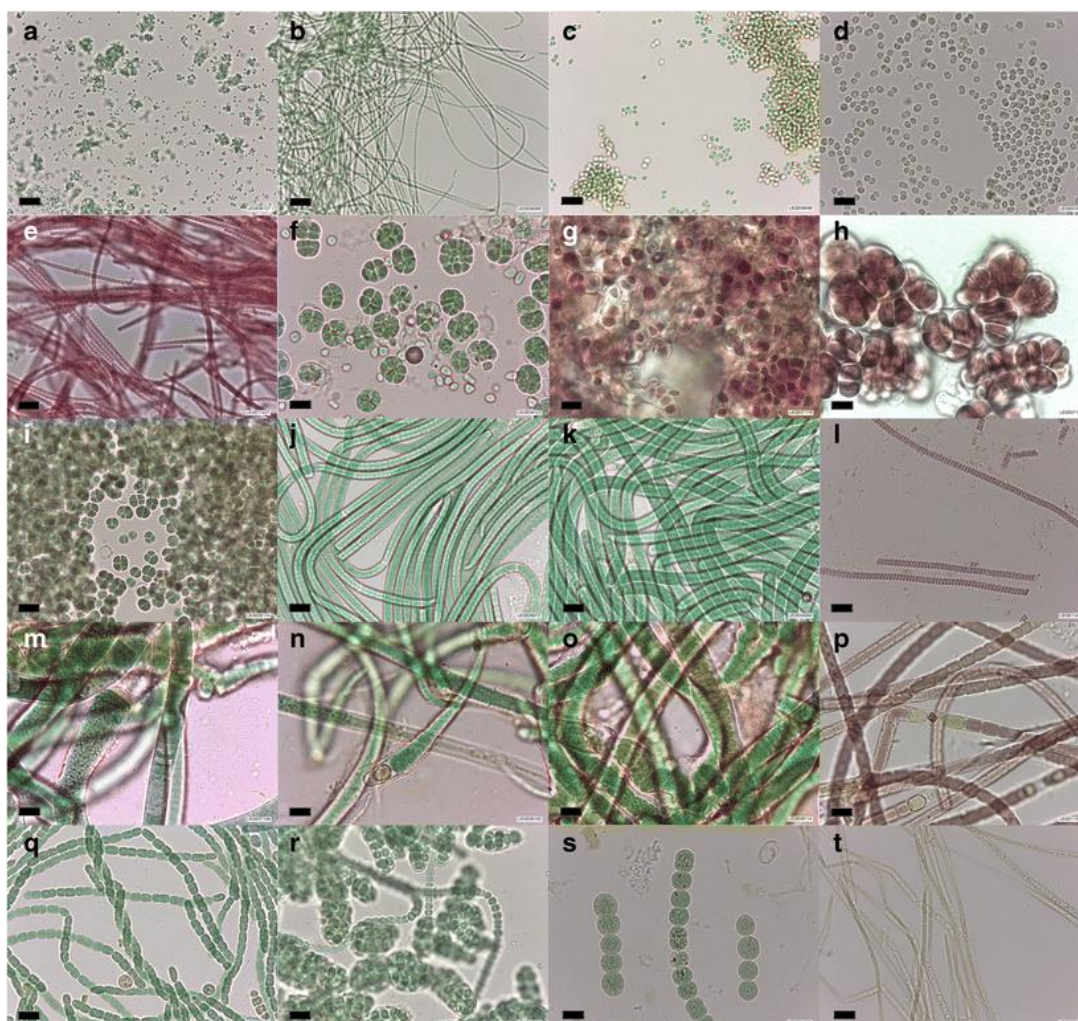


Figure A.7 – Cyanobacteria morphologic diversity in LEGE culture collection (LEGEcc). Micrographs of (A) *Cyanobium* sp. LEGE 06127, (B) *Nodosilinea* sp. LEGE 06069, (C) *Synechocystis salina* LEGE 06099, (D) *Microcystis aeruginosa* LEGE 91094, (E) unidentified *Chroococcales* LEGE 11427, (F) *Gloeocapsopsis crepidinum* LEGE 06123, (G) *Hyella patelloides* LEGE 07179, (H) *Chroococcopsis* sp. LEGE 07187, (I) *Chroococcidiopsis* sp. LEGE 06174, (J) cf. *Oxynema acuminatum* LEGE 06072, (K) *Phormidium* sp. LEGE 00064, (L) cf. *Spirulina* sp. LEGE 11439, (M) *Rivularia* sp. LEGE 07159 (N) *Calothrix* sp. LEGE 06100, (O) *Plectonema* cf. *radiosum* LEGE 06114, (P) *Tolypothrix* sp. LEGE 11397, (Q) *Nodularia* sp. LEGE 06071, (R) *Nostoc* sp. LEGE 07365, (S) *Dolichospermum flosaquae* LEGE 04289, (T) *Cylindrospermopsis raciborskii* LEGE 95046 (adapted from¹).

key players to maintain the biosphere homeostasis through carbon and nitrogen fixation.¹¹⁵

Cyanobacteria present unique complex morphologies and cellular differentiation.¹¹⁰ Apart from the vegetative cells, cyanobacteria can produce heterocysts to perform nitrogen fixation¹¹⁶ and akinetes for protection against unfavorable environmental conditions.¹¹⁷ Cyanobacteria present remarkable morphologic diversity including unicellular, colonial or filamentous forms^{118,119} (Fig. A.7) and different coloration due to unique combination of pigments like chlorophyll, carotenoids and phycobiliproteins.¹²⁰ Their diversity and adaptability allow them to inhabit a large range of ecosystems ranging from terrestrial, marine and freshwater, including the most extreme ones such as deserts, hypersaline or high-UV exposed regions.¹²¹⁻¹²³

2.2. Cyanobacterial natural products

During their long evolution, cyanobacteria have become rich producers of NPs^{124,125} with a wide range of ecological roles like toxins or UV protectors.^{126,127} Biotechnological applications for cyanobacterial NPs include uses as source of biofuels, food supplements and bioplastics.¹²⁸⁻¹³⁰ Generally, filamentous cyanobacteria are considered richer sources of NPs than unicellular cyanobacteria, with the orders of Oscillatoriales and Nostocales contributing 49% and 26%, respectively, of all cyanobacterial NPs known to date.¹³¹

Several cyanobacterial NPs also present pharmacologically-relevant bioactivities including antibacterial, antiparasitic, antifungal, antiviral, anti-inflammatory^{8,9} and mainly anticancer.¹²⁴ In fact, cyanobacteria are an exceptional source of lead compounds for drug discovery^{132,133} (Table A.2, Fig. A.8), representing the second richest marine biogenetic source of chemical entities approved for use in the clinic or undergoing clinical trials.¹³⁴

Perhaps the best example of cyanobacterial bioactive compounds is dolastatin 10¹³⁵ (Table A.2, Fig. A.8) whose anticancer activity allowed the development of Brentuximab vedotin, approved in 2011 by the FDA to treat Hodgkin's lymphoma, representing the first commercially available antibody drug conjugate based on a cyanobacterial NP. Additionally, several derivatives of dolastatin have been recently approved by FDA in other antibody-drug conjugates (Belantamab mafodotin, Polatuzumab vedotin and Enfortumab vedotin) and several more are currently undergoing clinical trials.

Apart from their recognized impact in the pharmacological area, cyanobacterial NPs present considerable structural diversity, which comprises extensive modifications, ranging from purely fatty-acid (FA), alkaloid, terpene, or peptide to a wide diversity of

hybrid PKs/NRPs^{124,136,137} (Fig A.8). Additionally, they present rare modifications that include unusual halogenations (trichlorination, iodination), various methylations, incorporation of α -hydroxy acids, unusual chain terminations¹³¹ and interesting cyclic moieties such as thiazoline, oxazole and resorcinol rings.¹³⁸⁻¹⁴¹ When compared to those of other chemically-rich bacteria, cyanobacterial NPs are particularly rich in fatty-acid moieties (Table A.3, Fig. A.8).^{141,142} This includes compounds such as the bartolosides,¹³⁹ cylindrocyclophanes,¹⁴³ hapalosins,^{144,145} microcarbonin,¹⁴⁶ jamaicamides,¹⁴⁶ nocuolin A,¹⁴⁷ chlorosphaerolactylates,¹⁴⁸ grenadadiene¹⁴⁹ and aranazoles.¹⁵⁰

Table A.2 - Relevant examples of cyanobacterial bioactive NPs

Bioactivity	Compound name	Type of compound	Cyanobacteria	Reference
Antibacterial	Noscomin	terpene	<i>Nostoc commune</i>	151
Anticancer	Apratoxin A	PK-NRP	<i>Moorea</i> (formerly <i>Lyngbya</i>) <i>bouillonii</i>	152
	Dolastatin 10	peptide	<i>Symploca</i> sp. <i>Lyngbya</i> sp.	135
Antifungal	Carriebowlinol	alkaloid	<i>Moorea bouillonii</i>	153
Antimitotic	Curacin A	PK-NRP	<i>Moorea producents</i>	154

Table A.3 - Relevant examples of fatty-acid containing cyanobacterial NPs

Compound name	Type of compound	Cyanobacteria	Reference
Cylindrocyclophane F	PK	<i>Cylindrospermum licheniforme</i>	143
Nocuolin A	others	<i>Nodularia</i> sp. HBU26	147
Hapalosin A	PK/NRP	<i>Hapalosiphon welwitschii</i>	137
Chlorosphaerolactylate A	PK/NRP	<i>Sphaerospermopsis</i> sp. LEGE 00249	148
Microcarbonin	Others (dialkylresorcinol)	<i>Microcystis</i> sp. MG-K	146
Jamaicamide B	PK/NRP	<i>Lyngbya majuscula</i> JHB	146
Grenadadiene	Others	<i>Lyngbya majuscula</i>	149
Bartoloside A	Others (dialkylresorcinol)	<i>Nodosilinea</i> sp. LEGE 06102	139
Aranazole C	PK/NRP	<i>Fischerella</i> sp. PCC 9339	150

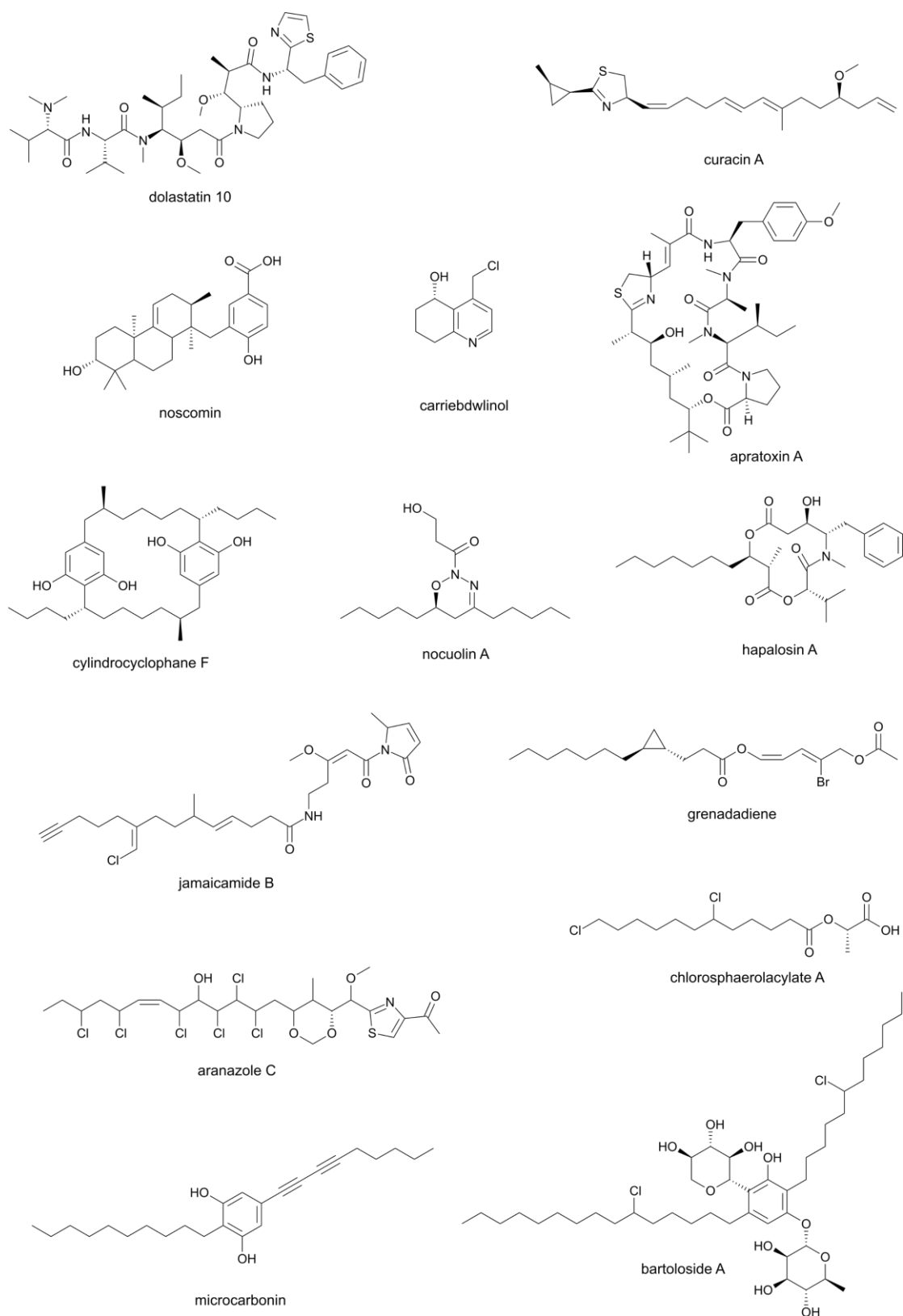


Figure A.8 - Chemical structures of cyanobacterial NPs including terpenes, PKs-NRPs, alkaloid, PKs, NRPs and fatty-acid rich compounds.

2.3. Cyanobacterial Biosynthetic enzymes

Several remarkable biosynthetic enzymes have been described from cyanobacteria in recent years. Among these, one can highlight CylK that performs an unprecedented alkylation reaction with biotechnological potential (Fig. A.9)^{78,155} CylK catalyzes a C-C bond formation via Friedel-Crafts alkylation with high promiscuity.^{78,155} Cyanobacteria are also a rich source of diverse halogenases.⁴⁰ For example, the nonheme iron/2OG-dependent halogenase WelO5 and homologues are able to perform aliphatic halogenation in site-selectively manner of fischerindole, ambiguine, and hapalindole alkaloids (Fig. A.9)¹⁵⁶. Another remarkable cyanobacterial halogenase is CylC, the only dimetal-carboxylate halogenase characterized to date (Fig. A.9).¹⁵⁷

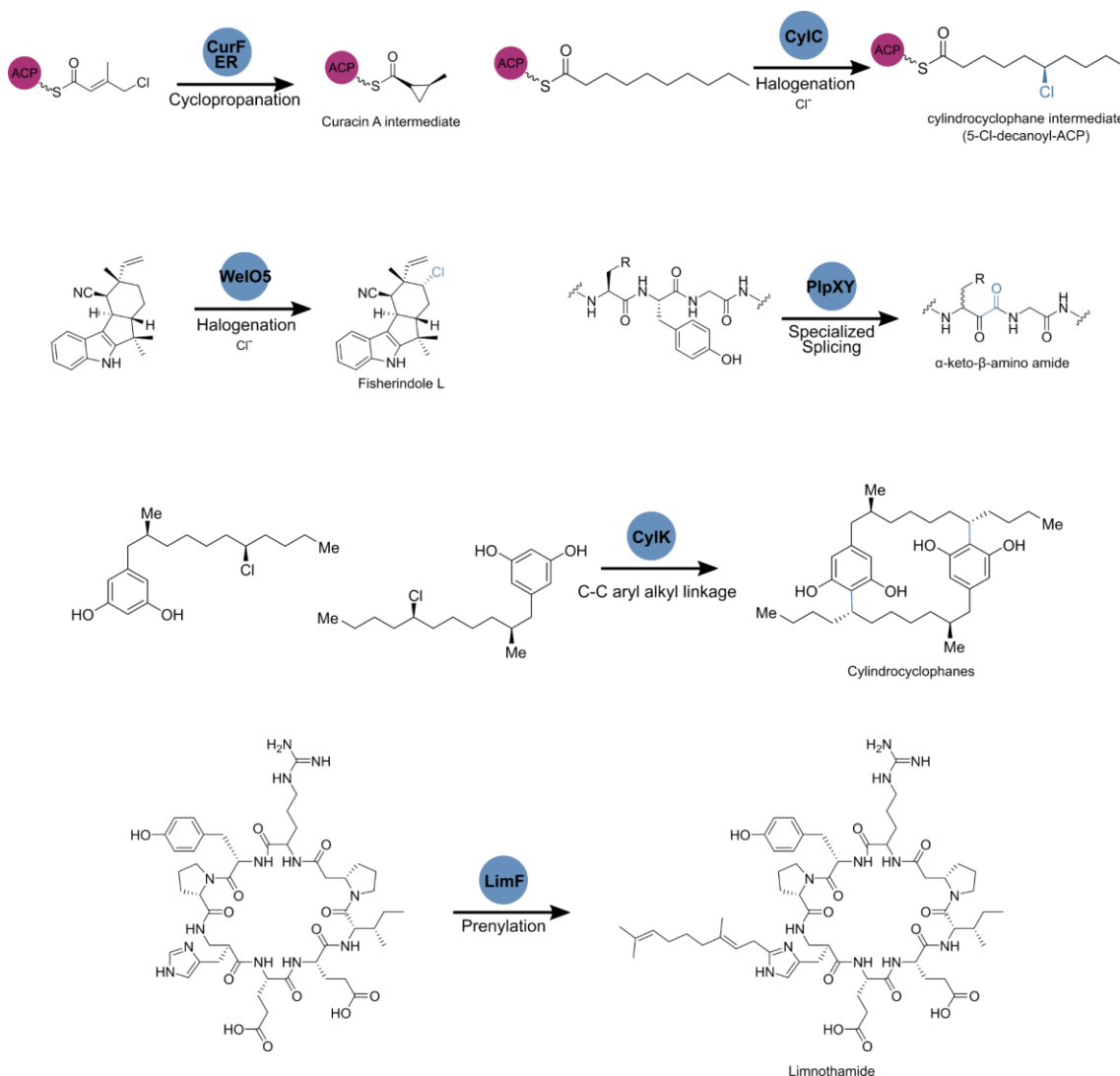


Figure A.9 – Examples of remarkable cyanobacterial biosynthetic enzymes and their unusual C-H activation, C-C formation, specialized splicing, halogenation and cyclopropanation, and prenylation including CurF ER, CylC, WelO5, PlpXY, CylK and LimF.

This enzyme performs aliphatic chlorination of FA-derived sp^3 carbons of the cylindrocyclophanes monomers with regio- and stereoselectively.⁷⁸ Other examples of

unusual cyanobacterial enzymes include the new family of radical SAM PlpXY,¹⁵⁸ CurF ER domain and the prenyltransferase LimF (Fig. A.9).¹⁵⁹ PlpXY catalyzes an unprecedented splicing reaction which leads to the loss of one tyramine equivalent from the peptide backbone, introducing diverse and multiple α -keto- β -amino acids into proteins.¹⁵⁸ On the other hand, CurF ER domain catalyzes an unprecedented cyclopropanation reaction on a chlorinated substrate, which represents a unique strategy to generate thioester enolates and subsequent ring formation.¹⁵⁹ Finally, LimF catalyzes an unprecedented His-C-geranylation in a wide diversity of cyclic and acyclic imidazole-containing peptides, showing promising potential as a versatile biocatalyst.¹⁶⁰

3. Nocuolin A and related metabolites: Chemistry, Bioactivity and Biosynthesis

3.1. Nocuolin A

The hydrazones class comprises compounds featuring the azomethine group ($C=N-N$).¹⁶¹ Natural hydrazones *per se* are a very restrict class of compounds,¹⁶¹ including the yellow pigment katorazone from *Streptomyces* sp. IFM 11299 and the mushroom toxin gyromitrin (Fig. A.10a). Hydrazones also include the even rarer oxadiazoles and oxadiazines, heterocycles containing two nitrogen and one oxygen atom as part of five and six membered rings, respectively. One of the few examples of natural oxadizoles includes the phidianidines from the opisthobranch mollusk *Phidiana militaris*, featuring a 1,2,4-oxadiazole (Fig. A.10a).¹⁶² However, most of the oxadiazoles known to date have been obtained by organic synthesis and not from natural sources.¹⁶³ Hydrazones and particularly 1,3,4-oxadiazines are known for their bioactive potential as antimicrobial, anticancer, anti-inflammatory, antimalarial, analgesic, antifungal, antiviral, anticonvulsant molecules,¹⁶⁴ representing a considerable platform for drug development.

Among these rare natural oxadiazines is the cyanobacterial nocuolin A (**1**) (Fig. A.10b), isolated by Hrouzek's group from *Nostoc* sp. CCAP1453/38 based on its cancer cell line cytotoxicity.¹⁴⁷ Simultaneously, the same compound and putative BGC (Fig. A.10c, A.10d) were found in *Nodularia* sp. LEGE 06071 by our group at CIIMAR (unpublished data). The unique structure of nocuolin A features a 1,2,3-oxadiazine moiety, unprecedented among reported NPs. In fact, the closest structural resemblance to the nocuolin A scaffold is found in geraldin D (Fig. A.10a), isolated from *Streptomyces* sp. LMA-545.¹⁶⁵ In nocuolin A, the 1,2,3-oxadiazine moiety is decorated with two alkyl moieties, forming a C_{13} aliphatic chain, and by a 3-hydroxypropanoyl moiety connected to N-2. Additionally, until nocuolin A discovery, oxadiazines were exclusively known from those generated through organic synthesis.¹⁶³ Despite this, still no organic synthesis methodology is available to date to produce 1,2,3-oxadiazines.¹⁶³

Like other oxadiazoles and oxadiazines compounds,^{162,166} nocuolin A exhibits antitumor activity on 2D and 3D colon cancer models caused by impairment of the mitochondrial oxidative phosphorylation.¹⁶⁷ IC₅₀ values for **1** in cancer cell line cytotoxicity assays ranged from the high nanomolar to low micromolar.^{147,167}

The unique structure of nocuolin A is created, in all probability, by uncommon biosynthetic logic or at least by a unique combination of known biochemistry. For instance, it is unclear whether the alkyl moiety derives from a single FA or if it is the product of C-C bond formation. Furthermore, the biochemical basis of the N-N bond is also elusive, like in most N-N bond containing NPs known to date^{82,168-171} due to the limited number of characterized N-N bond-forming enzymes. In fact, only in recent years progress has been made in the field and currently, only a total of six N-N bond forming enzymes have been characterized: SznF,^{82,169} KtzT,¹⁷⁰ CreM,⁸⁴ AzpL,¹⁷² AvaA6¹⁷¹ and PyrN.¹⁷³

The putative BGC for nocuolin A, *noc*, found initially by Hrouzek's group comprises a total of 19-22 genes encoding proteins from NocA to NocT, whose predicted annotation was revised by Martins et al.,¹⁷⁴ in a total of approximately 40 kb (Fig. A.10c), being also found in other cyanobacteria like *Anabaena* sp. PCC 7108, *Nodularia* sp. HBU26 and *Trichormus* sp. NMC-1.¹⁴⁷ The BGC is bidirectional, harboring genes *nocA* to *nocK* in one direction, and genes *nocL* to *nocT* in the opposite. The BGC was shown to contain genes coding for fatty acid incorporation and modification, including fatty acyl AMP ligases – FAALs – (*nocH*, *nocL*), acyl carrier proteins (*nocH*, *nocM*), ketosynthase (*nocG*), acyl-CoA dehydrogenase (*nocC*) and different N-processing/incorporating enzymes like aminotransferase (*nocF*) and monooxygenase (*nocE*) (Fig. A.10c, A.10d).¹⁴⁷ In addition, some of these N-related functionalities are associated with enzymes implicated in the biosynthesis of the diazo-group containing cremeomycin,¹⁷⁵ which further suggests that this could be in fact the nocuolin A BGC. However, no biochemical evidence was provided to unequivocally prove this connection. The candidate BGC is predicted to also encode uncommon functionalities such a Rieske oxygenase (NocJ), a hypothetical protein with no associated function (NocK) and other biosynthetic enzymes.

Because of the unique chemical structural, the study of nocuolin A biosynthesis is highly intriguing and represents an opportunity for uncovering unusual biochemistry and enzymatic function which can impact the future of biocatalysis.

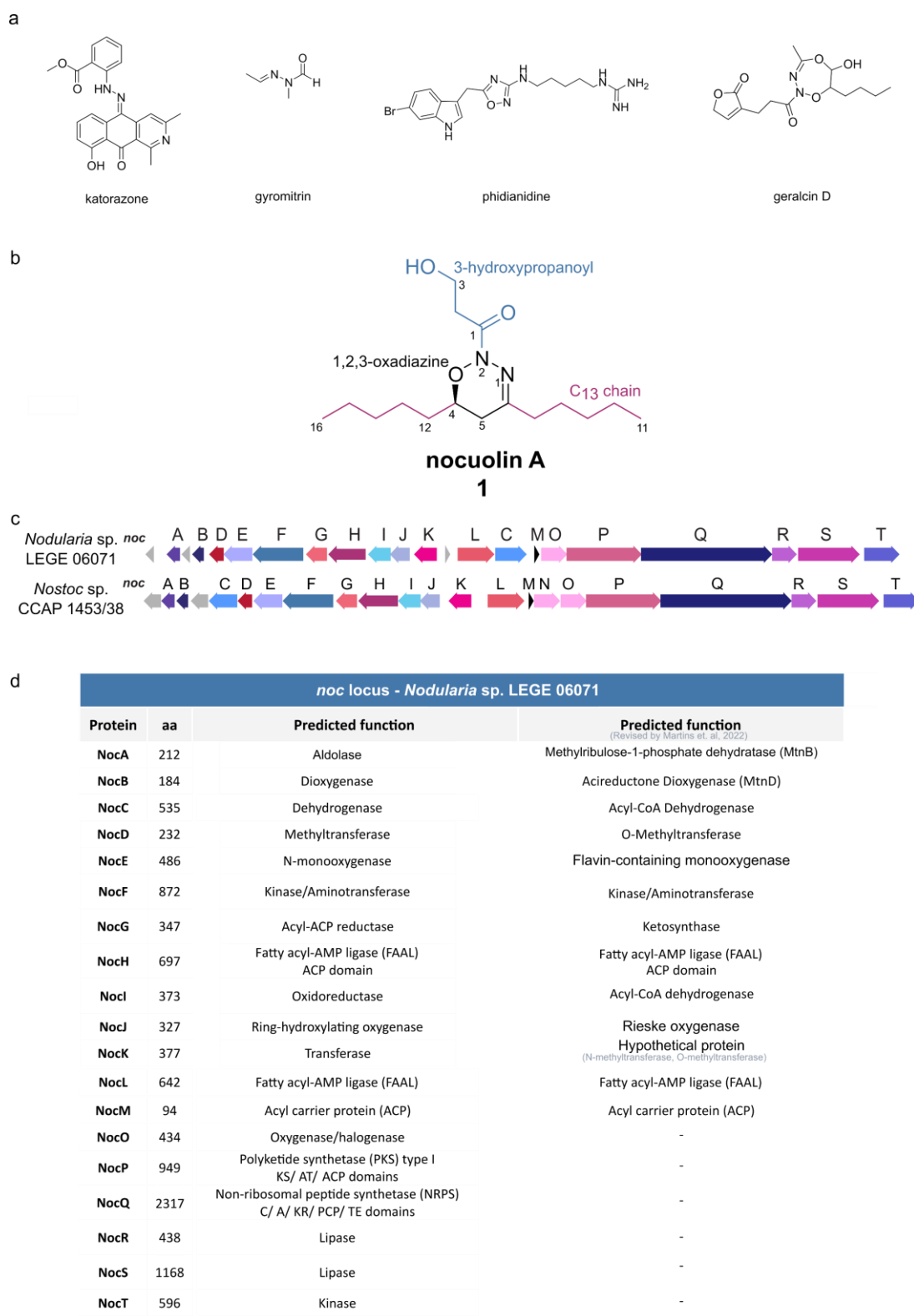


Figure A.10 - Nocuolin A is a 1,2,3-oxadiazine cyanobacterial NP encoded by the putative *noc* BGC. (a) Chemical structures of other hydrazones NPs. (b) chemical structure of nocuolin A highlighting the 1,2,3-oxadiazine (black), the 3- hydroxypropanoyl (blue) and C₁₃ chain (magenta). (c) Putative *noc* BGC in *Nostoc* sp. CCAP 1453/58 proposed by Voráčková et al.,¹⁴⁷ and the same cluster from *Nodularia* sp. LEGE 06071. (d) Detailed annotation of *Nodularia* sp. LEGE 06071 the *noc* BGC based on Voráčková et al.¹⁴⁷ and Martins et al.¹⁷⁴

3.1.1. Enzyme classes encoded in the *noc* BGC

3.1.1.1. Rieske oxygenases

Rieske oxygenases or Rieske non-heme iron oxygenases, are powerful oxidative enzymes with iron–sulfur clusters, capable of performing a wide diversity of challenging and selective reactions in the biodegradation of aromatic hydrocarbons and biosynthesis of NPs.^{30,176} The mechanism of reaction of these enzymes relies on the electron transfer from nicotinamide adenine dinucleotide phosphate (NADPH) to molecular oxygen (O_2). Depending on the enzyme, this can involve two or three separate proteins: a reductase, a ferredoxin and an oxygenase.¹⁷⁷ The reductase binds and oxidizes NAD(P)H to $NADP^+$, capturing two electrons which are shuttled by the ferredoxin to the non-heme iron center in the catalytic site of the Rieske oxygenase, generating a ferrous ion that activates O_2 , leading to substrate oxidation.¹⁷⁸

In NP biosynthesis, Rieske oxygenases are involved in an extensive catalytic repertoire of reactions (Fig. A.11), including oxidative carbocyclization like in prodiginine and marineosins biosynthesis, N-Oxidation (pyrrolnitrin and aurachin biosynthesis), C-Hydroxylation (promysalin and viomycin biosynthesis) and C–C bond desaturation (ambruticin and mupirocin biosynthesis).³⁰ This diversity precludes the reaction generalization of these enzymes and the high confidence prediction of its enzymatic activity in biosynthetic pathways.

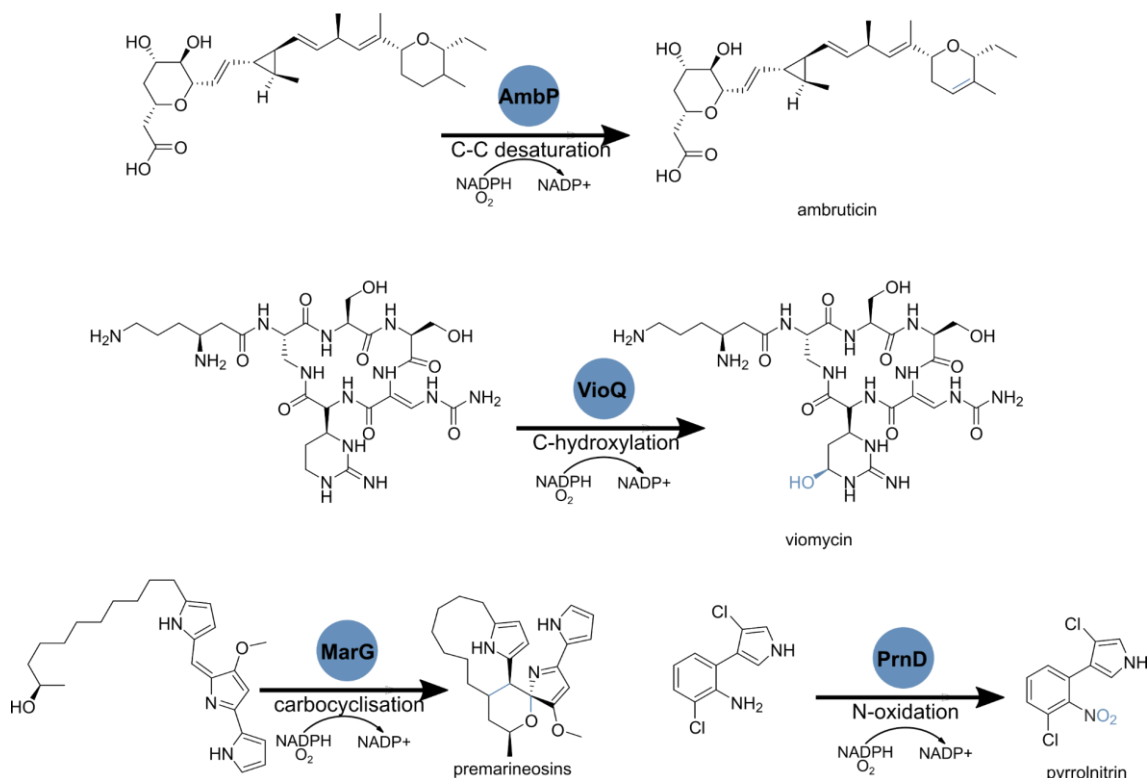


Figure A.11 – Rieske oxygenases with diverse reactivities: AmbP, VioQ, MarG and PrnD perform C-C desaturation, C-hydroxylation, carbocyclization and N-oxidation, respectively.

Rieske oxygenases selectivity, catalytic efficiency and reactivity are of main importance in the field of biocatalysis, but mechanistic studies of these enzymes have been limited by the oxygen-sensitive nature of the iron-sulfur clusters which precludes most of the *in vitro* enzymatic assays.¹⁷⁶

3.1.1.2. Acyl-CoA dehydrogenases

Acyl-CoA dehydrogenases (ACADs) are a large family of oxidoreductase flavoenzymes that desaturate fatty acid acyl-CoA conjugates.¹⁷⁹ These enzymes use flavin adenine dinucleotide (FAD) as a cofactor and play key roles in amino acid and fatty-acid metabolism, particularly in the β -oxidation cycle, catalyzing the first oxidative step of the cycle.¹⁸⁰ Most ACADs perform α,β -dehydrogenation (Fig. A.12a), which is initiated by α C-H deprotonation by a catalytic base, usually glutamate, while the β C-H hydrogen is transferred as a hydride to FAD, forming FADH₂.¹⁸¹ Based on their substrate specificity, most ACADs can be classified as short-chain acyl-CoA dehydrogenases (SCADs), medium-chain acyl-CoA dehydrogenases (MCADs), long-chain acyl-CoA dehydrogenases (LCADs), very long-chain acyl-CoA dehydrogenases (VLCADs), or isovaleryl-CoA dehydrogenases (IVDs), isobutyryl-CoA dehydrogenase (IBDs), short/branched-chain acyl-CoA dehydrogenases (SBCADs), and glutaryl-CoA dehydrogenases (GCADs).¹⁸² SCADs, MCADs, LCADs, VLCADs and other enzymes like ACAD8 and ACAD9 are involved in the β -oxidation while IVDs, IBDs, SBCADs, and GCADs are involved in leucine, valine, isoleucine, and lysine degradation.¹⁸²

In NPs biosynthesis, ACADs are responsible for the introduction of uncommon features like γ,δ -unsaturations in the biosynthesis of haliangicin and tacrolimus FK506^{183,184} or cyclopropane moieties like in the biosynthesis of kutznerides (Fig. A.12b)

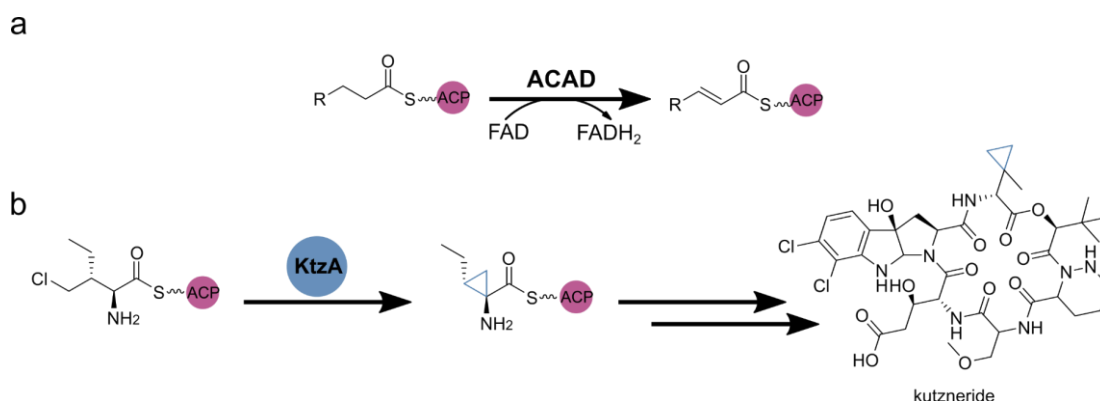


Figure A.12 – Acyl-CoA dehydrogenase enzymatic activity. (a) Schematic representation of the generic ACAD activity (for schematic simplification, thioesters are represented as ACP thioesters, but ACADs can be active with either ACP or CoA thioesters). (b) Schematic representation of KtzA role in the biosynthesis of kutznerides.

3.1.1.3. Fatty acyl-AMP ligases

Fatty acyl-AMP ligases (FAALs) are members of the acyl-adenylating enzyme superfamily and catalyze a two-step catalytic reaction: (1) formation of acyl-adenylate (acyl-AMP) intermediates from fatty-acids and adenosine triphosphate (ATP)^{186,187} (2) loading of the acyl-adenylates onto the Ppant arm of a downstream ACP, forming acyl-ACP thioesters (Fig. A.13a). The selected ACP, belonging to a specific secondary metabolism pathway, like PKS, NRPS and others, ensure the further modification of the thioesters into complex NPs. Typically, they activate single fatty-acid substrates, which are often elongated by ketosynthases, as part of polyketide synthases assembly lines.^{188,189} FAALs closely resemble fatty acyl-CoA ligases (FACLs). However, FACLs use the activated fatty-acids to form acyl-CoA thioesters instead of acyl-ACP thioesters.

FAALs are widely spread in the tree of life,¹⁹⁰ being quite abundant in cyanobacteria, with an average of one homologue per genome¹⁴⁵ and can be found as single enzymes, as part of FAAL-ACP complexes or even as part of multidomain enzymes.¹⁹¹ They are involved in the biosynthesis of several cyanobacterial NPs, including desmamides,¹⁹² microginins,¹⁹³ jamaicamides,¹⁹⁴ hectochlorines¹⁹⁵ and hapalosins (Fig. A.13b).¹⁹⁶

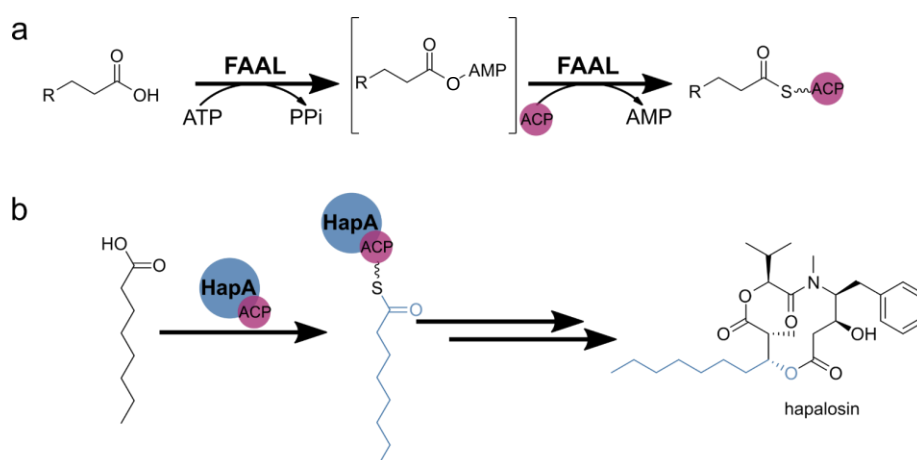


Figure A.13 – Fatty acyl-AMP ligases enzymatic activity. (a) Schematic representation of the generic FAAL activity. (b) Schematic representation of HapA (FAAL+ACP complex) role in the biosynthesis of hapalosins.

3.1.1.4. Ketosynthases

Ketosynthases (KSs) catalyze C-C bond formation via Claisen condensation as part of fatty acid synthases, polyketide synthases or as stand-alone enzymes.¹⁹⁷ Their reaction mechanism is based on 2 enzymatic steps: the transacylation and condensation steps. In the transacylation step, the acyl-ACP substrate is transferred onto the cysteine residue in the active site.¹⁹⁸ In the condensation step, malonyl-thioesters or similar are

decarboxylated, generating a carbanion tautomer that condenses with the previous acyl thioester, originating a β -ketoacyl-ACP (Fig. A.14a).¹⁹⁹ Most KSs can be classified as elongation KSs, performing the elongation of acyl-ACPs with malonylthioesters.¹⁹⁷ However, several new classes of KSs with acylation activity have been reported in recent years. This include examples like DarB, which condenses two fatty acid derived precursors to form 2,5-dialkylcyclohexane-1,3-diones (Fig. A.14b),²⁰⁰ XclC that performs the acetyl-CoA O-acylation of xenocycloin B to generate xenocycloin D²⁰¹ or OleA that condenses two long-chain acyl-thioesters to generate a long-chain β -ketoacid.²⁰²

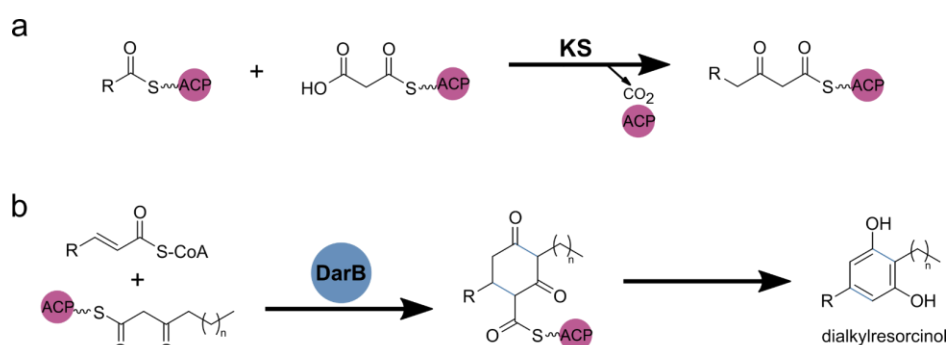


Figure A.14 – Ketosynthases enzymatic activity. (a) Schematic representation of the generic KS activity (elongation KS). (b) Schematic representation of DarB role (acylation KS) in the biosynthesis of dialkylresorcinols.

3.1.1.5. Aminotransferases

Aminotransferases (AmT) catalyze the reversible transfer of an amino group from an amino donor to a keto acid, ketone or aldehyde, generating an amide (Fig. A.15a).²⁰³ Based on the amino donor group, AT can be classified as amino acid AmTs (ATTs) or amine AmTs (ATAs).²⁰⁴

These enzymes usually form homodimers and are pyridoxal 5'-phosphate (PLP)-dependent enzymes, presenting high enantioselectivity and substrate specificity.²⁰⁴ They are ubiquitous to all life forms and play key roles in nitrogen metabolism, either in primary or secondary metabolism.²⁰⁵ The transamination reaction is performed in two enzymatic steps: (1) deamination of an amino acid/amine and release of the amino donor, (2) amination of a keto acid/ketone/aldehyde by the amino donor.²⁰⁶ In detail, the PLP binds covalently to the catalytic lysine, forming an internal aldimine that acts as a transient keeper of the amino group.²⁰⁶ By the end of the reaction, PLP is regenerated, enabling a new catalytic cycle.²⁰⁶ Additionally, due to their stereoselectivity,²⁰⁷ AmTs have been applied to access a wide range of pharmacologically relevant compounds like those

previously mentioned and others like naphterpin, furaquinocin D and harzianic acid (Fig. A.15b).^{208,209}

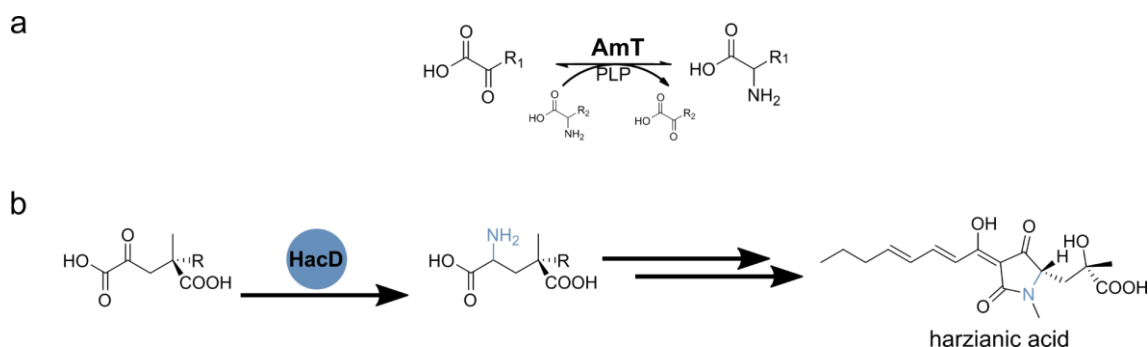


Figure A.15 – Aminotransferase enzymatic activity. (a) Schematic representation of the generic aminotransferase activity using amino acids as amino donors. (b) Schematic representation of HacD in the biosynthesis of harzianic acid.

3.1.1.6. Flavin containing monooxygenases

Flavin-containing monooxygenases (FMOs) are part of the flavoprotein monooxygenases (FPMOs) which also includes the *N*-hydroxylating monooxygenases (hydroxylate primary amines), and Baeyer–Villiger monooxygenases (ketones/aldehyde oxidation to ester or lactone).²¹⁰⁻²¹³ FMOs typically catalyze the monooxygenation of carbon-bound nucleophilic heteroatoms like nitrogen, phosphorus or sulfur (Fig. A.16a). They catalyze a wide variety of oxygenation reactions in several cellular processes like xenobiotics detoxification²¹⁴, primary metabolic processes like glucose homeostasis²¹⁵ and biosynthesis of NPs.²¹¹ The role of FMOs in detoxification processes is particularly important in eukaryotic systems allowing the conversion of lipophilic xenobiotic into more polar metabolites with easier excretion.²¹⁶ These enzymes use FAD as a prosthetic group and NADPH as a hydride donor to catalyze the binding of O₂ to a wide diversity of nucleophile-containing substrates.²¹⁷ In detail, NADPH binds to FAD, reducing it to FADH₂, which later binds O₂. The O-atom of the prosthetic group attacks the nucleophilic group of the substrate, which is oxygenated, releasing NADP⁺ and water.²¹⁷

FMOs have been frequently reported in NP biosynthesis, including the biosynthesis of antimicrobial lasalocid A and sorbicillactone A (Fig. A.16b).²¹¹

Because of their promiscuity,²¹⁸ FMOs are considered versatile catalysts with great biotechnological potential in the pharmacology field.²¹⁹

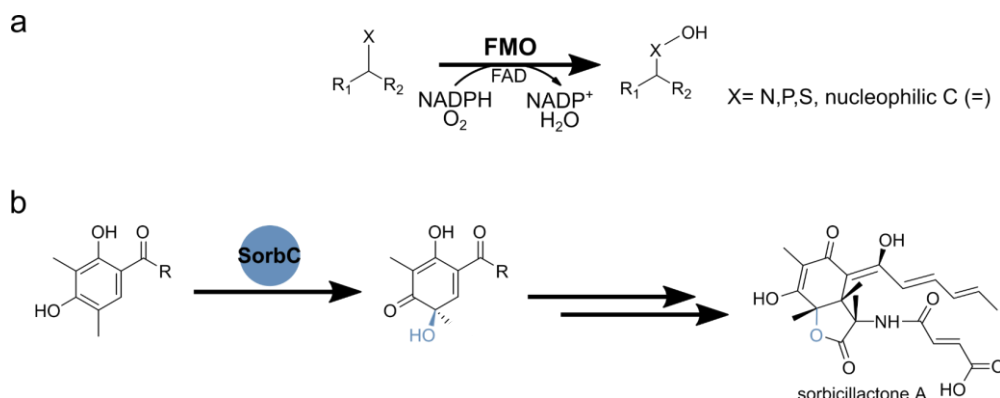


Figure A.16 – Flavin containing monooxygenases enzymatic activity. (a) Schematic representation of the generic FMO activity (X represent nitrogen, phosphate, sulfur, and nucleophilic carbons) (b) Schematic representation of SorbC role in the biosynthesis of sorbicillactone A.

3.1.1.7. Methyltransferases

Methylation is one of the most common reactions in all forms of life, playing an important part in primary and secondary metabolism. In primary metabolism, methyltransferases (MTs) can methylate sugars, proteins and nucleic acids, playing key roles in cellular processes like DNA function and cell signalling.²²⁰ In NP biosynthesis, MTs introduce important structural diversity that modifies its reactivity, bioavailability and bioactivity, with high specificity.²²¹ Depending on the methylated atom, MTs can be denominated as C-, O-, S-, and N-methyltransferases (Fig. A.17a).²²² O-MTs are the most abundant subclass of MT in NP biosynthesis, representing around 54% of all MTs, followed by N-MTs (23%), C-MTs (18%), S-MTs (3%) and others (2%).²²⁰ Most of the MTs require S-adenosylmethionine (SAM) as a co-factor.²²³ Mechanistically, the nucleophile of the substrate performs a nucleophilic attack to the sp^3 -hybridized methyl group of SAM, generating the methylated product and S-adenosyl-L-homocysteine (SAH), which is recycled through the methionine salvage pathway.²²⁴⁻²²⁶ The number of characterized MTs in secondary metabolites pathways is outstanding, including CobA in the biosynthesis of tetrapyrroles, PsmD in pyrroloindolines, XMT/MXMT/DXMT in caffeine biosynthesis (Fig. A.17b) and many more.^{221,227} Due to their chemo-, regio- and stereospecificity, MTs have been known for their potential as biocatalysts for the synthesis of pharmaceutical compounds,^{228,229} although such applications are challenged by the scalability issues of the SAM recycling systems.²²⁰

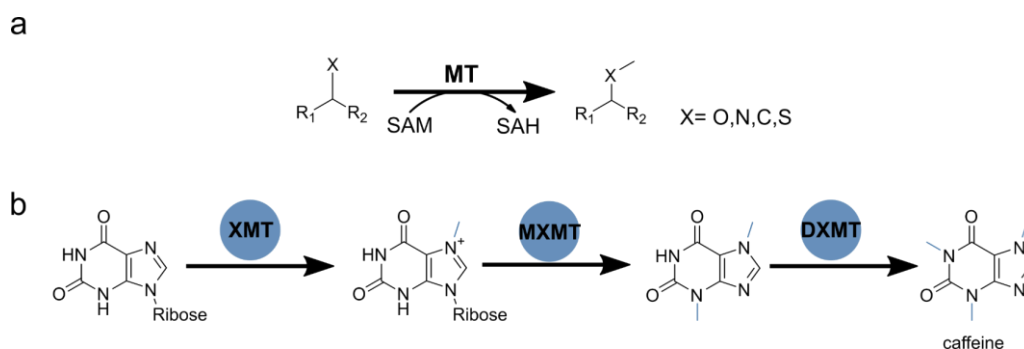


Figure A.17 – Methyltransferase enzymatic activity. (a) Schematic representation of the generic MT activity (X represent oxygen, nitrogen, sulfur, and carbon atoms) (b) Schematic representation of the several methylation steps performed by XMT (xanthosine 7-*N*-methyltransferase), MXMT (7-*N*-methylxanthine 3-*N*-methyltransferase) and DXMT (3,7-dimethylxanthine 1-*N*-methyltransferase) in the biosynthesis of caffeine.

3.1.1.8. Methylthioribulose-1-phosphate dehydratase, acireductone dioxygenase and the methionine salvage pathway enzyme

Methylthioribulose-1-phosphate dehydratase or MtnB and acireductone dioxygenase or MtnD are universal enzymes involved in the methionine salvage pathway, also known as 5'-methylthioadenosine (MTA) salvage pathway, which recycles methionine, and consequently SAM from MTA (Fig. A.18).²³⁰ In metabolic terms, the production of SAM is a costly process since it depends on methionine and ATP availability.²²¹ SAM plays a role as methyl and C₄N donor for methylation and polyamines biosynthesis, respectively, generating MTA as by-product.²³¹ To eliminate the toxicity of MTA accumulation in the cells and decrease SAM cost productions, organisms have developed the MTA salvage pathway.

In the “universal” MTA salvage pathway, the recycling to methionine is performed by the action of six enzymes: a nucleosidase/phosphorylase, an isomerase, a dehydratase, an enolase/phosphatase, a dioxygenase, and a transaminase (Fig. A.18). However, microbial MTA salvage pathways can present several divergent steps.²³²⁻²³⁴ In the “universal” pathway, MTA is first converted to 5'-methylthioribose-1-phosphate (MTR-1-P) either by MTA phosphorylase (MtnP) or by the combined action of MTA nucleosidase (MtnN) and 5'-methylthioribose kinase (MtnX).²³⁰ Subsequently, MTR-1-P is transformed to 5'-methylthioribulose 1-phosphate (MTRu-1-P) by the isomerase MtnA. MTRu-1-P is then dehydrated by MtnB (Methylthioribulose-1-phosphate dehydratase) and subsequently phosphorylated by the MtnC or by the combined action of the enolase MtnW and phosphorylase MtnX originating 1,2-dihydroxy-3-keto-5-methylthiopent(1)ene (DHK-MTPene). Finally, DHK-MTPene is oxidized by MtnD (DHK-MTPene dioxygenase/Acireductone dioxygenase) and aminated by the aminotransferase MtnE, regenerating L-methionine which can be converted to SAM, restarting the cycle.^{230,234}

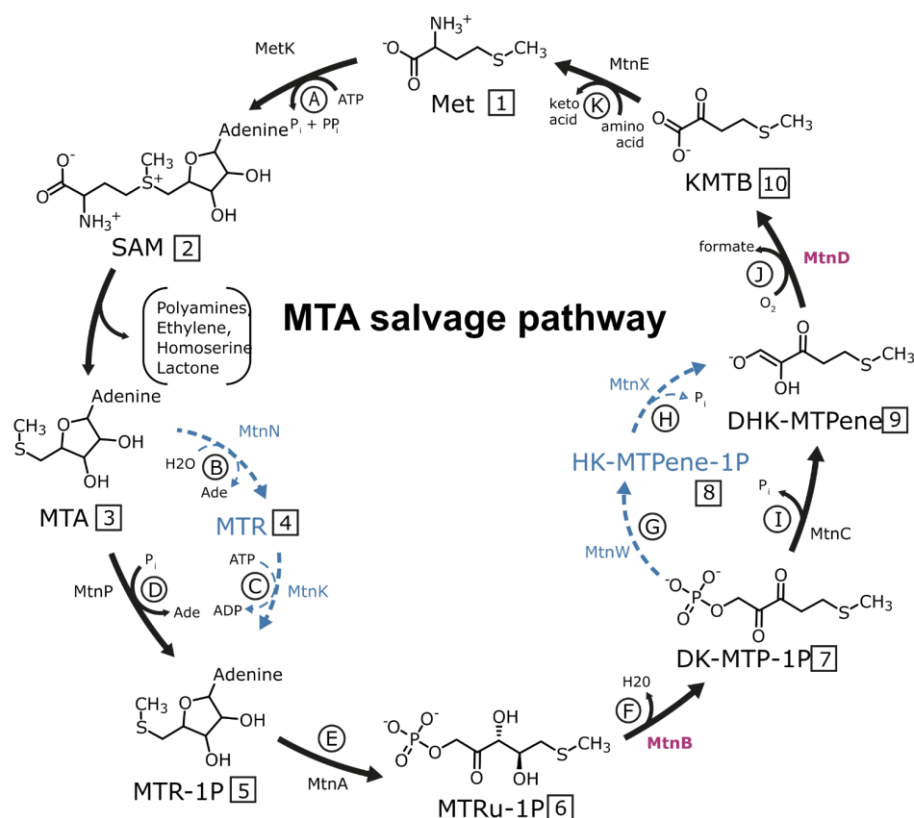


Figure A.18 – Schematic representation of the "universal" MTA salvage pathway (adapted from North et al., 2016²³⁴) highlighting some microbial variations of the pathway (in blue) and MtnB and MtnD which are homologous to NocA and NocB, respectively. (1: Met- L-Methionine; 2: SAM- S-adenosyl-L-methionine; 3: MTA - 5'-methylthioadenosine; 4: MTR - 5-methylthioribose; 5: MTR-1-P - 5'-methylthioribose-1-phosphate; 6: MTRu-1-P - 5-methylthioribulose-1-phosphate; 7: DK-MTP-1P - 2,3-diketo-5-methylthiopentyl-1-phosphate; 8: HK-MTPene-1P - 2-hydroxy-3-keto-5-methylthiopent(1)ene-1-phosphate; 9: DHK-MTPene - 1,2-dihydroxy-3-keto-5-methylthiopent(1)ene; 10: KMTB - 2-keto-4-methylthiobutyrate; A: MetK – SAM synthetase; B: MtnN – MTA nucleosidase; C: MtnK – MTR kinase; D: MtnP – MTA phosphorylase; E: MtnA – MTR-1P isomerase; F: MtnB -MTRu-1P dehydratase; G – MtnW – DK-MTP-1P enolase; H: MtnX – HK-MTPene-1P phosphorylase; I: MtnC - DK-MTP-1P enolase/phosphorylase; J: MtnD – DHK-MTPene dioxygenase (Acireductone dioxygenase), K: MtnE – KMTB transaminase)

3.2. Nocuolin A related metabolites

3.2.1. Chlorosphaerolactylates

The *noc* BGC annotation was revised when chlorosphaerolactylates (Fig. A.19a), described from *Sphaerospermopsis* sp. LEGE 00249 were connected to genes *nocL-nocT* and so, this part of the BGC was renamed as *cly* (Fig. A.19b, A.19c).²³⁵ These compounds are esters of (poly)chlorinated dodecanoic acid and L-lactic acid with weak antibacterial, antifungal, and antibiofilm properties.¹⁴⁸

With these studies, the biosynthesis of chlorosphaerolactylates was proposed to be initiated by the activation of dodecanoic acid and transference to the acyl-carrier protein

ClyB (previously NocM), by the FAAL ClyA (previously NocL). This was proposed to be followed by ClyC and ClyD single or poly-halogenation at the midchain and/or terminal positions, which differentiates the chlorosphaerolactylates A to D.²³⁵ The halogenated chain is transferred to the PKS module ClyE, which is inactive and transfers the intermediate to the PCP the NRPS module, ClyF. ClyF adenylating (A) and ketoreductase (KR) domains activate and reduce pyruvate, respectively, to form a lactyl thioester that is then condensed with the acyl moiety by the Condensation (C) domain.²³⁵ The generated dodecanoyl-lactyl-PCP thioester is hydrolyzed by the thioesterase (TE) domain of ClyF, forming the final chlorosphaerolactylates (Fig. 19d). The structures and proposed biosynthesis of chlorosphaerolactylates resembles that of the cyanobacterial lipids columbamides.²³⁶

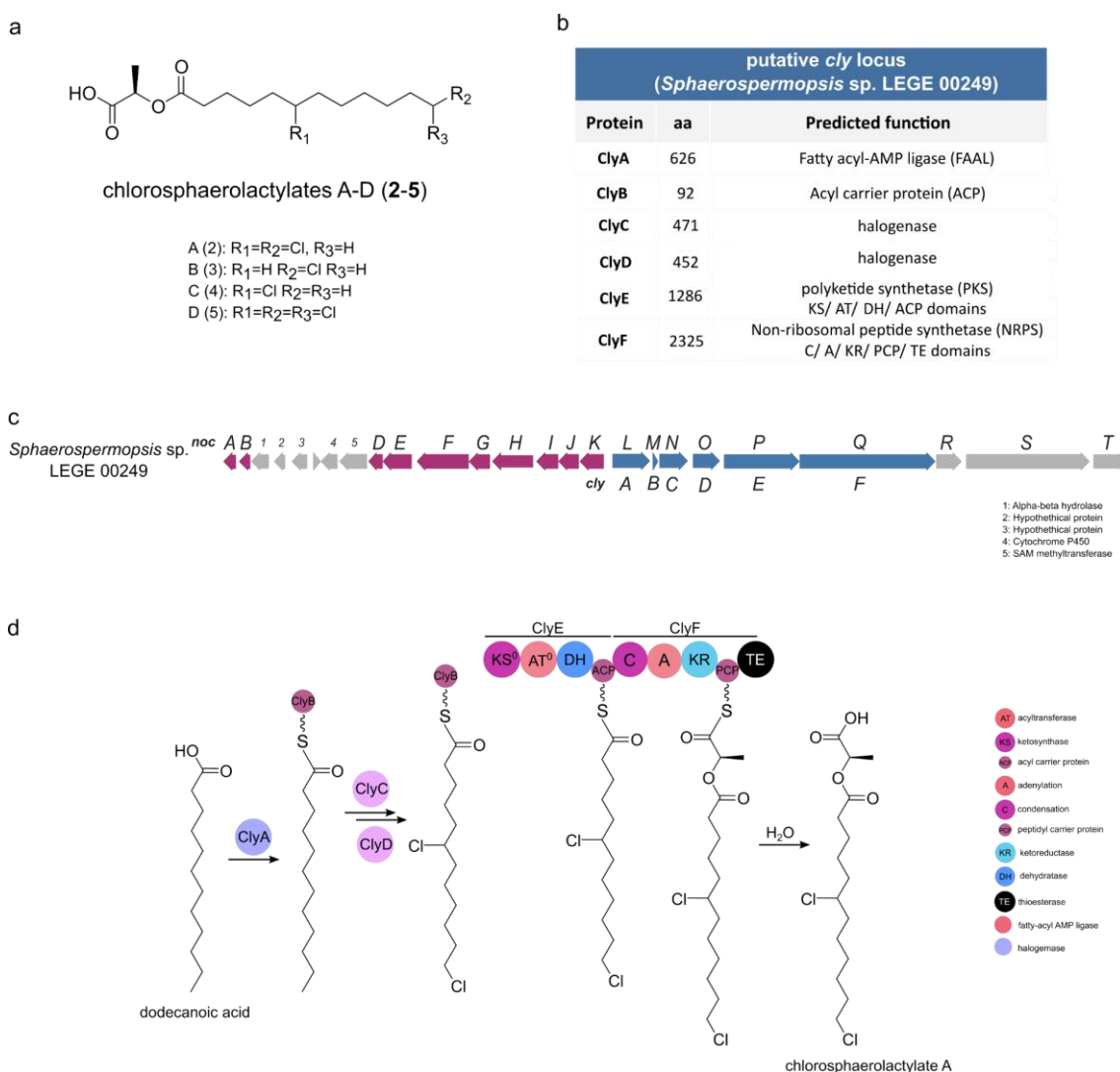


Figure A.19 – Chlorosphaerolactylates A-D are produced by the *cly* BGC in *Sphaerospermopsis* sp. LEGE 00249. (a) Chemical structures of chlorosphaerolactylates A-D. (b) Detailed annotation of the *cly* BGC from *Sphaerospermopsis* sp. LEGE 00249. (c) *noc* BGC renamed as *noc* (*nocA-nocK*) and *cly* (previous *nocL-nocT*). (d) Proposed biosynthetic pathway of Chlorosphaerolactylates (adapted from ²³⁵).

Interestingly, *cly* and *noc* BGCs were found to be co-localized in all strains identified so far, including *Sphaerospermopsis* sp. LEGE 00249 and *Nodularia* sp. LEGE06071.

3.2.2. Nocuolactylates

Recently, Figueiredo et al.¹⁴⁵ isolated and characterized a new family of compounds from *Nodularia* sp. LEGE 06071 – the nocuolactylates – which are hybrid molecules resulting from the esterification of nocuolin A and chlorosphaerolactylates (Fig. A.20a).¹⁴⁵ Hence, these compounds are likely the product of both *cly/noc* subclusters and can explain the co-localization of the two clusters. This further suggests that nocuolin A is likely encoded by genes *nocA-K*, chlorosphaerolactylates by *clyA-F* and nocuolactylates by both clusters, likely including genes *nocR*, *nocS* and *nocT*, which are present in all *noc* loci and encode two lipases and one kinase, respectively (Fig. A.20b). However, further evidence is required to connect the nocuolactylates with the *noc* BGC.

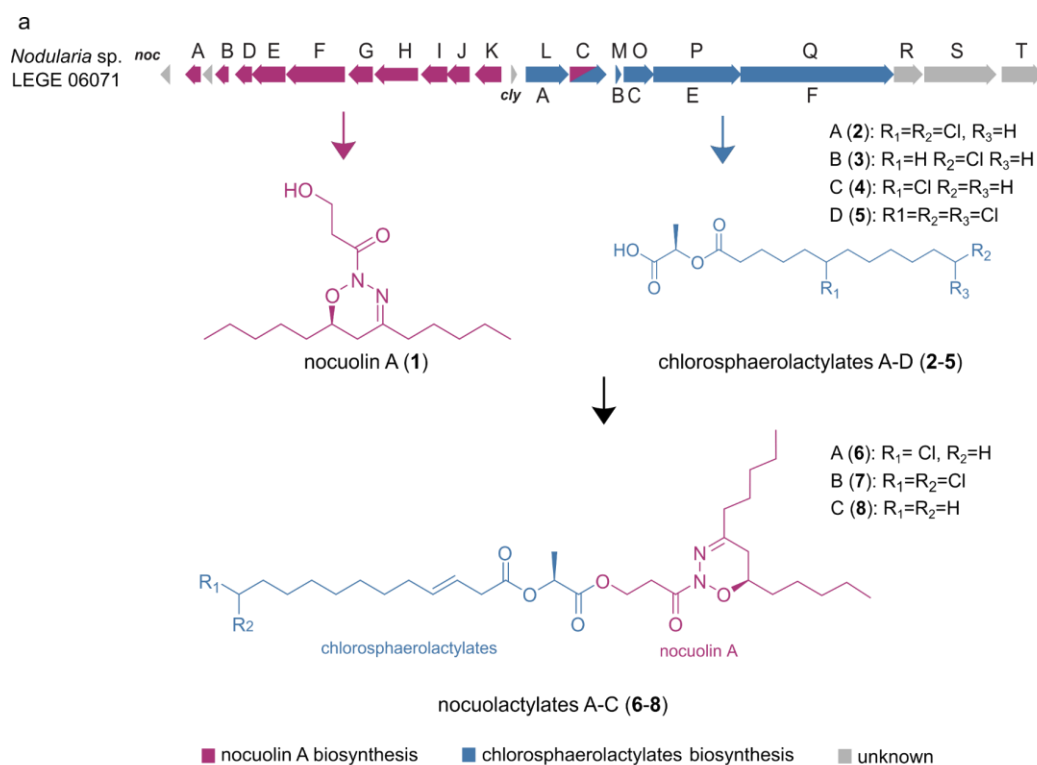
Nocuolactylates can be monohalogenated (nocuolactylate A), dihalogenated (nocuolactylate B) or non-halogenated (nocuolactylate C). These halogenations occur in the terminal carbon and are likely installed by the single halogenase encoded by the *Nodularia* sp. LEGE 06071 *cly* cluster - ClyC/NocO. This differs from the *cly* BGC in *Sphaerospermopsis* sp. LEGE 00249, which encodes two halogenases – ClyC/NocO and ClyD – and produces chlorosphaerolactylates with both midchain and terminal halogenations.

When tested for cytotoxicity against cancer cell lines in 2D and 3D models, the nocuolactylates were less potent than nocuolin A.¹⁴⁵ It is unclear if the nocuolactylates could act as a prodrug for the release of the potent nocuolin A.

4. Methodologies for biosynthetic studies in cyanobacteria

Biosynthetic studies can be performed by a combination of several methodologies like bioinformatics, precursor supplementation experiments, single/multiple gene inactivation, heterologous expression of BGCs or subsets of their genes, or *in vitro* enzymatic reconstitution (Fig. A.21).

BIOSYNTHESIS OF NOCUOLIN A AND RELATED METABOLITES

**b**

<i>noc</i> locus (<i>Nodularia</i> sp. LEGE 06071)			<i>cly</i> locus (<i>Nodularia</i> sp. LEGE 06071)		
Protein	aa	Predicted function	Protein	aa	Predicted function
NocA	212	Methylribulose-1-phosphate dehydratase (MtnB)	ClyA	642	Fatty acyl-AMP ligase (FAAL)
NocB	184	Acireductone Dioxygenase (MtnD)	ClyB	94	Acyl carrier protein (ACP)
NocD	232	O-Methyltransferase	ClyC	434	Oxygenase/halogenase
NocE	486	Flavin-containing monooxygenase	ClyE	949	Polyketide synthetase (PKS) type I KS/ AT/ ACP domains
NocF	872	Kinase/Aminotransferase	ClyF	2317	Non-ribosomal peptide synthetase (NRPS) C/ A/ KR/ PCP/ TE domains
NocG	347	Ketosynthase	other <i>noc/cly</i>? (<i>Nodularia</i> sp. LEGE 06071)		
NocH	697	Fatty acyl-AMP ligase (FAAL) ACP domain			
NocI	373	Acyl-CoA dehydrogenase			
NocJ	327	Rieske oxygenase			
NocK	377	Hypothetical protein			
			Protein	aa	Predicted function
			NocC	535	Acyl-CoA Dehydrogenase
			NocR	438	Lipase
			NocS	1168	Lipase
			NocT	596	Kinase

Figure A.20 – Nocuolactylates A-C are hybrid molecules formed by nocuolin A and chlorosphaerolactylates in *Nodularia* sp. LEGE 06071. (a) *noc* and *cly* BGCs and their relationship with the production of nocuolin A, chlorosphaerolactylates, and nocuolactylates (adapted from¹⁷⁴). (b) Detailed annotation of the *noc* and *cly* BGCs in *Nodularia* sp. LEGE 06071.

4.1. Bioinformatic tools

Bioinformatic platforms such as antiSMASH, NP.searcher, NaPDos and PRISM have facilitated the identification of BGCs in genome and metagenome data and the prediction of their encoded NPs.²³⁷⁻²⁴⁰ Additionally, remote homology detection bioinformatics tools like pBLAST, HHpred, InterPro and Swiss-Model²⁴¹ allow for protein functional prediction based on sequence homology. More recently, the Enzyme Function Initiative (EFI) has also developed tools to build sequence similarity networks (SSNs) that explores the sequence-function space in enzyme families and genome neighborhood networks (GNNs) that identifies neighbor proteins in each SSN cluster.^{242,243} Unlike the traditional phylogenetic trees, SSNs can use larger sequence sets, up to many thousands of proteins, and perform faster analysis in an easier visual network. Furthermore, bioinformatic protein analysis have benefited from the recent development of AlphaFold.²⁴⁴ This algorithm performs accurate predictions of protein structure, previously only possible with protein crystallography, facilitating research on evolution of protein structures and structure-function relations.²⁴⁴

4.2. Precursor supplementation experiments

Precursor supplementation experiments make use of putative biosynthetic precursors enriched in stable isotopes (e.g., ²H, ¹³C, ¹⁸O, ¹⁵N) that are added to the microbial culture media and consequently incorporated into the microbial metabolism,²⁴⁵ allowing the production of isotopically enriched NPs and NP intermediates. The labelling and labelling positioning of these compounds can be determined by mass spectrometry (MS)-based comparative metabolomics, which allows detection of mass shifts when compared with a non-supplemented control (Fig. A.21). Additionally, the labelling positioning can also be determined by Nuclear Magnetic Resonance (NMR). Consequently, this technique might enable the identification of putative intermediates, provide insight into reaction mechanisms, and help determine biosynthetic substrates. Radioactive isotopes like ³H, ¹⁴C and others can also be used in supplementation experiments, but require analysis by scintillation counting and very controlled working conditions.²⁴⁵

Supplementation experiments can be compromised by substrate metabolization inside the cells, causing considerable scrambling, i.e. the process of achieving an *equilibrium* distribution of isotopes within a specified set of atoms of a molecule. However, cyanobacteria seem to lack the β -oxidation pathway, with no full set of β -oxidative enzymes being found in their genomes.^{145,246} This indicates that supplementation experiments in cyanobacteria with FA can result in full-labeled metabolites without scrambling.¹⁴⁵

4.3. Gene/BGC inactivation (knockout)

The biosynthetic gene inactivation strategy includes the construction of a knockout strain by deleting one or several genes from the NP producing strain. By disrupting the production of the final target NP, this methodology enables the direct connection between the NP and the BGC where the inactivated gene(s) are located (Fig. A.21). The same strategy can lead to disruption of the pathway at different stages of the biosynthesis and facilitate the identification of pathway intermediates by comparative metabolomics between the knockout and the wild-type strains.

Although knockout generation is a widely used tool for biosynthetic studies in many genetically tractable organisms,²⁴⁷ in cyanobacteria such strategies are rare due to the scarce number of traditional genomic engineering tools like helper plasmids that encode specific methylases to protect the exogenous DNA from cyanobacterial endonucleases, particularly for non-model strains.²⁴⁸⁻²⁵⁰ Additionally, genetic manipulation in cyanobacteria is hampered by multiple number of genome copies that can go up to 225 in *Synechocystis* sp. PCC 6803²⁵¹ or by variable transformation efficiencies.^{252,253} In fact, reported knockout strains are limited to a very small number of model cyanobacterial species, most in *Synechocystis* sp. PCC 6803^{249,254}, *Synechococcus* sp. PCC 7002²⁵⁵ and *Anabaena* (*Nostoc*) sp. PCC 7120.²⁵⁶

Recent advances regarding RNA-guided genome editing system CRISPR/Cas9 technology²⁵⁷ represent an exciting new possibility to construct knockout strains in cyanobacteria. Despite the slower progress of this technique in prokaryotes compared with eukaryotic organisms, cyanobacterial genomic manipulation with CRISPR/Cas9 is a hot topic in the field, not only for knockout construction but also knock-ins and gene repression.²⁵⁸⁻²⁶¹

4.4. Heterologous expression of full and partial BGCs

Heterologous expression of an entire BGC is performed by cloning the BGC into a plasmid (self-replicative or integrative) and transformation into a suitable expression host.²⁴¹ Using comparative metabolomics with the wild-type organism, BGC/NP connection can be done (Fig. A.21). Like in gene inactivation strategies, this technique can also allow identification of pathway intermediates and be used also as a platform to generate knockout strains of target genes to dissect their biosynthetic role.

Expressing cyanobacterial BGCs in a heterologous host is also a limited task due to difficulties in cloning large BGCs, and finding a suitable host²⁵⁰ which might range from *Escherichia coli*²⁶²⁻²⁶⁴ to *Bacillus subtilis*,²⁶⁵ *Saccharomyces cerevisiae*,²⁶⁶ *Streptomyces coelicolor*²⁶⁷, *Aspergillus nidulans*²⁶⁸, *Anabaena* sp. PCC 7120²⁶⁹ and others. Several techniques have been developed to overcome the cloning issues like library-based

methods,²⁷⁰ direct cloning methods like transformation-associated recombination (TAR cloning)^{271,272} and different assembly methods.²⁷³ Assembly methods using smaller polymerase chain reaction (PCR) derived products have been particularly facilitated by the development of Gibson assembly.²⁷³ So far, over a dozen of cyanobacterial BGCs have been expressed in other cyanobacterial hosts or in *E. coli*.²⁵⁰ Recently, the direct pathway cloning (DiPaC) strategy^{263,264,274} based on Gibson assembly (or sequence and ligation-independent cloning - SLiC) , has been successfully used to clone BGCs up to 55 kb, including *sodorifen*²⁷⁴ and the cyanobacterial PKS/NRPS hybrids *hapalosins* and *microginins*.^{193,264}

4.5. *In vitro* enzymatic reconstitution

The *in vitro* enzymatic reconstitution approach is based on the heterologous expression of one or several target biosynthetic enzymes encoded by the BGC, followed by the characterization of its activity using *in vitro* enzymatic assays with predicted or surrogate substrate(s) (Fig. A.21).²⁷⁵ The detection of the product formed can be performed with different techniques including spectrophotometry, HPLC, MS, NMR and others.²⁷⁶⁻²⁷⁸

The production of the recombinant protein can be performed in several heterologous hosts, *E. coli* being one of the most common.²⁷⁹ This strategy allows the determination of the exact reaction conditions of a certain enzyme, including kinetics, necessary co-factors, substrate(s) and generated products. However, and apart from heterologous host selection, this strategy faces several challenges including possible protein toxicity to the host cell, protein insolubility caused by lack of specific codons, cofactors or chaperones, protein degradation by cell proteases, formation of inclusion bodies and others.^{279,280} Because of these difficulties, the strategy can be time consuming, but it is still a powerful approach to define gene functions and reconstitute entire biosynthetic pathways.

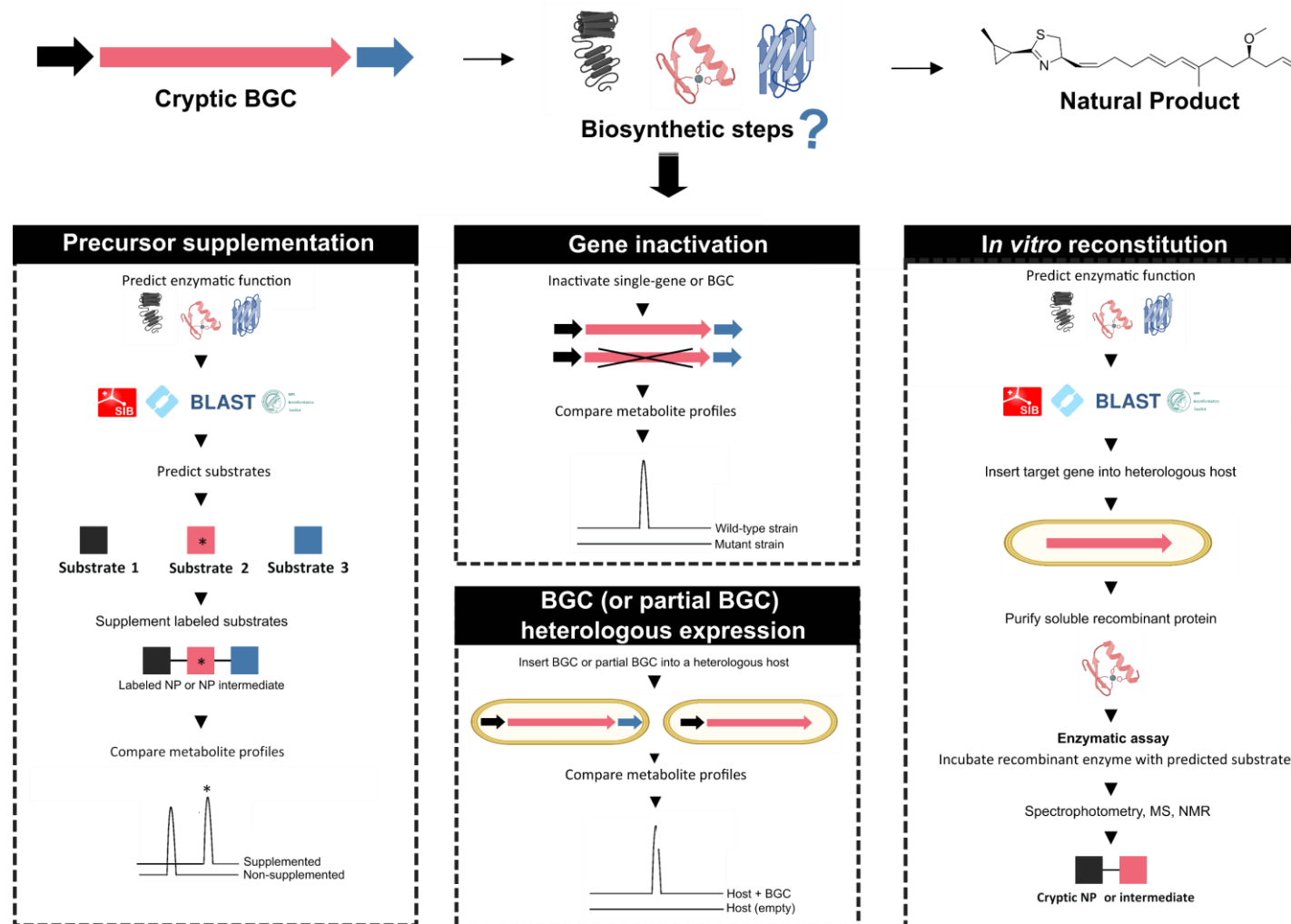


Figure A.21 - Overview of the most common techniques used in cyanobacterial biosynthetic studies, including precursor supplementation, Gene inactivation, BGC (and partial BGC) heterologous expression and *in vitro* enzymatic reconstitution.

Chapter I

Published in ACS ChemBio

<https://doi.org/10.1021/acscchembio.2c00464>.

Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A

Martins, T. P.; Glasser N. R.; Kountz, D. J.; Oliveira, P. N.; Baskus, E. P.; Leão, P. N. (2022). Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A. ACS Chemical Biology 17(9): 2528-2537 <https://doi.org/10.1021/acscchembio.2c00464>.

Biosynthesis of the unusual carbon skeleton of nocuolin A

Teresa P. Martins^{1,2}, Nathaniel R. Glasser³, Duncan J. Kountz³, Paulo Oliveira^{4,5,6}, Emily P. Balskus^{3*}, Pedro N. Leão^{1,*}

¹CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, 4450-208 Matosinhos, Portugal

²ICBAS – School of Medicine and Biomedical Sciences, University of Porto, 4050-313, Porto, Portugal

³Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA

⁴i3S – Institute for Research and Innovation in Health, University of Porto, 4200-135 Porto, Portugal

⁵IBMC – Institute of Molecular and Cell Biology, University of Porto, 4200-135 Porto, Portugal

⁶Department of Biology, Faculty of Sciences, University of Porto, 4169-00 Porto, Portugal

*corresponding authors: Emily P. Balskus (balskus@chemistry.harvard.edu); Pedro N. Leão (pleao@ciimar.up.pt).

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 an acyl carrier protein domain 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. I.1a), initially isolated by Voráčová 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} IC₅₀ values for **1** in cancer cell line cytotoxicity assays ranged from the high nanomolar to low micromolar.² 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 natural product with close structural resemblance to the nocuolin A scaffold is geraldin D (Fig. I.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 origins of the 3-hydroxypropanoyl moiety and the oxadiazine ring are also intriguing, and so is the basis for N-N bond formation, as there are few examples of characterized N-N bond-forming enzymes^{10,11} and no homologues of such enzymes are present in the *noc* biosynthetic gene cluster (BGC).

The initial report of the discovery of **1** by Voráčová and co-authors suggested that a 40 kb locus (*noc*, Fig. I.1a) shared by the genomes of cyanobacteria producing **1**, was the likely nocuolin 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. I.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 chlorosphaerolactylates⁶ but also hybrids of these two molecules named nocuolactylates (e.g., **3**, Fig. I.1b). 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.

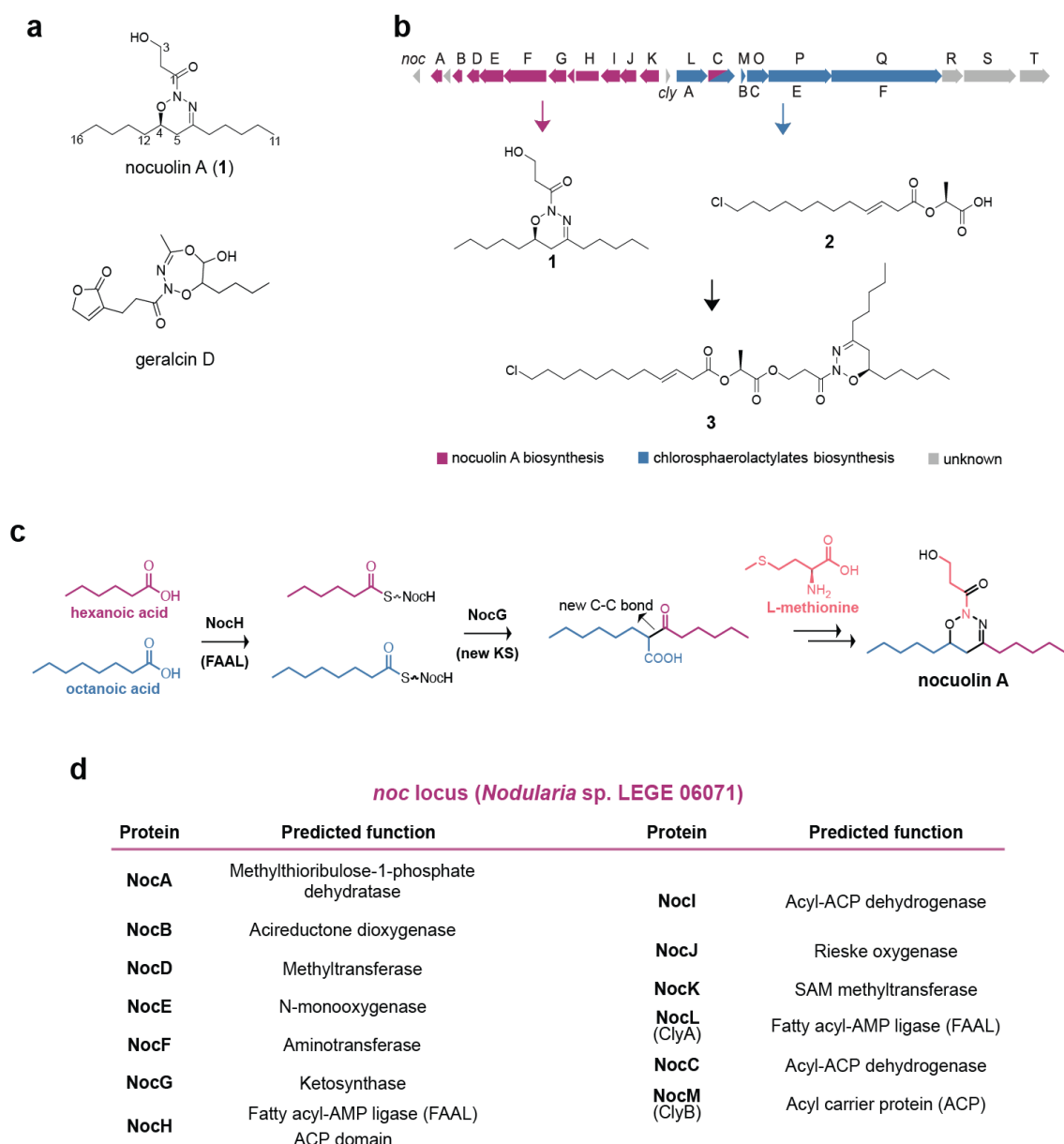


Figure I.1 - Schematic highlighting the structure of nocuolin A, the *noc* BGC and the proposed biosynthetic steps. (a) Chemical structures of nocuolin A (**1**) and Geraldine D, the natural product with the closest structural resemblance to the nocuolin A scaffold. (b) *noc* locus and its proposed relationship 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**.

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 ketosynthase (KS) NocG with high specificity. We show that NocG is part of a phylogenetically well-supported clade with no other characterized members and that the 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 C₁₃ alkyl moiety (Fig. I.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 Voráčová et al. was reviewed using the remote homologue 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. I.1d, Table SI.1). 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. SI.1), 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, Sfp¹⁵. This result confirmed the annotation of NocC and NocI as acyl-ACP dehydrogenases (Fig. SI.2, SI.3). 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. SI.4), 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) toward ClyB-bound fatty acyl

substrates (Fig. SI.2). 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 **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 FAs. 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 C₁₃ 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. I.2a). A shift of *m/z* 21.132 was also evident (Fig. I.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. SI.5, SI.6). These results indicate that the C₁₃ skeleton of **1** is formed from two FAs, implying that 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 FAs (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. I.2b, Fig SI.6). No incorporation from FAs with longer chains was observed (Fig. SI.7). We also supplemented *Nodularia* sp. LEGE 06071 with perdeuterated heptanoic acid (*d*₁₃-heptanoic acid) and found no deuterium incorporation from this substrate (Fig. SI.7). Overall, these experimental results indicate that one hexanoic and one octanoic acid unit are used to generate the C₁₃ moiety of **1**.

To pinpoint the positions of the hexanoic and octanoic acid-derived atoms in the C₁₃ 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. SI.5) 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. I.2c). With these data, we concluded that hexanoic acid is incorporated into positions C-6 to C-11 and

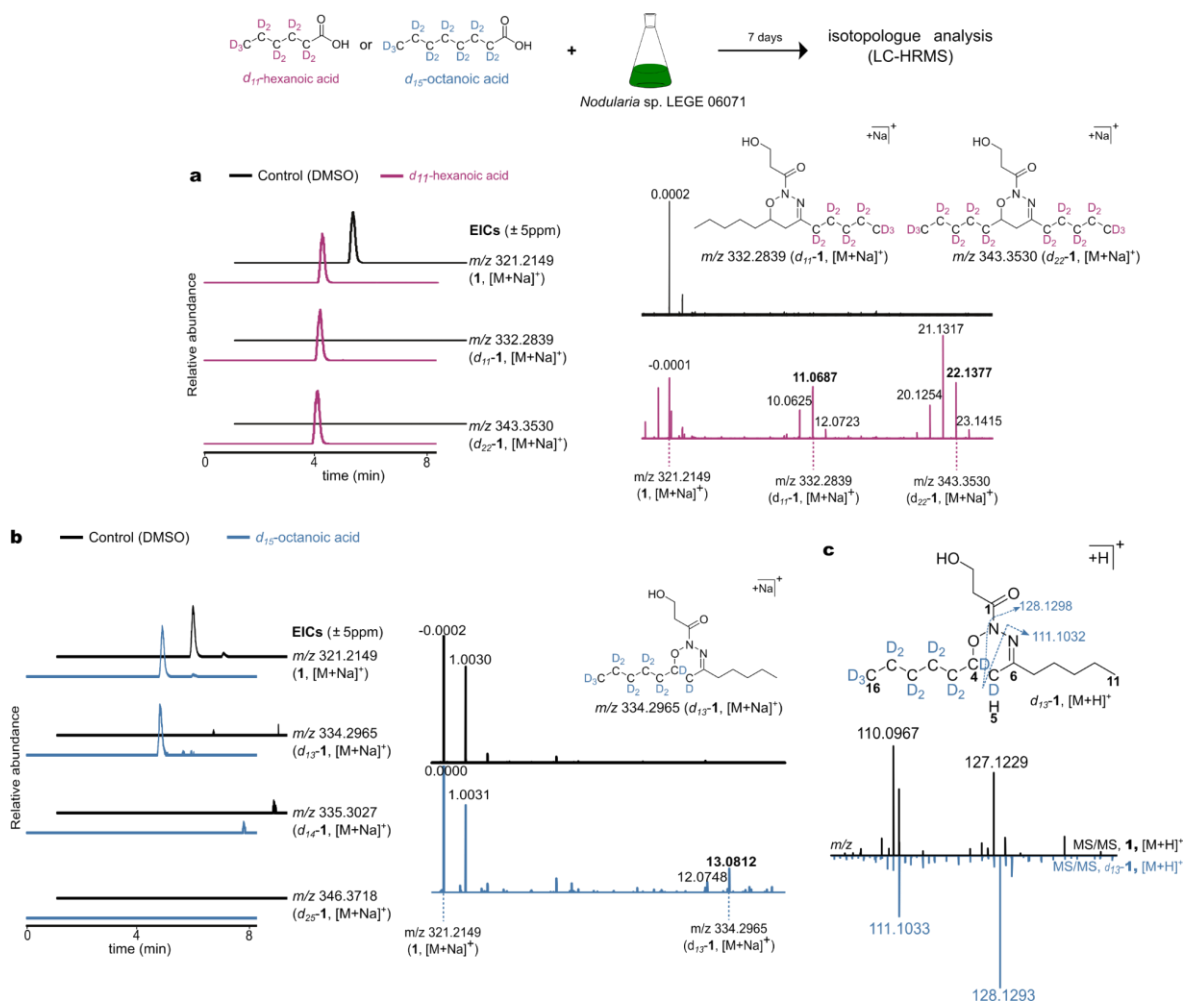


Figure I.2 - Supplementation of *Nodularia* sp. LEGE06071 with d_{11} -hexanoic acid and d_{15} -octanoic acid leads to $d_{11}1$ / $d_{22}1$ and $d_{13}1$, respectively. LC-HRMS-derived extracted ion chromatograms (EICs) of **1** and labeled version with (a) d_{11} -hexanoic acid and (b) d_{15} -octanoic acid 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.

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. SI.8, SI.9). 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 $^{13}\text{CO}_2$ recycling through the Calvin cycle) (Fig. SI.10). Overall, these supplementation experiments with stable isotope-labeled FAs 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* 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 and 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 co-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. I.3). NocH showed a clear preference for hexanoic and octanoic acids (Fig. I.3, SI.11, SI.12), 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. SI.12), further supporting the role of

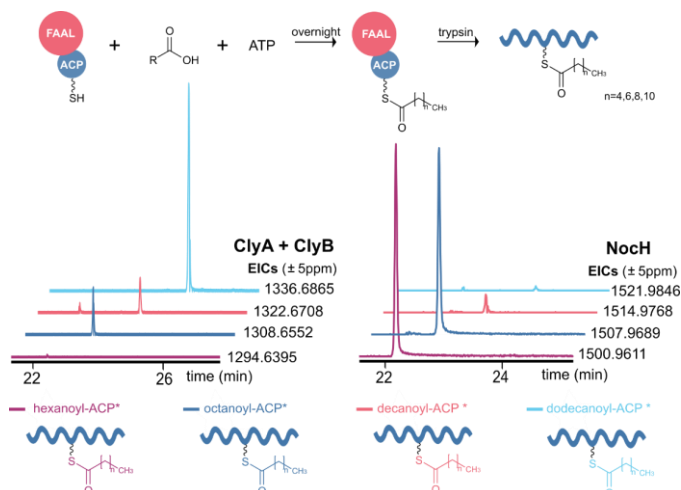


Figure I.3 - FAALs NocH and ClyA (NocM) activate hexanoic/octanoic acids and dodecanoic acid as acyl-ACP thioesters, respectively, in competition assays. Shown are LC-HRMS-derived EICs of a trypsinated ACP fragment loaded with each fatty-acid.

this enzyme in chlorosphaerolactylate biosynthesis. These experiments provided the first biochemical evidence supporting the connection between the *noc* locus and the synthesis of nocuolactylates, chlorosphaerolactylates and **1**.

NocG is a ketosynthase 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. I.1b, Table SI.1), 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. SI.1, SI.13, Table SI.2, SI.3). 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. I.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** was detected in control assays lacking NocG (Fig. I.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. I.4a). MS/MS analysis of **4** corroborated the proposed structure (Fig. SI.14). In addition, we repeated the NocG enzymatic assays with different combinations of ²H- and ¹³C-labeled FAs and observed the formation of products with the predicted *m/z* values for the corresponding labeled versions of **4** (Fig. SI.15). 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. I.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. SI.1, SI.2), 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. SI.16). C-C bond formation also did not occur with a single thioester and one free FA (Fig. SI.17a). Furthermore, in assay mixtures with hexanoyl-CoA and octanoyl-CoA, NocG failed to generate detectable

amounts of **4** (Figure SI.17b). In NocG assays with one of the FAs as a CoA thioester and the other as a NoCH thioester, compound **4** was produced but in slightly lower yield than in assays in which both FAs were activated as NoCH thioesters (Figure SI.17b). This provides further support to the hypothesis that NoCH activates both fatty FAs prior to NocG activity. 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. SI.18a) and no NoCH loading was observed for the hydroxylated FA (Fig. SI.18b).

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. I.2a). To obtain further insight into the activity of NocG, we carried out kinetic assays with octanoyl-SNAC and variable concentration of hexanoyl-SNAC and observed apparent K_m and V_{max} values of $135 \pm 29 \mu M$ (standard error, $n = 3$) and $0.224 \pm 0.016 \mu M \text{ min}^{-1}$ (standard error, $n = 3$) for the formation of **4**, respectively (Fig. SI.19).

Interestingly, despite the typical promiscuity of KSs regarding the alkyl chain length¹⁷⁻¹⁹ 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. SI.20a). Such reactivity contrasts with closely-related stand-alone KSs like OleA, which is able to accept C8 to C16 acyl-CoA substrates¹⁸ or PpyS, which can condense C8 to C14 (with different branching) thioesters to generate photopyrones A to H.^{17, 20} On the other hand, LstAB generates a C22 aliphatic skeleton with no longer or shorter homologues having been described.¹⁶ However, to our knowledge, the promiscuity of the LstAB heterodimer toward fatty-acyl substrates of different chain lengths was not tested.

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. SI.20b). 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.^{20,21} 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 KS 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 SI.4). The resulting phylogenetic tree shows that NocG clades separately from all other characterized KSs (Fig. I.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.²² PpyS (identity/similarity to NocG: 36/57%, respectively) and its closest homologues 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, unlike what is observed for NocG, this intermediate is not released and instead undergoes an intramolecular cyclization to generate the final pyrone product (Fig. I.4c).²² Previous phylogenetic studies have shown that PpyS homologues 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.²² It is within this second clade that NocG and its closest homologues 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. SI.21). 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. SI.22). Members of the SSN cluster that encompasses NocG and its closest homologues are most often associated with fatty acyl activation/tethering (NocH homologues) and amino acid transference (NocF homologues) (Fig. SI.21). Some of these clusters harbor additional homologues of the Noc enzymes, especially in *Nocardia* spp. (Fig. SI.23). On the other hand, PpyS homologues show much more diverse genomic contexts, mainly associated with sugar metabolism, and possess no obvious connection Noc-related enzymes (Fig. SI.22, Table I.1). NocG also clades separately from other reported stand-alone KSs that

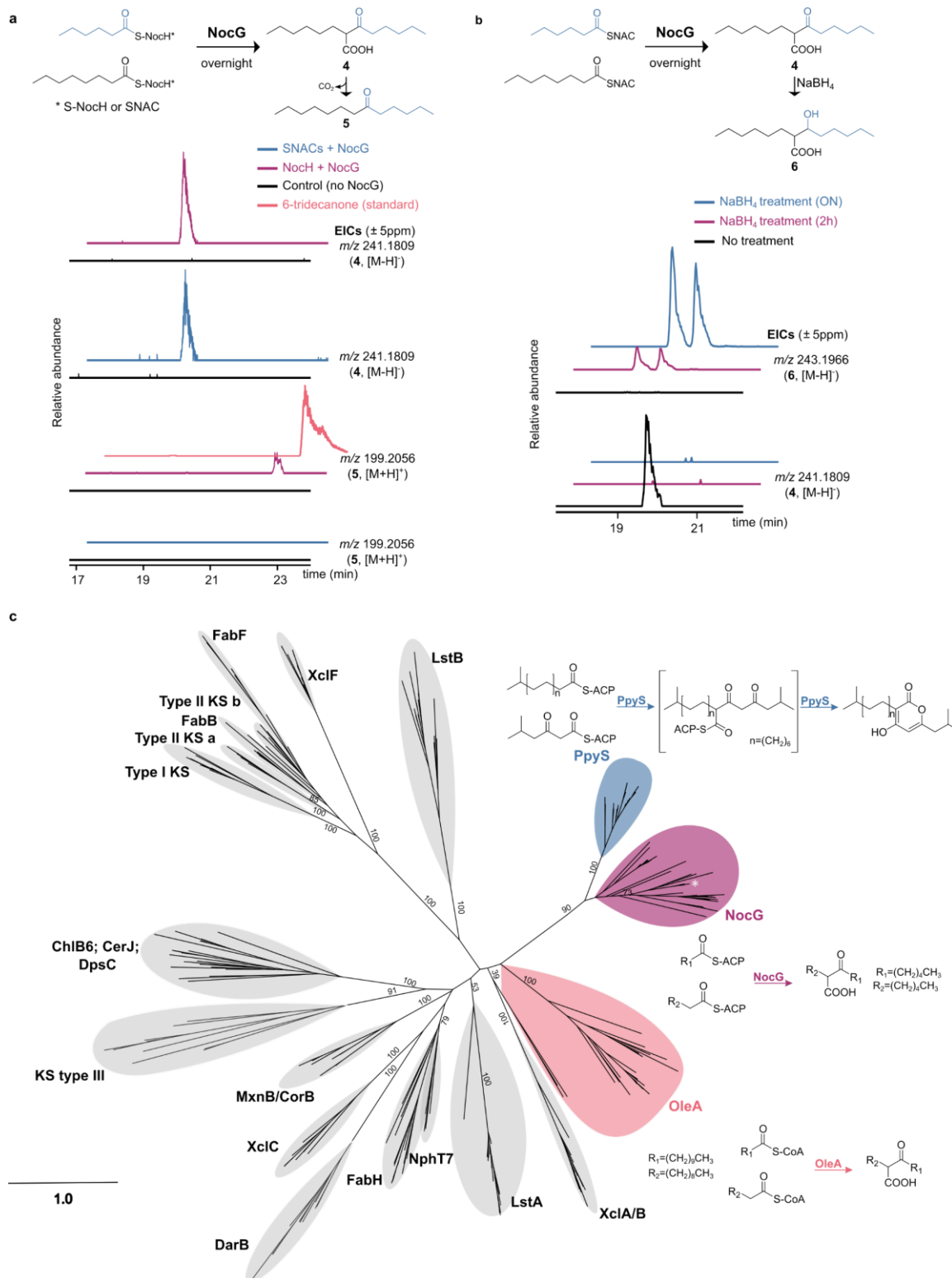


Figure I.4 - NocG belongs to a new class of KSs. (a) LC-HRMS-derived EICs of the β -ketoacid **4** and ketone **5** produced by NocG upon enzymatic assays with hexanoyl- and octanoyl-thioesters. (b) LC-HRMS-derived EICs of **6** upon reduction of **4** with sodium borohydride proving the non-decarboxylative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

use β -ketoacyl-CoAs as substrates, like MxnB (myxopyronin biosynthesis)²³ and CorB (coralloyronin 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 homologues – in NocG these are E100 (deprotonation of the acyl intermediate), C125, H275 and N304 (covalent binding of the precursor) (Fig. SI.24, SI.25). Based on this, and on the similar position of the predicted catalytic residues (Fig. SI.26), we hypothesize a similar reaction mechanism that starts with the deprotonation of the α -carbon of octanoyl-NocH thioester by Glutamate 100, which is able to form a hydrogen bond with the α -carbon of the substrates, creating nucleophile species that subsequently attacks the carbonyl carbon of hexanoyl-NocH thioester to form the new C-C bond (Fig. SI.27).²² To confirm this hypothesis, we constructed a E100A NocG mutant resulting in total loss of activity (Fig. SI.28). 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 I.1 – Comparison between NocG and its closest homologues, 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 branchings)	C ₁₉ to C ₃₂ alkyl- β -ketoacid

^aGenomic context of OleA was not investigated.

The 3-hydroxypropanoyl moiety of **1** is derived from L-methionine

Having elucidated the biosynthesis of the C₁₃ 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. I.1d, Table SI.1), 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 methylthioribulose-1-phosphate (MTRu-1P) by methylthioribose-1-phosphate isomerase (MtnA). MTR-1P is then converted to 2,3-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. SI.29, SI.30), 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 as the 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. SI.31).

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 d₈-L-methionine. LC-HRMS data showed a prominent M+4 peak in the isotope cluster of **1** ([M+H]⁺, Fig. I.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 ¹³C₅-L-methionine and ¹³C₅-¹⁵N-L-methionine (Fig. I.5a). Three ¹³C₅-L-methionine carbon atoms were found to be incorporated into the 3-hydroxypropanoyl moiety, as revealed by MS/MS analysis (Fig. I.5b), thereby establishing the origin of all carbon atoms in **1**. LC-HRMS/MS data from ¹³C₅-¹⁵N-L-methionine supplementation experiments revealed a much higher ¹⁵N incorporation into **1** in MS/MS fragments containing N-1, suggesting a direct incorporation of one nitrogen atom from methionine. Additionally, the calculated ratios

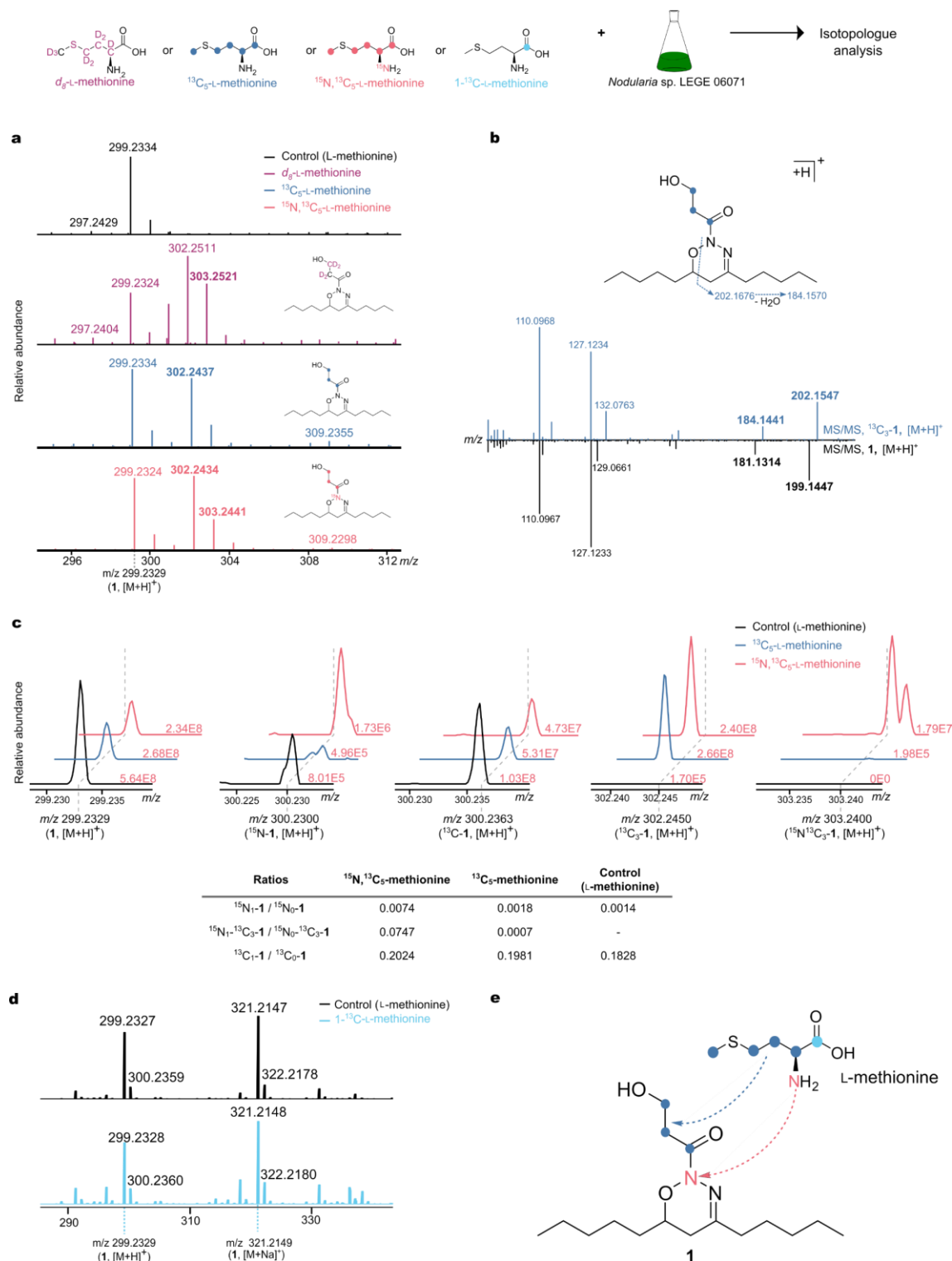


Figure I.5 - 3-Hydroxypropanoyl moiety of **1** is derived from C-2 to C-4 and N-1 of L-methionine. (a) HRMS spectra of **1** and its stable isotope labeled versions following supplementation with d_8 -L-methionine, $^{13}\text{C}_5$ -L-methionine and $^{15}\text{N}, ^{13}\text{C}_5$ -L-methionine, confirming incorporation of this precursor into **1**. (b) HRMS/MS spectrum of $^{13}\text{C}_3$ -**1** confirming methionine incorporation into the 3-hydroxypropanoyl moiety. (c) Fine structure of the M peak in the isotope clusters of $[\text{M}+\text{H}]^+$ 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 $1\text{-}^{13}\text{C-L-methionine}$, supporting a non-specific incorporation of C-1 from L-methionine. (e) Schematic representation of incorporation of C and N atom-derived L-methionine into the 3-hydroxypropanoyl moiety of **1**.

between direct ^{15}N incorporation ($^{15}\text{N}_1^{13}\text{C}_3\text{-}\mathbf{1}/^{15}\text{N}_0^{13}\text{C}_3\text{-}\mathbf{1}$) versus scrambled single ^{15}N incorporation in $\mathbf{1}$ ($^{15}\text{N}_1\text{-}\mathbf{1}/^{15}\text{N}_0\text{-}\mathbf{1}$) indicate a 10-fold increase as a result of direct incorporation from $^{13}\text{C}_5\text{-}^{15}\text{N}$ -L-methionine (Fig. I.5c, SI.32). Further feeding with $1\text{-}^{13}\text{C}$ -L-methionine also allowed us to conclude that the C-1 of L-methionine is not incorporated into $\mathbf{1}$, and, therefore, that the 3-hydroxypropanoyl moiety is formed by L-methionine carbons C-2 to C-4 (Fig. I.5d, SI.33).

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 groups²⁷. Recently, Cui and co-authors²⁸ and Barra *et al.*²⁹ have reported the transfer of a C_4N group derived from SAM to the muraymycin and altemicidin respectively, by a PLP-dependent enzymes, 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 $\mathbf{1}$ (Fig. I.5e). Overall, our findings strongly indicate that generation of the 3-hydroxypropanoyl scaffold in $\mathbf{1}$ involves unprecedented biochemical transformations.

Conclusions

This study shows that the carbon skeleton of the cytotoxic cyanobacterial metabolite $\mathbf{1}$ is formed from three different building blocks: hexanoic acid, octanoic acid and L-methionine (Fig. I.1c). Both FAs are ultimately condensed by NocG with remarkable selectivity, forming a new C-C bond between carbons 5 and 6. A detailed study of NocG shows that it belongs to a new class of KSSs, 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 $\mathbf{1}$ 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 $\mathbf{1}$ and reveals several instances of unusual biochemistry leading to a unique molecular scaffold. Further investigations on the biosynthesis of $\mathbf{1}$, 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.

Acknowledgments

P.N.L. 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. T.P.M. was supported by a scholarship SFRH/BD/138308/2018 from FCT and by the Fulbright Commission. E.P.B. acknowledges support from Harvard University and by the National Science Foundation (NSF) through grants CHE-1454007 and CHE-2003436. N.R.G. acknowledges the NSF Postdoctoral Research Fellowship in Biology (Grant No 1907240). We thank J. Morais for assistance with cyanobacterial cultures and M. Volpe for assistance with the synthesis of hexanoyl- and octanoyl-SNACs.

Author contribution

T.P.M. carried out all experimentation. N.R.G. assisted with the experimentation and D.J.K. assisted with the methionine salvage pathway enzymes. T.P.M., E.P.B. and P.N.L. designed research. T.P.M. and P.N.L. drafted the manuscript, with contributions from all co-authors.

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. In *Comprehensive Natural Products II: Chemistry and Biology, Vol 2: Natural Products Structural Diversity-II: Secondary Metabolites: Sources, Structures and Chemical Biology*, 2010; pp 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., et al., 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, 224-237.
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 I-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), W244-248.
14. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G., et al., 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), 923-936.
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. *Journal of the American Chemical Society* **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), 4108-4112.
20. 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), e00425.
21. 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.

22. 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.
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.; Kojima, C.; Yokota, A., Enzymatic characterization of 5-methylthioribulose-1-phosphate dehydratase of the methionine salvage pathway in *Bacillus subtilis*. *Biosci., Biotechnol., Biochem.* **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. *Biochemical Journal* **2013**, *451*, 145-154.
28. Cui, Z.; Overbay, J.; Wang, X. C.; Liu, X. D., et al., Pyridoxal-5'-phosphate-dependent alkyl transfer in nucleoside antibiotic biosynthesis. *Nat. Chem. Biol.* **2020**, *16* (8), 904-911.
29. Barra, L.; Awakawa, T.; Shirai, K.; Hu, Z. J.; Bashiri, G.; Abe, I., beta-NAD as a building block in natural product biosynthesis. *Nature* **2021**, *600* (7890), 754-758.

Chapter I – Supplementary Information

Published in ACS ChemBio

<https://doi.org/10.1021/acscchembio.2c00464>.

Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A

Methodology

1. General experimental procedures

Stable-isotope labeled (perdeuterated) fatty acids (d_{11} -hexanoic, d_{13} -heptanoic, d_{15} -octanoic, d_{19} -decanoic, d_{23} -dodecanoic and d_{27} -tetradecanoic acids) were obtained from CDN Isotopes. L-methionine, hexanoic acid-1- ^{13}C and octanoic acid-1- ^{13}C were obtained from Sigma-Aldrich. d_8 -L-methionine, $^{13}\text{C}_5$ -L-methionine, $^{13}\text{C}_5$ - ^{15}N -L-methionine and d_{12} -hexanal from Cambridge Isotope Laboratories. All solvents, purchased from VWR Chemicals, Sigma-Aldrich, Thermo Fisher Scientific, Carlo Erba and Chem Lab NV were ACS grade, except for HPLC solvents (HPLC gradient grade) and LC-MS solvents (MS-grade). NMR solvents were purchased from Cambridge Isotope Laboratories.

NMR spectra were acquired on a Varian-Mercury 400 (^1H 400 MHz) NMR spectrometer and the NMR data were analyzed in MestReNova 12.0.3 (MestrelabResearch).

LC-HRMS analyses were performed on either 1) an Agilent 1200 series LC system coupled with Agilent 6530 quadrupole time-of-flight (qTOF) or 2) Thermo Scientific UltiMate 3000 HPLC system, which consists of the following components: LPG-3400RS pump, WPS-3000RS autosampler, TCC-3x00RS column compartment, and MWD-3000RS UV/VIS detector coupled to a Q Exactive Focus Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Fisher Scientific), controlled by Q Exactive Focus (Exactive Series) 2.9 and Thermo Scientific Xcalibur 4.1.31.9 software. LC-HRMS/MS analyses were carried out on the latter system.

Luria-Bertani Lennox (LB) medium, used for NocH, NocI, NocC expressions and Terrific Broth (TB) medium, used for NocA, NocML and NocG expressions, were purchased from Thermo Fisher Scientific. IPTG was acquired from NZYTech. Oligonucleotides were synthesized by StabVida and Sigma-Aldrich. PCR reactions were carried in either a Veriti 96-well thermal cycler (Applied Biosystems) or a C1000 Touch Thermal Cycler (Bio-Rad). DNA polymerases, restriction enzymes and T4 DNA ligase were purchased from Thermo Fisher Scientific and New England Biolabs. DNA fragments' extractions from gel and clean-up of restriction reactions were performed using PCR clean & concentrator-5 (Zymo Research) and NZYGelpure kit (NZYTech). Plasmid purifications were performed using NZYMiniprep (NZYTech) and ZymoPURE Plasmid Miniprep kit (Zymo Research). DNA Sanger sequencing was performed by Eurofins Scientific and Genewiz.

Extracted DNA and PCR reactions were separated in 1 to 1.5% (w/v) agarose gels containing $1\times$ Tris/Acetic acid/EDTA (TAE) buffer (Bio-Rad), stained with GreenSafe Premium (NZYTech) and visualized and photographed on a gel documentation system (Gel

Doc EZ Imager, Bio-Rad). DNA Ladder 1 Kb Plus (Thermo Fisher Scientific) or NZYDNA Ladder III (NZYtech) and Precision plus all standards blue/Precision Plus (Biorad) were used as DNA and protein molecular weight markers, respectively. Proteins were separated in Mini-PROTEAN® TGX™ Precast Polyacrylamide gels – 4-15% (Biorad) containing 1x Tris/Glycine/SDS (TGS) buffer (Fisher Scientific) and stained with InstantBlue (Expedeon) or Bio-Safe Coomassie G-250 (Biorad). DNA and protein concentrations were determined using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix) and NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc.). Optical densities (600 nm) of *E. coli* cultures were determined using a Ultrospec 10 Spectrophotometer (GE Healthcare).

2. Plasmids, Strains and culture conditions

Nodularia sp. LEGE 06071 was acquired from the Blue Biotechnology and Ecotoxicology Culture Collection (LEGEcc) and grown in Z8 medium ¹ at 25 °C, under a 16 h light (25 μmol photons m⁻² s⁻¹)/8 h dark cycle and continuous shaking (190 rpm). In this work, the *noc* gene cluster of *Nodularia* sp. LEGE 06071 was re-annotated based on *in silico* analyses supported by HHPred and Swiss-Model (Table SI.1). To assess the functional role of selected proteins encoded by genes in the *noc* gene cluster, the respective coding sequences were cloned in specific expression plasmids, and the proteins purified. To that end, pET-28a and pET-29b were acquired from Merck. *E. coli* TOP10 (Life Technologies) was used for cloning and *E. coli* BL21 DE3 (Nzytech) and OverExpress C43(DE3) (Sigma) were used for recombinant protein expression.

3. Supplementation experiments

One hundred milliliters of *Nodularia* sp. LEGE 06071 culture were inoculated in 900 mL of Z8 medium and allowed to grow at 25 °C. After 4 days, 25 mL of culture were distributed evenly through 50 mL Erlenmeyer flasks (duplicates were performed for each condition). A pulse feeding protocol was used: right after inoculation (day 0) and at days 2 and 4, the cultures were supplemented with 0.17 mM of perdeuterated fatty acids: *d*₇-butanoic acid, *d*₁₁-hexanoic acid, *d*₁₃-heptanoic acid, *d*₁₅-octanoic acid, *d*₁₉-decanoic acid, *d*₂₃-dodecanoic acid, *d*₂₇-tetradecanoic acid, hexanoic acid-1-¹³C, octanoic acid-1-¹³C or DMSO (control) and the biomass harvested after 7 days of growth. For the methionine feeding experiments, pulse feeding was performed at days 0 and 2 with *d*₈-L-methionine, ¹³C₅-L-methionine ¹³C₅-¹⁵N-L-methionine or DMSO (control) and the biomass harvested after 4 days by centrifugation 4500 ×g, 10 min at 4 °C and washed with deionized H₂O. Experiments for supplementation of the synthetic intermediates *d*₁₂-**7** and *d*₁₂-**8** were performed for 14 days, with pulse feedings at days 0, 2, 4, 7, 9 and 11 in which the cultures

were supplemented with 0.1 mM of either compound. After harvesting, fresh biomass from each experimental condition was extracted with DCM:MeOH (2:1) with vigorous vortexing at room temperature. The resulting extracts were filtered with a Whatman No. 1 filter paper, concentrated using a rotary evaporator (R-210, Buchi) and transferred to pre-weighted vials. Extracts from the feeding experiments were dissolved in MS grade methanol to a final concentration of 2 mg mL⁻¹ and analyzed by LC-HRMS and LC-HRMS/MS.

4. LC-HRMS and LC-HRMS/MS analysis - Orbitrap

HRMS data were obtained in Full Scan positive and negative mode with a scan range of m/z 150-2000, with a capillary voltage of H set to -3.5 kV and the capillary temperature to 263 °C. Sheath gas flow rate was set to 50 units.

Cyanobacterial extracts were separated in a ACE(r) UltraCore™ SuperC18, 2.5 µm particle size, 95 Å pore size and dimensions 75 × 2.1 mm. The column oven was set to 40 °C. The samples were eluted at 0.4 mL/min with a mobile phase of 0.1% formic acid in MeOH/H₂O 1:1 (v/v) (eluent A) and in 0.1% formic acid in IPA (eluent B). The gradient program was as follows: gradient from 10 % B to 70 % B during 5 min, held at 70% B for 12 min, then a gradient to 85 % B over 2 min, and held again at 85% B for 9 min, before returning to the initial conditions.

Quenched enzymatic assays were separated in a Aeris peptide XB-C18 column (3.6 µm particle size, 150 × 4.6 mm). The column oven was set to 40 °C. The flow was set to 0.5 mL/min with min with a mobile phase of 0.1% formic acid in H₂O (eluent A) and in 0.1% formic acid in ACN (eluent B). The gradient program was as follows: isocratic 100 % A for 2 min, gradient from 100 % A to 100 % B over 16 min, held at 100 % B for 5 min, before returning to the initial conditions.

MS/MS parameters for the LC-HRMS/MS analysis were: resolution of 17500, with a 3 m/z isolation window, a loop count of 3, AGC target of 5 × 10⁴ and collision energy of 30 eV.

5. LC-MS and LC-MS/MS analysis - qTOF

LC-MS analysis were obtained in Full Scan positive and negative mode with a scan range of m/z 150-2000, with a capillary voltage of H set to 3.5 kV and the capillary temperature to 300 °C. Nebulizer pressure was set to 35 psi and drying gas (N₂) flow of 8 l min⁻¹.

Crude extracts and synthetically prepared compounds were separated in a Hypersil GOLD aQ C18 column (ThermoScientific), 3 µm particle size, and dimensions 150 × 3.0 mm. The column oven was set to 30 °C. The samples were eluted at 0.5 mL/min with a

mobile phase of 0.1% formic acid in ACN (eluent B) and in 0.1% formic acid in H₂O (eluent A). The gradient program was as follows: isocratic 1 % B for 2 min, gradient from 1 % A to 99 % B over 16 min, held at 99 % B for 5 min, before returning to the initial conditions.

Enzymatic assays were separated in an Aeris widepore XB-C18 column (3.6 μ m particle size, 150 $\times$ 4.6 mm). The samples were eluted at 0.5 mL/min with isocratic 95% solvent B (water with 0.1% v/v formic acid) for 5 min, followed by gradient to reach 95 % solvent B (acetonitrile with 0.1% v/v formic acid) during 23 min, followed by a gradient of 40 % solvent B to reach 95 % solvent A during 3 min, isocratic 95% solvent A for 1 min and return to the initial conditions followed by isocratic 95 % solvent B for 7 min.

6. Metabolomics analysis (nocuolin intermediates search)

Metabolomics analysis of LC-HRMS data from the *d*₁₁-hexanoic acid feeding experiments was carried out in MZmine 2.53.² The analysis was performed using the protocol defined by Figueiredo et al..³ The Adduct Search module was set with the following mass differences: 11.0681 m/z for *d*₁₁ labeling; 22.1369 m/z for *d*₂₂ labeling with 5 ppm tolerance. The *m/z* values corresponding to putative intermediates in the biosynthesis of **1** that were detected using this strategy were used to derive molecular formulas, by taking into account the atom composition of **1** and a 5 ppm tolerance.

7. Cloning, expression and purification of C-His₆-NocH

Cyanobacterial genomic DNA extraction was performed with the PureLink® Genomic DNA Mini Kit (Invitrogen) following the protocol described for Gram-negative bacteria. The *nocH* gene was amplified with primers containing restriction sites for XhoI and NdeI (Table SI.3). The reaction contained 1 $\times$ Phusion HF Buffer, 400 μ M dNTPs, 0.5 μ M of each primer, 0.4 U Phusion polymerase (Thermo Scientific) and was performed under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 60 °C for 30 seconds, extension at 72 °C for 1:30 min and the final extension at 72 °C for 7 min. PCR products were gel purified, restriction digested using XhoI and NdeI (New England Biolabs) for 1.5 h at 37 °C followed by 65 °C inactivation for 20 min and direct reaction purification. Digests were ligated into linearized C-His₆-Tag-containing pET-29b (+), using T4 DNA ligase (New England Biolabs), according to the manufacturer's instructions. An insert to vector molar ratio of 3:1 was used, and the reaction mixtures were incubated overnight at room temperature and the enzyme was then inactivated at 65 °C for 10 min.

Chemically competent *E. coli* TOP10 cells (Invitrogen) were transformed with 5 μ L of the ligation reaction using the heat shock method. After recovery, the mixture was plated

onto LB plates supplemented with 50 µg mL⁻¹ kanamycin. Colonies were screened by colony PCR using 2 µM T7 primers, 0.5 units of GoTaq Flexi DNA Polymerase (Promega), 1× Green GoTaq reaction buffer and 250 µM of each deoxyribonucleotide triphosphate (dNTP). The PCR program was defined as: initial denaturation at 95 °C for 5 min at 95 °C; 35 cycles with denaturation for 1 min at 95 °C, primer annealing at 48 °C for 1 min and extension at 72 °C for 3.5 min and final extension for 10 min at 72 °C. Individual positive colonies were inoculated in LB medium supplemented with 50 µg mL⁻¹ kanamycin and used for plasmid DNA extraction. The purified plasmids were sequenced to verify construct identity and amplification fidelity. pET-29b-C-His₆-NocH was transformed into chemically competent *E. coli* BL21(DE3) using the heat shock method.

Protein expression was optimized in small-scale studies with different temperatures (37 °C, 27 °C and 20 °C) and IPTG concentration (0.5 mM or 1 mM) (Table SI.2).

For the large-scale expression, glycerol stocks were used to inoculate in 100 mL of LB medium supplemented with 50 µg mL⁻¹ kanamycin and grown overnight at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 3 × 2 L of LB medium supplemented with 50 mg mL⁻¹ kanamycin at 37 °C, 175rpm. The cultures were grown to an OD of 0.2-0.25 and then placed at 27 °C. When OD_{600nm} reached 0.5-0.6, the cultures were induced with 0.5 mM of IPTG and allowed to grow overnight. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), transferred to 50 mL falcon tubes, re-centrifuged for 10 minutes at 5000 ×g and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl₂, 1 mM DTT, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (ThermoScientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi). The cell pellets were centrifuged for 30 min at 19 500 ×g and added to pre-washed Ni-NTA resin (1 mL). The resin and supernatant were allowed to incubate at 4 °C on a nutating mixer for 2 h. After incubation, the resin was collected by centrifugation (1000 ×g, 5 min) and transferred to a glass column. The column was washed with 5 mL lysis buffer and the protein was eluted in multiple 2 mL fractions of elution buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl₂, 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and fractions containing NocH were pooled and dialyzed overnight in Slide-A-Lyzer™ Dialysis Cassettes, 10K MWCO, 12 mL (Thermo Scientific) against 5 L of dialysis buffer (25 mM Tris-HCl, 50 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). The samples were then loaded onto Pierce™ Protein Concentrator tubes, 10 K MWCO (Thermo Scientific) and centrifuged twice at 7000

$\times g$ for 20 min to reach a final concentration of 354 μM (Fig. SI.1). The protein solution was added dropwise to liquid nitrogen and stored as pellets at $-80\text{ }^{\circ}C$.

8. Expression and purification of C-His₆-Sfp

Chemically competent *E. coli* TOP10 and BL21(DE3) cells were transformed with pET-29b containing the Sfp-encoding gene⁴ using the heat shock method. For large-scale expression, glycerol stocks were used to inoculate in 100 mL of LB medium supplemented with 50 $\mu g\text{ mL}^{-1}$ kanamycin and grown overnight at $37\text{ }^{\circ}C$. Twenty-milliliters of this overnight culture were then inoculated into $2 \times 2\text{ L}$ of LB medium supplemented with 50 $\mu g\text{ mL}^{-1}$ kanamycin at $37\text{ }^{\circ}C$, under 175 rpm shaking. The cultures were grown to an OD 0.5-0.6, before being induced with 0.5 mM of IPTG (Table SI.3). After 5 h, the cell pellets were recovered by centrifugation (10 min, $3000 \times g$), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 minutes at $5000 \times g$ and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi). The cell pellets were centrifuged for 30 min at $19\,500 \times g$ and added to previously washed Ni-NTA resin (1 mL). The resin and supernatant were allowed to incubate at $4\text{ }^{\circ}C$ on a nutating mixer for 2 h. After this incubation time, the resin was collected by centrifugation ($1000 \times g$, 5 min) and transferred to a glass column. The column was washed with 5 mL lysis buffer and the protein was eluted in multiple 2 mL fractions of elution buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM $MgCl_2$, 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and fractions containing the target protein were pooled and dialyzed overnight in Slide-A-Lyzer™ Dialysis Cassettes, 10K MWCO, 12 mL (Thermo Scientific) against 5 L of dialysis buffer (25 mM Tris-HCl, 50 mM NaCl, 10% (v/v) glycerol, 1mM DTT). The samples were then loaded onto Pierce™ Protein Concentrator tubes, 10 K MWCO (Thermo Scientific) and centrifuged at $7000 \times g$ for 2×20 min to reach a final concentration of 1.565 mM (Fig. SI.1). The protein was added dropwise to liquid nitrogen and stored at $-80\text{ }^{\circ}C$ as pellets.

9. Cloning, expression and purification of C-His₆-NocML

Vector pCDFDuet-1 (Novagen) was amplified by PCR using primers pCDFDuet-1-MCS1 (Table SI.3) using Q5 High-Fidelity Master Mix (NEB) with 0.5 μM of each primer and 0.04 ng/ μL template under the following conditions: initial denaturation at $98\text{ }^{\circ}C$ for 30

seconds, followed by 30 cycles of denaturation at 98 °C for 10 seconds, annealing at 65 °C for 10 seconds, extension at 72 °C for 2 min and a final extension at 72 °C for 2 min. The product was purified using a QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. The PCR-linearized vector was then digested using 2000 U/ml DpnI (NEB) for 1 hour at 37 °C, followed by heat-inactivation of DpnI at 80 °C for 20 minutes. The *nocM* gene was amplified from *Nodularia* sp. LEGE 06071 genomic DNA by PCR - Table S3: initial denaturation at 98 °C for 30 seconds, followed by 35 cycles of denaturation at 98 °C for 10 seconds, annealing at 60 °C for 10 seconds, extension at 72 °C for 30 seconds and a final extension at 72 °C for 2 min. The product was purified using a QIAquick PCR Purification Kit. The DpnI-digested pCDFDuet-1 fragment and *nocM* product were joined using Gibson assembly. The reaction contained 4 µl Gibson Assembly Master Mix (NEB), 0.041 pmol of DpnI-digested linearized pCDFduet-1, and 0.125 pmol of purified *nocM* PCR product in 8 µl total volume. The reaction was incubated at 50 °C for 30 minutes and transformed into BAP1 by electroporation. Transformants were selected on LB agar plates with 50 µg/ml spectinomycin. Individual colonies were grown overnight in LB with 50 µg/ml spectinomycin and the correct plasmid sequence - pCDF-His₆NocM-emptyMCS2 - was confirmed by plasmid isolation and Sanger sequencing. Vector pCDF-His₆NocM-emptyMCS2 was amplified by PCR using primers pCDFDuet-MCS2 (Table SI.3) and the amplification conditions as for pCDF-Duet-1. The product was purified and DpnI-digested. The *nocL* gene was amplified by PCR (Table SI.3) - as for *nocM*, except the extension time was 1 min and the annealing temperature was 58 °C. The product was purified using a QIAquick PCR Purification Kit. The DpnI-digested pCDF-His₆NocM-emptyMCS2 fragment and *nocL* product were joined using Gibson assembly as above. The reaction was transformed into TOP10 by electroporation and transformants were selected on LB agar plates with 50 µg mL⁻¹ spectinomycin. Individual colonies were grown overnight in LB with 50 µg mL⁻¹ spectinomycin and the correct plasmid sequence was confirmed by plasmid isolation and Sanger sequencing.

For the large-scale expression, glycerol stocks were used to inoculate in 100mL TB supplemented with 50 mg mL⁻¹ kanamycin and grown overnight at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 2 x 2 L of TB medium supplemented with 50 µg mL⁻¹ kanamycin and 5 mM of MgSO₄ at 37°C, 175rpm. The cultures were grown to an OD of 0.5-0.6 and then transferred to 25 °C induced with 0.5 mM of IPTG overnight. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL Falcon tubes, re-centrifuged for 10 minutes at 5000 ×g and frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl₂, 5 mM CaCl₂, 30mM imidazole, 10% (v/v) glycerol, pH 8.0)

supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19500 $\times g$ and the clarified supernatant loaded into 1 mL of Ni-NTA resin. The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM imidazole and the protein was eluted in 5 mL of elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction was dialyzed overnight using the Slide-A-Lyzer Dialysis Cassettes (ThermoFisher) and concentrated in the Spin-X-UF 10K, 6 mL (Corning) for 30 min at 3200 $\times g$ to reach a final concentration of 354.0 μM (Fig. SI.1). The protein was added dropwise to liquid nitrogen and stored at -80 °C.

10. Cloning, expression and purification of C-His₆-NocI

The *nocI* gene was amplified with primers containing restriction sites for XhoI and NdeI (Table SI.1). The reaction contained 1 $\times$ Phusion HF Buffer, 400 μM dNTPs, 0.5 μM of each primer, 0.4 U Phusion polymerase (Thermo Scientific) and was performed under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 58 °C for 30 seconds, extension at 72 °C for 1 minute and the final extension at 72 °C for 7 min. PCR products were gel purified, restriction digested using XhoI and NdeI (New England Biolabs) for 1.5 h at 37 °C followed by 65 °C inactivation for 20 min and direct reaction purification. Digests were ligated into linearized C-His₆-Tag-containing pET-29b, using T4 DNA ligase (New England Biolabs), according to the manufacturer's instructions. An insert to vector molar ratio of 3:1 was used, and the reaction mixtures were incubated overnight at room temperature and the enzyme was then inactivated at 65 °C for 10 min. Chemically competent *E. coli* TOP10 cells (Invitrogen) were transformed with 5 μL of the ligation reaction using the heat shock method. After recovery, the mixture was plated in LB agar plates supplemented with 50 $\mu g mL^{-1}$ kanamycin. Colonies were screened by colony PCR using 2 μM T7 primers, 0.5 units of GoTaq Flexi DNA Polymerase (Promega), 1 $\times$ Green GoTaq reaction buffer and 250 μM of each deoxyribonucleotide triphosphate (dNTP). The PCR program was defined as: initial denaturation at 95 °C for 5 min at 95 °C; 35 cycles with denaturation for 1 min at 95 °C, primer annealing at 52 °C for 1 min and extension at 72 °C for 1:30 min and final extension for 10 min at 72 °C. Individual positive colonies were inoculated in LB medium supplemented with 50 $\mu g mL^{-1}$ kanamycin and used for plasmid DNA extraction. The purified plasmids were sequenced to verify construct identity and amplification fidelity. pET-

29b-CHis₆-NocI was transformed into chemically competent *E. coli* BL21 (DE3) using the heat shock method. Protein expression was optimized in small-scale studies with different temperatures (30 °C, 25 °C and 20 °C) and IPTG concentration (0.5 mM or 1 mM) and growth time (Table SI.2). For large-scale expression, glycerol stocks were used to inoculate in 100 mL TB supplemented with 50 mg mL⁻¹ kanamycin and grown overnight at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 2 L of TB medium supplemented with 50 mg mL⁻¹ kanamycin and 5 mM of MgSO₄ at 37 °C, with 175 rpm shaking. The cultures were grown to an OD of 0.5-0.6 and then transferred to 30 °C induced with 0.5 mM of IPTG for 2 h. The cell pellets were recovered by centrifugation (10 min, 3000 × *g*), then transferred to 50 mL Falcon tubes, re-centrifuged for 10 minutes at 5000 × *g* and frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl₂, 1 mM DTT, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 × *g* and the clarified supernatant loaded into 1 ml of Ni-NTA resin. The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50mM imidazole and the protein was eluted in 5 mL of elution buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl₂, 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed overnight in Slide-A-Lyzer™ Dialysis Cassettes, 10K MWCO, 12 mL (Thermo Scientific) cassette against 5 L of dialysis buffer (25 mM Tris-HCl, 50 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). The samples were then loaded onto Pierce™ Protein Concentrator tubes, 10 K MWCO (Thermo Scientific) and centrifuged twice at 7000 × *g* for 20 min to reach a final concentration of 144 μM (Fig. SI.1). The protein was added dropwise to liquid nitrogen and stored as pellets at -80 °C.

11. Cloning, expression and purification of N-His₆-NocG

A *nocG* sequence codon-optimized for *E. coli* was synthesized by Genewiz and cloned in pET-28a using NdeI and XhoI (Fig. SI.2). Chemically competent *E. coli* TOP10 cells (Invitrogen) were transformed with 2 μL of plasmid and plated in LB agar plates supplemented with 50 μg mL⁻¹ kanamycin. Colonies were screened by colony PCR using 2 μM T7 primers, 0.5 units of GoTaq Flexi DNA Polymerase (Promega), 1× Green GoTaq reaction buffer and 250 μM of each deoxyribonucleotide triphosphate (dNTP). The PCR program was defined as: initial denaturation at 95 °C for 5 min at 95 °C; 35 cycles with denaturation for 1 min at 95 °C, primer annealing at 48 °C for 1 min and extension at 72 °C

for 1:30 min and final extension for 10 min at 72 °C. Individual positive colonies were inoculated in LB medium supplemented with 50 µg mL⁻¹ kanamycin and used for plasmid DNA extraction. The purified plasmids were sequenced to verify construct identity and amplification fidelity. The resulting construct - pET-28a-NHis₆-NocGsynthetic - was transformed into chemically competent *E. coli* C43 (DE3) using the heat shock method. Protein expression was optimized in small-scale studies with different temperatures, growth times and IPTG concentration (Table SI.3). For the large-scale expression, glycerol stocks were used to inoculate 100 mL TB medium supplemented with 50 µg mL⁻¹ kanamycin and grown overnight at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 2 × 2 L of TB medium supplemented with 50 µg mL⁻¹ kanamycin and 5 mM of MgSO₄ at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 20 °C, induced with 0.5 mM of IPTG and allowed to grow overnight. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 min at 5000 ×g and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl₂, 5 mM CaCl₂, 30 mM imidazole, 10 % (v/v) glycerol, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 ×g and the clarified supernatant loaded into 1 mL of HisPur Cobalt resin (Thermo Scientific). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM and 75 mM imidazole and the protein was eluted in 5 mL elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction was loaded into PD-10 desalting columns (GE Healthcare). The protein was eluted with 3.5mL of dialysis buffer (20 mM HEPES, 50 mM NaCl, 1 mM DTT, 10% (v/v) glycerol). The protein was then loaded onto Pierce™ Protein Concentrator tubes, 10 K MWCO (Thermo Scientific) and centrifuged at 7000 ×g for 2 × 20 min to reach a final concentration of 58.3 µM (Fig. SI.1). The protein was added dropwise to liquid nitrogen and stored as pellets at -80 °C.

12. Cloning, expression and purification of N-His₆-NocG E100A

N-His₆-NocG E100A was constructed using the Q5 Site-Directed Mutagenesis Kit (NewEngland Biolabs) according to the manufacture's instructions.

PCR reaction was performed on pET-28a-NocG with 1x Q5 Hot Start High-Fidelity, 10 µM of each primer, and was performed under the following conditions: initial denaturation

at 98 °C for 30 seconds, followed by 35 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 55 °C for 30 seconds, extension at 72 °C for 4 min and the final extension at 72 °C for 7 min. PCR products were gel purified, ligated with KLD treatment (NewEngland Biolabs), and transformed in TOP10 *E. coli*. The obtained constructs were confirmed by Sanger sequencing and then transformed into chemically competent *E. coli* C43 (DE3) using the heat shock method. For the large-scale purification, 100 mL TB medium supplemented with 50 µg mL⁻¹ kanamycin and grown overnight at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 2 × 2 L of TB medium supplemented with 50 µg mL⁻¹ kanamycin and 5mM of MgSO₄ at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 20 °C, induced with 0.5 mM of IPTG and allowed to grow overnight. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 min at 5000 ×g and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl₂, 5 mM CaCl₂, 30 mM imidazole, 10 % (v/v) glycerol, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 ×g and the clarified supernatant loaded into 1 mL of HisPur Cobalt resin (Thermo Scientific). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM and 75 mM imidazole and the protein was eluted in 5 mL elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed and concentrated using the using the Amicon Ultra Centrifugal Filters (Merck) with 3.5 mL of dialysis buffer (20 mM HEPES, 50 mM NaCl, 1 mM DTT, 10% (v/v) glycerol). The protein was added dropwise to liquid nitrogen and stored as pellets at -80 °C. Final protein concentration: 65.1 µM (Fig. SI.1).

13. Cloning, expression and purification of C-His₆-NocA

The *nocA* gene was amplified with primers containing restriction sites for XhoI and NdeI (Table SI.3). Applying the same cloning process as NocI, NocA was cloned in pET-29b and transformed in BL21(DE3). Protein expression was optimized in small-scale studies (Table SI.2). For the large-scale expression, glycerol stocks were used to inoculate 100 mL of LB medium supplemented with 50 µg mL⁻¹ kanamycin and grown overnight at 37 °C. Twenty-milliliters of this overnight culture were in turn inoculated into 1 L of LB medium supplemented with 50 µg mL⁻¹ kanamycin and 5 mM of MgSO₄ at 37 °C, with 175 rpm

shaking. The cultures were grown to an OD of 0.5-0.6 and then placed at 27 °C, induced with 0.5 mM of IPTG and allowed to grow overnight. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL Falcon tubes, re-centrifuged for 10 minutes at 5000 ×g and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl₂, 5 mM CaCl₂, 30 mM imidazole, 10% (v/v) glycerol, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (ThermoScientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 ×g and the clarified supernatant loaded into 1 mL of Ni-NTA resin. The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM imidazole and the protein was eluted in 5 mL of elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed overnight in Slide-A-Lyzer™ Dialysis Cassettes, 10K MWCO, 12 mL (Thermo Scientific) against 5 L of dialysis buffer (25 mM Tris-HCl, 50 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). The samples were then loaded onto concentrated in the Spin-X-UF 10K, 6 mL (Corning) for 30 min at 3200 ×g to reach a final concentration of 433 μM (Fig. SI.1). The protein was added dropwise to liquid nitrogen and stored at -80 °C as pellets.

14. Cloning and expression of N-His₆-MtnA and N-His₆-MtnK

Genomic DNA from *B. subtilis* 168 was extracted using the kit Genomic DNA Clean & Concentrator (Zymo Research), according to manufacturer's instructions. Genes encoding MtnA and MtnK were amplified with primers containing pET-28a Gibson tails (Table SI.3). The reaction contained 0.5 μM of each primer and 1× Q5 High-fidelity master mix (NewEngland Biolabs) and was performed under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 35 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 65 °C for 30 seconds, extension at 72 °C for 1 min and the final extension at 72 °C for 7 min. pET-28a was amplified using the same conditions and extension time of 4 min. The amplified pET-28a fragment was digested with DpnI at 37 °C for 10 min and assembled with *mtnA* or *mtnK* amplicons using the HiFi Gibson assembly kit (NewEngland Biolabs) at 50 °C for 15 min. Chemically competent *E. coli* TOP10 cells (Invitrogen) were transformed with 2 μL of plasmid and plated in LB plates supplemented with 50 μg mL⁻¹ kanamycin. Colonies were screened by colony PCR as described for NocG cloning. Individual positive colonies were inoculated in LB medium supplemented with

50 $\mu\text{g mL}^{-1}$ kanamycin and used for plasmid DNA extraction. The purified plasmids were sequenced to verify construct identity and amplification fidelity. pET-28a-CHis₆-MtnA and pET-28a-CHis₆-MtnK were transformed into chemically competent *E. coli* BL21(DE3) using the heat shock method. Large-scale expression and purification were performed as described for NocA (Table SI.2). MtnA and MtnK were purified with a final concentration of 463 μM and 70 μM , respectively (Fig. SI.1). The protein samples were added dropwise to liquid nitrogen and stored as pellets at -80 °C.

15. Small-scale expression tests and western blots

Briefly, 500 μL of overnight culture were inoculated in 50 mL LB or TB containing antibiotic and 5 mM of MgSO_4 at 37 °C, with 200 rpm shaking. The cultures were grown to an OD of 0.5-0.8 (~3h) and then transferred to the appropriate temperatures and induced with IPTG. One milliliter of culture was then centrifuged and frozen in liquid nitrogen. The cells were resuspended in BugBuster (Merck) containing a final concentration of 0.5 mg mL^{-1} of lysozyme (Sigma) and the suspension was incubated at 4 °C for 5 mins. Benzonase Nuclease (0.5 μL , Merck) and 10 mM MgSO_4 were added and the solution once again incubated at 4 °C for 5 min. The suspension was centrifuged at 4 °C at 16 000 $\times g$ for 30 min. The pellets were resuspended in 300 μL of 2% SDS were and warmed for 5 min at 50 °C. Protein expression was visualized in polyacrylamide gels. Due to undetectable amounts of soluble NocG, western blot analyses were performed to detect and optimize expression conditions. PVDF membranes (GE Healthcare) were soaked in methanol and placed in a sandwich form. The transfer was carried out in 25 mM Tris, 192 mM glycine, 20% methanol (v/v) at 100 V for 1 h. The membrane was blocked with 3% BSA for 1 h and then washed with 1 $\times$ Tris-Buffered Saline, 0.1% Tween 20 detergent (TBST) buffer. One microliter of antibody (6-His Tag monoclonal antibody, Invitrogen) was added in 1% BSA/TBST to the membrane and incubated for 1 h. The remaining antibody was then washed out of the membrane with TBST and the blot revealed with the CN/DAB kit (Thermo Scientific).

16. Enzymatic assays

In vitro enzymatic assays to test the function of Noc enzymes were carried out in Eppendorf tubes at room temperature.

16.1. NocC enzymatic assay

Briefly, holo-ClyB (holo-NocM) was prepared by incubating 100 μM apo-ClyB co-expressed with ClyA (NocL), 1 mM hexanoic, octanoic, decanoic- or dodecanoic-CoA, and 2 mM MgCl_2 in 50 mM HEPES pH 8.0 for 1 h at room temperature. The mixture was then

split into two conditions: a control with no enzyme addition (instead, we used 50 mM HEPES pH 8.0), and a test condition with the addition of 20 μ M NocC and 1.8 mM of ferrocenium hexafluorophosphate as electron acceptor. The reactions were incubated overnight at room temperature, digested with trypsin sequencing grade (Promega), and quenched with an equal volume of 25% formic acid (aq). The reactions were centrifuged at 16 000 $\times g$ for 5 min and analyzed by LC-HRMS.

16.2. NocI enzymatic assay

Briefly, holo-ClyB (holo-NocM) was prepared by incubating 100 μ M apo-ClyB co-expressed with ClyA (NocL), 1 mM hexanoyl, octanoyl, decanoyl- or dodecanoyl-CoA, and 2 mM $MgCl_2$ in 50 mM HEPES pH 8.0 for 1 h at room temperature. The mixture was then split into two conditions: a control with no enzyme addition (instead, we used 50 mM HEPES pH 8.0), and a test condition with addition of 20 μ M NocI and 1.8 mM of 1-methoxy-5-methylphenazinium methyl sulfate (Sigma-Aldrich) as electron acceptor. The reactions were incubated overnight at room temperature, digested with trypsin sequencing grade (Promega) and quenched with an equal volume of 25% formic acid. The reactions were centrifuged at 16 000 $\times g$ for 5 min and analyzed by LC-HRMS.

16.3. NocL and NocH competition assay

Briefly, 100 μ M of co-expressed holo-ClyB and ClyA(NocL) was incubated with 2 mM $MgCl_2$, 1 mM ATP and either 1 mM hexanoic, octanoic, decanoic or dodecanoic acid or 1 mM hexanoic, octanoic, decanoic, dodecanoic and tetradecanoic acid, in 50 mM HEPES pH 8.0. Reactions were incubated overnight at room temperature, digested with trypsin sequencing grade and quenched with an equal volume of 25% formic acid. The reactions were centrifuged at 16 000 $\times g$ for 5 min and analyzed by LC-HRMS.

For the NocH assay, 100 μ M of NocH was incubated with 10 μ M Sfp, 1 mM ATP, 200 μ M CoA, 2 mM $MgCl_2$ and either 1 mM hexanoic, octanoic, decanoic or dodecanoic acids or 1 mM hexanoic, octanoic, decanoic, dodecanoic and tetradecanoic acids, in 50 mM HEPES pH 8.0. Reactions were incubated overnight at room temperature, digested with trypsin sequencing grade, quenched with an equal volume of 25% formic acid. The reactions were centrifuged at 16 000 $\times g$ for 5 min and analyzed by LC-HRMS.

16.4. NocA enzymatic assay

Briefly, 5-methylthioribose was incubated with MtnK (25 μ M) in 50mM Tris-HCl buffer (pH = 7.5) supplemented with 5 mM $MgCl_2$, 5 mM ATP and 5 mM MTR at room temperature. After 1 h, the mixture was then split into two conditions: a control with no enzyme addition,

and a test condition with addition of 10 μ M MtnA. After an 1 h incubation, 50 μ M of NocA were added to the test condition and the reaction incubated overnight. The reactions were quenched with acetonitrile, centrifuged at 16 000 $\times g$ for 5 min and analyzed by LC-HRMS.

16.5. NocG enzymatic assay

Briefly, 100 μ M of NocH was incubated with 10 μ M Sfp, 1 mM ATP, 200 μ M CoA, 1 mM hexanoic and octanoic acids in 50 mM HEPES pH 8.0 and incubated for 1 h at room temperature. The mixture was then split into two conditions: a control with no enzyme addition, and a test condition with addition of 20 μ M NocG. The reactions were incubated at room temperature, overnight and were then quenched with the same volume of acetonitrile, centrifuged at 16 000 $\times g$ for 5 min and analyzed by LC-MS. The same assay was performed using acyl thioesters of N-acetylcysteamine (SNACs) substrate surrogates. In this case, 1 mM hexanoyl-SNAC and octanoyl-SNAC were incubated with 20 μ M of NocG in 50 mM HEPES pH 8.0 at room temperature, overnight.

Variants of the assays included the presence of NocI (10 μ M) and/or 1-methoxy-5-methylphenazinium methyl sulfate (100 μ M).

16.6. NocG E100A enzymatic assay

The same assay was performed using acyl thioesters of N-acetylcysteamine (SNACs) substrate surrogates. In this case, 1 mM hexanoyl-SNAC and octanoyl-SNAC were incubated with 20 μ M of NocG in 50 mM HEPES pH 8.0 at room temperature, overnight. A positive control was performed with NocG and NocH using the same conditions.

17. Determination of kinetic parameters for N-His₆-NocG

The determination of apparent K_m and V_{max} parameters for N-His₆-NocG was performed according to substrate consumption (octanoyl-SNAC) due to the lack of available standard for compound **4**.

Assays were carried in 30 μ L reaction buffer (50 mM HEPES pH 8.0) with 1 mM octanoyl-SNAC and 20 μ M N-His₆-NocG. A range of hexanoyl-SNAC concentrations (corresponding to 25, 50, 100, 175, 250, 500 and 750 μ M) were used to determine the initial reaction rates. Reactions were incubated at 25 °C and aliquots (5 μ L) were collected at different time points (30 min, 1 h, 2 h, 4 h, 6 h, 8 h), quenched with 45 μ L of a mixture of ACN/H₂O (1:1, v/v) and incubated in ice for 30min. Quenched reactions were then centrifuged at 17000 $\times g$, 4 °C, 3 min and analyzed by LC-HRMS/MS.

A calibration curve of octanoyl-SNAC was generated using dilutions from a 100 mM standard (0.05, 0.1, 0.5, 1, 5, 10, 50, 100 mM) and considering the integration of the EIC peak for m/z 246.1522 ($[M+H]^+$, C8-SNAC). The initial reaction rates (v_0) were calculated from each linear regression slope for the consumption of octanoyl-SNAC over time (some outlier time points were not considered) and were then used to calculate apparent K_m and V_{max} values using R.

18. Phylogenetic analysis

BlastP amino acid sequences of the NocG homologues with lowest e-value were retrieved from the NCBI database (43 homologues retrieved). Their amino acidic sequences were aligned, together with different groups of characterized ketosynthases and their homologues (266) based on previous literature data⁵ using MUSCLE alignment, from within Geneious R11 (Biomatters). The resulting alignment (with a total of 309 sequences) was trimmed to its core region and contained 742 positions. RaxML-HPC2 on XSEDE was ran in CIPRES⁶ to compute the maximum-likelihood phylogenetic tree, using pseudocounts and 1000 rate categories of sites.

19. Sequence similarity networks and Genomic Neighborhood Diagrams

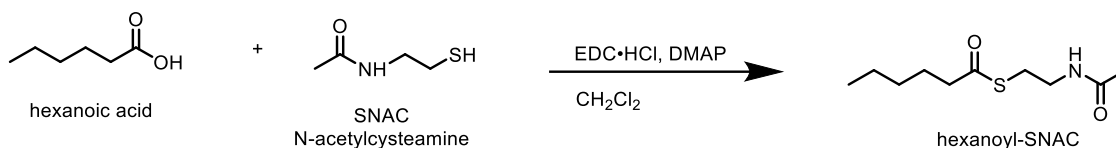
Sequence similarity networks (SSNs) were generated using the EFI-EST sever, following a “Sequence BLAST (Basic Local Alignment Search Tool)” of NocG (WP_194003077.1) as input⁷ using negative log e-values of 5 for UniProt BLAST sequence retrieval. The 1000 retrieved sequences and the query sequence were then used to generate the SSNs with an alignment score threshold of 110 and a minimum length of 200. The networks were visualized in Cytoscape (v3.90). The full SSN obtained in the previous step was used to generate Genomic Neighborhood Diagrams (GNDs) using the EFI-GNT tool.⁷ A Neighborhood Size of 10 was used and the Lower Limit for Co-occurrence was 20 %. The resulting GNDs were visualized in Cytoscape.

20. NocG structural model

NocG and PpyS sequences were submitted to Alphafold⁸ and the protein models generated following the standard parameters.

21. Chemical synthesis procedures and characterization data

21.1. Synthesis of hexanoyl-SNAC

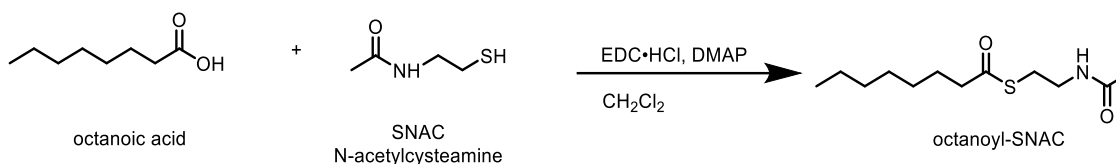


Hexanoic acid (100 mg, 0.86 mmol) was dissolved in dichloromethane (5 mL), and EDC·HCl (220 mg, 1.2 mmol) was added. The reaction mixture was cooled to 0°C and stirred for 20 min. The flask was then charged with *N*-acetylcysteamine (74 μL , 0.70 mmol) and 4-dimethylaminopyridine (10 mg, 0.080 mmol), and the reaction mixture was warmed to room temperature. After 3 h, the reaction mixture was quenched with water (10 mL) and extracted with ethyl acetate (3 $\times$ 10 mL). The combined organic layers were washed with brine (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 80-90% ethyl acetate in hexanes to afford hexanoyl-SNAC (70 mg, 0.32 mmol, 37%)⁹ (Figs. SI.34, SI.36).

HR-MS: calcd. for $\text{C}_{10}\text{H}_{19}\text{NO}_2\text{S}$ - $([\text{M}+\text{H}]^+)$ 218.1209, found 218.1227

^1H NMR (500 MHz, CDCl_3): δ 6.00 (br, 1H), 3.40 (q, J = 5.98 Hz, 2H), 3.00 (t, J = 6.35 Hz, 2H), 2.55 (t, J = 7.43 Hz, 2H), 1.95 (s, 3H), 1.64 (m, 2H), 1.29 (m, 4H), 0.87 (t, J = 6.6 Hz, 3H).

21.2. Synthesis of octanoyl-SNAC

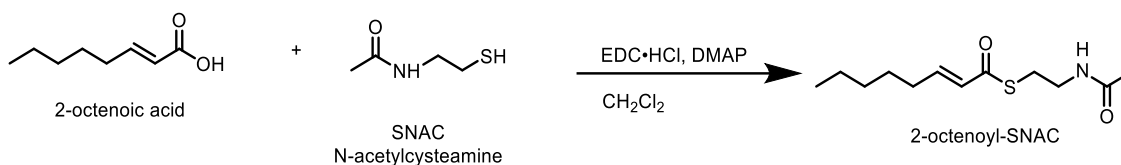


Octanoic acid (100 mg, 0.69 mmol) was dissolved in dichloromethane (5 mL), EDC·HCl (220 mg, 1.2 mmol) was added. The reaction mixture was cooled to 0 °C and stirred for 20 min. The flask was then charged with *N*-acetylcysteamine (74 μL , 0.70 mmol) and 4-dimethylaminopyridine (10 mg, 0.080 mmol), and the reaction mixture was warmed to room temperature. After 3 h, the reaction mixture was quenched with water (10 mL) and extracted with ethyl acetate (3 $\times$ 10 mL). The combined organic layers were washed with brine (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 80-90% ethyl acetate in hexanes to afford hexanoyl-SNAC (118 mg, 0.67 mmol, 73%) (Figs. SI.34, SI.37).

MS (500 MHz, CDCl₃): calcd. for C₁₂H₂₃NO₂S - ([M+H]⁺) 246.1522, found 246.1542

¹H NMR (500 MHz, CDCl₃): δ 6.05 (br, 1H), 3.39 (q, J = 6.2, 4.9 Hz, 2H), 2.99 (t, J = 6.6 Hz, 2H), 2.54 (t, J = 7.7 Hz, 2H), 1.94 (s, 3H), 1.62 (m, 2H), 1.26 (m, 8H), 0.85 (t, J = 6.9 Hz, 3H).

21.3. Synthesis of 2-octenoyl-SNAC



2-octenoic acid (100mg, 0.70 mmol) was dissolved in dichloromethane (5 mL), EDC•HCl (220 mg, 1.2 mmol) was added. The reaction mixture was cooled to 0 °C and stirred for 20 min. The flask was then charged with N-acetylcysteamine (74 µL, 0.70 mmol) and 4-dimethylaminopyridine (10 mg, 0.080 mmol), and the reaction mixture was warmed to room temperature. After 3 h, the reaction mixture was quenched with water (10 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic layers were washed with brine (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 80-90% ethyl acetate in hexanes to afford 2-octenoyl-SNAC (136 mg, 0.56 mmol, 80%)¹⁰ (Figs. SI.35, SI.38).

HR-MS: calcd. for $C_{12}H_{21}NO_2S$ - $([M+H]^+)$ 244.1366, found 244.1396

¹H NMR (500 MHz, CDCl₃): δ 6.96 (dt, J = 15.6, 7.0 Hz, 1H), 6.15 (d, J = 15.5 Hz, 1H), 3.48 (t, J = 6.6 Hz, 2H), 3.10 (t, J = 6.4 Hz, 2H), 2.20 (m, 2H), 2.04 (s, 3H), 1.47 (m, 2H), 1.30 (m, 4H), 0.89 (t, J = 6.7 Hz, 3H)

21.4. Synthesis of 5-methylthioribose

5-methylthioribulose (MTR) was synthesized from methylthioadenosine (MTA) via acid hydrolysis, as previously described in the literature.¹¹ In detail, methylthioadenosine (22.2 mg, 75 µmol) was dissolved in 50 mM sulfuric acid (25 mL) and heated to 100 °C for 3 h. The reaction was then allowed to cool to room temperature and neutralized with sodium hydroxide.

The compound was purified by flash chromatography, eluting with 10% MeOH (TLC: Tf=0.5, 100% EtOAc). The compound was visualized in TLC plates through the orcinol reaction (detection of reducing sugars) and concentrated *in vacuo*, yielding MTR (12.2 mg, 67 μ mol, 89%) (Figs. SI.35, SI.39).

HR-MS: calcd. for $C_6H_{12}O_4S$ - $([M-H]^-)$ 179.0384, found 179.0394

^1H NMR (500 MHz, DMSO): δ 2.10 (s, 1H), 2.57 (dd, $J = 13.6, 7.6$ Hz, 1H), 2.72 (dd, $J = 13.5, 4.1$, 1H), 3.63 (t, $J = 4.6$, 1H), 3.78 (m, 1H), 3.85 (dd, $J = 6.7, 4.8$ Hz, 1H), 4.82 (d, $J = 6.6$ Hz, 1H), 4.91 (m, 1H), 6.27 (d, $J = 5.4$ Hz, 1H)

Table SI.1 - Detailed annotation of the *noc* gene cluster in *Nodularia* sp. LEGE 06071. Annotation based on HHPred and Swiss-Model

Noc	Membrane	Previous annotation ¹²	HHPred			Swiss-Model		
			name	e-value	aa	name	identity	coverage
A	No	Aldolase	Aldolase II superfamily Fucose phosphate aldolase (MtnB)	2.8e-28	204	Methylthioribulose-1-phosphate dehydratase (MtnB) Putative aldolase class II	24.63 30.81	0.96 0.93
B	No	Dioxygenase	Acireductone Dioxygenase (MtnD)	1.1e-30	169	Acireductone dioxygenase (MtnD)	31.36	0.92
C	No	Dehydrogenase	Acyl-CoA dehydrogenase	9.2e-58	480	Short chain acyl-CoA dehydrogenase	26.81	0.70
D	Yes	Methyltransferase	Isoprenylcysteine O-methyltransferase	1.8e-25	191	Isoprenylcysteine carboxyl methyltransferase	27.37	0.77
E	No	N-monooxygenase	Flavin-containing monooxygenase	3.1e-62	475	Flavin-containing monooxygenase	34.90	0.96
F	Maybe the N terminal	Aminotransferase, kinase	Aminotransferase	1.9e-52	431	Aminotransferase class III	32.55	0.49
G	No	Acyl-ACP reductase	Ketosynthase (FabH homologue)	1.1e-36	314	Ketosynthase	22.04	0.90
H	No	Fatty-acyl AMP ligase, ACP	Amino acid adenylation domain protein	2.3e-61	697	Fatty-acyl AMP ligase	44.37	0.81
I	No	Oxidoreductase	Acyl-CoA dehydrogenase	1.4e-33	361	FAD-dependent oxidoreductase Glutaryl-CoA dehydrogenase	19.44 18.64	0.95 0.95
J	No	Ring-hydroxylating dioxygenase	3-ketosteroid 9 α -hydroxylase oxygenase	7.5e-45	322	3-ketosteroid-9- α -monooxygenase	25.57	0.94
K	No	Transferase	Glycine N-methyltransferase SAM-dependent methyltransferase	0.0037 0.0034	128 68	Phosphoethanolamine N-methyltransferase SAM methyltransferase	11.22 9.30	0.36 0.34

Table SI.2 - Expression conditions used for all proteins purified in this work

Protein	Predicted MW (kDa)	Expression condition					
		plasmid	<i>E. coli</i> strain	Media	Temp.	IPTG/ mM	time
NocA	24.82	pET-29b(+)	BL21(DE3)	LB	27 °C	0.5	overnight
NocC	60.52	pET-29b(+)	BL21(DE3)	TB	30 °C	0.5	2h
NocGsynthetic	40.89	pET-28a(+)	C43(DE3)	TB	20 °C	0.5	overnight
NocH	70.34	pET-29b(+)	BL21(DE3)	LB	27 °C	0.5	overnight
NocI	42.18	pET-29b(+)	BL21(DE3)	TB	25 °C	0.5	2h
NocML	61.64 11.54	pCDF-Duet1	BAP1	TB	25 °C	0.5	4h
MtnA	39.69	pET-28a(+)	BL21(DE3)	LB	37 °C	0.5	overnight
MtnK	45.91	pET-28a(+)	BL21(DE3)	LB	37 °C	0.5	overnight
Sfp	26.16	pET-29b(+)	BL21(DE3)	LB	37 °C	0.5	5h

Table SI.3 – Primers used in this work. Green: NdeI restriction site; Red: XhoI restriction site

Gene/ Vector	primer	Sequence	Reference
<i>nocA</i>	Fw	TAAACATATGACCAGCCCCACAACAGC	This work
	Rv	CGGGCTCGAGTTTATTGGGTAATACG	This work
<i>nocC</i>	Fw	CTTGCATATGCACCTCAATAGTCAAAGC	This work
	Rv	TCCGCTCGAGGGCATTATTACAAAG	This work
<i>nocH</i>	Fw	CATCCATATGAATCAGGTCATTACAACG	This work
	Rv	GATCCTCGAGTGATGCTACTTTAGC	This work
<i>nocI</i>	Fw	GCGCATATGCATTTTTTACACAAAG	This work
	Rv	AAACTCGAGAGGCATCATTAAAGG	This work
<i>nocG</i> E100A	Fw	GAAACCGCGACCAATGCCGG	This work
	Rv	ATCGCGCCAGCGGCCGCG	This work
<i>nocML</i>	NocM-Fw	CAGCCATCACCATCATCACCGGCAGCAGCAACTTTTGAAGACA AC	This work
	NocM-Rv	CAAGCTTGTCGACCTGCAGTTATTAGTTTCTCAGCCACATATTC	This work
	NocL-Fw	GTATATTAGTTAAGTATAAGAAGGAGATATACCATGAATGTTCTTTGA CCG	This work
	NocL-Rv	GTAGACGAGTCCATGTGCTTATTATTGAAGTGCATTAGCTTTTTTAC	This work
<i>mtnA</i>	Fw	GCAGCCATATGGCTAGCATGACCCATTCAATTTGCTG	¹³
	Rv	GGTGGTGGTGCTCGAGTGCAAATGAGCAAAGTCC	¹³
<i>mtnK</i>	Rv	GCAGCCATATGGCTAGCATGGGAGTCACAAAAACACC	¹⁴
	Rv	GGTGGTGGTGCTCGAGTGCAATCATTCCTTTACAAGT	¹⁴
pET-28a	Fw	GCTAGCCATATGGCTGC	This work
	Rv	GCACTCGAGCACCACCACC	This work
pCDF Duet-	MCS1-Rv	CTGCAGGTCGACAAGCTTG	This work
	MCS1-Fw	GTGGTGATGATGGTGATGGCTG	This work
	MCS2-Rv	GGTATATCTCCTTCTTATACTTAATAATACTAAGATGG	This work

Table SI.4 – Accession numbers of all sequences used in the KSs phylogenetic tree

WP_006034384.1	AXE28137	MBF2006620.1	WP_018779660.1	ABB69750.1
YP_132684.1	WP_179177256.1	WP_016876568.1	WP_040569064.1	AAN65231.1
WP_005305524.1	MBF6246144	WP_008198039.1	WP_021808728.1	AAG29787.2
WP_006644045.1	TCJ97066	TRT87724.1	WP_026096098.1	AFB35630.1
WP_005364526.1	THJ67643	WP_194003077.1	BAJ73499.1	ALE27503.1
WP_008989540.1	NLE79168	WP_187038421.1	KEP74827.1	ACJ24876.1
WP_007465048.1	WP_049708878.1	WP_016949109.1	AKF07269.1	AEI91069.1
YP_001102990.1	WP_073842843.1	WP_071187609.1	ZP_01906524.1	ZP_04991255.1
ABJ97437.1	AGO97060.1	AKL71640.1	ZP_05127044.1	AAF70109.1
AAB66504.1	CAE17216.1_1	WP_041033411.1	WP_057656623.1	BAB72048.1
AAP42874.1	WP_041380444.1	WP_090812699.1	MBH1494163.1	ABC00733.1
AAF82408.1	WP_132355917.1	WP_014267925.1	WP_108763677.1	AAA65208.1
ADC79637.1	WP_038246124.1	WP_114206319.1	WP_004085471.1	AAM21773.1
AAP85336.1	WP_012989958.1	AXC10748.1	NP_299252.1	AAK82824.1
NP_929153.1	WP_038180969.1	WP_018955464.1	pdb 3S21 A_1	sp Q9FUB7.1
NP_931374.1	WP_046337502.1	WP_015801934.1	NP_635607.1	AAA02827.1
AAF81728.1	WP_092513653.1	WP_190122337.1	WP_065469684.1	sp Q8SAS8.1
AAK06784.1	WP_010847197.1	WP_067355704.1	WP_101364527.1	NP_868579.1
CKK26894.1	WP_013184973.1	WP_005164847.1	WP_042820992.1	NP_216181.1
AAD20267.1	WP_095191442.1	WP_167474139.1	WP_104588541.1	NP_216176.1
YP_003086385.1	WP_102356958.1	WP_069160481.1	WP_034212769.1	NP_390087.1
YP_003121552.1	WP_150370633.1	WP_043690931.1	WP_022969495.1	BAB91443.1
ZP_00134992.2	WP_027611766.1	WP_028478668.1	WP_057949486.1	BAB91444.1
ZP_04562837.1	WP_041064001.1	WP_040795117.1	WP_139716540.1	NP_828307.1
NP_416826.1	WP_007967127.1	WP_195063367.1	TXI51133.1	pdb 3V7I A
YP_001881145.1	WP_134690235.1	WP_011207969.1	WP_164231769.1	WP_051546693
ACU62401.1	WP_104444175.1	MAZ80289.1	WP_197910318.1	WP_096413974
YP_143679.1	WP_023630065.1	MBA2673455.1	WP_007281261.1	WP_215840065
NP_344945.1	WP_062574518.1	WP_025596424.1	WP_030402327.1	WP_207525601
YP_001867862.1	WP_028692573.1	WP_027214818.1	WP_036407026.1	WP_057943777
NP_389016.1	WP_062359461.1	WP_013342411.1	WP_026404743.1	RYZ32449.1
NP_645683.1	WP_186662459.1	WP_014972441.1	WP_006975318.1	RKI20282.1
NP_930065.1	WP_025131331.1	WP_029966890.1	YP_007138278.1	4YUF_A
ZP_02902779.1	BAQ74133.1	WP_006049675.1	YP_007317906.1	WP_164933227
NP_415613.1	BAQ80432.1	WP_027802201.1	YP_001868566.1	WP_164933227.1
NP_287229.1	WP_060838760.1	WP_028204568.1	YP_007057764.1	ADI59524.1
BAC70003.1	WP_060843247.1	WP_042262981.1	YP_003899922.1	QVV57711.1
YP_001771620.1	WP_161759127.1	WP_028232503.1	YP_007155727.1	NP_287225.1
AAA67516.1	WP_166288705.1	WP_144270416.1	YP_001865657.1	WP_017748751.1
AAK06785.1_1	WP_046974710.1	WP_027819887.1	YP_001865628.1	WP_189969332.1
AAD20268.1_AAF81729.1	WP_036813671.1	WP_110327210.1	YP_007056099.1	WP_202203201
MBB4921493	WP_105397665.1	WP_074981949.1	YP_007130807.1	WP_109779861
WP_199828569.1	WP_046396397.1	WP_060611230.1	AAZ77679.1	WP_130513667
THA39683	WP_049583975.1	WP_028370407.1	AAZ77676.1	WP_229806990
KPI33618	GDY11852.1	WP_116122167.1	ADU85987.1_AAK83178.1	WP_011209988
MBW5486569	MBE1374984.1	WP_012428081.1	AAM70354.1	WP_067847169
WP_215140563.1	WP_097931105.1	ZP_01103251.1	ZP_04334033.1	WP_068065566
WP_125547774.1	WP_014890909.1	WP_017822397.1	ZP_01462124.1	WP_042260942.1
AIT38299.1_1	KAF0135624.1	WP_135541752.1	ACN64832.1	WP_203612468
WP_114059298.1	WP_018405831.1	WP_043453199.1	YP_001107471.1	WP_040901237
AGS77282.1	WP_029650722.1	KIA73109.1	CAA16179.1	WP_184933817
WP_217028965	EKB04829.1	YP_007911827.1	AAC18104.1	WP_053693089
WP_217028965.1	EKB07937.1	WP_010098042.1	ACI88883.1	WP_179177255.1
WP_012507654	EHP29910.1	YP_003947618.1	AAG30195.1	WP_114059297.1
WP_076792224	CAR66906.1	YP_003871436.1	BBB01291.1	AIT38298.1_1
EHB89432.1	EJL05977.1	WP_007431139.1	BAC73499.1	WP_125547776
EIG27057.1	AAN18032.1	YP_005077926.1	YP_001826619.1	WP_125547776.1
AFN79642.1	CCI22605.1	WP_010345468.1	AAV84077.1	WP_125652149.1
EJF43732.1	NP_347450.1	WP_007130623.1	NP_389015.1	YP_003714026.1
EJF37460.1	YP_006930640.1	sp D7URV0.1	NP_388898.1	WP_010848687.1
sp Q54206.1	CAM58805.1	ZP_07291840.1	NP_249690.1	WP_008043745.1
	AAQ08929.1	BAU79603.1		

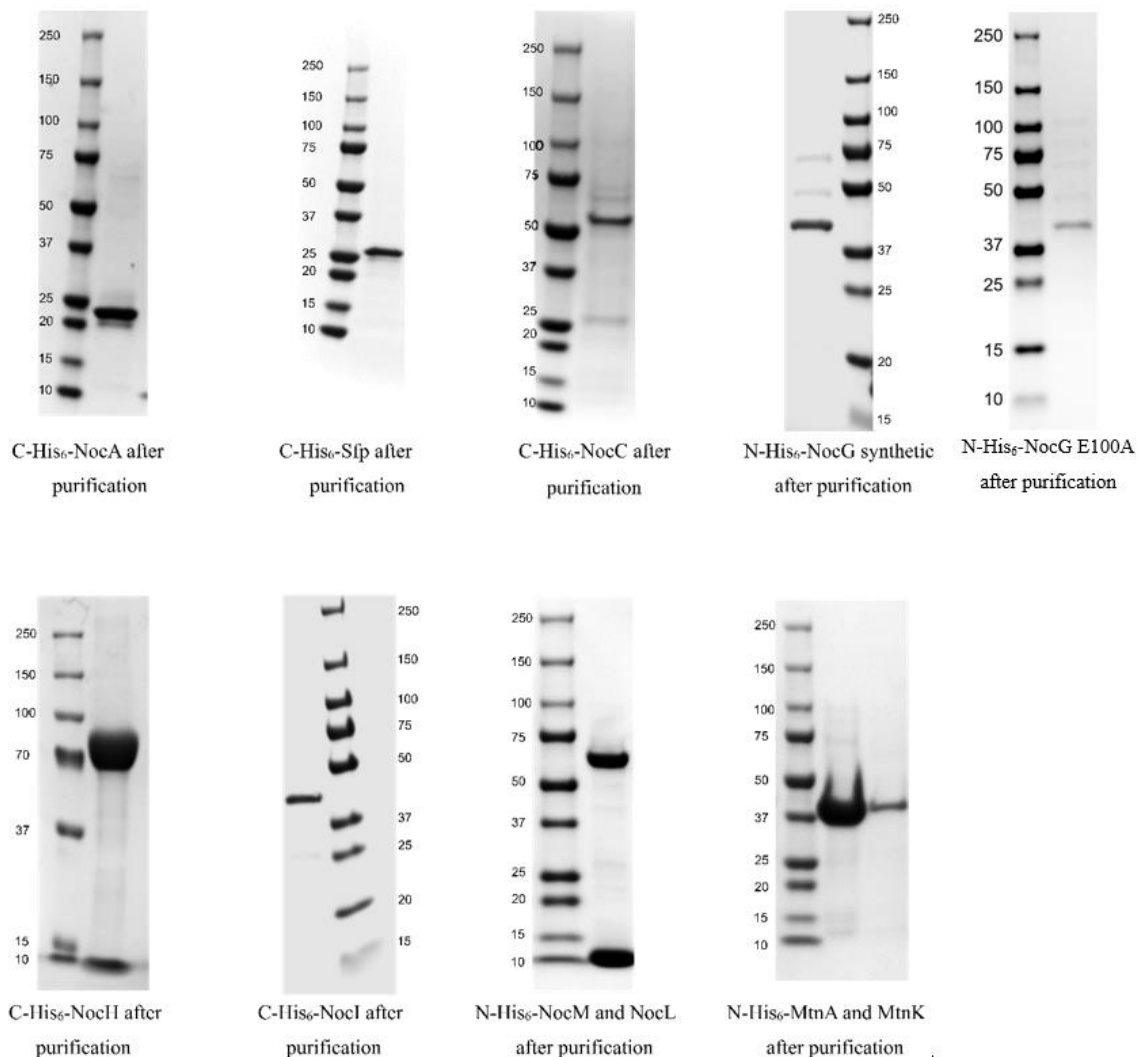


Figure SI.1– SDS-Page of NocA, NocB, NocC, NocG, NocI, NocML, MtnA and MtnK purifications. All proteins were purified by Ni-NTA resin except for NocG that used HisPur Cobalt resin.

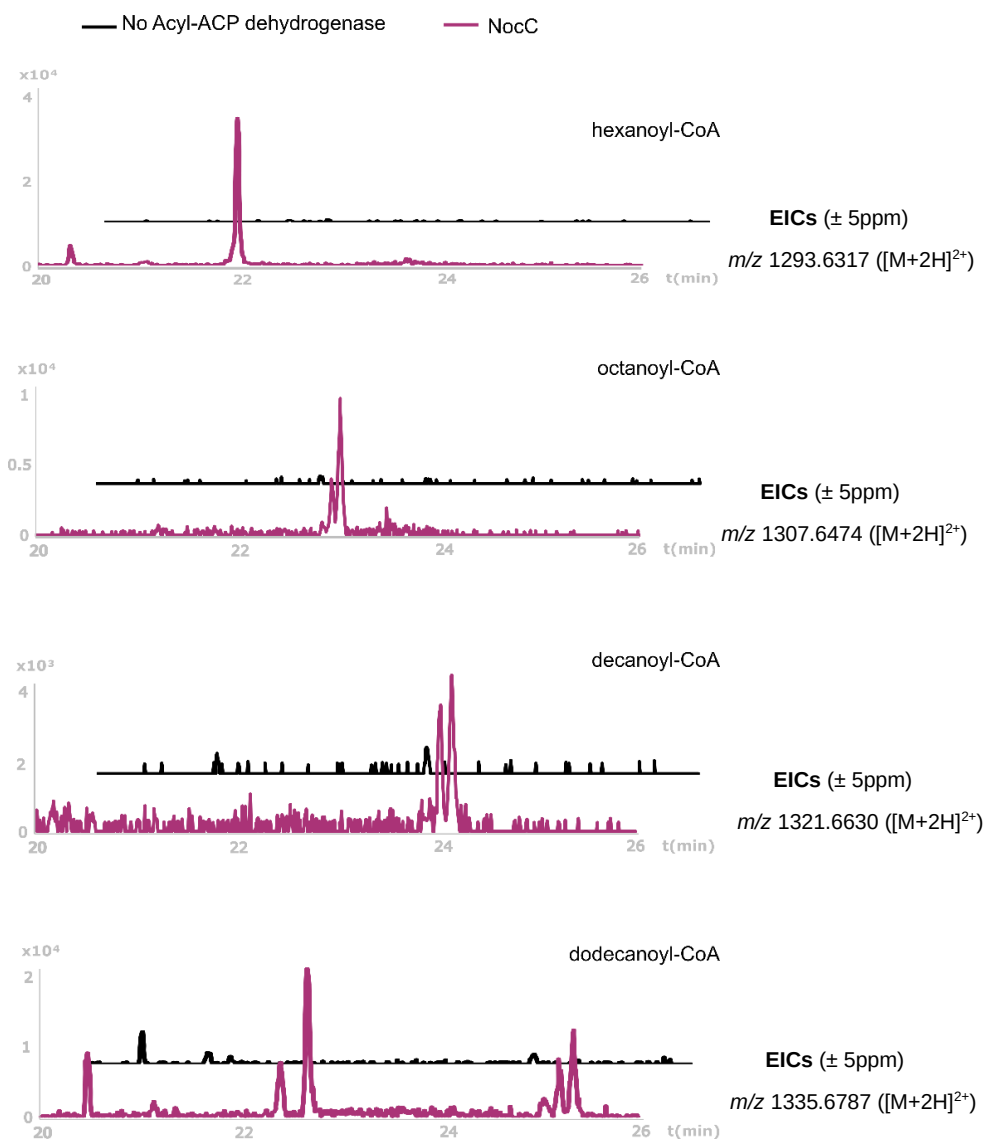


Figure SI.2 - Extracted ion chromatograms of the expected enoyl-ACP fragments predicted for NocM previously loaded with hexanoyl-, octanoyl-, decanoyl- and dodecanoyl-CoA (m/z 1293.6317, 1307.6474, 1321.6630 and 1335.6787, respectively). The assay confirmed NocC activity as acyl-ACP dehydrogenase by identifying the expected enoyl-ACP fragments for all thioesters tested. NocC recognizes NocM and is able to reduce all its possible substrates. Thus, we propose that this enzyme is involved in the biosynthesis of the chlorosphaerolactylates.

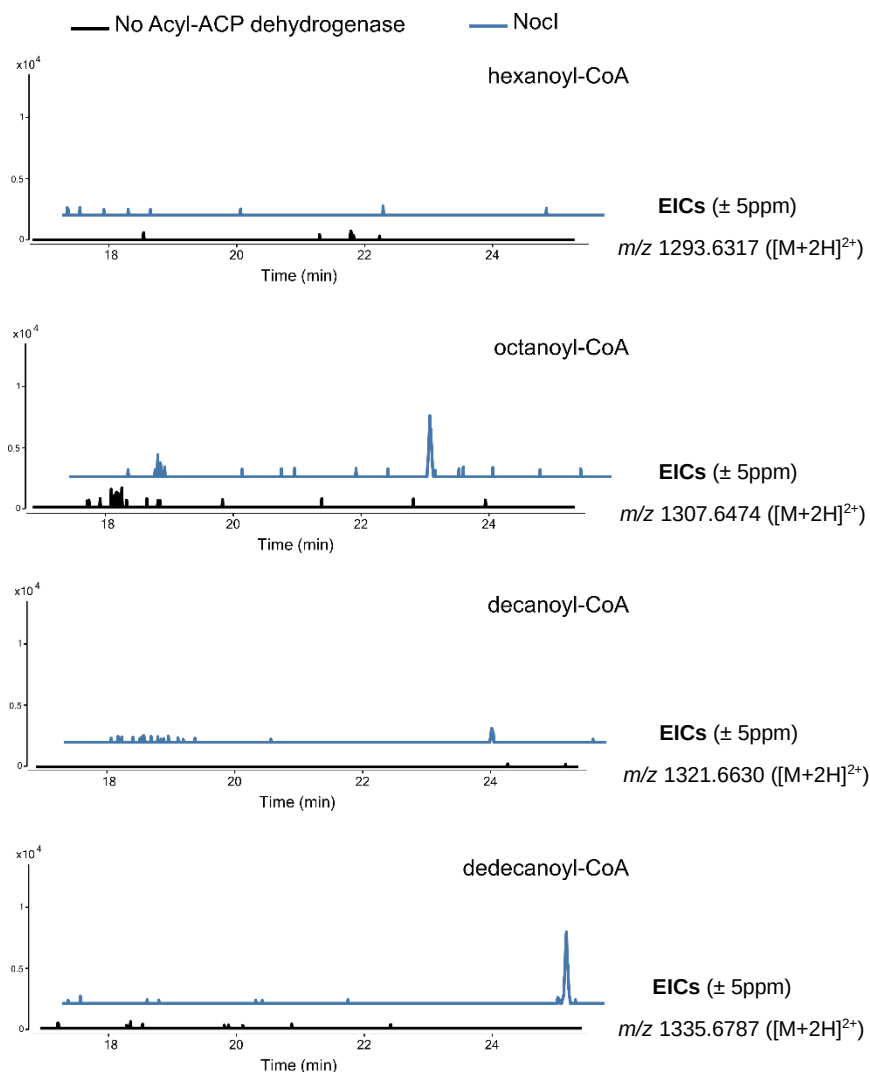


Figure SI.3 - Extracted ion chromatograms of the expected enoyl-ACP fragments predicted for NocM previously loaded with hexanoyl-, octanoyl-, decanoyl- and dodecanoyl-CoA (m/z 1293.6317, 1307.6474, 1321.6630 and 1335.6787, respectively). NocI dehydrogenase activity using octanoyl-, decanoyl- and dodecanoyl-CoA. Unlike NocC, which was active with all tested thioesters (C6 to C12), NocI seems to be less promiscuous and does not act on hexanoic acid. Additionally, its activity with decanoic acid is also very reduced. NocI activity was significant with dodecanoic acid and octanoic acid.

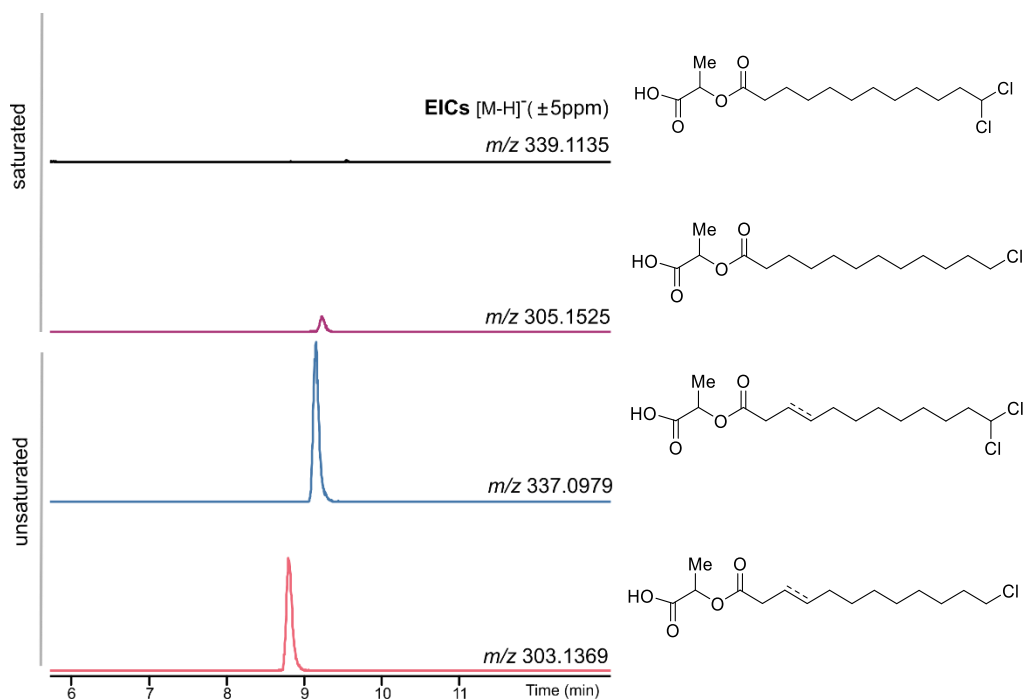


Figure SI.4 - Extracted ion chromatograms of saturated and unsaturated versions of chlorosphaerolactylates A and B in *Nodularia* LEGE sp. 06071. The higher abundance of the unsaturated versions and the acyl-ACP dehydrogenase activity of NocC supports its involvement in the biosynthesis of the chlorosphaerolactylates. The dashed double bond indicates the putative positioning of the unsaturation based on the structure of the nocuolactylates.

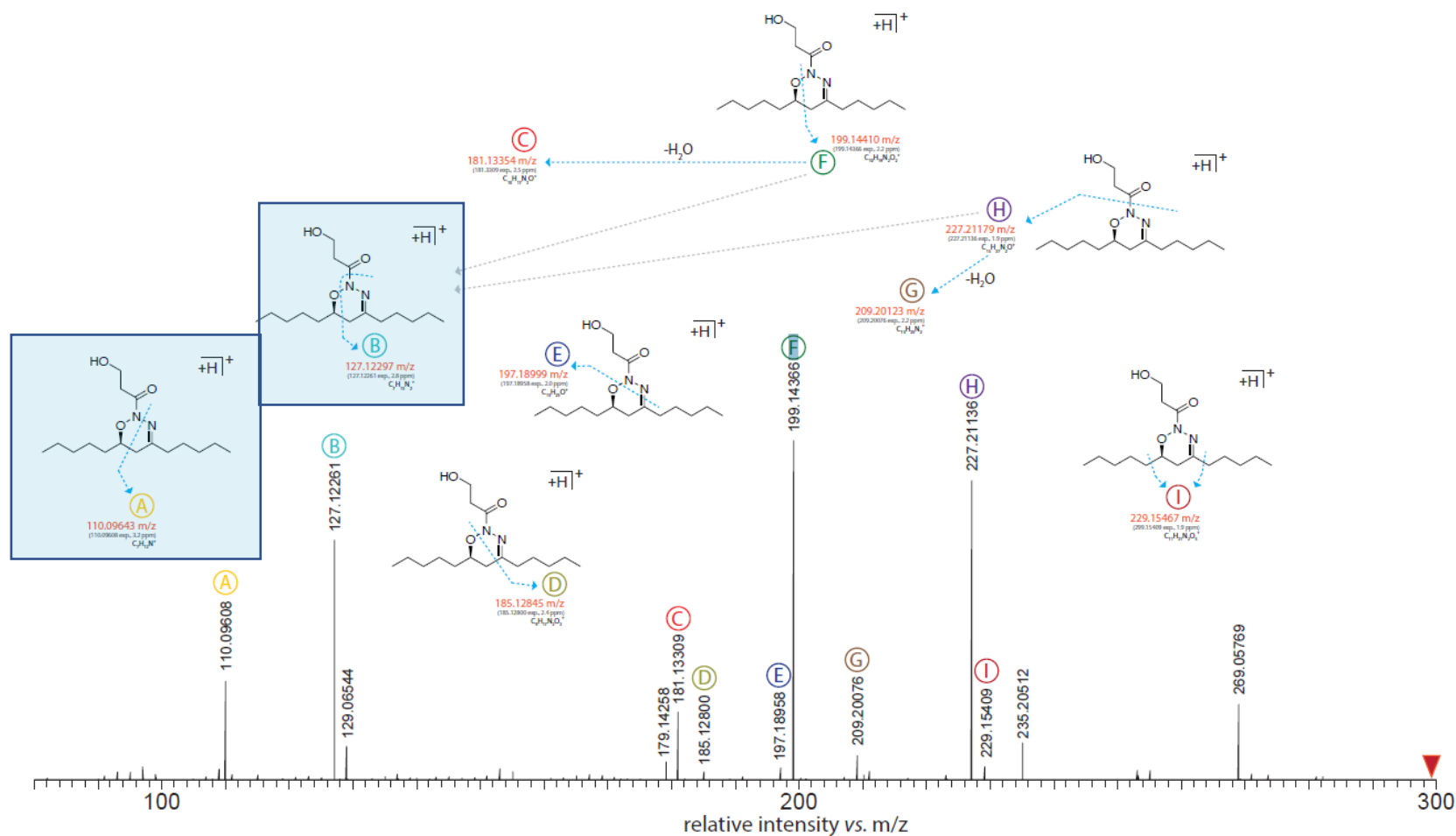


Figure SI.5– MS/MS fragmentation of nocuolin A, highlighting key fragments that include position C-5

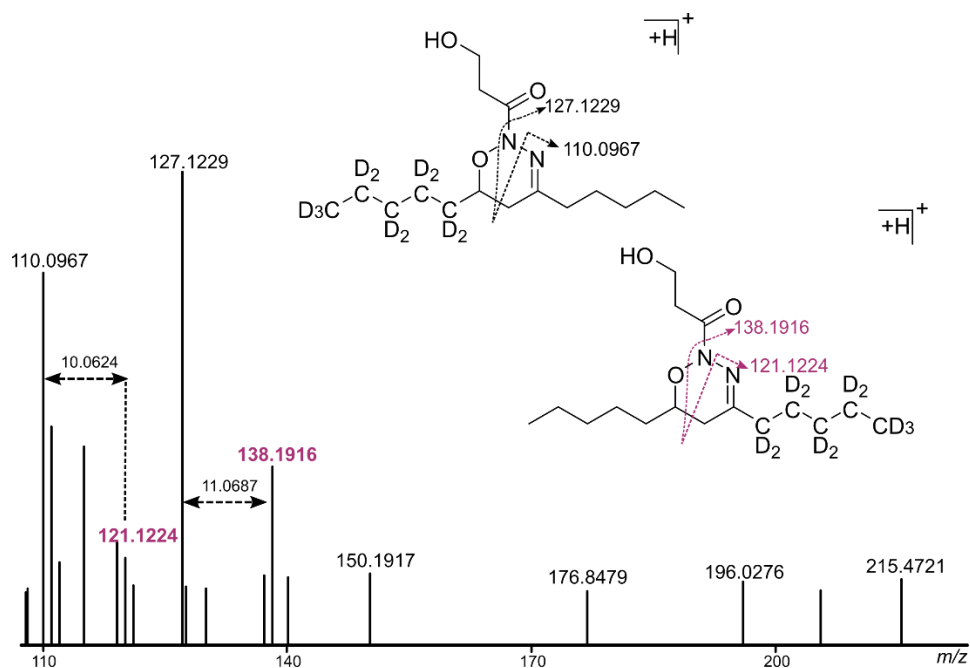


Figure SI.6– HRMS/MS spectrum of *d*₁₁-1 (m/z 332.2839) highlighting the diagnostic fragmentations that support the incorporation of *d*₁₁-hexanoic acid in the C₁₃ carbon skeleton of nocuolin A.

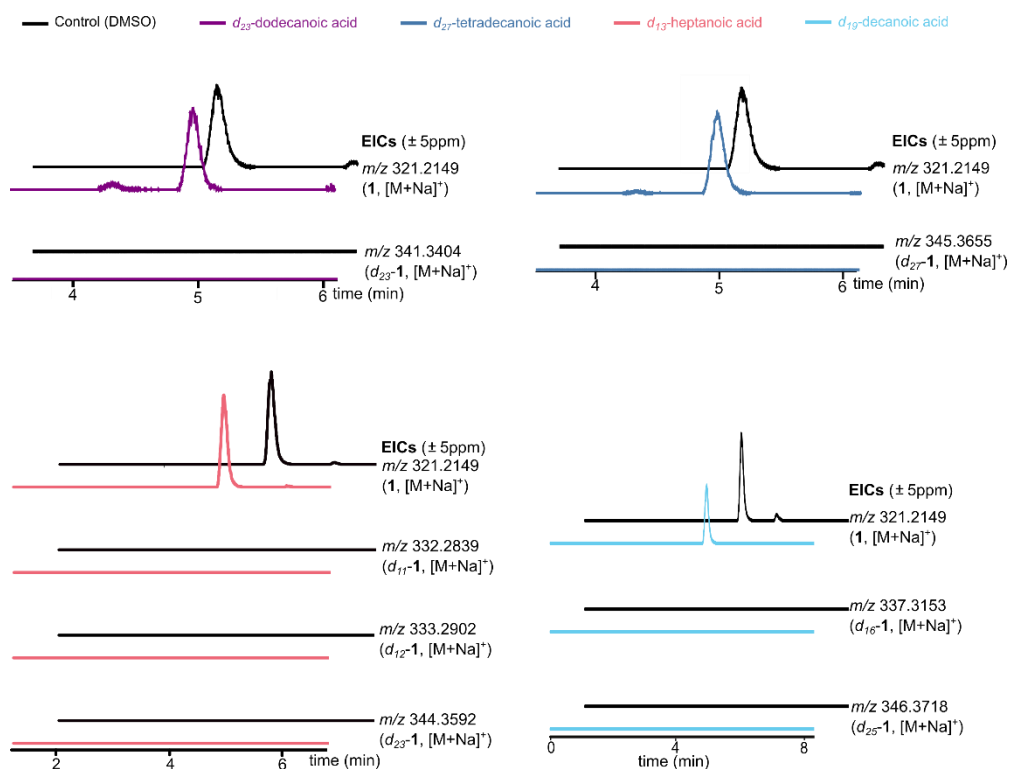


Figure SI.7 – LC-EICs of *1* and labeled version upon supplementation of *Nodularia* sp. LEGE06071 with *d*₁₃-heptanoic acid, *d*₁₉-decanoic acid *d*₂₃-dodecanoic acid and *d*₂₇-tetradecanoic acids. No incorporation of any of the labeled fatty acids was detected.

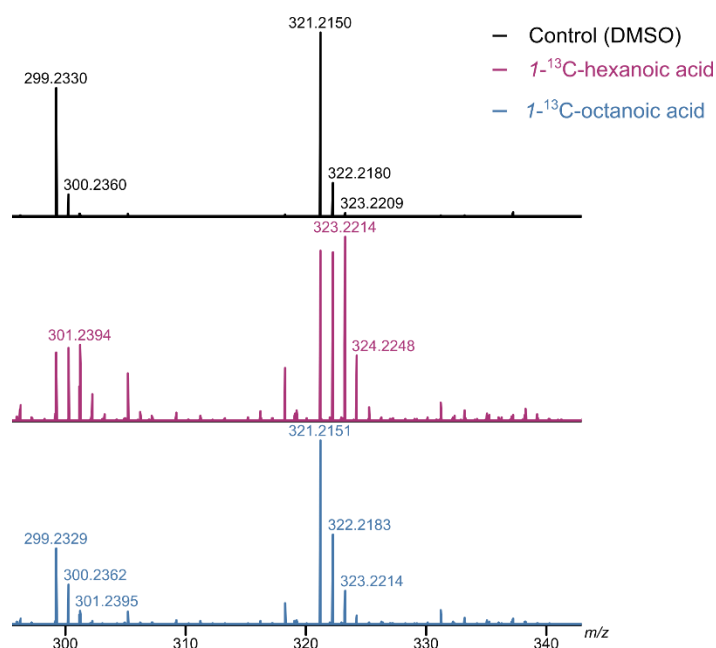


Figure SI.8 - HRMS/MS spectrum of **1** ($[\text{M}+\text{H}]^+$ and $[\text{M}+\text{Na}]^+$ isotope clusters are shown) in supplementation experiments with $1\text{-}^{13}\text{C}$ -FAs. Incorporation of one or two $1\text{-}^{13}\text{C}$ -hexanoic acids was observed by the increase in $\text{M}+1$ and $\text{M}+2$ isotope peaks.

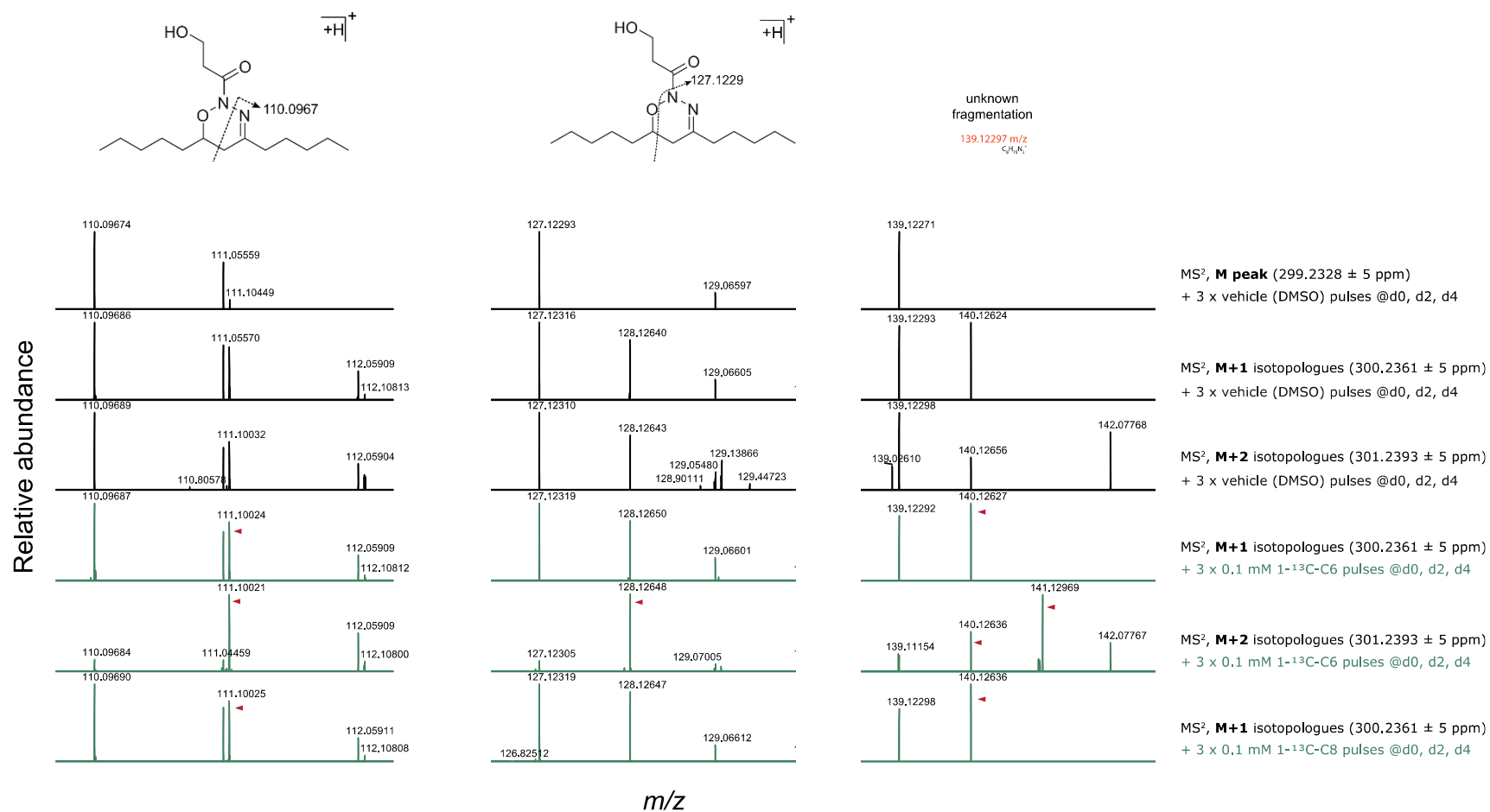


Figure SI.9 – HRMS/MS analysis of the M, M+1 and M+2 isotope peaks for the $[M+H]^+$ species of **1** after supplementation with 1-¹³C-labeled hexanoic and octanoic acids. Shown are selected portions of the MS/MS spectra in which differences could be observed between control (DMSO) and FA-supplemented treatments.

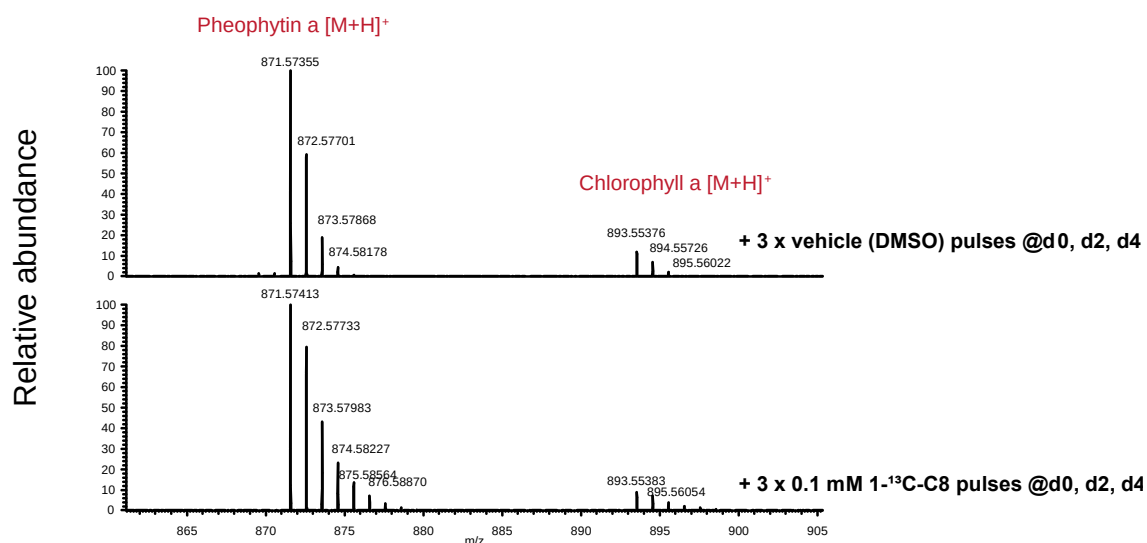


Figure SI.10 – HRMS analysis of chlorophyll a and pheophytin a in the crude extracts of *Nodularia* sp. LEGE 06071, showing non-specific incorporation of ^{13}C into these non-FA-derived pigments when the cyanobacterium was supplemented with $1\text{-}^{13}\text{C}$ -octanoic acid.

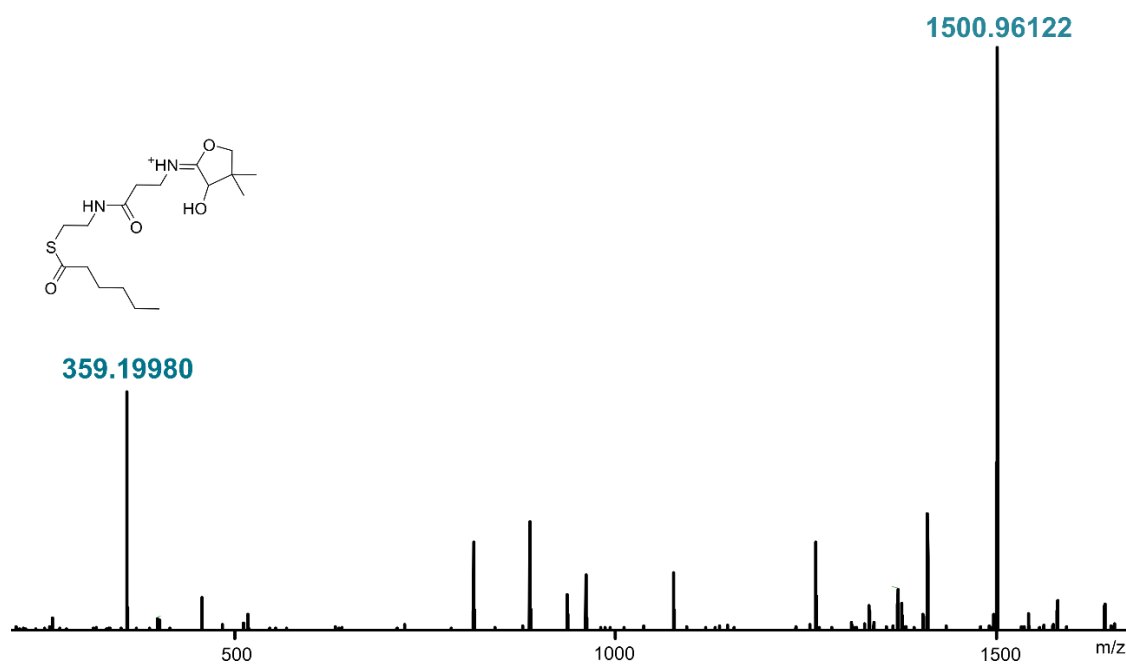


Figure SI.11 - LC-HRMS² spectra of the NoCH fragment loaded with hexanoic acid highlighting the diagnostic C6:0-Ppant fragment (m/z 359).

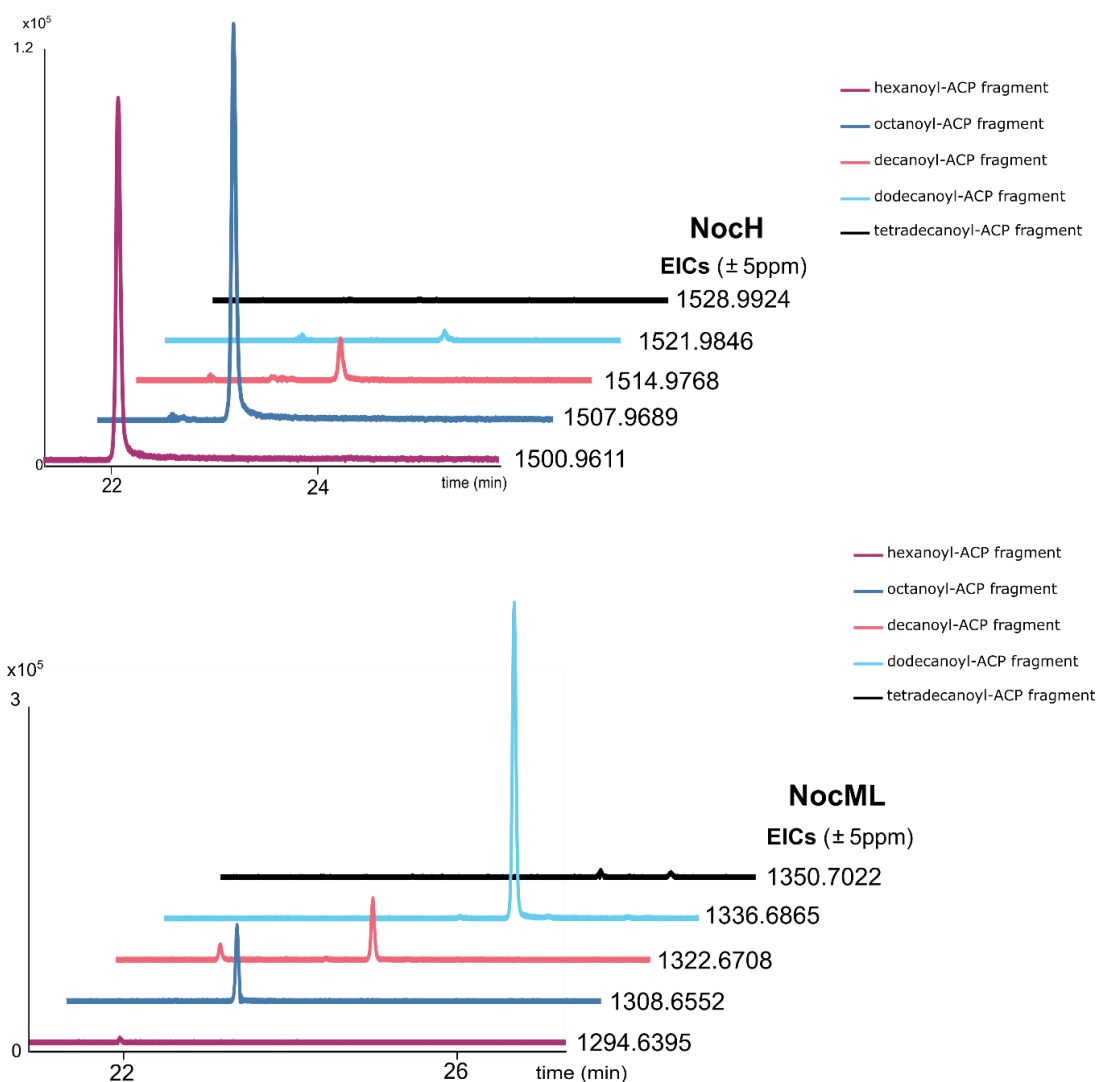


Figure SI.12 – LC-HRMS analysis of the fatty-acyl AMP ligases – NocH and NocL – competition assay with hexanoic, octanoic, decanoic, dodecanoic and tetradecanoic acids. Shown are EICs for the trypsinated ACP fragment loaded with each fatty-acid: $[M+H]^{4+}$ for NocH and $[M+H]^{2+}$ for NocML.

CATATGATGAAATCGAAGCCATCAAAGCGGCGCTCCCGACCAAGAAGCTCACGAACGACGATATTCTGGCGCTGATC
GAGGAATACAGCAAACCGACCTTCACGGGCGATCTGAAGGCGGCCCTCAGCAAGATCAACTTCTTCTCCGCTACAGCGGCAG
CGATACCCGCCACTGGCTGGACAAAGATGAACGCCCAATCGACCTCGTTGCCCGTGCCTTCCAAGAAGCCATCACCGAAGCGA
ACTGCGACAAGGACGACATCGATCTCCTCGTTTACACCGGCATTGGTCGCGGTTTCATCGAACCAGCGGCCGCTACCATATT
GCGGCGAGTCTGGGTCTCCAGAAAGCCGAATGCTTCGACATTCTGGATGCGTGCATGAGCTGGACCCGCACGCTGAATATTGT
GCACAGTCTGCTGCAGACGGGCCGCTATAAGCGCGCGCTGATCGTGAATGCCGAGTGCAATATGCGCTTCGGCGGCCCGGTT
TTCCGACGGTGTTCAGCTCCCAAATGTGGAAGCGATTGAGTGGAGCTTCCAGCCTATACGCTGGGCGAAGCGGCGACGG
CCACCATTCTGAGCTGCGATCCGGACCGCGAGTGGGAGTTCCACTTCAAAGTCGCCAGATCTGGCCGATCTGTGCAACGTT
CCGCTGGATGGCTACGAGGGCTACTGCACCCCAAGTCAGCGCATCGGTCTGAACGGCATCAACAAGTTCACCAGCTTCGGCGT
GGAAGTGCACGCCAATGCCTTCCCGGAATTATCAACATCTTCAAGTGTCTGAAGGCCCGATTGAAGAGATCCGCGCGATCTT
TCCGCATGCGAGCAGCAAACGCGAATGGGATAAAGTGGCGGAAGCGCTGAATCTGAAGCATCTGCTCTGGCATATCTACCCGA
CCTACGGCAATCTCGTGAGCGCGAGCGTTCAACCGGTATTGCGAGCGCGATCGCCGCGGGTCAGATTAAACGCGGTGATCG
TATTGCCGCTGGGTTGGTAGCGCGGGCATGAGTTTTTGC GCGTACAGCTTCATCTATTAAC**CTCGAG**

Figure SII.13 – *E. coli* codon-optimized sequence (5' to 3') of the gene encoding NocG (NocG synthetic gene).
NdeI and XhoI restriction sites are highlighted in green and red, respectively.

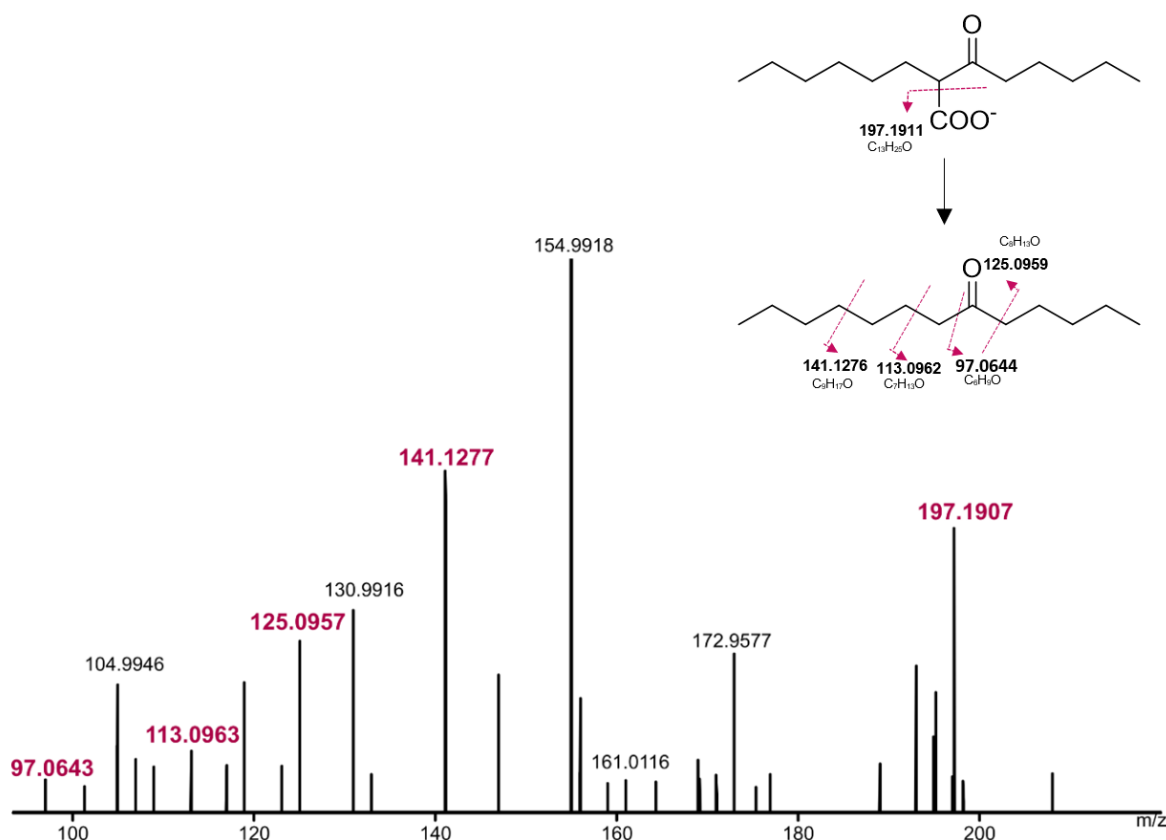


Figure SI.14 - HRMS/MS spectrum of 4 [M-H]⁻ (collision energy 30 eV), supporting its identification as 2-hexyl-3-oxooctanoic acid.

BIOSYNTHESIS OF NOCUOLIN A AND RELATED METABOLITES

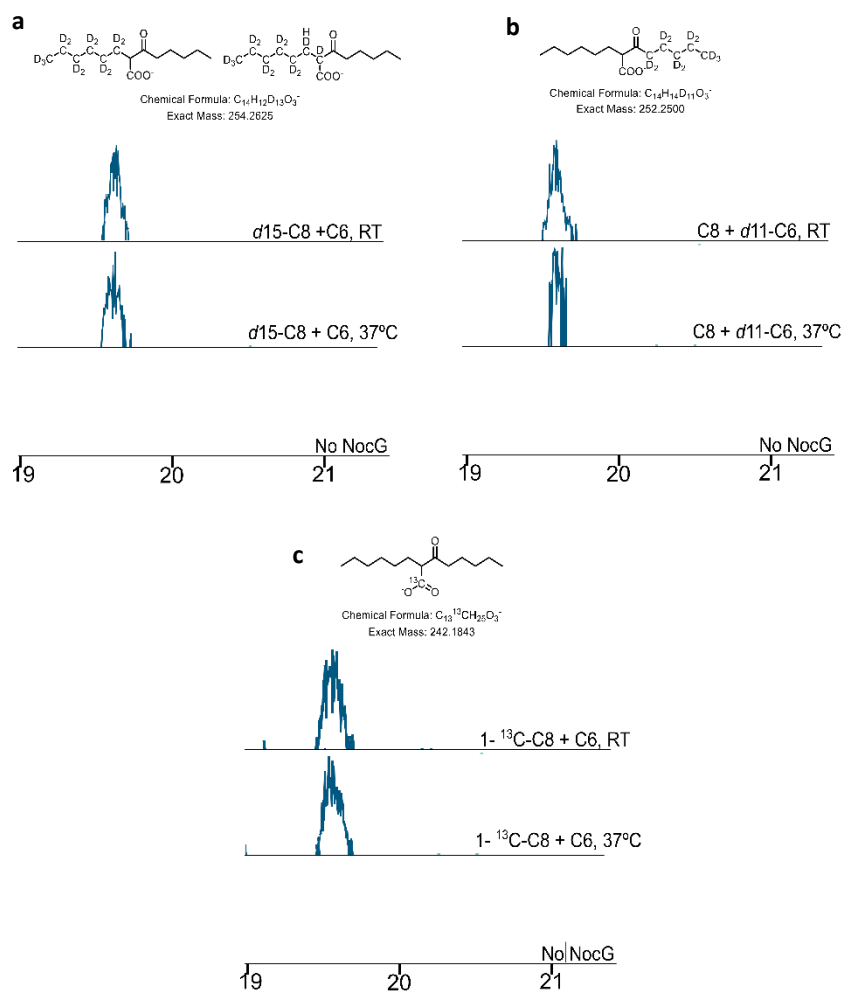


Figure SI.15 – LC-HRMS analysis of NocG and NocH coupled enzymatic assays with different combinations of ^2H - and ^{13}C -labeled FAs. Shown are EICs for d_{13} -4 formed by d_{15} -octanoic acid and unlabeled hexanoic acid (a), d_{11} -4 formed by d_{11} -hexanoic acid and unlabeled octanoic acid (b) and $1-^{13}C$ -4 formed by $1-^{13}C$ -octanoic acid and unlabeled hexanoic acid (c).

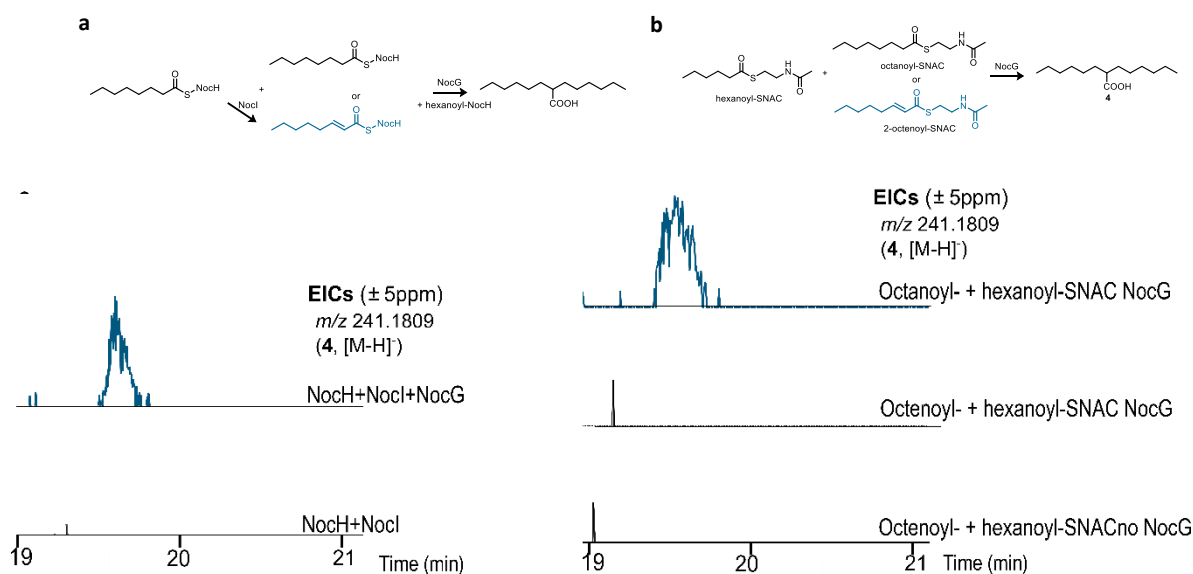


Figure SI.16 – LC-HRMS analysis of NocG enzymatic assays to test the potential role of NocI in the C-C bond forming reaction. Shown are EICs for **4** in NocG enzymatic assays with and without NocI (a) and with octanoyl-SNAC or 2-octenoyl-SNAC (b). These assays allowed to conclude that NocG activity is independent of NocI and does not proceed with an unsaturated substrate analogue and so, NocH must activate hexanoic and octanoic acids, with the corresponding thioesters being directly used by NocG to produce **4**.

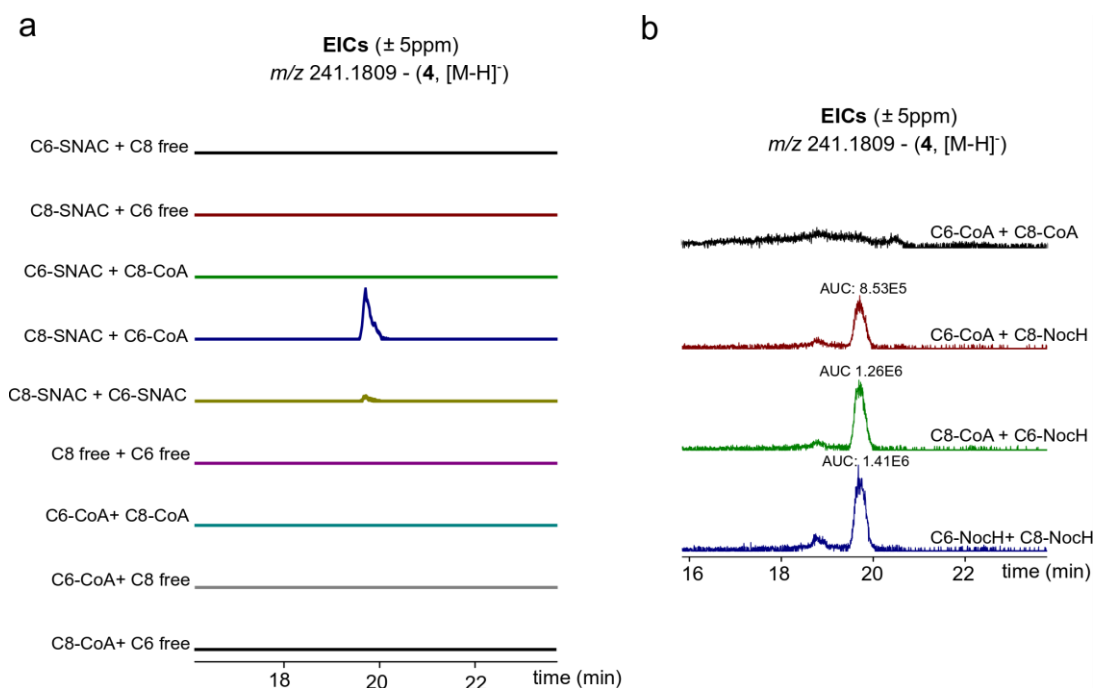


Figure SI.17 – LC-HRMS-derived analysis of NocG enzymatic assays with free fatty-acids – hexanoic and octanoic acids – and/or activated thioesters – hexanoyl-SNAC, octanoyl-SNAC, hexanoyl-NocH, octanoyl-NocH, hexanoyl-CoA and octanoyl-CoA. EICs are shown for compound **4**, showing its detection exclusively when using both thioesters as NocG substrates (a). EICs are shown for compound **4**, showing that activation of both FAs by NocH provides slightly higher production of compound **4** than using the corresponding CoA thioesters, providing further support to the hypothesis that NocH activates both fatty acids prior to NocG activity (b).

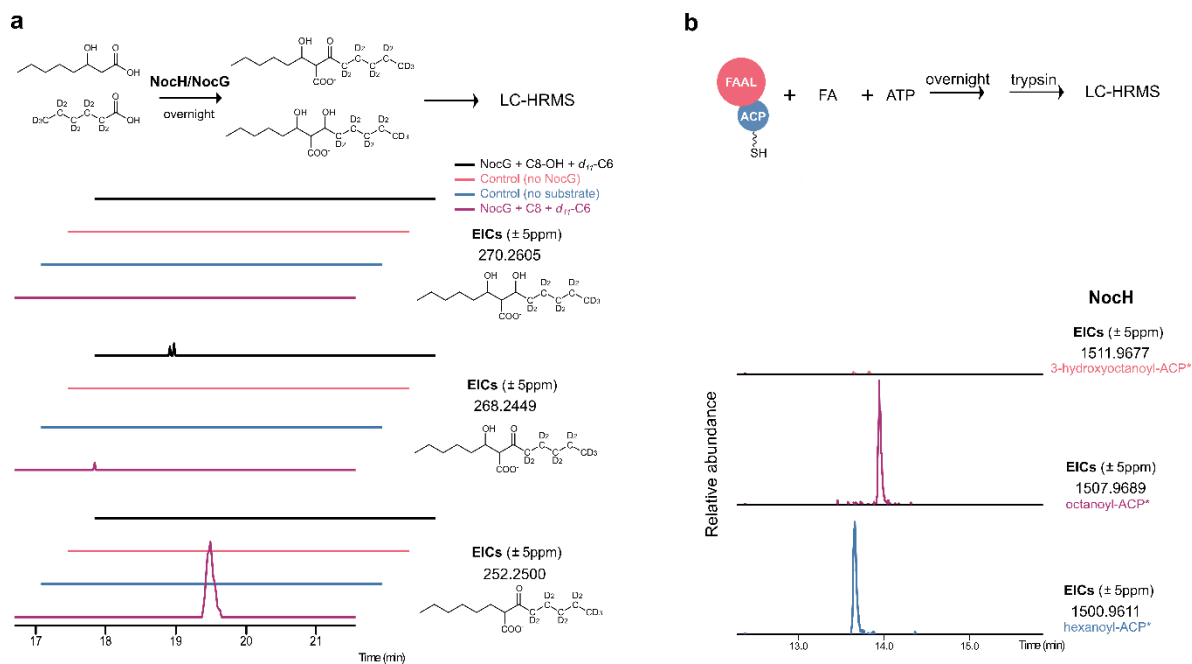


Figure SI.18 – 3-hydroxyoctanoic acid is not a NocG and NocH substrate. (a) EICs for the putative β -ketoacids formed by NocG using 3-hydroxyoctanoic and d_{11} -hexanoic acid in coupled assays with NocH. (b) EICs for the trypsinated ACP fragment loaded with 3-hydroxyoctanoic acid, octanoic acid and hexanoic acid from the NocH competition assay with the 3 fatty-acids: $[M+H]^4+$ for NocH.

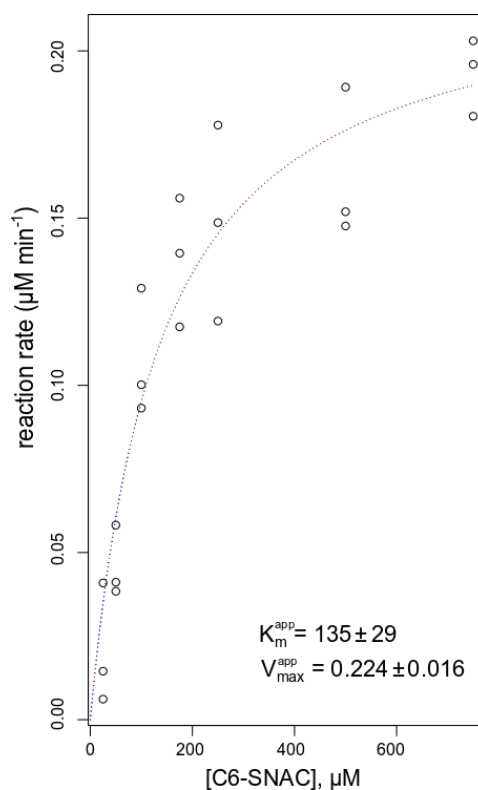


Figure SI.19 - Michaelis-Menten kinetic data for N-His₆-NocG. Apparent K_m and V_{max} were calculated from the non-linear least squares fit of the Michaelis-Menten equation to the kinetic data obtained ($\pm$ standard error, $n=3$ independent assays) for different concentrations of hexanoyl-SNAC and constant octanoyl-SNAC concentration (10 μ M).

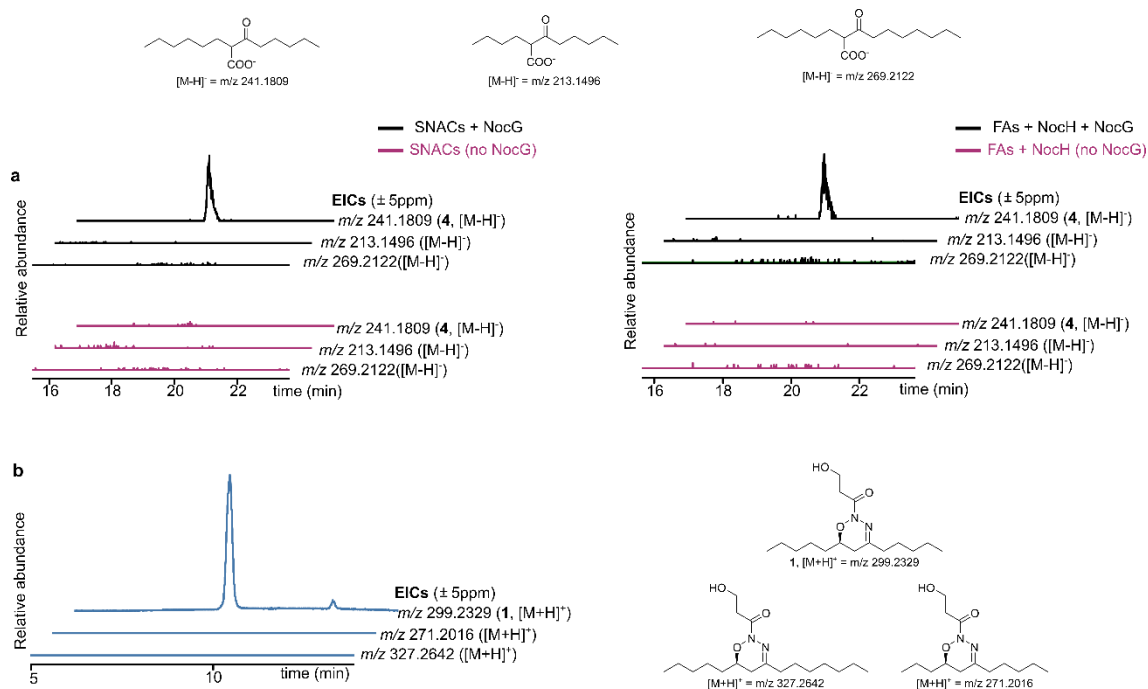


Figure SI.20 – NocG performs a tight control of the β -ketoacid chain length. (a) EICs for different chain lengths of the β -ketoacid putatively formed by NocG (C_{13} , C_{11} and C_{15}) using SNACs or hexanoyl/octanoyl-NocH thioesters, in *in vitro* assays. (b) EICs for **1** and putative shorter (m/z 271) and longer (m/z 327) versions of **1**, from LC-HRMS data of a crude organic extract of *Nodularia* sp. LEGE 06071.

**b**

BIOSYNTHESIS OF NOCUOLIN A AND RELATED METABOLITES

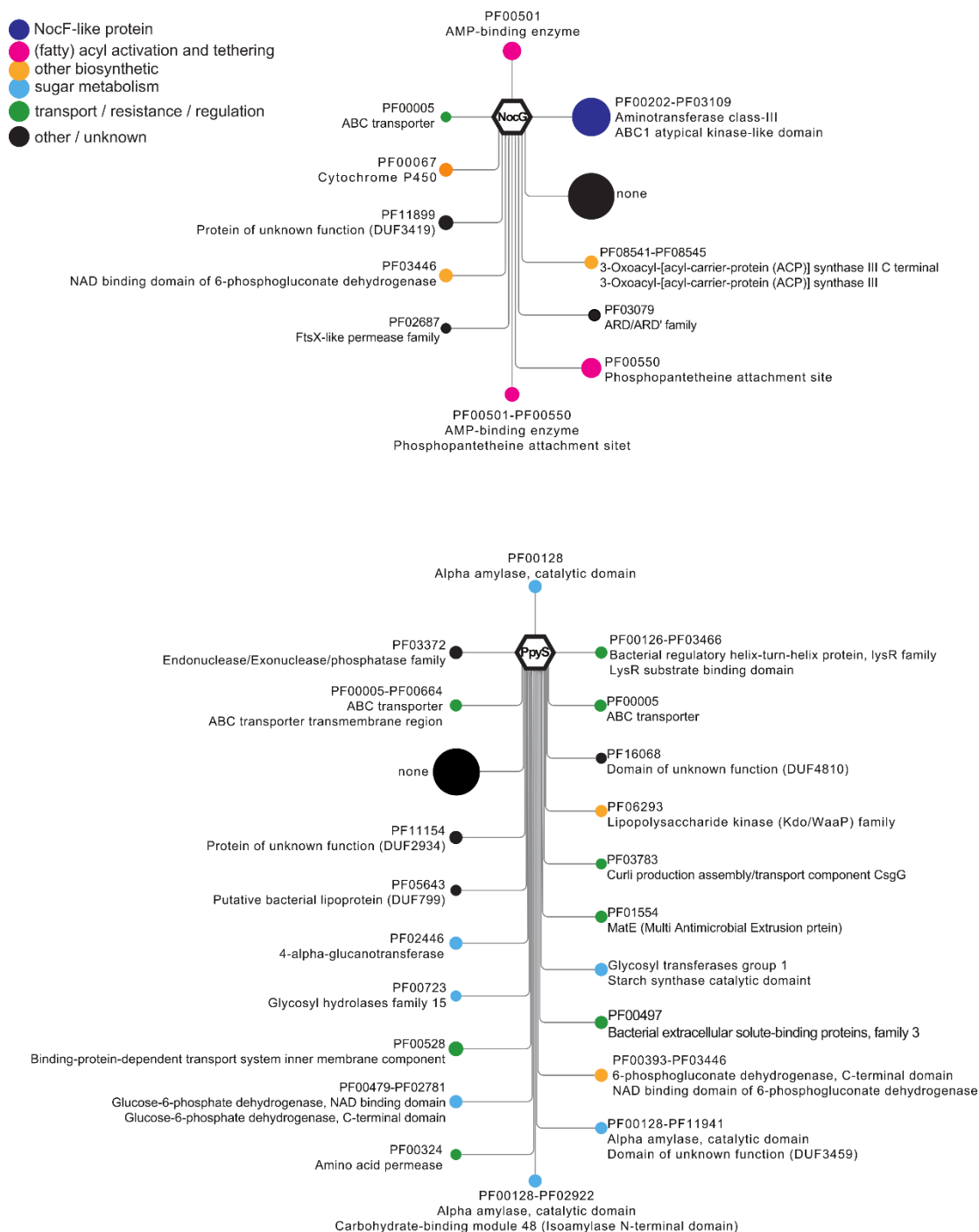


Figure SI.22 - Genome Neighborhood Network (GNN) hub-and-spoke clusters depicting the Pfam domains associated with NocG and PpyS clusters from the initial SSN. The size of each node is proportional to the prevalence of the Pfam domain within the genomic context of the NocG/PpyS homologues from each SSN cluster. The different colors represent a categorization based on common genes found within the associated biosynthetic gene clusters (see legend).

ICBAS
BIOSYNTHESIS OF NOCUOLIN A AND RELATED METABOLITES



Figure SI.23 - Diversity and genomic context of NocG-like enzymes BGCs. Genomic Neighborhood Diagrams (GNDs) showing examples of the different BGC architectures found among the SSN cluster encoding NocG homologues. *noc*-like genes are highlighted in color (see legend), evidencing the frequent co-occurrence of *nocH*, *nocG* and *nocF*.

BIOSYNTHESIS OF NOCUOLIN A AND RELATED METABOLITES

```

NocG      1  ----MKIEAIKALP[KK]TNDIDIALIEYS[KPTFTG]KAA[SK]----NFFLRYS[SD]RRLDMDERPHDLVARNFOEATTEANCDKDTLLIMVTSGRGEFTEPAAHYHAA[SK]
OleA_ZP_0512704 1  -----HFEVWLSLAADAPHSITSK[TS]SD[PT]MDRLGVRENHPE[SH]SGIARR[IN]P[TS]PSDAATHAERKALDSDGIDSRIGVITISVSRDLEPSPACV[SH]ENI
OleA_CA17216 1  MREHFMKITGIAGLPFRFVSNCEVILLVHYYSKELIFH[GT]PTKTLGK[SK]----LNKAGISRR[IL]ADNEK[PM]LIENAFNRAACANIKEDHLLITVSVA[RE]FTEHANS[IV]SK[IL]
OleA_ZP_0110325 1  -----MSGNAKFTLIDTATVSYTAHRAPEVITSASD[DD]SMRHTERUGTQPGHESLAGISERRWPEGHFTF[AR]ABAGRKALAAANIKED[NG]LLINTVS[RD]LEPSSAVT[VH]HLL
OleA_NF_635607 1  -----NFFDVSAGLAHIDAPHILTSK[IN]ERLQCFTHORUGIKTDV[IG]VAGISARELDDQVQASDAATCAARKALIDANIGIKGLLINTVS[RD]LEPSTASTVSGNL
OleA_NF_299252 1  -----NFFDVSAGLAHIDAPHILTSK[IN]ERLQCFTHORUGIKTDV[IG]VAGISARELDDQVQASDAATCAARKALIDANIGIKGLLINTVS[RD]LEPSTASTVSGNL

NocG      112  ST-KQAECDLLDACS[IR]TIRTNVHSH[LT]OTERYKRALVNE[CN]MRFGGVFPTF[FQ]LP-N[E]AIEWS[EP]NTLGP[ER]H[IT]SCDPDREWEP[HK]SRPDLADLCNVPLDG[EG]YTPS
OleA_ZP_0512704 109  GLASICNFDVANACLFINGMDIAR[MT]ERSPIDYALV[SE]SRP[IT]BAT[ER]LDCVGAAC[TR]EFASLTLC[SG]AARAM[IT]RE[AT]P[SE]HTIR[ES]-----VRSAT[IN]NLL[CG]N
OleA_CA17216 117  ST--TCNMDVVDAC[ET]N[MD]I[NS]K[AS]ET[ER]AVN[PE]TMSQNGP[IF]FTF[TS]N--SHSVAIR[ES]ST[IT]SE[AV]T[IS]NESFENK[EM]INRPDLSDL[ST]HS[PT]GR[IT]NDI
OleA_ZP_0110325 115  STSSCN[EP]N[AR]C[LG]F[IN]AQ[RM]N[ES]CHD[AL]V[SE]SSRQPC[KT]T[AR]ASO[AT]A[LA]FA[PA]T[IS]S[AR]CTV[SE]STENAGSH[IT]SE-----VRSAT[IN]NLL[CG]N
OleA_NF_635607 109  STSDHCNFDVANACLFINGMDIAR[MT]ERSPIDYALV[SE]SRP[IT]BAT[ER]LDCVGAAC[TR]EFASLTLC[SG]AARAM[IT]RE[AT]P[SE]HTIR[ES]-----VRSAT[IN]NLL[CG]N
OleA_NF_299252 109  STSGCCNFDVANACLFINGMDIAR[MT]ERSPIDYALV[SE]SRP[IT]BAT[ER]LDCVGAAC[TR]EFASLTLC[SG]AARAM[IT]RE[AT]P[SE]HTIR[ES]-----VRSAT[IN]NLL[CG]N

NocG      230  ORIG[NG]IN---KTFISGV[EH]AN[FPE]R[NT]FKC[KAP]EE[RA]IF[AS]S[RE]W[KAE]D[N]K-HL[NH]I[PT]Y[GN]V[SASV]R[IG]SAIAA[CP]K[KG]DRIAGV[GS]A[CV]SFCAYSF
OleA_ZP_0512704 223  LDRMVTDR---NLSAC[GL]---NAC[TE]ASCST[IT]D[VS]E[HE]DQFTHQVSR[PH]T[SK]K[LG]N[PK]H[AY]P[HN]G[IG]PASV[IV]LAK[EP]EAC[IL]KKGDRIAL[LG]SG[NG]AN[AA]V
OleA_CA17216 223  DVERIGF[EG]KYQ[NS]SAL[H]H[EP]E[ET]N[LF]H[K]IS-HDI[SH]F[IT]H[EG]S[KT]MAR[CP]K[GL]G[IL]D[KH]H[YG]G[OV]G[NI]D[AS]L[EP]G[DA]IN[CH]E[IL]N[OL]CM[GW]G[SG]V[FA]AMTIF
OleA_ZP_0110325 229  LDRMVTDR---NLSAC[GL]---NAC[TE]ASCST[IT]D[VS]E[HE]DQFTHQVSR[PH]T[SK]K[LG]N[PK]H[AY]P[HN]G[IG]PASV[IV]LAK[EP]EAC[IL]KKGDRIAL[LG]SG[NG]AN[AA]V
OleA_NF_635607 223  LDRMVTDR---NLSAC[GL]---NAC[TE]ASCST[IT]D[VS]E[HE]DQFTHQVSR[PH]T[SK]K[LG]N[PK]H[AY]P[HN]G[IG]PASV[IV]LAK[EP]EAC[IL]KKGDRIAL[LG]SG[NG]AN[AA]V
OleA_NF_299252 223  LDRMVTDR---NLSAC[GL]---NAC[TE]ASCST[IT]D[VS]E[HE]DQFTHQVSR[PH]T[SK]K[LG]N[PK]H[AY]P[HN]G[IG]PASV[IV]LAK[EP]EAC[IL]KKGDRIAL[LG]SG[NG]AN[AA]V

NocG      346  [V]
OleA_ZP_0512704 337  [V]
OleA_CA17216 352  [K]
OleA_ZP_0110325 338  [A]
OleA_NF_635607 337  [V]
OleA_NF_299252 337  [V]

```

Figure SI.24 - Protein alignment (Geneious alignment - Global alignment with free ends and gaps) of NocG, OleA (CAE17216) and OleA homologues. Catalytic residues are highlighted with a *.

```

PpyS      1  ----MKI[GI]ALP[FR]FV[N]C[CV]ILV[BY]SE[ET]E[SD]PKT[KL]CKM[KN]KAGIS[RY]WLADNE[EM]EL[EA]F[NH]A[AC]ANIKEDIDLLIY[SV]ARGFTEPANS[SV]S
PyrS      1  LVPRGSHMLFQNSHAGAHIDAPHILASREINERQ[SK]----PYDR[IG]IKTDV[LG]---DVAGIHARELWQCYQASDAATCAARKALIDANIGIEH[GL]PLNFSVSRDYEPSTASIVS
NocG      1  ----MKIEAIKALP[KK]TNDIDIALIEYS[KPTFTG]KAA[SK]----NFFLRYS[SD]RRLDMDERPHDLVARNFOEATTEANCDKDTLLIMVTSGRGEFTEPAAHYHAA[SK]

PpyS      109  KALGL[IT]---CRNMDVVDAC[ET]N[MD]I[NS]K[AS]ET[ER]AVN[PE]TMSQNGP[IF]FTF[TS]N--SHSVAIR[ES]ST[IT]SE[AV]T[IS]NESFENK[EM]INRPDLSDL[ST]HS[PT]GR[IT]NDI
PyrS      113  SNLGVSDHCNFDVANACLFINGMDIAR[MT]ERSPIDYALV[SE]SRP[IT]BAT[ER]LDCVGAAC[TR]EFASLTLC[SG]AARAM[IT]RE[AT]P[SE]HTIR[ES]-----VRSAT[IN]NLL[CG]N
NocG      109  STSGCCNFDVANACLFINGMDIAR[MT]ERSPIDYALV[SE]SRP[IT]BAT[ER]LDCVGAAC[TR]EFASLTLC[SG]AARAM[IT]RE[AT]P[SE]HTIR[ES]-----VRSAT[IN]NLL[CG]N

PpyS      226  NLDVPRIGP[RG]SKYQ[NS]SAL[H]H[EP]E[ET]N[LF]H[K]IS-HDI[SH]F[IT]H[EG]S[KT]MAR[CP]K[GL]G[IL]D[KH]H[YG]G[OV]G[NI]D[AS]L[EP]G[DA]IN[CH]E[IL]N[OL]CM[GW]G[SG]V[FA]AMTIF
PyrS      227  RGNLDRMVTDR---NLSAC[GL]---NAC[TE]ASCST[IT]D[VS]E[HE]DQFTHQVSR[PH]T[SK]K[LG]N[PK]H[AY]P[HN]G[IG]PASV[IV]LAK[EP]EAC[IL]KKGDRIAL[LG]SG[NG]AN[AA]V
NocG      227  --LPSQRTGLN[CG]--INKFTS[EG]VD[EH]AN[FPE]R[NT]FKC[KAP]EE[RA]IF[AS]S[RE]W[KAE]D[N]K-HL[NH]I[PT]Y[GN]V[SASV]R[IG]SAIAA[CP]K[KG]DRIAGV[GS]A[CV]SFCAYSF

```

Figure SI.25 – Protein alignment (Geneious alignment - Global alignment with free ends and gaps) of NocG, PpyS (AGO97060) and PyrS (WP_007967127). Catalytic residues are highlighted with a *.

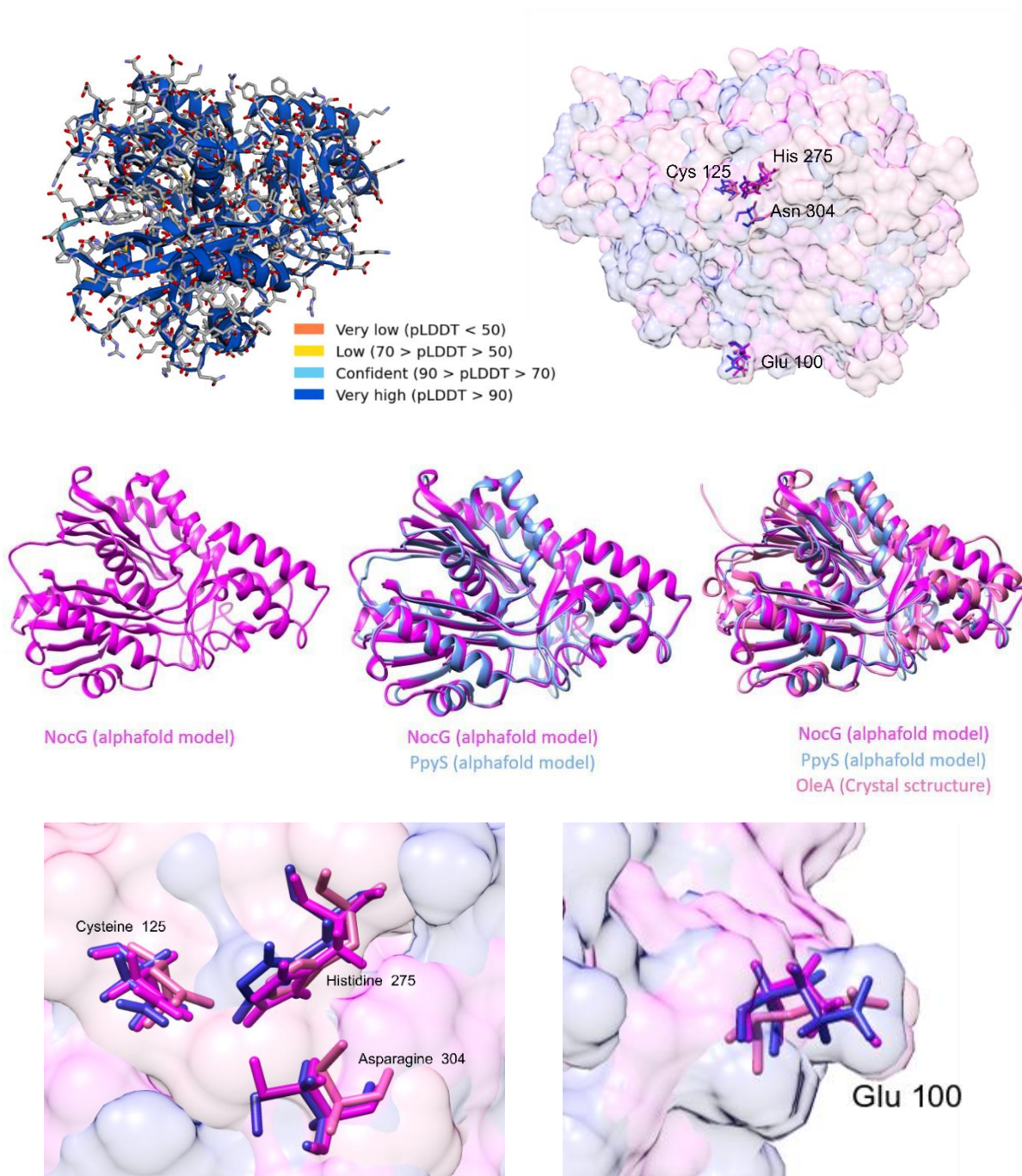


Figure SI.26 – Structural model of NocG (LEGE06071) and comparison with PpyS model and OleA (crystal structure). (a) AlphaFold model of NocG presented very high confidence level. (b) Key catalytic residues of NocG. Glutamate 100 will interact with chain B catalytic site of the homodimer, performing the deprotonation of octanoyl-thioester and initialing the enzymatic reaction. (c) Comparison of NocG with PpyS and OleA (NocG – magenta, PpyS – blue, OleA – pink). (d) Close-up of NocG active site evidencing the similar position of the catalytic residues. (e) Close-up of Glutamate 100 that act as a base, initializing the reaction mechanism.

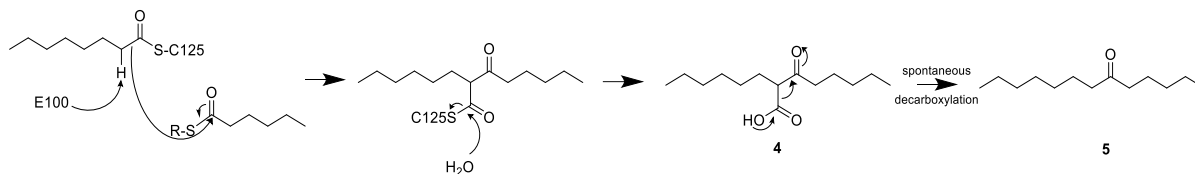


Figure SI.27 – Proposed mechanism for NocG activity.

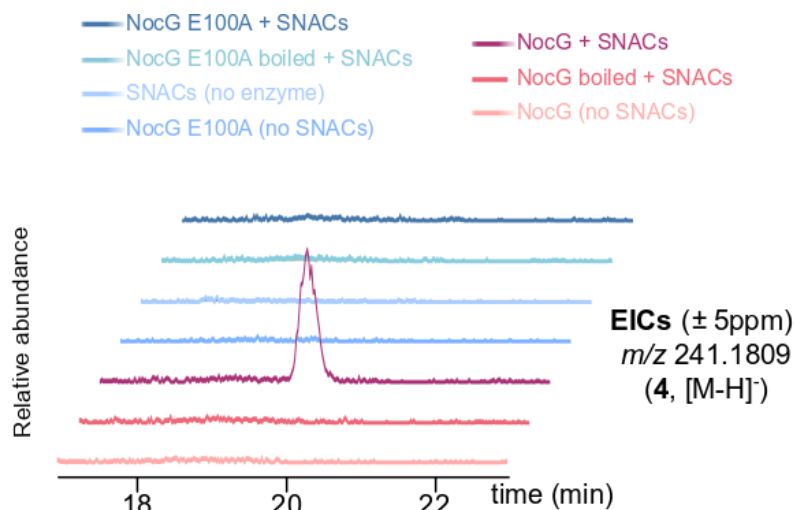


Figure SI.28 – NocG E100A mutation results in loss of detectable product formation. LC-MS/MS-derived analysis of NocG and NocG E100A enzymatic assays with hexanoyl-SNAC and octanoyl-SNAC. EICs are shown for compound 4.

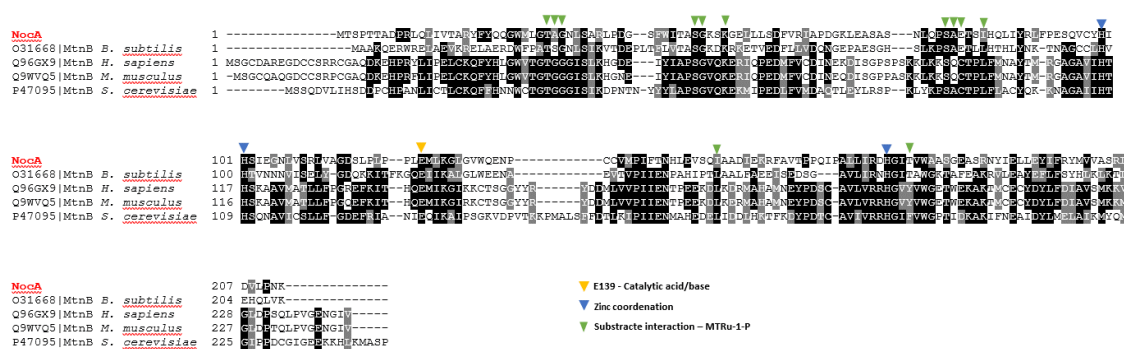


Figure SI.29 – Protein alignment (Geneious alignment – Global alignment with free ends and gaps) of NocA and characterized MtnB enzymes. Catalytic, metal coordination and substrate interaction residues are highlighted with yellow, blue and green arrows, respectively. Key residues were identified based on the literature 15.

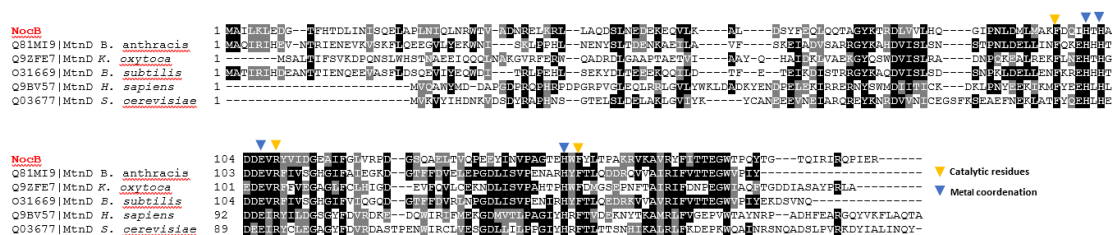


Figure SI.30 – Protein alignment (Geneious alignment – Global alignment with free ends and gaps) of NocB and characterized MtnD enzymes. Catalytic and metal coordination residues are highlighted with yellow and blue arrows, respectively.

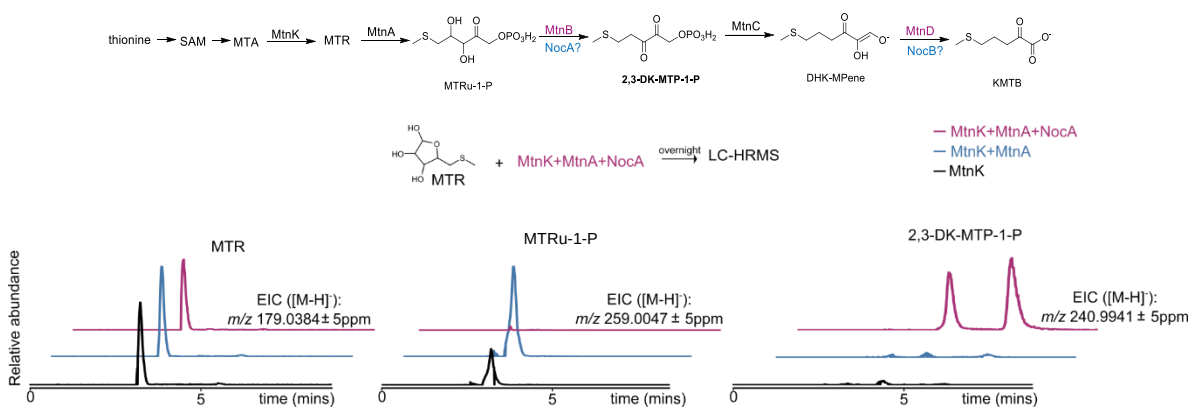


Figure SI.31 – LC-HRMS-derived analysis of NocA coupled enzymatic assay with MtnK and MtnA from *B. subtilis*. Shown are EICs of 2,3-DK-MTP-1-P produced by NocA activity as a methylthioribulose-1-phosphate dehydratase and reaction intermediates.

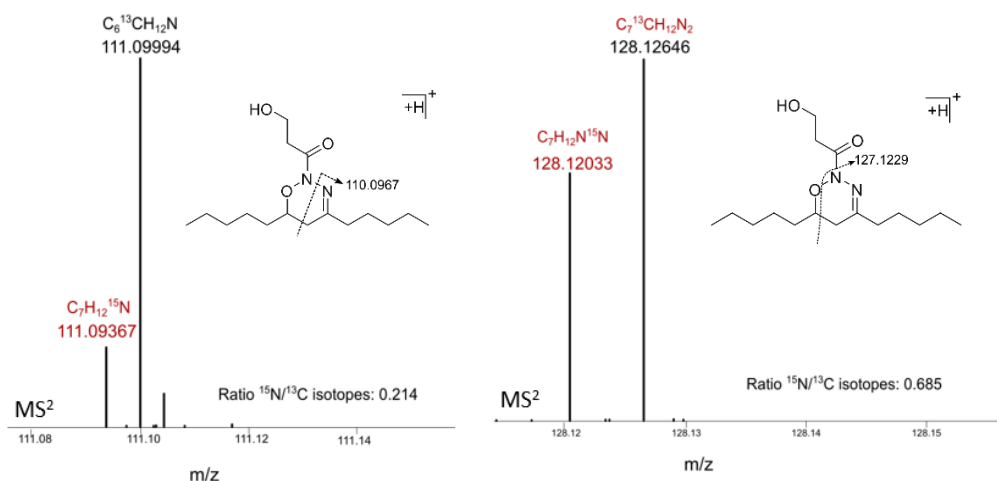
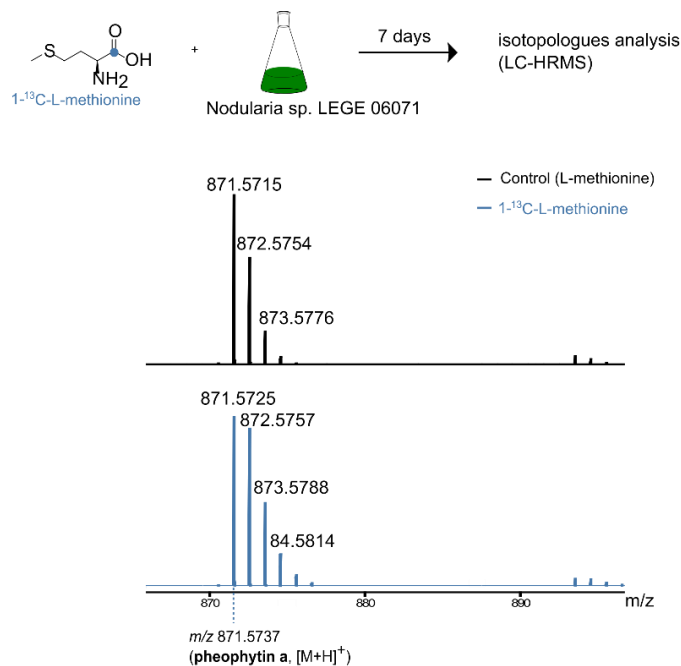


Figure SI.32 – HRMS/MS data of peak 303.2459 ($C_{12}^{13}C_4H_{31}O_3N^{15}N$) in the ^{15}N , $^{13}C_5$ -L-methionine feeding experiment from *Nodularia* sp. LEGE 06071. $^{15}N/^{13}C$ ratio in fragments containing both nitrogen atoms (m/z 128.123) is 3.2 fold higher than the same ratio in fragments containing N-1 (m/z 110.097) supporting the incorporation of ^{15}N from ^{15}N -methionine in N-2.



Control (nocuolin A)			Control (pheophytine a)			Control (MGDG 36:4)		
321.2149	322.2182	322.2182/321.2149	871.5737	872.5800	872.5800/871.5737	845.5327	846.5389	846.5389/845.5327
1.16E+08	1.99E+07	0.171	3.55E+10	2.17E+10	0.612	1.07E+08	5.19E+07	0.483
8.36E+07	1.44E+07	0.172	2.68E+10	1.66E+10	0.619	1.14E+08	5.59E+07	0.492
		0.172			0.616			0.488
Feeding (nocuolin A)			Feeding (pheophytine a)			Feeding (MGDG 36:4)		
321.2149	322.2182	322.2182/321.2149	871.5737	872.5800	872.5800/871.5737	845.5327	846.5389	846.5389/845.5327
4.93E+07	1.20E+07	0.243	1.75E+10	1.62E+10	0.923	7.20E+07	7.45E+07	1.036
8.59E+07	2.14E+07	0.249	2.75E+10	2.5E+10	0.906	4.89E+07	5.58E+07	1.139
		0.246			0.915			1.088
Feeding/Control (averages)		1.431			1.486			2.230

Figure SI.33 - The 3-hydroxypropanoyl moiety of **1** is derived from carbons C-2, C-3 and C-4 from L-methionine. HRMS spectra of pheophytine a and its stable isotope labeled versions following supplementation with $1\text{-}^{13}\text{C}$ -L-methionine, confirming the non-incorporation of C-1 of methionine. Calculations of the ^{13}C scrambling upon $1\text{-}^{13}\text{C}$ -L-methionine feeding. The areas of the **1**, $d_1\text{-1}$, pheophytine a, d_1 -pheophytine a, monogalactosyldiacylglycerol (MGDG) 36:4 and d_1 -MGDG peaks.

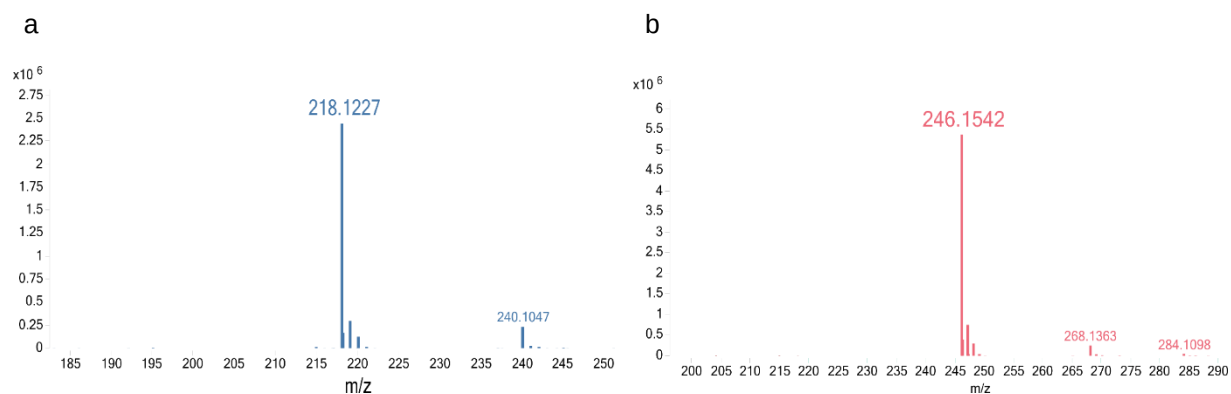


Figure SI.34 - HRMS spectrum of synthetic compounds hexanoyl-SNAC (a) and octanoyl-SNAC (b) confirming the expected masses (calc. m/z octanoyl-SNAC: 246.1522; calc. m/z hexanoyl-SNAC: 218.1209)

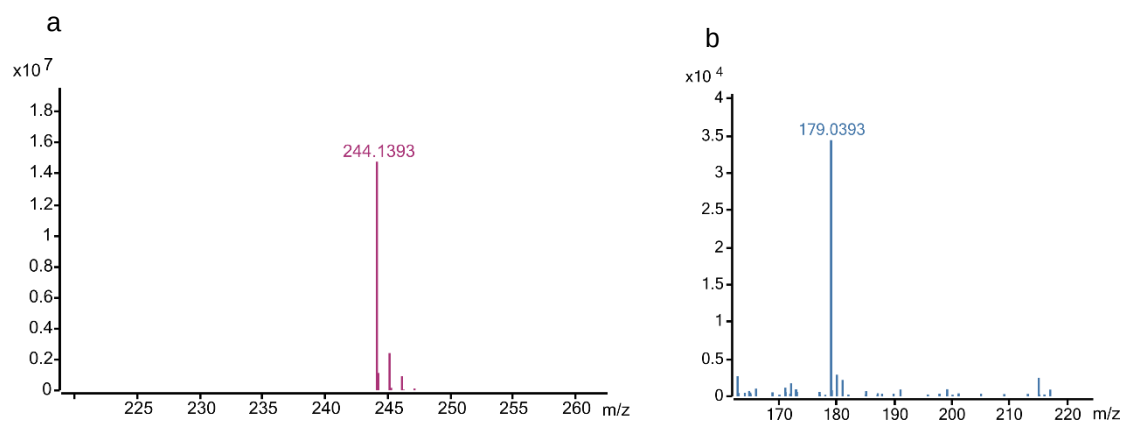


Figure SI.35 - HRMS spectrum of synthetic compounds of 2-octenoyl-SNAC (a) and MTR (b) confirming the expected masses (calc. m/z 2-octenoyl-SNAC: 244.1366; calc. m/z MTR: 179.0384)

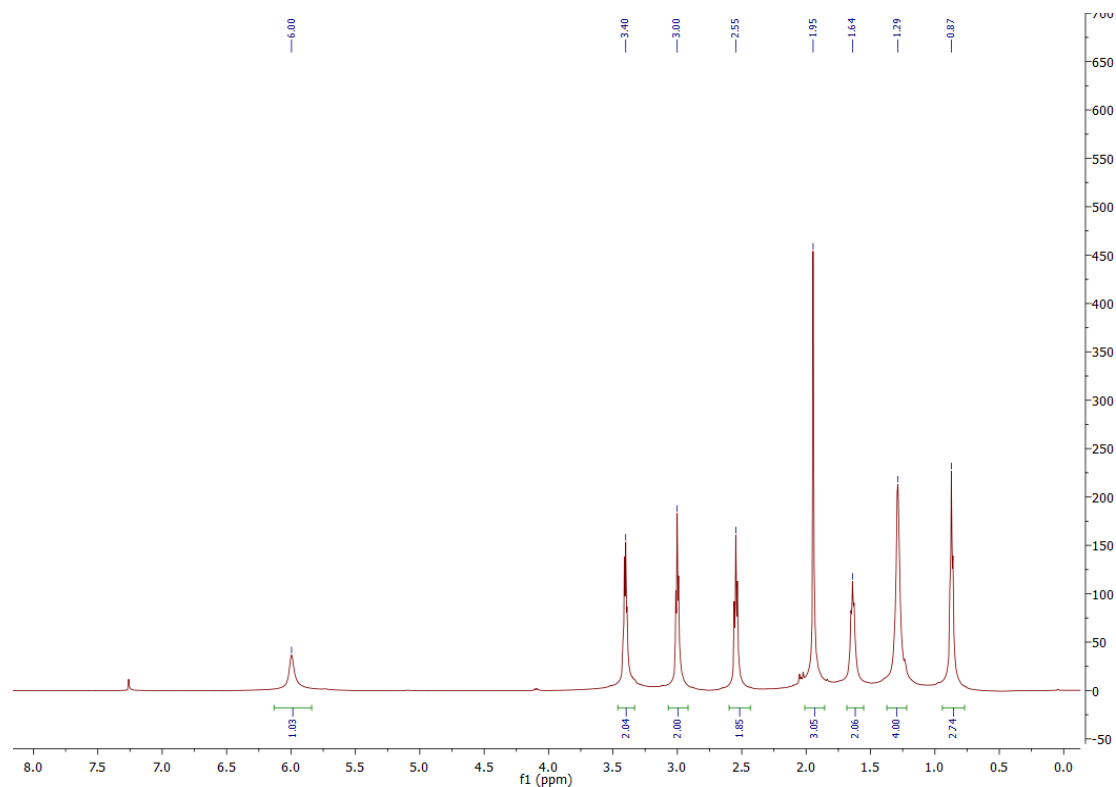


Figure SI.36 - ^1H NMR (400 MHz) of hexanoyl-SNAC in CD_3Cl .

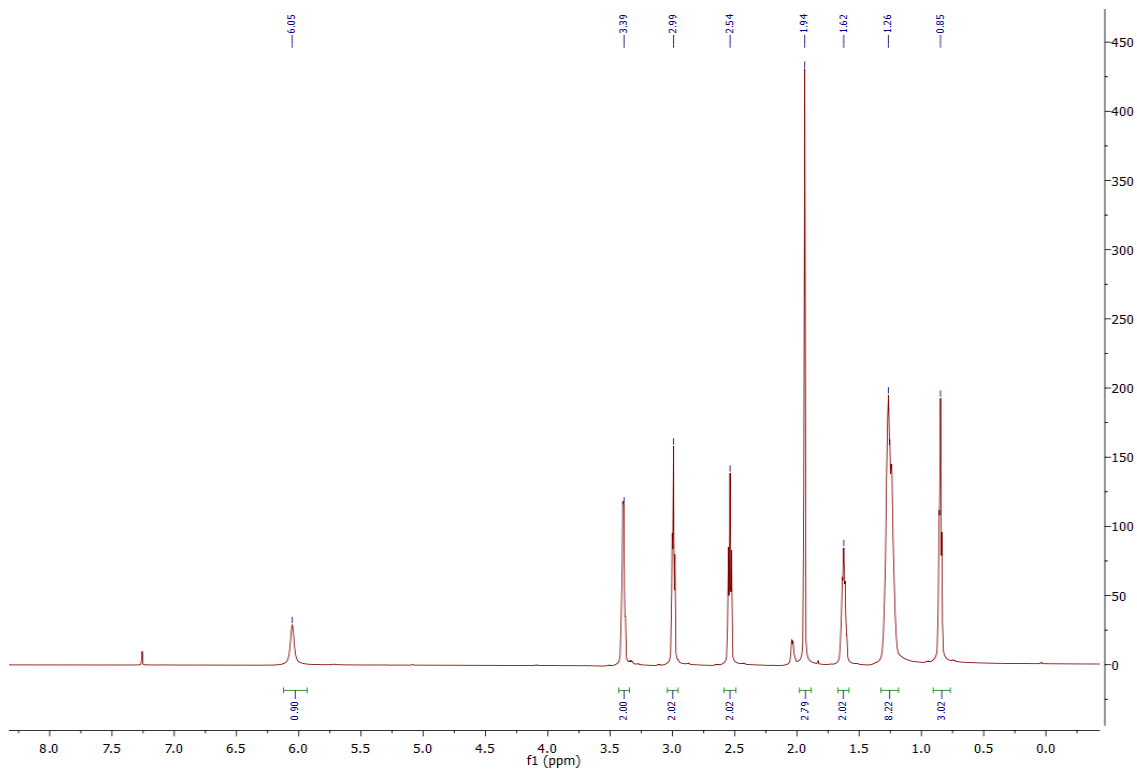


Figure SI.37- ¹H NMR (400 MHz) of octanoyl-SNAC in CD₃Cl.

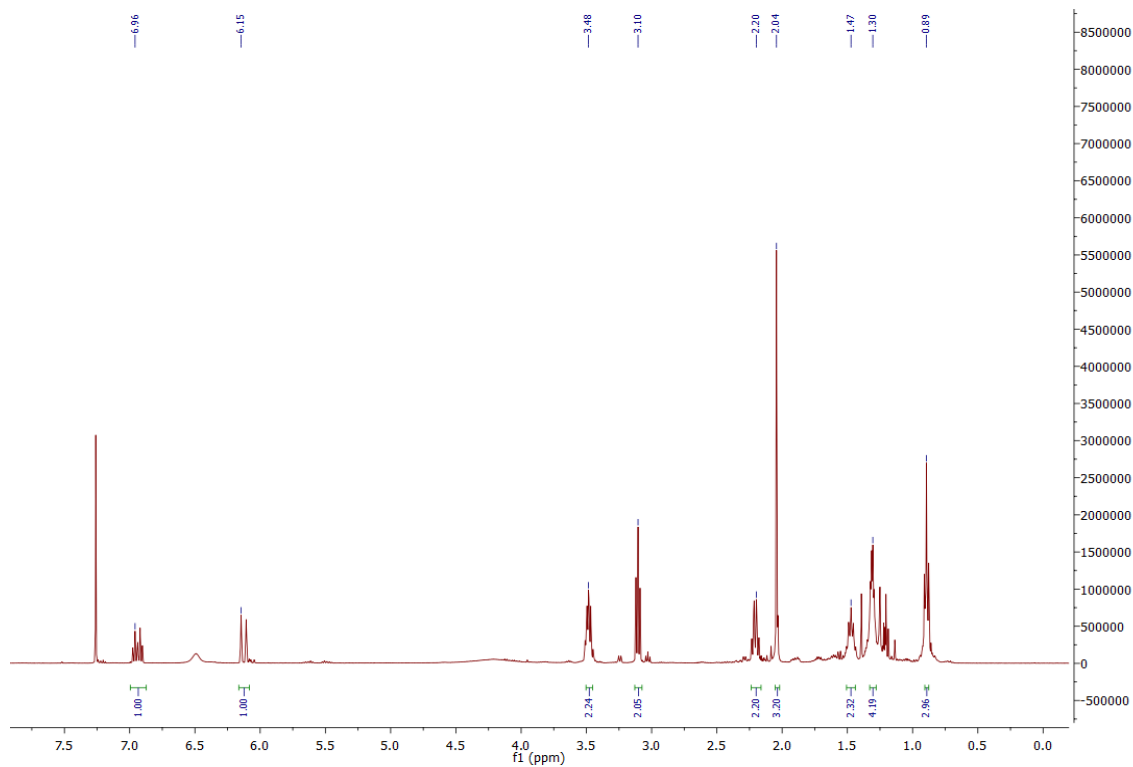


Figure SI.38 - ¹H NMR (400 MHz) of 2-octenoyl-SNAC in CD₃Cl.

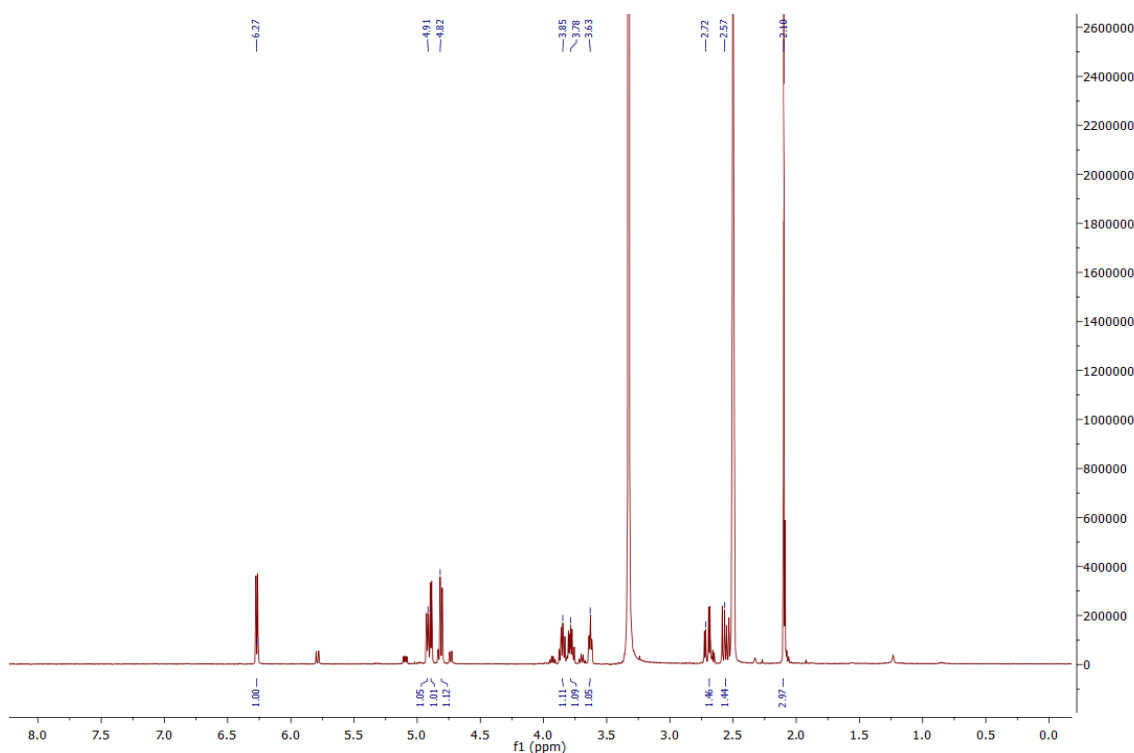


Figure SI.39 - ^1H NMR (400 MHz) of MTR in DMSO.

References

1. Kotai, J., *Instructions for Preparation of Modified Nutrient Solution Z8 for Algae*. Oslo: Norwegian Institute for Water Research: 1972.
2. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010**, *11*, 395-406.
3. 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.
4. Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T., Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998**, *37* (6), 1585-1595.
5. 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.
6. Miller, M. A.; Pfeiffer, W.; Schwartz, T., Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 2010; pp 1-8.

7. Zallot, R.; Oberg, N.; Gerlt, J. A., The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, 58 (41), 4169-4182.
8. Jumper, J.; Evans, R.; Pritzel, A.; Green, T., et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583-589.
9. Sun, Y.; Feng, Z.; Tomura, T.; Suzuki, A., et al., Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Sci. Rep.* **2016**, 6, 22091-22102.
10. Quade, N.; Huo, L.; Rachid, S.; Heinz, D. W.; Muller, R., Unusual carbon fixation gives rise to diverse polyketide extender units. *Nat. Chem. Biol.* **2011**, 8 (1), 117-124.
11. 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), 286-90.
12. Voráčová, K.; Hájek, J.; Mareš, J.; Urajová, P., et al., The cyanobacterial metabolite nocuolin a is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLoS One* **2017**, 12 (3), e0172850.
13. Ashida, H.; Saito, Y.; Kojima, C.; Yokota, A., Enzymatic characterization of 5-methylthioribulose-1-phosphate dehydratase of the methionine salvage pathway in *Bacillus subtilis*. *Biosci., Biotechnol., Biochem.* **2008**, 72 (4), 959-967.
14. 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.
15. 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.

Chapter II

Manuscript

Diversification of the 6-tridecanone scaffold during nocuolin A biosynthesis

Diversification of the 6-tridecanone scaffold during nocuolin A biosynthesis

Teresa P. Martins^{1,2}, Nathaniel R. Braffman³, Marco Pretro¹, Nathaniel Glasser³, Paulo Oliveira^{1,4}, Emily P. Balskus^{3,5}, Pedro N. Leão¹

¹*Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n 4450-208 Matosinhos, Portugal*

²*School of Medicine and Biomedical Sciences, University of Porto, Rua Jorge de Viterbo Ferreira 228, 4050-313, Porto, Portugal*

³*Department of Chemistry & Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, 02138 MA, USA*

⁴*Department of Biology, Faculty of Sciences, University of Porto, 4169-00 Porto, Portugal*

⁵*Howard Hughes Medical Institute, Harvard University, Cambridge, MA USA*

* corresponding authors

Abstract

Nocuolin A is a cyanobacterial natural product with an unprecedented alkylated 1,2,3-oxadiazine moiety and a 3-hydroxypropanoyl moiety. The unique structure of nocuolin A is likely the result of enzymatic steps of uncommon biosynthetic logic or a unique combination of known biochemistry. In our previous study, we have established the first biosynthetic steps in nocuolin A biosynthesis, leading to the key intermediate 6-tridecanone. Here, we report our efforts to elucidate the following enzymatic reactions that diversify this scaffold, including the incorporation of the first nitrogen and second oxygen atoms. Supplementation experiments and biochemical assays revealed that NocE (a C-monooxygenase) hydroxylates 6-tridecanone through direct C-H bond activation. This substrate undergoes an unusual transamination reaction by NocF, which likely results in the formation of an imine intermediate. Both NocE and NocF exhibit unique reactivities and have low sequence similarity to characterized enzymes. We further suggest that NocD is a methyltransferase involved in the methylation of nocuolin A to generate nocuolin B, a previously unreported nocuolin analogue. Additionally, we have cloned the entire *noc* pathway in *E. coli* and characterized two putative pathway intermediates from its heterologous expression. These findings pave the way for the future elucidation of the enzymatic basis of 1,2,3-oxadiazine ring formation.

Introduction

Cyanobacteria generate a variety of natural products (NPs) with pharmacologically-relevant bioactivities¹ and high structural diversity ranging from small alkaloids like carriebowlinol² to large post-translationally modified peptides like microviridins.³ Many of these cyanobacterial NPs feature rare modifications including unusual halogenations (trichlorination, iodination)^{4,5}, various methylations,⁶ unusual chain terminations⁷, hydroxylations^{8,9} as well as uncommon cyclic moieties such as thiazoline, oxazole, oxadiazine and resorcinol rings.^{8,10-12} These modifications result from the action of biosynthetic enzymes, and are often challenging to achieve by current synthetic methodology. These rare modifications have guided the discovery of several remarkable cyanobacterial biosynthetic enzymes in recent years and firmly established the potential of cyanobacteria as a source of novel enzymes.¹³⁻¹⁶

Nocuolin A (**1**) is a structurally unusual cyanobacterial NP featuring an unprecedented 1,2,3-oxadiazine,¹⁷ a heterocycle unique among NPs and which has not been prepared synthetically to date (Fig. II.1a).^{17,18} In **1**, this ring structure incorporates the middle portion of a linear C₁₃ alkyl chain and is decorated with a 3-hydroxypropanoyl moiety.¹⁷ In our previous study,¹⁹ we have established the formation of the C₁₃ alkyl chain from hexanoic and octanoic acids. Each of these fatty acids was found to be activated by a fatty acyl-AMP ligase, NocH, and condensed by NocG, the first characterized representative of a new group of ketosynthases (Fig. II.1b).¹⁹ The action of NocG gives rise to a C₁₃- β -ketoacid which is predicted to undergo spontaneous decarboxylation, yielding 6-tridecanone. Additionally, we proved that the 3-hydroxypropanoyl moiety derives from L-methionine through yet unknown enzymatic steps.¹⁹ These results confirmed that the unique structure of **1** is generated by previously unrecognized enzymatic chemistry. Still, many enzymes encoded by the *noc* BGC are still uncharacterized and the enzymatic steps leading to **1** from the 6-tridecanone intermediate are still unknown. Currently uncharacterized Noc enzymes include a predicted N-monooxygenase (NocE), an aminotransferase (NocF), a Rieske oxygenase (NocJ), a hypothetical enzyme (NocK) and a methyltransferase (NocD).^{17,19} Based on the structure of **1**, we hypothesized that the monooxygenase or Rieske oxygenase could be involved in the activation of C-4 and the aminotransferase in the transfer of a nitrogen atom onto C-6 to diversify the key 6-tridecanone intermediate. However, no clear biosynthetic hypothesis could be rationalized for the methyltransferase NocD and the hypothetical enzyme NocK.

In this study, we focused on the biosynthetic steps that might occur following 6-tridecanone formation. Using stable-isotope labeled precursor supplementation experiments and *in vitro* enzymatic assays, we show that 6-tridecanone is hydroxylated in

position C-4 through direct C-H bond activation performed by the monooxygenase NocE. The generated 8-hydroxytridecan-6-one is then transformed by the aminotransferase NocF, which likely catalyzes the unusual conversion of the ketone to an imine, introducing the first nitrogen atom into **1**. We obtained additional evidence suggesting that NocD is a methyltransferase involved in the methylation of nocuolin A to generate nocuolin B, the first nocuolin analogue to be reported. Additionally, using Direct Pathway Cloning (DiPaC)^{20,21} we have cloned the entire *noc* pathway into *E. coli*, which led to the detection and isolation of two metabolites that can provide guidance for future dissection of the *noc* pathway.

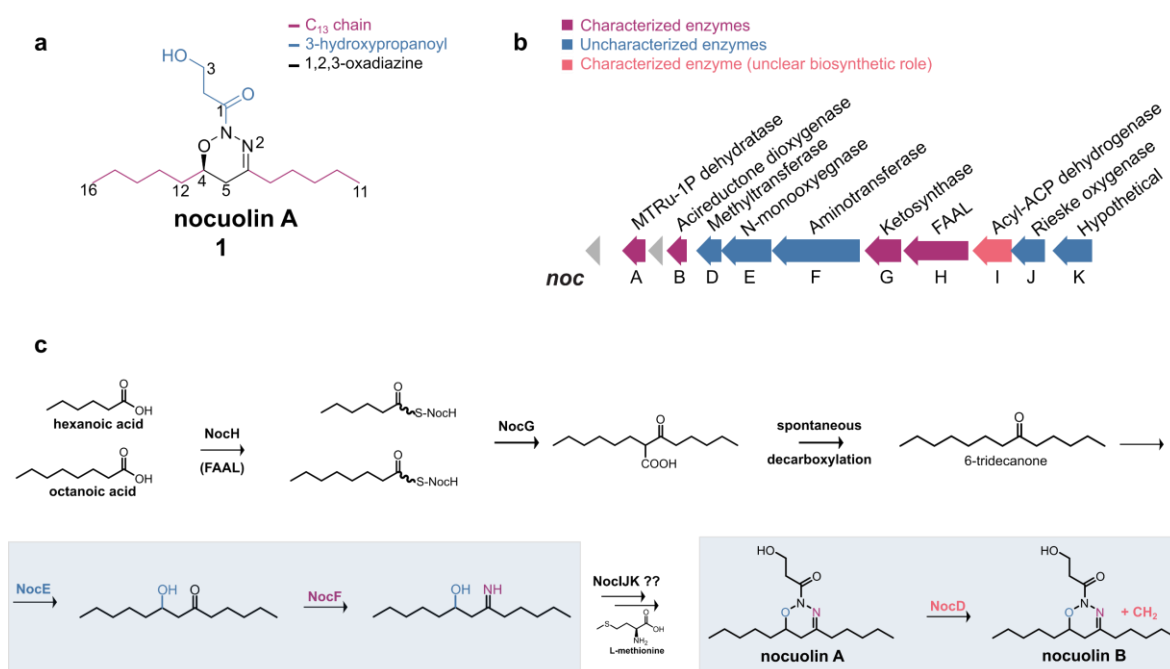


Figure II.1 - Schematic highlighting the Nocuolin A structure, the *noc* BGC and the biosynthetic steps previously characterized to form nocuolin A C₁₃ alkyl chain. (a) Chemical structure of nocuolin A. (b) The *noc* locus and predicted enzyme annotation. (c) Representation of the known biosynthetic steps in the biosynthesis of **1**.

Results and Discussion

Nocuolin A biosynthesis requires CH activation by NocE

Considering the formation of the β -ketoacid by NocG and its likely spontaneous decarboxylation,¹⁹ we turned our attention to how the C-4 position in **1** could be functionalized. The incorporation of one deuterium atom in this position and in C-5 upon feeding *d*₁₅-octanoic acid was observed in the previous supplementation experiments with perdeuterated fatty acids.¹⁹ Because C-C bond formation by NocG requires abstraction of one deuterium atom from C-5, and C-4 retains two deuterium atoms, loss of one deuterium at C-4 could only be reconciled with functionalization through direct C-H bond activation.

We envisaged that this might occur early in the biosynthesis of **1** and supplemented *Nodularia* sp. LEGE 06071 with perdeuterated hexanoic acid (d_{11} -C₆) to search for putative biosynthetic intermediates containing the C₁₃ alkyl moiety of **1** with a functionalized C-4 position. Considering the labeling observed for deuterated versions of **1** after d_{11} -C₆ feeding (d_{21} or d_{22}), the possible mass range and molecular composition, we found two mass features matching putative intermediates (Fig. SII.1). Molecular structures were rationalized for the corresponding m/z values with both C-4 and C-6 positions functionalized. One of the detected intermediates is proposed to correspond to an enolized ketone in position C-6 (8-hydroxytridecan-6-one, **2**), while the other intermediate is proposed to be the related tridecane-6,8-diol (**3**, Fig. SII.1). A matching diol was also found upon *Nodularia* sp. LEGE 06071 supplementation with d_{15} -C₈ (Fig. SII.1).

To clarify if **2** or **3** were actual intermediates in the biosynthesis of **1**, we synthesized the two compounds with a d_{12} -label (d_{12} -**2**, d_{12} -**3**, see SI methods) and used these to supplement *Nodularia* sp. LEGE 06071 cultures. Both compounds were incorporated into **1** (either as d_{11} -**1** or d_{12} -**1**) confirming that the C-4 position is functionalized to an alcohol during the biosynthesis of **1** with loss of a single deuterium atom (Fig. II.2a). LC–HRMS/MS analysis of d_{11} -**1** revealed diagnostic fragments with m/z 127.1231 and 110.0970 featuring the expected m/z shifts of +11.069 and +10.062, respectively (Fig. II.2b). Fragmentation of d_{12} -**1** also proved to be in accordance with the expected product structure since none of the diagnostic fragments were labeled (Fig. II.2b). Furthermore, our previous study confirmed that this C-4 functionalization is introduced after NocG activity¹⁹ when testing 3-hydroxyoctanoic acid as substrate, as it did not lead to activation by NocH nor was it condensed with hexanoic acid by NocG. Considering the annotation of the remaining uncharacterized Noc enzymes (Table SII.1), we hypothesized that the putative Rieske oxygenase, NocJ, could perform the predicted C-H activation step. Rieske oxygenases are well-known for their wide range of oxidative transformations, including extremely challenging reactions, which include C-hydroxylation, N-oxygenation, oxidative carbocyclization, C-C desaturation and dealkylation reactions.^{22,23} The monooxygenase NocE was also considered a candidate, although it had been bioinformatically-predicted by others and us to correspond to a putative N-monooxygenase.¹⁹ To investigate whether any of these enzymes could perform the proposed C-4 functionalization, recombinant NocJ and NocE were expressed in *E. coli* (Fig. SII.2). Both enzymes were particularly challenging to obtain in soluble form (Fig. SII.3-SII.6). However, NocJ was purified as N-Strep-SUMO-NocJ-C-His₆ and NocE was purified as N-His₆-SUMO-NocE, both from *E. coli* C43(DE3) cells. Both enzymes were assayed *in vitro* with 6-tridecanone as a substrate. In assay mixtures with NocE, we detected a prominent mass feature of m/z 215.2006 (Fig. II.2c,

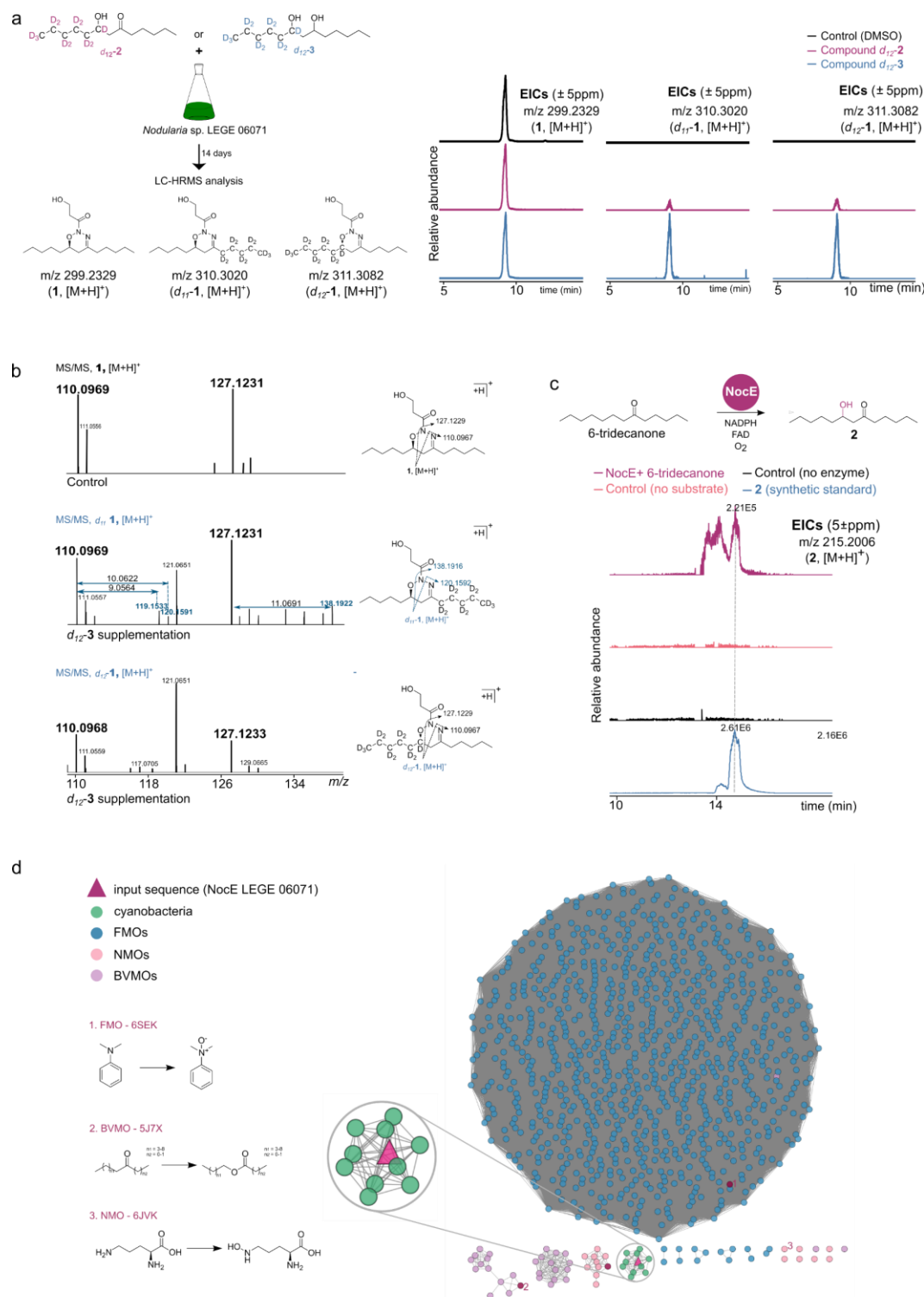


Figure II.2—Supplementation of *Nodularia* sp. LEGE 06071 with $d_{12}-2$, $d_{12}-3$ leads to the formation of $d_{11}-1$ and $d_{12}-1$. (a) LC-HRMS-derived EICs of 1 and labeled versions with $d_{12}-2$ and $d_{12}-3$ in *Nodularia* sp. LEGE 06071. (b) HR-MS/MS spectrum of 1 (control) $d_{11}-1$ and $d_{12}-1$. (c) LC-HRMS-derived EICs of 2 in NocE enzymatic assays with 6-tridecanone, proving that role of the enzyme as the monooxygenase acting after NocG activity. Additional isomeric peaks might be caused by NocE inespecificity. (d) Sequence similarity network of NocE and 1000 homologues, highlighting the NocE cluster composed exclusively by cyanobacterial strains. Characterized enzymes belonging to the 3 subclasses are highlighted (450-625 aa, alignment score of 143 and %ID of 42%).

SII.7), compatible with the formation of **2**. When compared with the synthetic standard of **2**, the generated m/z 215.2006 feature matched the retention time and was not detected in NocJ assays or control assays lacking either NocE or 6-tridecanone (Fig. II.1c, Fig. SII.8). Mass features with the same m/z value but different retention times were observed in the NocE assay, suggesting that the enzyme might lack regioselectivity under the assay conditions (Fig. II.1c). Additionally, features compatible with **3** were not detected in the assays (Fig. SII.9), supporting that **2** is the actual intermediate in the biosynthesis of **1**.

The role of NocE as a C-monooxygenase was not in accordance with the previous annotation of this enzyme as an N-monooxygenase.¹⁷ We performed a more detailed bioinformatic investigation of NocE using HHPred and Swiss-model, which led us to classify NocE as a member of the flavoprotein monooxygenases group B family (Table SII.1). These flavoproteins catalyze a wide variety of oxygenation reactions using flavin adenine dinucleotide (FAD) as a prosthetic group, NADPH as a hydride donor, and oxygen as a co-substrate, oxygenating a wide diversity of nucleophile-containing substrates.²⁴ The group B of this protein family is divided into three classes: flavin-containing monooxygenases (FMOs), N-hydroxylating monooxygenases (NMOs), and Baeyer–Villiger monooxygenases (BVMOs).^{25–27} Despite the high diversity of oxygenation reactions performed by these enzymes,^{26,28} the reactivity of NocE does not directly fall into any of the described groups since the hydroxylation is performed on an unactivated carbon with no direct connection with nucleophilic groups. BlastP analysis of NocE led to the retrieval of a total of 11 homologues with high sequence similarity (>84%), all uncharacterized proteins from cyanobacteria, with the homology level then dropping to 57% (e-value: $1e^{-173}$). The lack of close characterized homologues was also supported by a sequence similarity network (SSN) analysis of NocE (Fig II.1d), in which NocE clusters exclusively with cyanobacterial sequences, separately from all other homologues. Despite this, NocE homologues are mainly annotated as FMO type 5, specifically N,N-dimethylaniline monooxygenases which catalyze N-oxide formation (Fig. II.1d). This suggests that NocE might catalyze a new type of hydroxylation of less nucleophilic substrates, unlike previously characterized FMOs. Furthermore, the characteristic hydride transfer catalyzed by FMOs is compatible with the single deuterium in the C-4 position of **1**, observed in our feeding experiments. The rigorous characterization of the NocE product and the novelty of NocE catalysis will be the subject of future studies.

Overall, these results indicate that NocE act as a monooxygenase responsible for the direct functionalization of 6-tridecanone in the C-4 position, performing a hydroxylation that originates **2**, which is likely key for the formation of the 1,2,3-oxadiazine ring. However, more

experimental data, including NMR of **2** and *in vitro* assays with deuterated substrates, are necessary to undoubtedly confirm NocE activity and the identity of **2**.

NocF introduces N-2 in the biosynthesis of nocuolin A

Having established the formation of **2** by NocE, we hypothesized that the next biosynthetic step might introduce the first nitrogen atom of the 1,2,3-oxadiazine ring. We considered candidate enzymes for this C-N bond formation. Based on the bioinformatic analysis detailed in our previous work,⁵ NocF was annotated as an aminotransferase (Table SII.1). These enzymes typically catalyze the transamination reaction between an amino acid and a ketone using the pyridoxal-5'-phosphate (PLP) as cofactor,²⁹ thus being the most obvious candidate enzyme to perform this reaction encoded in the *noc* cluster.

NocF is an 872 amino acid PLP-dependent enzyme composed of two domains. The first domain (1-412 aa) shows homology towards kinases, and the second domain (413-872 aa) is predicted to be an aminotransferase (Fig. SII.10). Additionally, bioinformatics analyses of NocF in InterPro and Swiss-Model suggest that the enzyme is periplasmic and contains two small N-terminal transmembrane motifs (Fig. SII.11). To test the activity of NocF, *nocF* was cloned and heterologously expressed in *E. coli*, but obtaining soluble expression of the enzyme was particularly challenging. To overcome the solubility issues, we performed codon optimization, cloning into pET-22b which encodes a periplasmic signal sequence (*pelB*) and truncating the first 165 aa of the protein to remove the transmembrane domain (Fig. SII.12-SII.17). After several expression attempts which included testing several solubility tags, *E. coli* strains, full and truncated versions (165 aa N-terminal transmembrane domain removed – termed NocF165), growth temperatures and induction times, we were able to obtain soluble recombinant truncated NocF as N-Strep-SUMO-*pelB*-NocF165-His₆ in *E. coli* C43(DE3) (Fig. SII.12-SII.17).

Recombinant N-Strep-SUMO-*pelB*-NocF165-His₆ was purified (Fig. SII.2) and tested *in vitro* with the putative substrates 6-tridecanone, **2** or **3** (and the deuterated versions *d*₁₂-**2** or *d*₁₂-**3**) (Fig. II.3a, II.3b). In assay mixtures with **2**, we detected a low-abundance feature with *m/z* 214.2165 (Fig. II.3a, SII.18) compatible with the formation of 8-aminotridecan-6-one or 8-iminotrican-6-ol (**4**). However, considering that position C-4 in **1** retains a deuterium upon supplementation of cultures with *d*₁₅-C8,¹⁹ only 8-iminotrican-6-ol is compatible with our observations. Furthermore, in the corresponding assay with *d*₁₂-**2**, a low abundance feature of the corresponding *d*₁₂-**4** (*m/z* 226.2919) was also detected, with the same retention time, indicating that no deuterium loss was caused by NocF activity which is compatible with the formation of 8-iminotrican-6-ol and not 8-aminotridecan-6-one (Fig. II.3b, SII.19). Neither **4** or *d*₁₂-**4** were detected in control assays lacking NocF

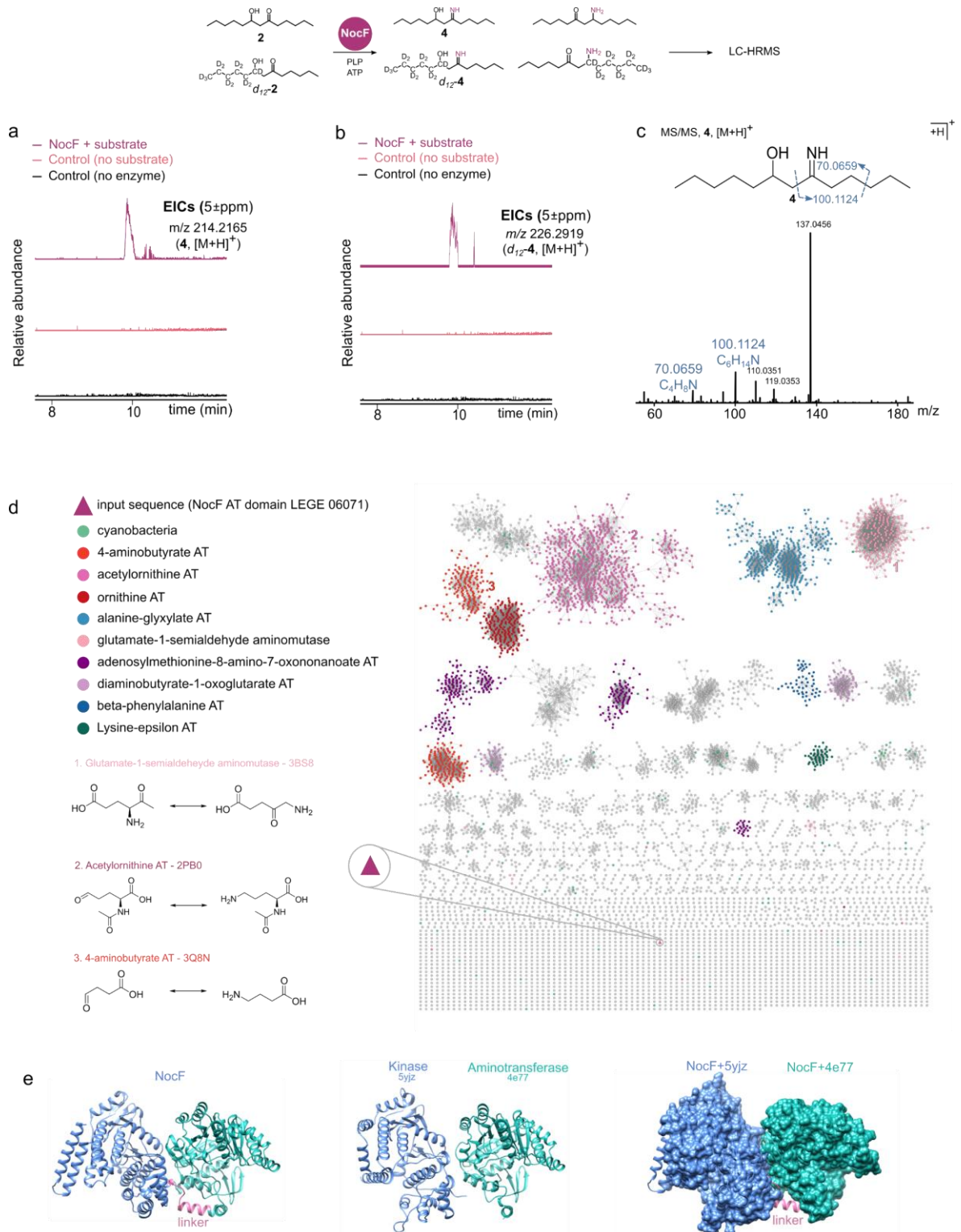


Figure II.3 – NocF is predicted to be an unusual aminotransferase (a) LC-HRMS-derived EICs of the 8-iminotrican-6-ol (4) produced by NocF upon enzymatic assays with 8-hydroxytridecan-6-one, 2. (b) LC-HRMS-derived EICs of d₁₂-8-iminotrican-6-ol d₁₂-4 produced by NocF upon enzymatic assays with d₁₂-8-hydroxytridecan-6-one, d₁₂-2. (c) HRMS/MS-based structural deduction of 4. (d) Sequence similarity network (SSN) of NocF, 1000 closest homologues and the IPR005814 InterPro family, indicating a separation of NocF from the remaining sequences (length 200-500aa, alignment score 54, %ID 53%) (e) AlphaFold model of NocF and comparison with crystal structures of characterized aminotransferase (in blue), 4e77, and kinase (in green), 5yjl highlighting the linker (in pink) between the two NocF domains.

or the substrate (Fig. II.3a, II.3b). Additionally, no other potential nitrogen-containing products were found in the assays with 6-tricadenone and **3** (Fig. SII.20). LC–HRMS/MS analysis of **4** was also compatible with the proposed structure as 8-iminotrican-6-ol (Fig. II.3c), but an unequivocal confirmation of the identity of the metabolite could not be obtained. The formation of **4** was also detected in an *in vitro* assay using N-His₆-SUMO-pelB-NocF and in *E. coli* lysates enriched in this protein, when supplemented with **2** (Fig. SII.21, SII.22). It is also important to notice that **4** could also be compatible with the formation of an enamine, although HRMS/MS fragments are in accordance with the presence of an imine group.

The transformation of a ketone into an imine group like observed in **4** is, to our knowledge, unprecedented among aminotransferases, which typically convert ketones to amine groups.²⁹ The uniqueness of this enzyme is also demonstrated by the sequence similarity network (SSN) analysis, in which NocF aminotransferase domain clusters exclusively with cyanobacterial sequences, separately from all other homologues and members of the IPR005814 InterPro family (aminotransferases class-III) (Fig. II.3d, SII.23). Additionally, aminotransferases are not usually associated with kinase domains. The presence of this domain in NocF is intriguing and further studies need to be performed to clarify its role. However, detailed bioinformatic analysis of the domain indicates an ABC1 atypical kinase-like domain – IPR004147 InterPro family. Members of the ABC1 family have been suggested to be novel chaperonins.^{30–32} On the other hand, the closest structural homologue of this domain in the protein data bank (PDB) is Rv3197 from *Mycobacterium tuberculosis* (5yjl), an antibiotic ABC transporter with a hypothetical role as cooperative enzyme to export antibiotics upon interaction with efflux pumps.^{33,34} This could indicate a non-catalytic role of the NocF kinase domain regarding protein folding or substrate/product transport which may be consistent with the predicted location of NocF in the periplasm. Furthermore, the AlphaFold predicted structure of the NocF kinase domain closely matches the crystal structure of Rv3197. In NocF, this domain is connected to the aminotransferase domain, which has a predicted structure that closely resembles other aminotransferases structures, by a linker formed by a short alpha helix of 18 aa (Fig. II.3e).

Overall, these results indicate that NocF is likely the aminotransferase responsible for transamination of **2** and formation of **4**. However, due to the low abundance of **4** in the enzymatic assays and the lack of precedence for imine group formation by an aminotransferase, further studies are necessary to confirm the identity of **4**, study the reaction mechanism of the enzyme, identify the amino group donor, and characterize the role of the kinase domain.

Further intermediates in the biosynthesis of **1**

To support the enzymatic role of NocF as an aminotransferase and unveil new nocuolin A biosynthetic intermediate, we attempted to generate knockout mutants of the gene in *Nodularia* sp. LEGE 06071 and express the entire *noc* BGC in *E. coli*. *Nodularia* sp. LEGE 06071 is a non-model strain with no engineering tools available and in all our attempts to generate knockout mutants, no colonies were obtained (Fig. SII.24). On the other hand, we successfully assembled the roughly 16 kb *noc* BGC from *Nodularia* sp. LEGE 06071 into a pET-28-ptetO:gfpv2 vector backbone, encoding a C-terminal *gfp* fusion for the last gene in the BGC, using the DiPaC strategy.^{20,21} The construct pET28-ptetO::*noc*ABDEFGHIJK-gfpv2 (hereafter pET-28::*noc*) was transformed onto BAP1 *E. coli*, which harbors the promiscuous 4'-phosphopantetheinyltransferase Sfp, essential to generate holo-acyl carrier proteins (ACPs).⁵ Crude extracts from BAP1 pET-28::*noc* transformants in several growth conditions did not yield detectable levels of **1** (Fig. SII.25). However, comparison of the LC–HRMS profiles of BAP1 pET28::*noc* transformants with those of empty vector transformants, revealed two mass features of *m/z* 214.2165 (**5**) and 272.2584 ([M+H]⁺) that were only present in cells containing the pET28::*noc* vector (Fig. II.4a). Based on the accurate masses, molecular formulae for **5** and **6** were deduced to be C₁₃H₂₈NO and C₁₆H₃₄O₂N, respectively (Fig. SII.26).

To clarify whether **5** and **6** were associated with the *noc* BGC, pET28::*noc* transformant cultures were supplemented with *d*₁₁-C6 and *d*₈-L-methionine, building blocks of the C₁₃ alkyl chain and 3-hydroxypropanoyl moiety of **1**, respectively.¹⁹ Following LC–HRMS analysis of cultures that underwent *d*₁₁-hexanoic acid supplementation, we detected versions of **5** and **6** 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 (Fig. II.4a, II.4b, SII.27). On the other hand, under *d*₈-L-methionine supplementation, a mass shift of *m/z* 4.0255, indicating the incorporation of one *d*₈-L-methionine moiety like previously observed for **1**,¹⁹ was detected only for compound **6** (Fig. II.4a, II.4b, SII.27). Based on this, we rationalized three chemical structures for **5** and **6** (Fig. SII.25), featuring a C₁₃ alkyl chain and an additional nitrogen and oxygen (**5**) or a 3-hydroxypropanoyl moiety or similar (**6**).

A detailed HRMS/MS analysis of **5** revealed several prominent peaks that supported the proposed chemical structures (Fig. II.4c). However, the presence of an abundant fragment at *m/z* 116.1070 supports the identity of **5** as 6-nitrosotridecane due to the facilitated fragmentation of the labile nitroso bond instead of C-C bond fragmentation, as required in the remaining hypothetical structures. Additionally, **5** presents a different retention time than **4** (Fig. SII.28), which is in accordance with the identity of **5** as 6-nitrosotridecane. HRMS/MS analysis of **6** revealed two prominent fragments with *m/z*

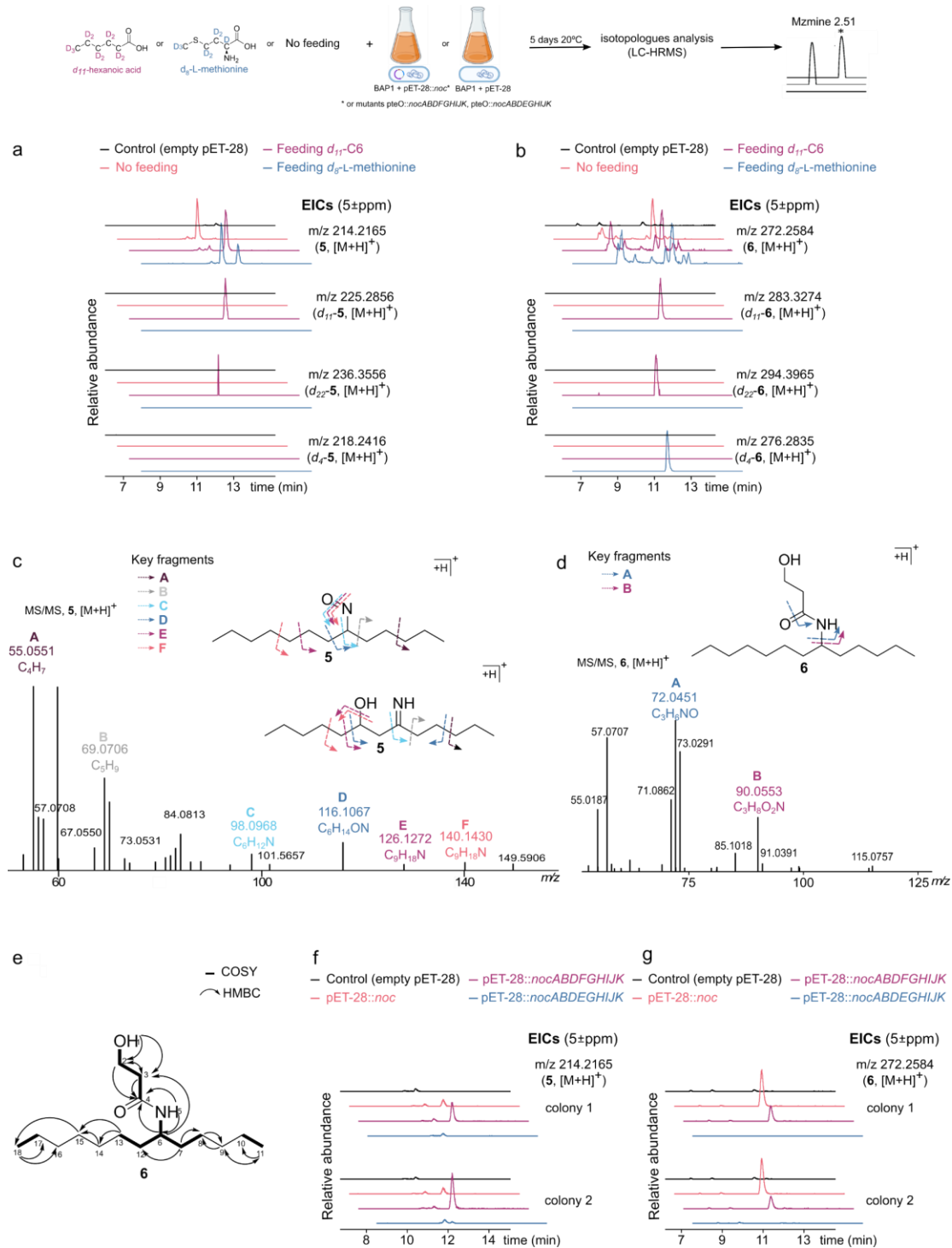


Figure II.4 – Expression of the *noc* BGC in *E. coli* yield the putative intermediates 3-hydroxy-N-(tridecan-6-yl)propenamide (**6**) and putative 6-nitrosotridecane (**5**). (a) LC-HRMS-derived EICs of **5**, d_{12} -**5**, d_{22} -**5** and d_4 -**5** in control and supplementation experiments with d_{11} -**6** and d_8 -L-methionine. (b) LC-HRMS-derived EICs of **6**, d_{12} -**6**, d_{22} -**6** and d_4 -**6** in control and supplementation experiments with d_{11} -**6** and d_8 -L-methionine. (c) HRMS/MS-based structural deduction of **5**, highlighting key fragments. (d) HRMS/MS-based structural deduction of **6**, highlighting key fragments. (e). Key NMR correlations supporting the structural elucidation of **6**. (f) LC-HRMS-derived EICs of **5** in *E. coli* harboring the wild-type pET-28::noc and the mutants pET-28::nocABDFGHIJK and pET-28::nocABDEGHIJK. (g) LC-HRMS-derived EICs of **6** in *E. coli* harboring the wild-type pET-28::noc and the mutants pET-28::nocABDFGHIJK and pET-28::nocABDEGHIJK.

72.0451 and 90.0553, compatible with the structures of either 3-hydroxy-*N*-(tridecan-6-yl)propenamide or 1-(tridecan-6-ylideneamino)propane-1,3-diol (Fig. II.4d). To confirm the identity of **5** and **6**, we focused next on the isolation of this metabolite to enable NMR-based structure elucidation. A 20 L fermentation of BAP1 pET28::*noc* yielded 1.9 g of biomass (d.w.), which was extracted and fractionated using vacuum liquid chromatography (VLC), normal phase flash chromatography and high-performance liquid chromatography (HPLC) steps to afford pure **5** (0.1 mg) and **6** (0.3 mg). Due to the low amount of **5**, we were unable to obtain good-quality NMR data for this metabolite. However, compound **6** was subjected to 1D and 2D NMR-based structure elucidation (Fig. II.4e, SII.29-S33, Table SII.2). The ¹H NMR data for **6** showed two isolated spin systems. HMBC and COSY correlations established a spin system connecting the hydroxyl proton at δ3.06 to the quaternary carbon with resonance at δ172.1, characteristic of amide bonds. In the second spin system, the NH proton resonating at δ5.31 was connected by HMBC and COSY correlations to methine H-6 with a coupling constant of ~ 8.8 Hz. Further COSY and HMBC correlations were used to interconnect the C₁₃ chain. The two spin systems were connected via an amide bond, supported by HMBC correlations from H-5 to carbons C-4 (δ172.1) and C-6 (δ49.5). The connection was further confirmed by the HMBC correlation observed between the H-6 and C-3 carbons. This established the structure of **6** as 3-hydroxy-*N*-(tridecan-6-yl)propenamide (Fig. II.4e).

Despite the clear connection of **6** with the *noc* BGC, its role in the biosynthesis of **1** is unclear. It is possible that **6** represents a shunt product of the *noc* pathway in *E. coli* due to inability of this host to properly express and fold certain Noc enzymes, as suspected by the lack of detectable **1** in crude extracts of pET28::*noc* transformants. The lack of the second nitrogen atom in this structure is particularly intriguing. Considering the bond between 3-hydroxypropanoyl and the C₁₃ alkyl chain in **6**, we questioned the ability of NocF to directly use 3-hydroxypropanamide to perform the transamination reaction. However, NocF *in vitro* assays with 3-hydroxypropanamide were unable to generate a 3-hydroxypropanoyl product (Fig. SII.34). The same unclear role in the biosynthesis of **1** is also valid for **5**, whose structure is not consistent with the observed intermediates from NocE and NocF assays.

To understand the putative role of NocF and NocE in generating **5** and **6** in *E. coli*, we assembled versions of pET28-ptetO::*noc* lacking *nocE* or *nocF*. The constructed pET28-ptetO::*nocABDFGHIJK* and pET28-ptetO::*nocABDEGHIJK* were transformed in BAP1 and expression tests were performed with the necessary control conditions. NocF mutants were unable to produce both compounds which corroborates the role of NocF as the aminotransferase that introduces the first nitrogen atom in **1** and intermediates (Fig. II.4f, II.4g). NocE mutants were able to produce both compounds, with substantial accumulation

of **5**, suggesting that this compound could be a NocE substrate, in which case NocE could act after NocF (Fig. II.4f, II.4g). This could also simply indicate that NocE is relatively promiscuous, accepting both 6-tridecanone and **5**, which is in accordance with the broad substrate acceptance range of many FMOs.³⁵

Overall, these results support the role of NocF as an aminotransferase and suggest direct incorporation of 3-hydroxypropanamide at a later stage in the biosynthesis of **1**. Further inferences are precluded by the likely inability of *E. coli* to fully reconstitute this cyanobacterial pathway. Future studies will focus on the structural elucidation of compound **5** and construction of further mutant strains, by deleting genes with unknown functions, namely NocJ, NocI and NocK.

NocD is a putative methyltransferase involved in the biosynthesis of nocuolin B

Having established the role of most enzymes from the *noc* BGC, only a few remain to be characterized. Among these is NocD, predicted to be a methyltransferase, despite the absence of methylation in **1**. On the other hand, the *noc* BGC in *Nodularia* sp. HBU26¹⁷, a known producer of **1**, does not harbor a NocD homologue, indicating that this enzyme might not be required to generate **1**. Consequently, we hypothesized that *nocD*-harboring cyanobacteria could produce methylated analogues of nocuolin A, which could have been overlooked in previous studies. In crude extracts of *Nodularia* sp. LEGE 06071, we detected a mass feature with *m/z* 313.2486, around 200-400 fold less abundant than **1**, compatible with the formation of a methylated version of **1**, which we named nocuolin B (**7**) (Fig. II.5a). The same feature was labeled with the expected 11 deuterons upon *d*₁₁-C6 supplementation experiments (Fig. II.5a). LC–HRMS/MS analysis of **7** corroborated the identity of the compound as a methylated version of **1** (Fig. II.5b), suggesting that methylation should occur on the eastern side of the molecule. However, NMR analysis will be required to determine the exact position of this methylation. The production of **7** was also confirmed in other *nocD*-harboring cyanobacteria, namely *Anabaena aphanizomenoides* LEGE 00250 and *Nodularia* sp. LEGE 04288 (Fig. SII.35).

To test the putative role of NocD as the methyltransferase responsible for the methylation of **1**, NocD from *Anabaena* sp. PCC 7108 and *Nodularia* sp. LEGE 06071 were heterologously expressed in *E. coli*. Because NocD is predicted to be a membrane protein, purification of soluble recombinant NocD was challenging, yielding poor protein preparations unsuitable for *in vitro* assays (Fig. SII.36). Instead, we performed *in vitro* assays with pET-28a::*nocD*(PCC) lysates and clarified lysates but **7** was not detected (Fig. SII.37). To overcome purification issues, the activity of NocD was probed in *in vivo* assays by supplementing *E. coli* C43(DE3) pET-28a::*nocD*(PCC) with **1**. In supplemented

conditions a low abundance feature with m/z 313.2486, compatible with **7**, was detected, in contrast with the controls with no supplementation and with the empty pET-28a (Fig. 5c). This feature matched the retention time of **7** detected in the extracts of *Nodularia* sp. LEGE 06071, supporting that it should correspond to compound **7**. However, due to its low abundance, it was not possible to obtain MS/MS data for this metabolite.

Overall, these results suggest that NocD is responsible for the methylation of **1**, generating **7**, the first reported nocuolin A analogue. Due to the low-quality MS data and lack of *in vitro* assays with the purified enzyme, further studies are required to firmly establish the structure of **7** and the role of NocD in its generation.

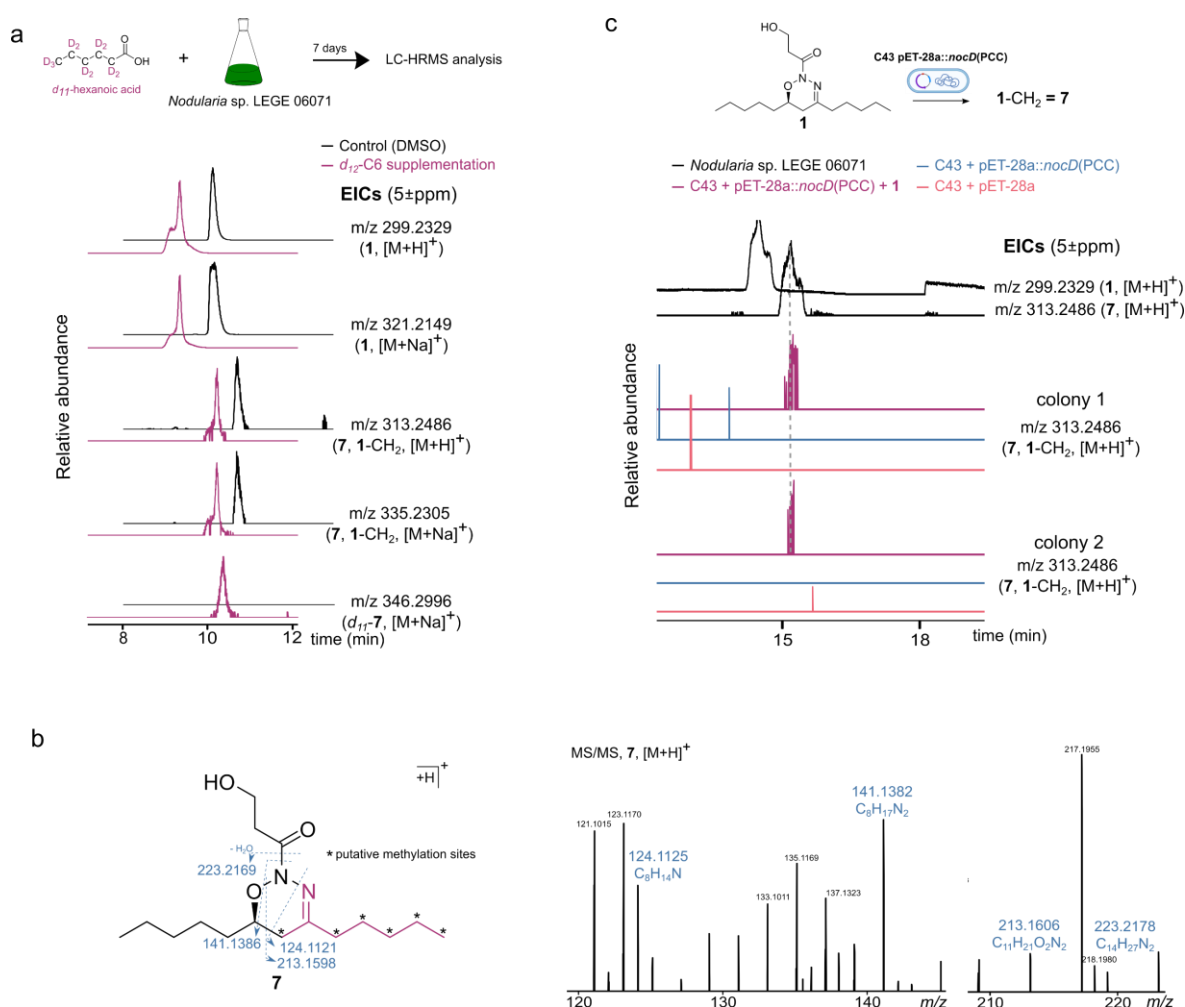


Figure II.5 – Nocuolin B is the methylated analogue of **1**, likely resulting from NocD activity. (a) LC-HRMS-derived EICs of **1** and **7** in the crude extracts of *Nodularia* sp. LEGE 06071 with and without d_{11} -C6 supplementation. (b) HR-MS/MS spectrum of **7**, highlighting the indicative fragments of an additional methylation on the eastern side of the C_{13} alkyl chain, compared with **1**. (c) LC-HRMS-derived EICs of **7** in *in vivo* assays of C43 pET-28a::*nocD*(PCC) supplemented with **1**.

Unknown roles for NocJ and Nock in the biosynthesis of **1**

The roles of NocJ, putative Rieske oxygenase, and the hypothetical enzyme Nock, in the biosynthesis of **1** are still unclear. Recombinant NocJ was tested in several enzymatic assays with 6-tridecanone and in coupled assays with NocF but no activity was observed (Fig. SII.38). Like NocE and NocF, a NocJ SSN also indicates a lack of sequence similarity to characterized enzymes (Fig. SII.39).

From all Noc proteins, Nock is the only hypothetical protein encoded by the *noc* gene cluster, belonging to a family with no characterized homologues (DUF3419). No InterPro family is predicted for this protein and its closest non-*noc* homologue, (hypothetical protein from *Acidobacteriota* bacterium) presents only 59% identity (e-value: $1e^{-154}$). Additionally, Nock clusters separately from most of its closest homologues in our SSN analysis (Fig. SII.40). However, Swiss-model was able to identify N-methyltransferase and SAM-dependent methyltransferase, as Nock structural homologues (11.22 identity, 0.36 coverage; 9.30 identity, 0.34 coverage). Despite the low homology score, this might indicate a Nock role regarding 3-hydroxypropanoyl formation. To test this hypothesis, we have successfully expressed and purified N-His₆-SUMO-Nock from *E. coli* (Fig. SII.2, SII.41). We hypothesized that Nock could directly incorporate 3-hydroxypropanamide into **4**. Coupled assays with Nock and NocF were carried out with **2**, but no *m/z* features compatible with the formation of 3-hydroxypropanoyl were observed (Fig. SII.41a). To investigate the putative formation of intermediate **6**, Nock was incubated with 6-tridecanone in an *in vitro* assay, but **6** was not detected (Fig. SII.41b).

We hypothesize that NocJ, Nock and potentially NocI could also cooperate in the later biosynthetic steps regarding the formation of the 1,2,3-oxadizine ring. The roles of these enzymes will be subject of further future studies.

Conclusion

In this study, we have reconstituted two proposed steps following 6-tridecanone formation in the biosynthesis of **1** (Fig. II.6). Our study shows that 6-tridecanone is converted to 8-iminotrican-6-ol by the consecutive enzymatic actions of the monooxygenase, NocE and the hybrid kinase/aminotransferase, NocF. In the first step, the monooxygenase NocE hydroxylates 6-tridecanone to ketone 8-hydroxytridecan-6-one by direct C-H bond activation. In the second step, this intermediate undergoes an unusual transamination reaction by NocF, which likely results in the formation of 8-iminotrican-6-ol. Both NocE and NocF may exhibit unique reactivities that do not resemble any characterized members of their enzyme families, which is further confirmed by the formation of isolated clusters in

SSNs, separately from the remaining protein family homologues. However, further studies are necessary to confirm and characterize their reactivity and reaction mechanisms, especially confirm the identity of **4**, determine the amino group donor and characterize the role of the kinase domain of NocF.

Additionally, the study suggests that NocD is a methyltransferase involved in the methylation of **1** to generate **7** (Fig. II.6).

Further investigations on the biosynthesis of **1**, namely of the transformations leading to the 1,2,3-oxadiazine moiety, which will likely involve NocI, NocJ and NocK, are now facilitated by the identification of these intermediates and biosynthetic steps.

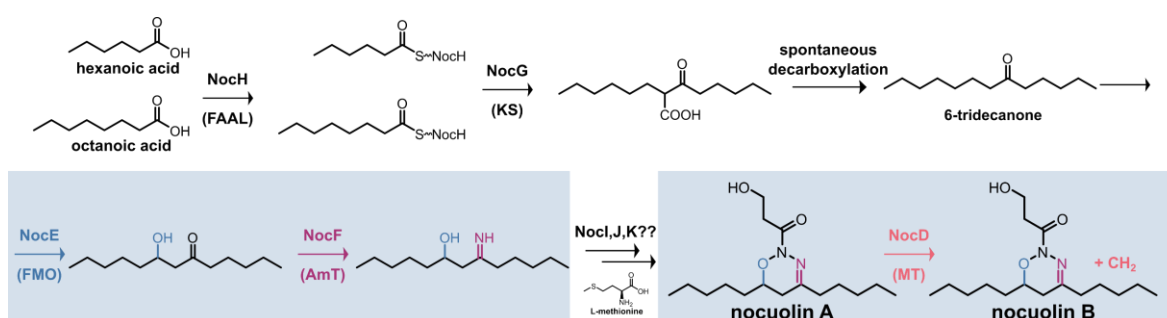


Figure II.6 – Proposed biosynthesis of **1** and **7**, highlighting the biosynthetic steps described in this study (in blue). Previous 6-tridecanone formation steps were described at Martins et al., 2022.¹⁹

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 and UIDP/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 (LEGEcc, CIIMAR) for assistance with cyanobacterial cultures.

Author contribution

T.P.M. carried out experimentation. N.R.B., M.P. and N.R.G. assisted with organic synthesis, NP isolation and protein expression, respectively. T.P.M., E.P.B. and P.N.L. designed research. T.P.M. and P.N.L. drafted the manuscript, with contributions from all co-authors.

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. Soares, A. R.; Engene, N.; Gunasekera, S. P.; Sneed, J. M.; Paul, V. J., Carriebowlinol, an antimicrobial tetrahydroquinolinol from an assemblage of marine cyanobacteria containing a novel taxon. *J. Nat. Prod.* **2015**, 78 (3), 534-538.
3. Amaral, S. C.; Monteiro, P. R.; Neto, J.; Serra, G. M.; Goncalves, E. C.; Xavier, L. P.; Santos, A. V., Current Knowledge on Microviridin from Cyanobacteria. *Mar. Drugs* **2021**, 19 (1), 17-46.
4. Gutierrez-Del-Rio, I.; Brugerolle de Fraissinette, N.; Castelo-Branco, R.; Oliveira, F., et al., Chlorosphaerolactylates A-D: Natural Lactylates of Chlorinated Fatty Acids Isolated from the Cyanobacterium *Sphaerospermopsis* sp. LEGE 00249. *J. Nat. Prod.* **2020**, 83 (6), 1885-1890.
5. Tidgewell, K.; Clark, B. R.; Gerwick, W. H., The Natural Products Chemistry of Cyanobacteria. In *Comprehensive Natural Products II: Chemistry and Biology, Vol 2: Natural Products Structural Diversity-II: Secondary Metabolites: Sources, Structures and Chemical Biology*, 2010; pp 141–188.
6. Kopycki, J. G.; Stubbs, M. T.; Brandt, W.; Hagemann, M.; Porzel, A.; Schmidt, J.; Schliemann, W.; Zenk, M. H.; Vogt, T., Functional and structural characterization of a cation-dependent O-methyltransferase from the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Biol. Chem.* **2008**, 283 (30), 20888-20896.
7. Jones, A. C.; Monroe, E. A.; Eisman, E. B.; Gerwick, L.; Sherman, D. H.; Gerwick, W. H., The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat. Prod. Rep.* **2010**, 27 (7), 1048-1065.
8. Martins, T. P.; Rouger, C.; Glasser, N. R.; Freitas, S.; de Fraissinette, N. B.; Balskus, E. P.; Tasdemir, D.; Leao, P. N., Chemistry, bioactivity and biosynthesis of cyanobacterial alkylresorcinols. *Nat Prod Rep* **2019**, 36 (10), 1437-1461.
9. Malmvarn, A.; Zebuhr, Y.; Kautsky, L.; Bergman, K.; Asplund, L., Hydroxylated and methoxylated polybrominated diphenyl ethers and polybrominated dibenzo-p-dioxins in red alga and cyanobacteria living in the Baltic Sea. *Chemosphere* **2008**, 72 (6), 910-916.
10. Dahiya, R.; Dahiya, S.; Fuloria, N. K.; Jankie, S., et al., Natural Thiazoline-Based Cyclodepsipeptides from Marine Cyanobacteria: Chemistry, Bioefficiency and Clinical Aspects. *Curr. Med. Chem.* **2021**, 28 (38), 7887-7909.
11. Leao, P. N.; Nakamura, H.; Costa, M.; Pereira, A. R.; Martins, R.; Vasconcelos, V.; Gerwick, W. H.; Balskus, E. P., Biosynthesis-Assisted Structural Elucidation of the Bartolosides, Chlorinated Aromatic Glycolipids from Cyanobacteria. *Angew. Chem.* **2015**, 5 (48), 14895-14895.
12. Nagatsu, A.; Kajitani, H.; Sakakibara, J., Muscoride A: A new oxazole peptide alkaloid from freshwater cyanobacterium *Nostoc muscorum*. *Tetrahedron Lett.* **1995**, 36 (23), 4097-4100.
13. Ng, T. L.; Rohac, R.; Mitchell, A. J.; Boal, A. K.; Balskus, E. P., An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **2019**, 566 (7742), 94-118.
14. Reis, J. A.; Figueiredo, S.; Leao, P. N., A Novel Enzymatic Esterification Involved in the Biosynthesis of Branched Bartolosides. *Mar. Drugs* **2020**, 18 (1), 1458-1267.

15. Nakamura, H.; Schultz, E. E.; Balskus, E. P., A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **2017**, *13* (8), 916-921.
16. Morinaka, B. I.; Lakis, E.; Verest, M.; Helf, M. J., et al., Natural noncanonical protein splicing yields products with diverse beta-amino acid residues. *Science* **2018**, *359* (6377), 779-782.
17. Voráčková, K.; Hájek, J.; Mareš, J.; Urajová, P., et al., The cyanobacterial metabolite nocuolin a is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLoS One* **2017**, *12* (3), e0172850.
18. 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.
19. Martins, T. P.; Glasser, N. R.; Kountz, D. J.; Oliveira, P.; Balskus, E. P.; Leao, P. N., Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A. *ACS Chem. Biol.* **2022**, 2528-2537.
20. Greunke, C.; Duell, E. R.; D'Agostino, P. M.; Glockle, A.; Lamm, K.; Gulder, T. A. M., Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **2018**, *47*, 334-345.
21. D'Agostino, P. M.; Gulder, T. A. M., Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth. Biol.* **2018**, *7* (7), 1702-1708.
22. Barry, S. M.; Challis, G. L., Mechanism and Catalytic Diversity of Rieske Non-Heme Iron-Dependent Oxygenases. *ACS Catal* **2013**, *3* (10), 2362-2370.
23. Perry, C.; de los Santos, E. L. C.; Alkhalaf, L. M.; Challis, G. L., Rieske non-heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. *Nat. Prod. Rep.* **2018**, *35* (7), 622-632.
24. Eswaramoorthy, S.; Bonanno, J. B.; Burley, S. K.; Swaminathan, S., Mechanism of action of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (26), 9832-9837.
25. Deng, Y. M.; Zhou, Q.; Wu, Y. Z.; Chen, X.; Zhong, F. R., Properties and Mechanisms of Flavin-Dependent Monooxygenases and Their Applications in Natural Product Synthesis. *Int. J. Mol. Sci.* **2022**, *23* (5), 2622-2657.
26. Furst, M. J. L. J.; Gran-Scheuch, A.; Aalbers, F. S.; Fraaije, M. W., Baeyer-Villiger Monooxygenases: Tunable Oxidative Biocatalysts. *Acs Catalysis* **2019**, *9* (12), 11207-11241.
27. Badiyan, S.; Bach, R. D.; Sobrado, P., Mechanism of N-Hydroxylation Catalyzed by Flavin-Dependent Monooxygenases. *J. Org. Chem.* **2015**, *80* (4), 2139-2147.
28. Landry, Z.; Swan, B. K.; Herndl, G. J.; Stepanauskas, R.; Giovannoni, S. J., SAR202 Genomes from the Dark Ocean Predict Pathways for the Oxidation of Recalcitrant Dissolved Organic Matter. *Mbio* **2017**, *8* (2), e00413.
29. Slabu, I.; Galman, J. L.; Lloyd, R. C.; Turner, N. J., Discovery, Engineering, and Synthetic Application of Transaminase Biocatalysts. *ACS Catalysis* **2017**, (7), 8263-8284.
30. Brasseur, G.; Tron, P.; Dujardin, G.; Slonimski, P. P.; BrivetChevillotte, P., The nuclear ABC1 gene is essential for the correct conformation and functioning of the cytochrome bc(1) complex and

the neighbouring complexes II and IV in the mitochondrial respiratory chain. *Eur. J. Biochem.* **1997**, *246* (1), 103-111.

31. Bousquet, I.; Dujardin, G.; Slonimski, P. P., Abc1, a Novel Yeast Nuclear Gene Has a Dual Function in Mitochondria - It Suppresses a Cytochrome-B Messenger-Rna Translation Defect and Is Essential for the Electron-Transfer in the Bc1 Complex. *EMBO J.* **1991**, *10* (8), 2023-2031.

32. Jasinski, M.; Sudre, D.; Schansker, G.; Schellenberg, M.; Constant, S.; Martinoia, E.; Bovet, L., AtOSA1, a member of the Abc1-like family, as a new factor in cadmium and oxidative stress response. *Plant Physiol.* **2008**, *147* (2), 719-731.

33. Zhang, Q. Q.; Liu, H. J.; Liu, X.; Jiang, D. Q., et al., Discovery of the first macrolide antibiotic binding protein in Mycobacterium tuberculosis: a new antibiotic resistance drug target. *Protein & Cell* **2018**, *9* (11), 971-975.

34. Nunez-Samudio, V.; Chesneau, O., Functional interplay between the ATP binding cassette Msr(D) protein and the membrane facilitator superfamily Mef(E) transporter for macrolide resistance in Escherichia coli. *Res. Microbiol.* **2013**, *164* (3), 226-235.

35. Oesch-Bartlomowicz, B.; Oesch, F., Mechanisms of Toxification and Detoxification which Challenge Drug Candidates and Drugs. In *Comprehensive Medicinal Chemistry II*, Taylor, J. B.; Triggle, D. J., Eds. 2007; Vol. 5, pp 193-214.

Chapter II – Supplementary Information

Manuscript

Diversification of the 6-tridecanone scaffold during nocuolin A biosynthesis

Methodology

1. General experimental procedures

1.1. Chemistry

Stable-isotope labeled fatty acids (d_{11} -hexanoic, d_{15} -octanoic, acids) were obtained from CDN Isotopes. L-methionine was obtained from Sigma-Aldrich, d_8 - L-methionine and d_{12} -hexanal from Cambridge Isotope Laboratories and 3-hydroxypropanamide from Fluorochem. All solvents, purchased from VWR Chemicals, Sigma-Aldrich, Thermo Fisher Scientific, Carlo Erba and Chem Lab NV were ACS grade, except for HPLC solvents (HPLC gradient grade) and LC-MS solvents (MS-grade). NMR solvents were purchased from Cambridge Isotope Laboratories.

NMR spectra were acquired on a Varian-Mercury 400 (400 MHz, 100 MHz) or on Bruker Avance III HD, 600 MHz, the latter equipped with a 5 mm cryoprobe and controlled by TopSpin 3.6.1.(Materials Center of the University of Porto - CEMUP) NMR spectrometers. NMR data was analyzed in MestReNova 12.0.3 (MestrelabResearch).

LC-MS analysis was performed on an Agilent 1200 series LC system coupled with Agilent 6530 quadrupole time-of-flight (qTOF). LC-HRMS and LC-HRMS/MS analyses were obtained on a Thermo Scientific UltiMate 3000 HPLC system, which consists of the following components: LPG-3400RS pump, WPS-3000RS autosampler, TCC-3x00RS column compartment, and MWD-3000RS UV/VIS detector coupled to a Q Exactive Focus Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Fisher Scientific), controlled by Q Exactive Focus (Exactive Series) 2.9 and Thermo Scientific Xcalibur 4.1.31.9 software.

1.2. Molecular Biology

Terrific Broth (TB) media were purchased from Condalab, Luria-Bertani Lennox (LB) medium from Thermo Fisher Scientific, M9 minimal medium salt from MP Biomedicals, IPTG from NZYTech and tetracycline from ThermoScientific. Oligonucleotides were synthesized by StabVida and Sigma-Aldrich. PCR reactions were carried in Veriti 96-well thermal cycler (Applied Biosystems). DNA polymerases, restriction enzymes and T4 DNA ligase were purchased from Thermo Fisher Scientific and New England Biolabs. Gel extractions of DNA fragments and restriction reactions clean ups were performed using NZYGelpure kit (NZYTech). Plasmid purifications were performed using NZYMiniprep (NZYTech). DNA sequencing was performed by Eurofins Scientific. Cyanobacterial genomic DNA extraction was performed with the PureLink® Genomic DNA Mini Kit (Invitrogen) following the protocol described for Gram-negative bacteria.

Chemically competent *E. coli* cells were transformed with plasmid DNA using the heat shock method (5 μ L of a given ligation reaction or 1 μ L of purified plasmid DNA).

Electrocompetent cells were prepared from overnight cultures. The cells (5 mL) were centrifuged at 5000 $\times g$ 5mins, washed with sterile 1ml of ice-cold 10% glycerol and centrifuged again for 1 minute at 3200 $\times g$ at 4°C. The process was repeated 4 times and aliquots of 3 x 150 μ l were prepared. Electroporation was performed with 0.5 μ l of DNA, gently mixed with the electrocompetent cells aliquots and transferred to 0.1 mm path length electroporation cuvette. The electroporation was performed with EC1 setting (observed time constant $\sim$ 5 ms) and 1 ml of LB medium was immediately added to the electroporation cuvette. Cells were allowed to recover at 37 °C for 1h and then plated into LB plates with antibiotic.

Extracted DNA and PCR reactions were separated in 1% to 1.5% (w/v) agarose gels containing 1 $\times$ Tris/Acetic acid/EDTA (TAE) buffer (Bio-Rad), stained with GreenSafe Premium (NZYTech) and visualized and photographed on a gel documentation system (Gel Doc EZ Imager, Bio-Rad). DNA Ladder 1 Kb Plus (Thermo Fisher Scientific)/NZYDNA Ladder III (NZYtech) and Precision plus all standards blue (Biorad) or PageRuler Plus Prestained (Thermo Scientific) were used as DNA and protein molecular weight markers, respectively. Proteins were separated in Mini-PROTEAN® TGX™ Precast Polyacrylamide Gel – 4-15% (Biorad) containing 1xTris/Glycine/SDS (TGS) buffer (Fisher Scientific) and stained with or Bio-Safe Coomassie G-250 (Biorad). DNA and protein concentrations were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc.). Optical densities (600 nm) of *E. coli* cultures were determined using an Ultrospec 10 Spectrophotometer (GE Healthcare).

2. Plasmids, Strains and culture conditions

pET-28a(+), pET-29b(+), pET-43.1a(+), pET-51b(+), pET-32a(+), pET-22b(+) were acquired from Merck. pET-His₆-Sumo-TEV-LIC and pMAL-2cx, were acquired from Addgene.

Escherichia coli TOP10 (Life Technologies) was used for cloning and *E. coli* BL21 DE3 (Nzytech), OverExpress C43(DE3) (Frilabo), RosettaTM(DE3) (Merck Millipore), SoluBL21(DE3) (Genlatis), Tuner(DE3) (Novagen) were used for recombinant protein expression.

Nodularia sp. LEGE 06071, *Nodularia* sp. LEGE 04288 and *Anabaena aphanizomenoides* LEGE 00250 were acquired from LEGEc and grown in Z8 medium ¹ at 25 °C, under a 16 h light (25 μ mol photons m⁻² s⁻¹)/8 h dark cycle and continuous shaking (190 rpm). *Anabaena* sp. PCC 7108 was acquired from Pasteur Culture Collection and grown under the same conditions.

3. Supplementation experiments in *Nodularia* sp. LEGE 06071

One-hundred milliliters of a *Nodularia* sp. LEGE 06071 culture were inoculated in 900 mL of Z8 medium and allowed to grow at 25 °C. After 4 days, 25 mL of culture were distributed evenly through 50 mL erlenmeyers (duplicates were performed for each condition). A pulse feeding protocol was used: at days 0, 2, 4, 7, 9 and 11 in which the cultures were supplemented with 0.1 mM of d_{12} -**2**, d_{12} -**3** or DMSO (control) and the biomass harvested after 14 days by centrifugation 4500 $\times g$, 10 min at 4 °C and washed with deionized H₂O. For fatty acid supplementation experiments, pulse feeding was performed at days 0, 2 and 4 with 0.17mM d_{11} -hexanoic acid, d_{15} -octanoic acid or DMSO (control) and the biomass harvested after 4 days by centrifugation 4500 $\times g$, 10 min at 4 °C and washed with deionized H₂O.

4. Organic extraction of cyanobacteria

The fresh biomass of *Nodularia* sp. LEGE 06071, *Nodularia* sp. LEGE 04288 and *Anabaena aphanizomenoides* LEGE 00250 were extracted with dichloromethane:methanol (2:1) with vigorous vortex at room temperature. The extracts were filtered with a Whatman No. 1 filter paper, concentrated using a rotary evaporator (R-210, Buchi) and transferred to pre-weighted vials. Extracts from the feeding experiments were dissolved in MS-grade methanol to a final concentration of 2.0 mg mL⁻¹ and analyzed by LC-HRMS and LC-HRMS/MS.

5. LC-HRMS and LC-HRMS/MS analysis – Orbitrap

HRMS data were obtained in Full Scan positive and negative mode with a scan range of m/z 150-2000, with a capillary voltage of HESI set to -3.5 kV and the capillary temperature to 263 °C. Sheath gas flow rate was set to 50 units.

Cyanobacterial extracts were separated in an ACE(r) UltraCore™ SuperC18, 2.5 μ m particle size, 95 Å pore size and dimensions 75 $\times$ 2.1 mm. The column oven was set to 40 °C. The samples were eluted at 0.4 mL/min over a gradient of 90 % solvent A (water/methanol 1:1 with 0.1% v/v formic acid) to reach 70 % solvent B (isopropanol with 0.1% v/v formic acid) during 5 min, followed by isocratic 70% solvent B for 12 min, and a gradient of 30 % solvent A to reach 80 % solvent B during 2 min, followed by isocratic 85% solvent B for 9 min and return to the initial conditions followed by isocratic 90 % solvent A for 2 min.

Enzymatic assays were separated in an Aeris peptide XB-C18, 3.6 μ m particle size, and dimensions 150 $\times$ 4.6 mm. The column oven was set to 40 °C. The samples were eluted at 0.5 mL/min with isocratic 100 % solvent A (water with 0.1% v/v formic acid), followed by

gradient to reach 100 % solvent B (acetonitrile with 0.1% v/v formic acid) during 16 min, followed by isocratic 100 % solvent B for 5 min and return to the initial conditions followed by isocratic 100 % solvent A for 2 min. In alternative, NocF enzymatic assays were analyzed in a similar fashion: 0.5 mL/min with a linear gradient from 100% A to reach 100 % solvent B during 16 min, followed by isocratic 100 % solvent B for 5 min and return to the initial conditions followed by isocratic 100 % solvent A for 2 min (total of 23mins).

MS/MS parameters for the LC-HRMS/MS analysis were: resolution of 17500, with a 3 m/z isolation window, a loop count of 3, AGC target of 5×10^4 and collision energy of 30 eV.

6. LC-MS and LC-MS/MS analysis – qTOF

LC-MS analysis were obtained in Full Scan positive and negative mode with a scan range of m/z 150-2000, with a capillary voltage of H set to 3.5 kV and the capillary temperature to 300 °C. Nebulizer pressure was set to 35 psi and drying gas (N₂) flow of 8 L min⁻¹.

Crude extracts and synthetically prepared compounds were separated in a Hypersil GOLD aQ C18 column (ThermoScientific), 3 µm particle size, and dimensions 150 × 3.0 mm. The column oven was set to 30 °C. The samples were eluted at 0.5 mL/min with a mobile phase of 0.1% formic acid in ACN (eluent B) and in 0.1% formic acid in H₂O (eluent A). The gradient program was as follows: isocratic 1 % B for 2 min, gradient from 1 % A to 99 % B over 16 min, held at 99 % B for 5 min, before returning to the initial conditions.

7. Metabolomic analysis (nocuolin intermediates search)

Metabolomic analysis of LC-HRMS data from the *d*₁₁-C6 and *d*₁₁-C8 feeding experiments was carried out in MZmine 2.53.² The analysis was performed using the standard protocol defined by Figueiredo et al..³ The Adduct Search module was set with the following mass differences: 11.0681 m/z for *d*₁₁ labeling; 22.1369 m/z for *d*₂₂ labeling with 5 ppm tolerance. Molecular formulas were calculated for the putative nocuolin intermediates using Xcalibur 4.1.31.9 with 5ppm tolerance.

8. Small-scale expression tests and western blots

E. coli transformants with Noc enzymes were tested to evaluate protein expression and solubility. Briefly, 500 µl of overnight culture were inoculated in 50 mL TB supplemented with 5 mM of MgSO₄ or M9 minimal supplemented with 10 mL L⁻¹ of glucose solution (40% (w/v)), 2 mL L⁻¹ of filtered Vitamin mix (500×) with 50 µg mL⁻¹ antibiotic at 37 °C, 200rpm. The cultures were grown to an OD of 0.5-0.8 (~3h) and then transferred to the appropriate

temperatures and induced with IPTG. One milliliter of culture was then centrifuged and frozen in liquid nitrogen. The cells were resuspended in BugBuster (Merck) containing a final concentration of 0.5 mg/mL of lysozyme (Sigma) and the suspension was incubated at 4 °C for 5 min. 0.5 µl of Benzonase Nuclease (Merck) and 10 mM of MgSO₄ were added and the solution once gain incubated at 4 °C for 5 min. The suspension was centrifuged at 4 °C at maximum speed for 30 min. The pellets were resuspended in 300 µl of 2% SDS were and warmed for 5 min at 50 °C. Protein expression was visualized in polyacrylamide gels.

Due to undetectable amounts of soluble protein, western blots analyses were performed to detect and optimize expression conditions. PVDC membranes (GE Healthcare) were soaked in methanol and placed in a sandwich form. The transfer was carried out in 25 mM Tris, 192 mM glycine, 20% methanol (v/v) at 100V for 1h. The membrane was blocked with 3% BSA for 1h and then washed with 1X Tris-Buffered Saline, 0.1% Tween 20 detergent (TBST) buffer. One microliter of antibody (Penta-His HRP Conjugate Kit, Qiagen, or Anti-MPB Tag Mouse Monoclonal Antibody (9Y5) HRP, Abbkine Scientific) was added in 1% BSA/TBST to the membrane and incubated for 1h. After this, the remaining antibody was washed out of the membrane with TBST and the results revealed with the CN/DAB kit (Thermo Scientific).

9. Cloning, expression and purification of NocE

9.1. Cloning of *nocE* in pET-28a/pET-29b

The *nocE* gene was amplified from *Nodularia* sp. LEGE 06071 gDNA with 1 × Phusion HF Buffer, 400 µM dNTPs, 0.5 µM of each primer (Table SII.3), 0.4 U Phusion polymerase (Thermo Scientific) and was performed under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 58 °C for 30 seconds, extension at 72 °C for 2 min and the final extension at 72 °C for 7 min. PCR products were gel purified, restriction digested using NdeI and XhoI (New England Biolabs) for 1:30 h at 37 °C followed by 65 °C inactivation for 20 min and direct reaction purification. Digests were ligated into linearized into the corresponding plasmids using T4 DNA ligase (New England Biolabs), according to the manufacturer's instructions. An insert-to-vector molar ratio of 3:1 was used and the reaction mixtures were incubated overnight at room temperature and the enzyme was then inactivated at 65 °C for 10 min.

Chemically competent *E. coli* TOP10 cells (Invitrogen) were transformed with 5 µL of each ligation reaction using the heat shock method. After recovery, the mixture was plated onto LB plates supplemented with 50 µg mL⁻¹ of antibiotic. Colonies were screened

by colony PCR using T7 primers and GoTaq Flexi (Promega): 1x Green GoTaq® Flexi Buffer, 1.0 mM of MgCl₂ solution, 0.20 mM of PCR Nucleotide Mix, 0.20 µM of each primer, molecular biology grade water and template DNA picked directly from the colony. The PCR conditions were: initial denaturation step at 95 °C for 5 min, followed by 35 cycles of a denaturation step at 95°C for 45 sec, annealing at 48 °C for 30 sec and extension at 72 °C for 2 min, followed by a final extension step at 72 °C for 7 min. Clones that led to amplicons with the expected size were grown overnight in liquid LB medium supplemented with 50 µg mL⁻¹ of kanamycin, at 37 °C. The NZYMini-prep kit (NZYTech) was used to isolate plasmids from overnight cultures, which were sent for Sanger sequencing to verify construct identity and amplification fidelity. The correctly assembled plasmids were transformed in BL21(DE3), C43(DE3), Rosetta(DE3), Tuner(DE3), SoluBL21(DE3) and Artic Express by heat shock.

Protein expression was optimized in small-scale studies like described above. Different temperatures (37 °C, 25 °C, 16 °C, 15 °C, 12° C), incubation time (overnight, 1 day, 1.5 days, 2 days) and culture medium composition (terrific broth, M9 minimal) were tested (Fig. SII.3, SII.4).

9.2. Cloning of *nocE* in pET-His₆-SUMO-TEV and pET-51b

Cloning of *nocE* into pET-His₆-SUMO-TEV and pET-51b was performed using *Nodularia* sp. LEGE 06071 gDNA as template and Gibson assembly to ligate vector and insert. PCR reactions were performed using 1x Q5 High-fidelity master mix (NewEngland Biolabs) and 0.5 µM of each primer (Table SII.3). PCR conditions were: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 55-57 °C for 30 seconds (inserts) or 56 °C for 30 seconds (vector), extension at 72 °C for 1:30 min (inserts) or 4 min (vector) and the final extension at 72 °C for 10 min. The PCR products were gel purified and treated with DpnI (NewEngland Biolabs) at 37 °C for 30mins followed by at 80 °C for 20 min. For each construct, the two fragments (vector and insert) were all mixed in the HiFi Gibson assembly (NewEngland Biolabs) according to manufacturers' instructions. Transformation and colony PCR were performed as described previously for NocE cloning in pET-28a/pET-29b. The constructs were then transformed in C43(DE3) and tested in small-scale as previously above.

9.3. Purification of N- His₆-SUMO-NocE

Based on the results of the protein expression in small-scale and due to unsuccessful purification of C-His₆-NocE despite the soluble protein detectable by the western blot, we purified N-SUMO-NocE-His₆. C43 harboring N-SUMO-NocE-His₆ was

supplemented with 50 $\mu\text{g mL}^{-1}$ ampicillin and grown overnight at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 2 $\times$ 2 L of TB medium supplemented with 50 $\mu\text{g mL}^{-1}$ ampicillin, and 5 mM of MgSO_4 at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 12 °C, induced with 0.5 mM of IPTG and allowed to grow for 6 days. The cell pellets were recovered by centrifugation (10 min, 3000 $\times g$), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 minutes at 5000 $\times g$ and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl_2 , 5 mM CaCl_2 , 30mM imidazole, 10 % (v/v) glycerol, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 $\times g$ and the clarified supernatant loaded into 1mL Ni-IDA packed columns (Macherey Nagel). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM imidazole and the protein was eluted in 5mL elution buffer (20 mM HEPES, 300mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed and concentrated using the Amicon Ultra Centrifugal Filters (Merck) with 3.5 mL of dialysis buffer (20 mM HEPES, 50 mM NaCl, 1 mM DTT, 10% (v/v) glycerol) through centrifugation for 30 min at 7000 $\times g$ to reach a final concentration of 42.1 μM (Fig. SII.2). The protein was added dropwise to liquid nitrogen and stored as pellets at -80 °C.

10. Cloning, expression and purification of C-His₆-NocJ

10.1. Cloning of *nocJ* in pET-28a/pET-29b

The *nocJ* gene was amplified from *Nodularia* sp. LEGE 06071 gDNA with 1 $\times$ Phusion HF Buffer, 400 μM dNTPs, 0.5 μM of each primer (Table SII.4), 0.4 U Phusion polymerase (Thermo Scientific) and was performed under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 58 °C for 30 seconds, extension at 72 °C for 1 min and the final extension at 72 °C for 7 min. PCR products were gel purified, restriction digested using NdeI and XhoI, ligated and transformed in *E. coli* TOP10 as previously described for NocE cloning in pET-28a/pET-29b. The colonies were screened by PCR as previously described and purified plasmids sent for Sanger sequencing to verify construct identity and amplification fidelity.

The correctly assembled plasmids were transformed in BL21(DE3), C43(DE3), Tuner(DE3), Rosetta(DE3), SoluBL21(DE3) and Artic Express by heat shock.

Protein expression was optimized in small-scale studies like described previously. Different temperatures (37 °C, 25 °C, 16 °C, 15 °C 12 °C), incubation time (overnight, 1 day, 1.5 days, 2 days), culture medium composition (terrific broth, M9 minimal) and supplementation with 0.2 mM FeSO₄ were tested (Fig. SII.5, SII.6).

10.2. Cloning of *nocJ* in pET-51b-SUMO

Cloning of *nocJ* in pET-51b-SUMO was performed using *Nodularia* sp. LEGE 06071 gDNA (Table SII.4) as template and Gibson assembly to ligate vector and insert as described above for *nocE* cloning in pET-His₆-SUMO-TEV. pET-51b-SUMO was constructed using pET-51b and adding the original SUMO tag from pET His₆ SUMO TEV as described below in N-peIB-SUMO-NocF-His₆ cloning. The constructs were then transformed in C43(DE3), Rosetta(DE3), BL21(DE3), SoluBL21(DE3) and tested in small-scale as described above.

10.3. Purification of N-Strep-SUMO-NocJ-His₆

Based on the results of the protein expression in small-scale and due to unsuccessful purification of C-His₆-NocJ from Artic Express, which led to contaminated protein preparations, we successfully purified N-Strep-SUMO-NocJ-His₆.

C43 harboring N-Strep-SUMO-NocJ-His₆ was grown overnight in LB medium supplemented with 50 µg mL⁻¹ ampicillin at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 2 × 2 L of TB medium supplemented with 50 µg mL⁻¹ ampicillin, and 5 mM of MgSO₄ at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 15 °C, induced with 0.5 mM of IPTG and allowed to grow at 15°C for 1 day. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 min at 5000 ×g and flash frozen in liquid nitrogen. The protein purification was performed as previously described for NocE purification, yielding a final concentration of 42.8 µM (Fig. SII.2).

11. Cloning, expression and purification of NocF

11.1. Cloning of *nocF* in pET-28a/pET-29b

Cloning of NocF in pET-28a and pET-29b was performed as described for NocE using the primers described in Table SII.5 (Table SII.5, SII.6).

11.2. Cloning of *nocF* in pET-22b

A codon-optimized (*E. coli*) *nocF* gene was synthesized by GeneUniversal and cloned into pET-22b. A truncated version of the gene (starting in amino acidic position 165) was obtained by PCR with primers containing restriction sites for EcoR1 and XhoI (Table SII.5). PCRs were carried out using Q5 High-Fidelity 2× Master Mix (NEB) according to manufacturer's instructions: 98 °C for 30 seconds, followed by 35 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 65°C for 30 seconds, extension at 72°C for 2:30 min and the final extension at 72°C for 5 min. PCR products were gel purified, restriction digested using XhoI and EcoR1 (New England Biolabs) for 1 h at 25 °C followed by 65 °C inactivation for 20 min and direct reaction purification. Digests were using T4 DNA ligase (New England Biolabs), according to the manufacturer's instructions. An insert to vector molar ratio of 3:1 was used, and the reaction mixtures were incubated overnight at room temperature and the enzyme was then inactivated at 65 °C for 10 min.

11.3. Cloning of *nocF/nocG* in pET-29b

The genes were amplified from *Nodularia* sp. LEGE 06071 gDNA with 1 × Phusion HF Buffer, 400 µM dNTPs, 0.5 µM of each primer (Table SII.5), 0.4 U Phusion polymerase (Thermo Scientific) and was performed under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 55 °C for 30 seconds, extension at 72 °C for 3 min and the final extension at 72 °C for 7 min. PCR products were gel purified, restriction digested using BamHI and HindIII (New England Biolabs) for 1.5 h at 37 °C followed by 65 °C inactivation for 20 min and direct reaction purification. Digests were ligated into linearized into the corresponding plasmids using T4 DNA ligase (New England Biolabs), according to the manufacturer's instructions. An insert to vector molar ratio of 3:1 was used, and the reaction mixtures were incubated overnight at room temperature and the enzyme was then inactivated at 65 °C for 10 min.

11.4. Cloning of *pelB-nocF* in pET-41a

Cloning of *pelB-nocF* in pET-41a was performed using pET-22b-NocF as template and using Gibson assembly. The incorporation of *pelB* proved to be essential for NocF soluble expression in previous expression tests and so, primers were designed to amplify a fused *pelB-nocF* fragment that was assembled in the pET-41a backbone (Table SII.5). PCR reactions were performed using 1x Q5 High-fidelity master mix (NewEngland Biolabs) and 0.5 µM of each primer. PCR conditions were as follows: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing

at 55-57 °C for 30 seconds (inserts) or 56 °C for 30 seconds(vector), extension at 72 °C for 2:30 min (inserts) or 4 min (vector) and the final extension at 72 °C for 10 min. The PCR products were gel purified and treated with DpnI (NewEngland Biolabs) at 37 °C for 30min followed by inactivation at 80 °C for 20 min. For each construct, the two fragments (vector and insert) were all mixed in the HiFi Gibson assembly (NewEngland Biolabs) according to manufacturer's instructions.

11.5. Cloning of *pelB-nocF* in pET-51b, pMAL-2cx, pET-His₆-SUMO-TEV-LIC, pET-32a and pET-43.1a

Cloning of *pelB-nocF* in pET-51b, pMAL-2cx, pET-His₆-SUMO-TEV-LIC, pET-32a and pET43.1a was performed as described for pET-41a-*pelB-NocF* using the primers in table SII.5.

11.6. Cloning of N-SUMO-*pelB-NocF165-His₆* and N-SUMO-*pelB-NocF-His₆* in pET-51b

N-SUMO-*pelB-NocF165-His₆* and N-SUMO-*pelB-NocF1-His₆* were cloned in pET-51b as described for pET-41-*pelB-NocF*. The N-SUMO-*pelB-NocF-His₆* and N-SUMO-*pelB-NocF165-His₆* sequences were amplified from the previously constructed pET-His₆-SUMO-TEV-LIC harboring *nocF* or *nocF165* and using the primers of Table SII.5.

11.7. Expression tests

Chemically competent *E. coli* TOP10 cells (Invitrogen) were transformed with 5 µL of each ligation reactions using the heat shock method. After recovery, the mixture was plated onto LB plates supplemented with 50 µg mL⁻¹ of antibiotic. Colonies were screened by colony PCR as described previously (see *NocE* purification). Individual positive colonies were inoculated in LB medium supplemented with 50 µg mL⁻¹ of the respective antibiotic and used for plasmid DNA extraction. The purified plasmids were sent for Sanger sequencing to verify construct identity and amplification fidelity.

The correctly assembled plasmids were transformed in BL21(DE3), C43(DE3), Rosetta(DE3), Tuner(DE3), SoluBL21(DE3) and Artic Express by heat shock.

Protein expression was optimized in small-scale studies like described previously (Table S6). Different temperatures (37 °C, 25 °C, 20 °C, 16°C, 15°C, 14°C, 12°C), IPTG concentration (0.5 mM, 0.1 mM or 0.02 mM), incubation time (overnight, 1 day, 2 days, 4 days, 5 days, 6 days), culture media (Terrific broth and M9 minimal medium), and supplementation with 0.2 mM of the predicted co-factor and substrate (PLP and 6-tricadenone, respectively) were tested (Fig. SII.12-SII.15)

11.8. Protein Purification

11.8.1. N-Strep-SUMO-pelB-NocF165-His₆

Several attempts were made to purify NocF using different resins (nickel or cobalt) and washing steps (Fig. SII.16), using the protocol described for NocE purification. However, the best preparation was obtained for N-Strep-SUMO-pelB-NocF165-His₆ using a Cobalt resin. *E. coli* C43 cells harboring N-Strep-SUMO-pelB-NocF165-His₆ was supplemented with 50 µg mL⁻¹ ampicillin and grown overnight at 37 °C.

Twenty-millilitres of this overnight culture were then inoculated into 2 × 2 L of TB medium supplemented with 50 µg mL⁻¹ ampicillin, and 5 mM of MgSO₄ at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 12 °C, induced with 0.5 mM of IPTG and allowed to grow for 5 days. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 minutes at 5000 ×g and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl₂, 5 mM CaCl₂, 30 mM imidazole, 10 % (v/v) glycerol, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 ×g and the clarified supernatant loaded into 1mL HisPur™ Cobalt resin (Thermo Scientific). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50mM imidazole and the protein was eluted in 5 mL elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed and concentrated the using the Amicon Ultra Centrifugal Filters (Merck) with 3.5mL of dialysis buffer (20 mM HEPES, 50 mM NaCl, 1 mM DTT, 10% (v/v) glycerol) for 30 min at 7000 ×g to reach a final concentration of 38.6 µM (Fig. SII.2). The protein was added dropwise to liquid nitrogen and stored as pellets at -80 °C. This protein was used for all NocF enzymatic assays, except when mentioned otherwise.

11.8.2. N-His₆-SUMO-pelB-NocF

N-His₆-SUMO-pelB-NocF was purified as described for N-Strep-SUMO-pelB-NocF165-His₆. C43 harboring N-His₆-SUMO-pelB-NocF165 were grown at 12 °C for 4 days in TB medium supplemented with 50 µg mL⁻¹ ampicillin, and 5 mM of MgSO₄ at 37 °C, 175 rpm. After concentration, N-SUMO-pelB-NocF-His₆ was at a final concentration of 46.4 µM (Fig. SII.2).

12. Cloning, expression and purification of C-His₆-NocD(LEGE) and N-His₆-NocD(PCC)

The *nocD* gene was amplified from *Nodularia* sp. LEGE 06071 or *Anabaena* sp. PCC 7108 gDNA with 1 × Phusion HF Buffer (Table SII.7) as described for *nocJ* cloning in pET-28a/pET-29b. The correctly assembled plasmids were transformed into *E. coli* C43(DE3) by heat shock. Protein expression was optimized in small-scale studies like described previously. Different temperatures (25 °C, 15 °C) were tested.

C43(DE3) *E. coli* cells carrying the pET-28a-NHis₆-NocD(PCC) or pET-29b-C-His₆-NocD(LEGE) constructs which provided the best results in the expression tests, were inoculated from frozen stocks into TB medium containing 50 µg mL⁻¹ kanamycin, grown at 37 °C with shaking (200 rpm) When the culture (2 L) reached the OD 0.6, they were induced with 0.5 mM IPTG and incubated at 25 °C with shaking. After overnight growth the cell pellet was recovered by centrifugation, flash frozen with liquid N₂ and stored at -80 °C until further processing.

Briefly, the cell pellet was thawed on ice and resuspended in freshly prepared chilled Resuspension Buffer (50 mM Tris HCl pH 8.5, 100 mM NaCl, 10 mM MgCl₂) supplemented with protease inhibitors (Pierce Protease Inhibitor Tablets, Thermo Scientific). The cell suspension was lysed by two rounds of passage through a cell disruptor (~20000 psi) and centrifuged at 107000 ×g for 1 h on a Beckman 45Ti rotor at 4 °C. The supernatant was discarded and the pellet resuspended in 50 mL Resuspension Buffer supplemented with protease inhibitors and re-centrifuged at 107000 ×g for 1h. The supernatant was discarded, and the pellet resuspended in freshly prepared Solubilization Buffer (50 mM Tris HCl pH 8.5, 100 mM NaCl, 2% (w/v) Triton X-100). The suspension was placed in a nutating mixer for 30 minutes before centrifugation at 107000 ×g for an additional hour. The supernatant containing the solubilized membranes was then loaded into 1 mL Ni-IDA packed columns (Macherey Nagel). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM imidazole and the protein was eluted in 5mL elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed and concentrated the using the Amicon Ultra Centrifugal Filters (Merck) with 3.5mL of dialysis buffer (20 mM HEPES, 50 mM NaCl, 1 mM DTT, 10% (v/v) glycerol) for 30 min at 7000 ×g which led to low concentration and impure protein preparations, unsuitable to perform *in vitro* assays (Fig. SII. 35).

13. Cloning, expression and purification of NocK

13.1. Cloning of *nocK* in pET-28a/pET-29b

The *nocK* gene was amplified from *Nodularia* sp. LEGE 06071 gDNA with 1 × Phusion HF Buffer (Table SII.8) as described for *nocJ* cloning in pET-28a/pET-29b (Fig. SII.39).

13.2. Cloning of *nocK* in pET-51b-SUMO

Cloning of *nocK* in pET-51b-SUMO was performed using *Nodularia* sp. LEGE 06071 gDNA (Table SII.8) as template and Gibson assembly to ligate vector and insert as previously described for NocE cloning in pET His₆ SUMO TEV. pET-51b-SUMO was constructed using pET-51b and adding the original SUMO tag from pET His₆ SUMO TEV as described above for N-SUMO-peIB-NocF-His₆ cloning. The constructs were then transformed in *E. coli* C43(DE3), Rosetta(DE3), BL21(DE3) and SoluBL21(DE3) cells and tested in small-scale as described above.

13.3. Purification of N-Strep-SUMO-NocK-His₆

E. coli C43(DE3) harboring N-Strep-SUMO-NocK-His₆ were grown overnight in LB medium supplemented with 50 µg mL⁻¹ ampicillin at 37 °C. Twenty-milliliters of this overnight culture were then inoculated into 2 × 2 L of TB medium supplemented with 50 µg mL⁻¹ ampicillin, and 5 mM of MgSO₄ at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 15 °C, induced with 0.5 mM of IPTG and allowed to grow at 15 °C for 1 day. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 minutes at 5000 ×g and flash frozen in liquid nitrogen. The protein purification was performed as previously described for NocE purification.

14. Enzymatic assays

In vitro enzymatic assays to test the function of the Noc enzymes were carried out in Eppendorf tubes at 37 °C, unless mentioned otherwise.

14.1. NocF enzymatic assays

Briefly, 1 mM of compound **2**, *d*₁₂-**2**, **3**, *d*₁₂-**3** or 6-tridecanone was incubated with 20mM NocF, 1mM ATP, 1mM amino acid mixture (glutamate, methionine, alanine, aspartate, glycine and tryptophan), 1 mM PLP, and 2 mM MgCl₂ in 50 mM HEPES pH 8.0 overnight at 37°C. The necessary reaction controls with no enzyme and no substrate were prepared simultaneously. The reactions were quenched with an equal volume of

acetonitrile, centrifuged at 16000 $\times g$ for 5 min and analyzed by LC-HRMS. All assays were carried out with N-Strep-SUMO-pelB-NocF165-His₆, unless otherwise stated.

NocF was also tested using **2** and 3-hydroxypropanamide. Briefly, 1 mM of compound **2**, was incubated with 20 mM NocF, 1 mM ATP, 1 mM PLP, 1 mM amino acid mixture (glutamate, methionine, alanine, aspartate, glycine and tryptophan), 1 mM 3-hydroxypropanamide and 2 mM MgCl₂ in 50 mM HEPES pH 8.0 overnight at 37 °C. The remaining procedure was performed as described above.

14.2. NocE enzymatic assay

Briefly, 1 mM of 6-tridecanone was incubated with 20 μ M NocE, 1 mM ATP, 1 mM FAD, 1 mM NADPH and 2 mM MgCl₂ in 50 mM HEPES pH 8.0 overnight at 30°C. The necessary reaction controls with no enzyme and no substrate were prepared simultaneously. The reaction was quenched with an equal volume of acetonitrile and centrifuged at 16000 $\times g$ for 5mins and analyzed by LC-HRMS.

14.3. NocJ enzymatic assay

Briefly, 1 mM of 6-tricadenone was incubated with 20 μ M NocJ, 1mM NADPH, 100 μ g/mL ferredoxin and 100 mU/mL ferredoxin reductase (Sigma), 1 mM ATP and 2 mM MgCl₂ in 50 mM HEPES pH 8.0 overnight at 30 °C. The necessary reaction controls with no enzyme and no substrate were prepared simultaneously. The reactions were quenched with an equal volume of acetonitrile, centrifuged at 16000 $\times g$ for 5min and ultimately analyzed by LC-HRMS.

NocJ was also tested using compound **2** as substrate in a coupled assay with NocF. Briefly, 1 mM of **2** was incubated with 20 μ M NocJ, 20 μ M NocF, 1 mM NADPH, 1 mM amino acid mixture 100 μ g/mL ferredoxin and 100 mU/mL ferredoxin reductase (Sigma), 1 mM ATP and 2 mM MgCl₂ in 50 mM HEPES pH 8.0 overnight at 30°C. The necessary reaction controls with no enzyme and no substrate were prepared simultaneously. The reactions were quenched with equal volume of acetonitrile, centrifuged at 16000 $\times g$ for 5mins and ultimately analyzed by LC-HRMS.

14.4. NocK enzymatic assay

NocK was tested in a coupled assay with NocF. Briefly, 1 mM of **2** was incubated with 20 μ M NocK, 20 μ M NocF, 1 mM amino acid mixture, 1mM PLP, 1 mM 3-hydroxypropanamide, 1 mM ATP and 2 mM MgCl₂ in 50 mM HEPES pH 8.0 overnight at 30 °C. The necessary reaction controls with no enzyme and no substrate were prepared

simultaneously. The reactions were quenched with an equal volume of acetonitrile, centrifuged at 16000 ×g for 5mins and analyzed by LC-HRMS.

NocK was tested using 6-tridecanone as substrate. Briefly, 1 mM of **2** was incubated with 20 μM NocK, 1mM ATP, 1 mM 3-hydroxypropanamide and 2 mM MgCl₂ in 50 mM HEPES pH 8.0 overnight at 30°C. The necessary reaction controls with no enzyme and no substrate were prepared simultaneously. The reactions were quenched with equal volume of acetonitrile, centrifuged at 16000 ×g for 5min and ultimately analyzed by LC-HRMS.

14.5. NocF enzymatic assay with lysate and clarified lysate

E. coli C43 cells carrying pET-SUMO::pelB::NocF grown overnight were re-inoculated in 50 mL TB supplemented with 5 mM of MgSO₄, 50 μg mL⁻¹ ampicillin and 0.2mM of **2**. The cultures were grown at 37 °C (200 rpm) to an OD of 0.5-0.8 (~3h) and then transferred to 25 °C and induced with 0.5mM IPTG. After overnight incubation, one millilitre of culture was centrifuged and frozen in liquid nitrogen. The cells were resuspended in BugBuster (Merck) containing a final concentration of 0.5mg/mL of lysozyme (Sigma) and the suspension was incubated at 4 °C for 5 min. 0.5 μL of Benzonase Nuclease (Merck) and 10mM of Mg SO₄ were added and the solution once gain incubated at 4°C for 5mins (non-clarified lysate). The suspension was centrifuged at 4 °C at maximum speed for 30 min and the supernatant recovered (clarified lysate). Non-clarified and clarified lysates were filtered and analyses by LC-HRMS. The necessary controls (C43 pET-SUMO::pelB::NocF with no supplementaion with **2** and C43 pET-SUMO) were prepared simultaneously.

14.6. NocD enzymatic assay with lysate and clarified lysate

E. coli C43 cells harboring either pET-28a::nocD(PCC) or pET-29b::nocD(LEGE) were grown overnight and were re-inoculated in 50 mL TB supplemented with 5 mM of MgSO₄, 50 μg mL⁻¹ ampicillin and 0.2 mM of **1**. The cultures were grown at 37 °C (200 rpm) to an OD of 0.5-0.8 (~3h) and then transferred to 25 °C and induced with 0.5 mM IPTG. The lysates preparation was performed as described above for “NocF enzymatic assay with lysate and clarified lysate”.

15. Sequence Similarity networks

The sequence similarity networks (SSN) were generated using the EFI-EST server,⁴ following a “Sequence BLAST (Basic Local Alignment Search Tool)” using negative log e-values of 5 for UniProt BLAST sequence retrieval, retrieving 1000 sequences. Specific parameters of each generated SSN are listed in Table SII.9. The network was visualized in Cytoscape (v3.90).

16. Attempt to generate knockout mutants in *Nodularia* sp. LEGE 06071

16.1. Cloning

To generate *nocE* and *NocF* deletion strains by double homologous recombination, two plasmids were constructed: one targeting for the deletion of *nocE* and the other targeting for the deletion of *nocF*. Primers targeting approximately 1 kb flanking regions in respect to the gene/locus were designed with the specific overhangs to allow for Gibson reaction (Table SII.10). The flanking regions were amplified by PCR using genomic DNA of *Nodularia* sp. LEGE 06071 as template. The mobilizable plasmid pRL278⁵ was amplified using the primers in Table SII.10. PCR reactions were performed using 1x Q5 High-fidelity master mix (NewEngland Biolabs) and 0.5 μ M of each primer. PCR conditions were: initial denaturation at 98 °C for 30 seconds, followed by 35 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 60 °C for 30 seconds (inserts) or 56 °C for 30 seconds (vector), extension at 72 °C for 1 min (inserts) or 3 min (vector) and the final extension at 72°C for 5 min. The PCR products were gel purified and treated with DpnI (NewEngland Biolabs) at 37 °C for 30 min and then at 80 °C for 20 min. For each construct, the two fragments (vector and insert) were all mixed in the HiFi Gibson assembly (NewEngland Biolabs) according to the manufacturer's instructions and the ligation product was transformed into TOP10 *E. coli*. Screening of positive colonies was performed with OneTaq Quick-Load 2X Master Mix with Standard Buffer (NewEngland Biolabs) and 0.2 μ M of each primer. PCR conditions were as follow: initial denaturation at 94 °C for 30 seconds, followed by 35 cycles of a denaturation step of 94 °C for 30 seconds, annealing at 57 °C for 45 seconds, extension at 68 °C for 2:30 min and the final extension at 68°C for 5 min. Positive colonies were grown overnight and the plasmids extracted using the ZymoPURE Plasmid Miniprep kit (Zymo Research). The fragment identities were confirmed by Sanger sequencing. Sequencing data showed a medium-size insertion in the pRL278::*nocF* plasmid and so we proceeded only with the attempt to generate *nocE* knockouts.

16.2. Sucrose sensibility test

The mobilizable plasmid pRL278, which was used to construct the cargo plasmid pRL278::*nocE* contains the *sacB* gene, encoding for levansucrase. This enzyme converts sucrose to products that are toxic for cyanobacteria ⁶. Nevertheless, as the mutation rate of the *sacB* gene has been experimentally shown to be relatively high, *E. coli* strains harboring the constructed cargo plasmids were tested for their sensitivity to medium supplemented with 5% sucrose. Recombinant *E. coli* TOP10 cells were inoculated in LB solid medium with antibiotic and 5% sucrose and in LB solid medium with antibiotic for viability comparison. Plates were incubated overnight at 37 °C and the *E. coli* cells presenting higher sucrose

sensitivity (warranting the function of SacB) were selected for conjugation. This selection step is essential to select for the double recombinants generated in cyanobacteria and not mutants that incorporated the entire cargo plasmid.

16.3. Transformation in *Nodularia* sp. LEGE 06071

The resulting cargo plasmids, pRL278::*nocE* was transformed into electrocompetent *Escherichia coli* HB101 harboring the helper plasmid pRL623 encoding for specific *Anabaena* sp. PCC 7120. Finally, plasmid pRL278::*nocE* was transferred to *Nodularia* sp. LEGE 06071 by electroporation or conjugation, using triparental mating of the cyanobacterial strain, *E. coli* harboring the helper plasmid pRL623 and pRL278::*nocE* and *E. coli* harboring the conjugative plasmid (either RP4 or pRL443).⁶

16.3.1. Triparental mating

E. coli HB101 cells containing the helper/cargo plasmids and *E. coli* TOP10 cells containing conjugative plasmids (RP4 and pRL443) were grown overnight at 37 °C in liquid LB selective media according with Table SII.11, on an orbital shaker at approximately 120 rpm. To obtain cultures in the exponential phase of growth, 1 mL of the overnight cultures was inoculated in fresh LB medium with the same antibiotic(s) and allowed to grow for 4 h at 37 °C, on an orbital shaker at approximately 120 rpm. The cultures were centrifuged at 3000 $\times g$ for 10 min and washed twice with LB medium. *E. coli* harboring the helper/cargo plasmids were mixed with *E. coli* containing the conjugative plasmids and incubated for 1 h at room temperature.

Meanwhile, *Nodularia* sp. LEGE 06071 culture was centrifuged at 4000 $\times g$ for 10 min and suspended in the least medium volume possible of BG11. To break the filaments and increase the number of single cells, cyanobacteria were passed through a syringe needle several times and for pRL443 sonicated for 10s. The suspension was incubated with the *E. coli* mixture, incubated at room temperature for 1h and finally plated onto BG11 plates. After 48h of incubation at 25°C in low light conditions, the plates were supplemented with 25 $\mu g mL^{-1}$ neomycin (Nm). The plates were incubated at 25 °C in a 16 h light/8 h dark regimen.

16.3.2. Electroporation

The electroporation protocol was based on the *Synechocystis* sp. PCC 6407 electroporation. Briefly, 50mL of *Nodularia* sp. LEGE06071, $OD_{750nm} = 0.2$, were centrifuged at 5000 $\times g$ for 10min and resuspended in 10 mL of HEPES solution 1 mM pH=7.5. This washing step was repeated 4 times and ultimately, the cells were suspended in 0.4 mL of

HEPES and 60 μL of cells transferred to electroporation cuvettes. The cells were mixed with approximately 1 μg of DNA keep on ice for 3 min and electroporated with 2.5 kV (MicroPulser Electroporated, Biorad). The cells were recovered immediately with 400 μL of BG11 medium and plated on BG11 plates. The plates were incubated at 25 $^{\circ}\text{C}$, under low light for 24 h and then the cells were transferred to new BG11 plates supplemented with 25 $\mu\text{g mL}^{-1}$ neomycin.

17. Direct Pathway Cloning -DiPaC

17.1. Cloning

The vector backbone, pET28-ptetO::gfpv2, was linearized by PCR amplification using 0.5 μM of forward and reverse primer, 1 $\times$ Q5 High-Fidelity 2 $\times$ Master Mix (NEB) and 30-50 ng of template DNA. Primers were designed to split the *nocA-K* into two fragments avoiding the intergenic terminators identified using the Arnold webtool (Table SII.12). Thermal cycling program was as follows: initial denaturation step at 98 $^{\circ}\text{C}$ for 30 seconds, followed by 35 cycles of a denaturation step at 98 $^{\circ}\text{C}$ for 10 seconds, annealing at 57 $^{\circ}\text{C}$ for 30 seconds, and extension at 72 $^{\circ}\text{C}$ for 4 min, followed by a final extension step at 72 $^{\circ}\text{C}$ for 10 min. The resulting PCR product was treated with DpnI (NewEngland Biolabs) at 37 $^{\circ}\text{C}$ for 1 h followed by at 65 $^{\circ}\text{C}$ for 20 min.

The two *noc* fragments without Gibson overhangs (*nocA-NocF* and *nocG-NocK*) were amplified using the same PCR conditions. Thermal cycling program was as follows: initial denaturation step at 98 $^{\circ}\text{C}$ for 30 seconds, followed by 35 cycles of a denaturation step at 98 $^{\circ}\text{C}$ for 10 seconds, annealing at 52/55 $^{\circ}\text{C}$ for 30 seconds, and extension at 72 $^{\circ}\text{C}$ for 4 min, followed by a final extension step at 72 $^{\circ}\text{C}$ for 10 min. After gel purification, the DNA was used as template for a Nested PCR to introduce the Gibson overhangs using Q5 High-Fidelity 2 $\times$ Master Mix (NEB). PCR conditions was as follows: initial denaturation step at 95 $^{\circ}\text{C}$ for 30 seconds, followed by 35 cycles of a denaturation step at 95 $^{\circ}\text{C}$ for 15 seconds, annealing at 55 $^{\circ}\text{C}$ for 30 seconds, and extension at 72 $^{\circ}\text{C}$ for 4 min, followed by a final extension step at 72 $^{\circ}\text{C}$ for 10 min. All fragments and vector backbone were purified by gel band excision using NZYGelpure kit (NZYTech).

Gibson assembly was prepared in 10 μL with the NEBuilder HiFi DNA Assembly Master Mix, following the manufacturer's instructions. The reaction was incubated at 50 $^{\circ}\text{C}$ for 15 min and then used to transform *E. coli* TOP10 cells by heat shock. Colony PCR was used to screen positive clones using GoTaq Flexi (Promega): 1 $\times$ Green GoTaq® Flexi Buffer, 1.0 mM of MgCl_2 solution, 0.20 mM of PCR Nucleotide Mix, 0.20 μM of each primer, molecular biology grade water and template DNA picked directly from the colony. The PCR conditions were: initial denaturation step at 95 $^{\circ}\text{C}$ for 5 min, followed by 35 cycles of a

denaturation step at 95°C for 45 seconds, annealing at 52 °C for 30 seconds and extension at 72 °C for 90 seconds, followed by a final extension step at 72 °C for 5 min. Clones that led to amplicons with the expected size were grown overnight in liquid LB medium supplemented with 50 µg mL⁻¹ of kanamycin, at 37 °C. NZYMini-prep kit was used to isolate plasmids from overnight cultures (NZYTech) that were sent for Sanger sequencing.

For the insertion of a new fragment, the vector, pET28b-ptetO::orf1-*nocGHIJK*-gfpv2 was digested with PmlI (NewEngland Biolabs) at 37 °C for 90 min followed by a deactivation step at 65 °C for 20 min. Purification by gel excision and Gibson reaction were carried out as detailed above. Positive transformants were screen as described above and sequenced again. The final pET28b-ptetO::orf1-*nocABCDEFGHIJK*-gfpv2 (referred to as pET-28::*noc*) was sent for full plasmid sequencing (Plasmidsaurus), confirming the correct assembly of the fragments.

17.2. Heterologous expression and LC-HRMS analysis

For heterologous expression experiments, *E. coli* BAP1 (Kerafast) cells were transformed by electroporation with pET28::*noc*. *E. coli* BAP1 cells previously transformed with pET28-ptetO::gfpv2 (from now on pET-28) (empty vector) were used as control. A starter culture grown overnight in LB media supplemented with 50 µg mL⁻¹ of kanamycin at 37 °C was inoculated in 25 mL of M9 minimal supplemented with 10.0 mL L⁻¹ of filtered glucose solution (40% (w/v)), 2 mL L⁻¹ of filtered vitamin mix (500×). Cultures were grown at 37°C to an OD600 of 0.4-0.5 and then induced with 0.5 or 0.1 µg mL⁻¹ tetracycline at 20 °C or 25°C. After 5 days of growth, the cultures were incubated with Amberlite XAD-16 resin (Alfa Aesar) for 30 min and harvested by centrifugation (5000 ×g, 10 min at 4 °C). The cell pellets and resin were freeze dried and extracted with MeOH/Acetone 1:1 followed by ethyl acetate (EtOAc) with shaking. The supernatants were concentrated using a rotary evaporator (R-210, Buchi) and transferred to a previously weighted vial.

Extracts were weighted and dissolved in MeOH to a concentration of 5 mg mL⁻¹. The resulting solutions were filtered (0.2 µm) and submitted to LC-HRMS analysis.

17.3. Supplementation experiments in *E. coli*

BAP1 pET28::*noc* was grown in 25 mL M9 medium as described previously for 5 days at 20 °C, induced with 0.5 µg mL⁻¹ tetracycline. The cultures were supplemented with 0.2 mM of *d*₁₁-C6 and *d*₈-L-methionine. Before harvesting, the cultures were incubated with Amberlite XAD-16 for 30min. Cell pellets and resin were collected at 4 °C, 5000 ×g for 10min. Organic extraction and LC-HRMS analysis were performed as described above.

17.4. Metabolomics analysis (nocuolin A intermediate search in *E.coli*)

Comparative metabolomics analysis of LC-HRMS data from BAP1 pET-28::*noc* and BAP1 pET28b was carried out in MZmine 2.53.² The analysis was performed using the standard protocol defined by Figueiredo et al.,³ but no adduct search was performed. Nocuolin A intermediates were identified based on their *m/z* value (190-300), retention time and putative molecular formulas fitting $C_xH_yN_zO_a$ ($13 \leq x \leq 17$, $25 \leq y \leq 32$, $0 \leq z \leq 3$, $0 \leq a \leq 4$). Molecular formulas were generated using Xcalibur 4.1.31.9 with 5 ppm tolerance.

17.5. Heterologous expression of *noc* in large scale

BAP1 pET28::*noc* was grown in 10 x 2 L M9 medium as described previously for 5 days at 20 °C, induced with 0.5 µg mL⁻¹ tetracycline. Before harvesting, the cultures were incubated with Amberlite XAD-16 for 30 min. Cell pellets and resin were collected by centrifugation at 4 °C, 5000 ×g for 10 min and freeze dried.

17.6. Isolation of 5 and 6

17.6.1. Organic extraction

The biomass and resin present during culture growth were extracted with EtOAc, followed by an extraction with DCM:MeOH mixture (2:1), using and a magnetic stirrer in the extraction vessel. Each extraction was carried with enough volume of extraction solvent to fully submerge the biomass and repeated 5 times. The organic extract was concentrated using a rotary evaporator (R-210, Buchi) and transferred to a previously weighted vial, yielding 1.94 g of dry mass.

17.6.2. Flash chromatography

The crude extract was fractioned by normal phase flash chromatography in a Pure C-850 Flash Prep chromatography system (Buchi) using an 80 g normal phase silica (40-63 µm) cartridge. The UV-detector was set up at 254, 280 and 330 nm wavelengths and the sample was eluted at 20 mL/min using hexanes (solvent A), EtOAc (solvent B), MeOH (solvent C), and DCM (solvent D). The gradient initiated with 60% solvent A to reach 100% solvent B over 20 min, followed by isocratic 100% solvent B for 25 min, a gradient to reach 100% solvent C over 30 min and isocratic 100% C for 10min. The final washing step was performed with a gradient to reach solvent D during 5 min, followed by isocratic 100% solvent D for 10 min. The fractionation resulted in 15 fractions (total mass 1.7g; 88% recovery).

17.2.1. High Performance Liquid Chromatography (HPLC) – Compound 5

The isolation of compound **5** from fraction 2 (18.3 mg) of the initial flash chromatography required several HPLC steps. The last of these steps was performed using Vanquish HPLC system (Thermo Scientific) equipped with a PDA detector and an Excel 3 μ m C18-AR 7.5x4.6mm (ACE). The sample was dissolved in MeOH and eluted at 0.5 mL/min with isocratic 50% water/50% acetonitrile over 27 min, followed by a gradient to reach 100% acetonitrile over 3 min and isocratic 100% acetonitrile for 5min. The fractionation yielded 0.1 mg of **5** (32.8 min).

17.2.2. High Performance Liquid Chromatography (HPLC) – Compound 6

Compound **6** from fraction 5 (3.98 mg) of the initial flash chromatography was purified using a Thermo Ultimate 3000 HPLC (Thermo Scientific) and a Synergy Hydro-RP 250x4.6 mm column (Phenomenex). The sample was dissolved in MeOH and eluted at 0.8mL/min with isocratic 30% water/70% acetonitrile over 6mins, followed by a gradient to reach 50% acetonitrile/50% isopropanol over 6mins and isocratic 50% acetonitrile for 12 min, before returning to the initial conditions. The fractionation yielded 0.22 mg of **6** (13.5 min).

17.2.3. Structural elucidation - NMR analysis

The isolated compounds from HPLC purification were dissolved in 400 μ L of deuterated chloroform and placed in a Shigemi tube (Shigemi, inc.). 1D and 2D NMR experiments (^1H , ^{13}C APT, HSQC, HMBC, COSY, NOESY) were carried out at CEMUP, in a 600MHz spectrometer equipped with a 5 mm cryoprobe. NMR spectra were processed in the Mnova software.

17.2.4. Construction of pET-28::*noc* mutants

pET28::*noc* was amplified as two fragments skipping genes *nocE* or *nocF*, respectively (Table SII.12). PCR was performed as described previously using Q5 High-Fidelity 2 $\times$ Master Mix (NEB). Thermal cycling program was as follows: initial denaturation step at 98 $^{\circ}\text{C}$ for 30 seconds, followed by 40 cycles of a denaturation step at 98 $^{\circ}\text{C}$ for 10 seconds, annealing at 57 $^{\circ}\text{C}$ for 30 seconds, and extension at 72 $^{\circ}\text{C}$ for 5 min, followed by a final extension step at 72 $^{\circ}\text{C}$ for 10 min. The resulting PCR products were treated with DpnI (NewEngland Biolabs), purified, and assembled using the NEBuilder HiFi DNA Assembly Master Mix (NEB), following the manufacturer's instructions. Positive colonies were selected by colony PCR and the sequences confirmed by Sanger sequencing. The final pET28b-ptetO::*orf1-nocABDFGHIJK-gfpv2* and pET28b-ptetO::*orf1-nocABDEGHIJK-*

gfpv2 constructs were fully sequenced (Plasmidsaurus) and transformed into BAP1 *E. coli* as previously described. Heterologous expression tests were performed as described before for 5 days at 20 °C, followed by organic extraction and LC-HRMS analysis.

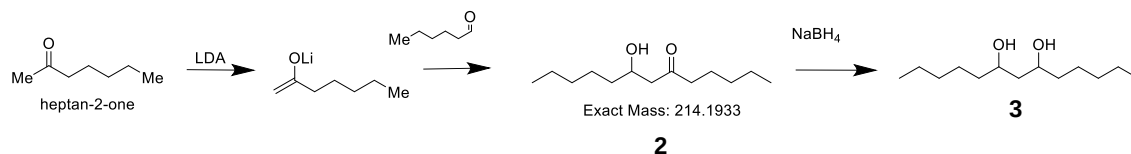
18. Structural models

A structural model of NocF was generated by AlphaFold⁷ and visualized in Chimera. Crystal structures of the aminotransferase 4e77, and kinase 5yz were obtained from the PDB.

19. Organic synthesis

19.1. Synthesis of 2

A solution of LDA (1 M in hexane, 1.05 eq., 3.15 mmol) in THF (1.5 mL) was treated at -78 °C with a solution of the ketone (3.0 mmol) in THF (3 mL). After 60 min, a solution of hexanal (3.0 mmol) in THF (3 mL) was added.⁸ After 10 min at the same temperature, the reaction mixture was cannulated with Tris-HCl pH-7 buffer (6 mL)⁹. The organic phase was separated, and the aq. phase thoroughly extracted with Et₂O. Drying of the combined organic extracts (MgSO₄), evaporation of the solvent and flash chromatography using mixtures of diethyl ether/ethyl acetate. Compound **2** was purified by flash chromatography, eluting with 10% Et₂O in hexanes (36.4 mg, 0.17 mmol, 56%) (Fig. SII.43, SII.45).



MS (ESI): calcd. for C₁₃H₃₀O₂ - ([M+H]⁺) 215.2006, found 215.2009

¹H NMR (500 MHz, CDCl₃): δ 4.02 (br, 1H), 2.60 (dd, J=17.6, 2.8, 1H), 2.51 (q, J=8.9, 8.0 2H), 2.42 (q, J=7.7 Hz, 2H), 1.64 (dt, J= 4H), 1.47 (m, 4H), 1.30 (m, 10H), 0.89 (m, 6H)

19.2. Synthesis of 3

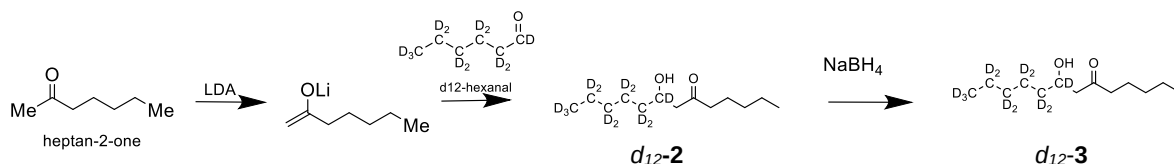
Compound **2** was reduced to compound **3** using sodium borohydride. In detail, 20 mg of the compound were dissolved in 3.5mL of MeOH and cooled down to 0°C, followed by the addition of NaBH₄ (1.2 eq.). After 1h, the reaction mixture was quenched with Tris-HCl pH=7. Compound **3** was purified by flash chromatography, eluting with 10% Et₂O in hexanes (16.9 mg, 0.078 mmol, 84%) (Fig. SII.43, SII.46).

MS (ESI): calcd. for $C_{13}H_{28}O_2$ - $([M+H]^+)$ 217.2162, found 217.2174

^1H NMR (500 MHz, $CDCl_3$): δ 4.18 (s, 1H), 3.91 (d, 1H), 1.59 (m, 2H), 1.44 (m, 2H), 1.31 (m, 16H), 0.88 (m, 6H)

19.3. Synthesis of d_{12} -2

A solution of LDA (1 M in hexane, 1.05 eq., 3.15 mmol) in THF (1.5 mL) was treated at -78 °C with a solution of the ketone (1.0 mmol) in THF (3 mL). After 60 min, a solution of d_{12} -hexanal (1.0 mmol) in THF (3 mL) was added. After 10 min at the same temperature, the reaction mixture was cannulated with Tris-HCl pH-7 buffer (6 mL) ⁹. The organic phase was separated, and the aq. phase thoroughly extracted with Et_2O . Drying of the combined organic extracts ($MgSO_4$), evaporation of the solvent and flash chromatography using mixtures of diethyl ether/ethyl acetate. Compound d_{12} -2 was purified by flash chromatography, eluting with 10% Et_2O in hexanes (15.4 mg, 0.068 mmol, 56%) (Fig. SII.44, SII.47).



MS (ESI): calcd. for $C_{13}H_{14}D_{12}O_2$ - $([M+H]^+)$ 227.2759, found 227.2770

^1H NMR (500 MHz, $CDCl_3$): δ 3.55 (br, 1H), 2.60 (d, $J=6.4$, 1H), 2.55 (m, 1H), 2.43 (dt, $J=1.8$, 7.5 Hz, 2H), 1.64 (t, $J=7.3$ 2H), 1.45 (m, 1H), 1.29 (m, 4H), 0.89 (m, 3H)

19.4. Synthesis of d_{12} -3

Compound d_{12} -2 was reduced to compound d_{12} -3 using sodium. In detail, 10 mg of the compound were dissolved in 3.5mL of MeOH and cooled down to 0°C, followed by the addition of $NaBH_4$ (1.2 eq.). After 1h, the reaction mixture was quenched with Tris-HCl pH=7. Compound d_{12} -3 was purified by flash chromatography, eluting with 10% Et_2O in hexanes (8.2 mg, 0.036 mmol, 81%) (Fig. SII.44, SII.48).

MS (ESI): calcd. for $C_{13}H_{16}D_{12}O_2$ - $([M+H]^+)$ 229.2915, found 229.2922

^1H NMR (500 MHz, $CDCl_3$): δ 3.40 (m, 1H), 1.55 (dd, $J=1.8$, 7.5 2H), 1.40 (m, 2H), 1.31 (m, 6H), 0.89 (m, 3H)

Table SII.1 - Detailed annotation of the *noc* gene cluster in *Nodularia* sp. LEGE 06071. Annotation based on HHPred and Swiss-Model (adapted from ¹⁰)

Noc	Membrane	Previous annotation ¹¹	HHPred			Swiss-model		
			Name	e-value	aa	name	identity	coverage
A	No	Aldolase	Aldolase II superfamily Fucose phosphate aldolase (MtnB)	2.8e-28	204	Methylthioribulose-1-phosphate dehydratase (MtnB) Putative aldolase class II	24.63 30.81	0.96 0.93
B	No	Dioxygenase	Acireductone Dioxygenase (MtnD)	1.1e-30	169	Acireductone dioxygenase (MtnD)	31.36	0.92
C	No	Dehydrogenase	Acyl-CoA dehydrogenase	9.2e-58	480	Short chain acyl-CoA dehydrogenase	26.81	0.70
D	Yes	Methyltransferase	Isoprenylcysteine O-methyltransferase	1.8e-25	191	Isoprenylcysteine carboxyl methyltransferase	27.37	0.77
E	No	N-monooxygenase	Flavin-containing monooxygenase	3.1e-62	475	Flavin-containing monooxygenase	34.90	0.96
F	Maybe the N terminal	Aminotransferase, kinase	Aminotransferase	1.9e-52	431	Aminotransferase class III	32.55	0.49
G	No	Acyl-ACP reductase	Ketosynthase (FabH homologue)	1.1e-36	314	Ketosynthase	22.04	0.90
H	No	Fatty-acyl AMP ligase, ACP	Amino acid adenylation domain protein	2.3e-61	697	Fatty-acyl AMP ligase	44.37	0.81
I	No	Oxidoreductase	Acyl-CoA dehydrogenase	1.4e-33	361	FAD-dependent oxidoreductase Glutaryl-CoA dehydrogenase	19.44 18.64	0.95 0.95
J	No	Ring-hydroxylating dioxygenase	3-ketosteroid 9 α -hydroxylase oxygenase	7.5e-45	322	3-ketosteroid-9- α -monooxygenase	25.57	0.94
K	No	Transferase	Glycine N-methyltransferase SAM-dependent methyltransferase	0.0037 0.0034	128 68	N-methyltransferase O-methyltransferase	10.29 11.36	0.06 0.09

Table SII.2 - NMR data of compound **6** in CD₃Cl (600 MHz)

Number	δC^a	Group	δH^a / mult., J (Hz)		HMBC [δC] ^b	COSY [δH]
1	-	OH	3.06	m	59.3	3.88
2	59.3	CH ₂	3.88	q, 5.5 Hz	172.1, 38.1	3.06, 2.42
3	38.1	CH ₂	2.42	m	172.1, 59.3	3.88
4	172.1	C	-	-		
5	-	NH	5.31	d, 8.8 Hz	172.1	3.92
6	49.5	CH	3.92	m	172.1	5.31, 1.49
7	35.4	CH ₂	1.49 1.34	m m	35.5, 25.7	3.92, 1.34, 1.49, 1.30
8	25.7	CH ₂	1.30	m	32.0	3.92, 1.49
9	32.0	CH ₂	1.27	m	25.7, 14.2	1.27, 0.87
10	22.7	CH ₂	1.27	m	25.7, 14.2	0.87, 1.27
11	14.2	CH ₃	0.87	m	32.0, 22.7	1.26, 1.30
12	35.5	CH ₂	1.48 1.32	m m	35.4, 29.7, 26.1	1.48, 3.92, 1.32, 1.49, 1.30
13	26.1	CH ₂	1.30	m	29.7, 31.9	1.48, 3.92
14	29.7	CH ₂	1.27	m	29.7, 29.4, 31.9, 14.3	1.27, 0.87
15	29.4	CH ₂	1.27	m	29.7, 29.4, 31.9, 14.3	1.27, 0.87
16	31.9	CH ₂	1.27	m	29.7, 29.4, 31.9, 14.3	1.27, 0.87
17	22.8	CH ₂	1.27	m	29.7, 29.4, 31.9, 14.3	1.27, 0.87
18	14.3	CH ₃	0.87	m	31.9, 22.8	1.27

Table SII.3 – Primers used to clone *NocE* in the several expression vectors, using restriction cloning or Gibson assembly (restriction sites: green - NdeI, red – XhoI, Gibson overhangs underlined)

Primer name	Primer Sequence	Description	Reference
NocE-28/29-F	CGGCCATATGAACTTACTAAAATCTTCC	Amplification of <i>nocE</i> for pET-28a cloning	This work
NocE-28-R	CTTTCTCGAGTTACTTCCGAAGTAGCTTG		
NocE-28/29-F	CGGCCATATGAACTTACTAAAATCTTCC	Amplification of <i>nocE</i> for pET-29b cloning	This work
NocE-29-R	CTTTCTCGAGTTACTTCCGAAGTAGCTTG		
NocE-SUMO-F	<u>GAAACCTGTACTTCCAAATGA</u> ACTTACTAAAATCTTC C	Amplification of <i>nocE</i> for pET His ₆ SUMO TEV LIC cloning	This work
NocE-SUMO-R	<u>GGCGCGCCGCGATCG</u> TTACTTCCGAAGTAGCTTG		
pET-SUMO-F	TTGGAAGTACAGGTTTTC	Amplification of pET His ₆ SUMO TEV	This work
pET-SUMO-R	CGATCGCGGCGCGCC		
NocE-51b-F	<u>GGTACCGGATCCGACGT</u> CGATGAAGTACTAAAATCT TCC	Amplification of <i>nocE</i> for pET-51b cloning	This work
NocE-51b-R	<u>GGAAATTGAGGAGAAGC</u> TTACTTCCGAAGTAGCTTG		
pET-51b-F	CGACGTGCGATCCGGTACC	Amplification of pET-51b	This work
pET-51b-R	GGCTTCTCCTCAATTTC		

Table SII.4 – Primers used to clone NocJ in the several expression vectors, using restriction cloning or Gibson assembly (restriction sites: green NdeI, red – XhoI, Gibson overhangs underlined)

Primer name	Primer Sequence	Description	Reference
NocJ-28a-F	CTGCCATATGTGCTTGAATCAAAATTTCC	Amplification of <i>nocJ</i> for pET-28a cloning	This work
NocJ-28a-R	CGGTCTCGAGTTAGTAAAATTGTTCAACC		
NocJ-29b-F	CTGCCATATGTGCTTGAATCAAAATTTCC	Amplification of <i>nocJ</i> for pET-29b cloning	This work
NocJ-29b-R	CGGTCTCGAGGTAAAATTGTTCAACC		
NocJ-SUMO-Strep-Fw	<u>ACAGAGAACAGATTGGTATGTGCTTGAATCAAAATT</u> TCC	Amplification of <i>nocJ</i> for pET-51b-SUMO cloning	This work
NocJ-SUMO-Strep-Rv	<u>AAATTGAGGAGAAGCCGTAAAATTGTTCAACCC</u>		
51-SUMO-Fw	ACCAATCTGTTCTCTGTGAGC	Amplification of pET-51b-SUMO cloning	This work
51-SUMO-Rv	GGCTTCTCCTCAATTTCCGC		

Table SII.5 – Primers used to clone NocF in the several expression vectors, using restriction cloning or Gibson assembly (restriction sites: green NdeI, red – XhoI, orange - EcoR1, blue – BamH, purple- HindIII; Gibson overhangs underlined)

Primer name	Primer Sequence	Description	Reference
NocF-28/29-F	CTTCCATATGTTGTTGCAACTTTTCAGG	Amplification of <i>nocF</i> for pET-28a cloning	This work
NocF-28-R	CGGTCTCGAGTTAAAGTGTAAGAAACC		
NocF-28/29-F	CTTCCATATGTTGTTGCAACTTTTCAGG	Amplification of <i>nocF</i> for pET-29b cloning	This work
NocF-29-R	CGGTCTCGAGAAAGTGTAAGAAACC		
NocF-165-F	CTAAGAATTCACGAGAACTAGGTTTACCAAAGC	Amplification of <i>nocF165*</i> for pET-22b cloning	This work
NocF-pet22b-R	CGGTCTCGAGAAGTGTAAGAAACC		
NocGcoexpress-F	TTATGGATCCATGAAAATTGAAGCAATC	Amplification of <i>nocGF</i> for pET-29b cloning	This work
NocFcoexpress-R	GATAAAGCTTAAGTGTAAGAAACC		
pet-41-F	GTGCTGGCGTTTCAATTTAGC	Amplification of pET-41a	This work
pet-41-R	GACAAGAGTCCCATGGGATATCGG		
NocF-F-41	GCTAAATTCGAACGCCAGATGAAATACCTGCTGCCGACC	Amplification of <i>nocF</i> for pET-41a cloning	This work
NocF-R-term-41	CCGATATCCCATGGGACTTTACAGGGTGAAAAAGCCATCTTCC		
pMAL-F	AACGACGGCCAGTGCCAAGCTTGC	Amplification of pMAL-2cx	This work
pMAL-R	CTTAATCGCCTTGCAGCACATCCC		
pET-51-F	AAGCTTGTCGACCGACGTCGG	Amplification of pET-51b	This work
pET-51-R	GCTCCGGGCTTCTCCTCAATTTCC		
pET-SUMO-F	ATTGGATTGGAAGTACAGGTTTTC	Amplification of pET-SUMO ^a	This work
pET-SUMO-R	CGATCGCGGCGCGCC		
NocF-F-pMAL	GCAAGCTTGGCACTGGCCATGAAATACCTGCTGCCGACC	Amplification of <i>nocF</i> for pMAL-2cx cloning	This work
NocF-R-term-pMAL	GGGATGTGCTGCAAGGCGTTACAGGGTGAAAAAGCCATCTTCC		
NocF-F-51	CCGACGTCGGTCGACAAGATGAAATACCTGCTGCCGACC	Amplification of <i>nocF</i> for pET-51b cloning	This work
NocF-R-term-51	GGAAATTGAGGAGAAGCCTTACAGGGTGAAAAAGCCATCTTCC		
NocF-F-SUMO	GAAAACCTGTACTTCCAAATGAAATACCTGCTGCCGACC	Amplification of <i>nocF</i> for pET-SUMO ^a	This work
NocF-R-term-SUMO	GGCGCGCCGCGATCGTTACAGGGTGAAAAAGCCATCTTCC		
NocF165-F-SUMO	GAAAACCTGTACTTCCAAATGCGCGAACTGGGCCTGCGC	Amplification of <i>nocF165*</i> for pET-SUMO ^a	This work
NocF-R-term-SUMO	GGCGCGCCGCGATCGTTACAGGGTGAAAAAGCCATCTTCC		
pET-43-F	TCCACGAGCTCCCGG	Amplification of pET-43.1a	This work
pET-43-R	GAAGCTTGC GGCCGC		
F-43-F	CCGGGAGCTCGTGAATGCTGCTGCAGCTGTTTCGC	Amplification of <i>nocF</i> for pET-43.1a cloning	This work
F-43-R	GCGGCCGCAAGCTTCTTACAGGGTGAAAAAGCC		
F165-43-F	CCGGGAGCTCGTGAATGCGCGAACTGGGCCTGCCGAAACG	Amplification of <i>nocF165*</i> for pET-43.1a cloning	This work
F-43-R	GCGGCCGCAAGCTTCTTACAGGGTGAAAAAGCC		
pET-32-F	GTCGGTACCCAGATCTGG	Amplification of pET-32a	This work
pET-32-R	CGTCGACAAGCTTGCGGC		
F-32-F	CCAGATCTGGGTACCGACATGCTGCTGCAGCTGTTTCGC	Amplification of <i>nocF165*</i> for pET-32a cloning	This work
F-32-R	GCCGCAAGCTTGTGACAGTTACAGGGTGAAAAAGCC		
F165-32-F	GCCGCAAGCTTGTGACAGATGCGCGAACTGGGCCTGCCGAAACG	Amplification of <i>nocF165*</i> for pET-32a cloning	This work
F-32-R	GCCGCAAGCTTGTGACAGTTACAGGGTGAAAAAGCC		
NocF-Strep-F	GGTACCGGATCCGACGTCGATGAAATACCTGCTGCCGACC	Amplification N-pelB-SUMO-NocF(165)-His ₆ for pET-51b cloning	This work
NocF-Strep-R	GGAAATTGAGGAGAAGCCGGCTTTGTTAGCAGCCGG		
pET-51-F-SUMO	CGACGTCGGATCCGGTACC	Amplification of pET-51b cloning (for N-pelB-SUMO-NocF(165)-His ₆)	This work
pET-51-R-SUMO	GGCTTCTCCTCAATTTCC		

Table SII.6 – Expression plasmids used for NocF cloning, their antibiotic resistance and expected recombinant protein and size

Plasmid	Resistance	Tag	Recombinant NocF protein	Protein MW
pET-41a	Km	N-GST (Glutathione S-transferase), N-His ₆	N-GST- His ₆ -pelB- NocF	112.3
pMAL-2cx	Amp	N-MBP (maltose binding protein), C-His ₆	MBP-pelB- NocF	123.1
pET-51b	Amp	C-His ₆ , N-Strep II (Streptavidin)	N-Strep-pelB- NocF-His ₆	105.5
			N-Strep-SUMO-pelB- NocF-His ₆	124.3
			N-Strep-SUMO-pelB- NocF165-His ₆	105.5
pET His6 SUMO TEV	Amp	N-His ₆ , N-SUMO (Small ubiquitin-like modifier)	N-His ₆ SUMO-pelB-NocF	118.8
			N-His ₆ -SUMO-pelB-NocF165	100.0
pET-29b	Km	C-His ₆	C-His ₆ -SUMO-NocF	100.0
pET-28a	Km	N-His ₆	N-His ₆ -SUMO-NocF	100.8
pET-22b	Amp	C-His ₆	pelB- NocF-C-His ₆	100.1
pET-43.1a	Amp	N-His ₆ , C-His ₆ , Nus.tag (N-utilization substance)	N-His ₆ -Nus-pelB- NocF	167.2
pET-32a	Amp	N-Trx (Thioredoxin), C-His ₆ ,	N-Trx-pelB- NocF- His ₆	127.5

Table SII.7 – Primers used to clone NocD from LEGE 06071 and PCC 7108 in the several expression vectors, using restriction cloning (restriction sites: green NdeI, red – XhoI)

Primer name	Primer Sequence	Description	Reference
NocD-LEGE-28/29-F	CGGCCATATGATGATTATAACCAATTT	Amplification of <i>nocD</i> (LEGE) for pET-28a cloning	This work
NocD-LEGE-28-R	TAAGCTCGAGCTAATAAATCAACGGAATTATC		
NocD-LEGE-28/29-F	CGGCCATATGATGATTATAACCAATTT	Amplification of <i>nocD</i> (LEGE) for pET-29b cloning	This work
NocD-LEGE -29-R	TAAGCTCGAGATAAATCAACGGAATTA		
NocD-PCC-28/29-F	TGGCCATATGAGCATATTCAAAGCTATTG	Amplification of <i>nocD</i> (PCC) for pET-28a cloning	This work
NocD- PCC -28-R	TCAGCTCGAGCTAATAAATCAGAGGAATT		
NocD- PCC-28/29-F	TGGCCATATGAGCATATTCAAAGCTATTG	Amplification of <i>nocD</i> (PCC) for pET-29b cloning	This work
NocD-PCC -29-R	GCAGCTCGAGATAAATCAGAGGAATT		

Table SII.8 – Primers used to clone NocK in the several expression vectors, using restriction cloning or Gibson assembly (restriction sites: purple - BamHI. red – XhoI, Gibson overhangs underlined)

Primer name	Primer Sequence	Description	Reference
NocK-28a-F	TCCAGGATCCATGCCAAATCTAATCAAG	Amplification of <i>nocK</i> for pET-28a cloning	This work
NocK-28a-R	CATTCTCGAGCTAACTAATCCCAGCTAC		
NocK-29b-F	TCCAGGATCCATGCCAAATCTAATCAAG	Amplification of <i>nocK</i> for pET-29b cloning	This work
NocK-29b-R	CATTCTCGAGCTAATCCCAGCTAC		
NocK-SUMO-Strep-Fw	ACAGAGAACAGATTGGTATGCCAAATCTAATCAAG	Amplification of <i>nocK</i> for pET-51b-SUMO cloning	This work
NocK-SUMO-Strep-Rv	AAATTGAGGAGAAGCCACTAATCCCAGCTACTG		
51-SUMO-Fw	ACCAATCTGTTCTCTGTGAGC	Amplification of pET-51b-SUMO cloning	This work
51-SUMO-Rv	GGCTTCTCCTCAATTTCCGC		

~

Table SII.9 – Parameters used for the construction of NocE, NocF, NocJ and NocK SSNs

Protein	BLAST sequences	InterPro family	Sequence length	Alignment score	%ID
NocE	1000	-	450-625	140	42
NocF	1000	IPR005814	200-500	54	53
NocJ	1000	-	200-600	70	49
NocK	1000	IPR021829	275-460	50	55

Table SII.10 - Primers used to clone NocE and NocF flanking regions for knockout construction (Gibson overhangs underlined)

Primer name	Primer Sequence	Description	Size (bp)	Reference
pRL278-F	AACTCCAGCATGAGATCC	Amplification of the pEL278 vector backbone	4929	This work
pRL278-R	CTGTCAGACCAAGTTTACTC			
nocE-KO-5F	<u>GGATCTCATGCTGGAGTTGATTTTCAGACCA</u> AGGAATG	Amplification of the 5' <i>nocE</i> flanking region	1004	This work
nocE-KO-5R	<u>CAAAATATTAGTCTATACTATTAACAATAGAG</u> AATCAGCTTCTTATC			
nocE-KO-3F	<u>GATAAGAAGCTGATTCTCTATTGTTAATAGTA</u> TAGGACTAATATTTG	Amplification of the 3' <i>nocE</i> flanking region	1163	This work
nocE-KO-3R	<u>GAGTAAACTTGGTCTGACAGCCAATAATCGT</u> TTGAGTTTCG			
nocF-KO-5F	<u>GGATCTCATGCTGGAGTTGAGGAATATAGCA</u> AACCAACC	Amplification of the 5' <i>nocF</i> flanking region	1105	This work
nocF-KO-5R	<u>CTGTTTAACCTCTTGTTAATTAATGCTGTTTA</u> ACCTCTTGTTAATTAATG			
nocF-KO-3F	<u>CATTAATTAACAAGAGGTTAAACAGGTTTAGT</u> TAGTGAATCAGAGATAAAG	Amplification of the 3' <i>nocF</i> flanking region	1104	This work
nocF-KO-3R	<u>GAGTAAACTTGGTCTGACAGCGATATCTACT</u> AATTCGCTGC			

Table SII.11 - List of Plasmids used for the knockout creation and its antibiotic resistances

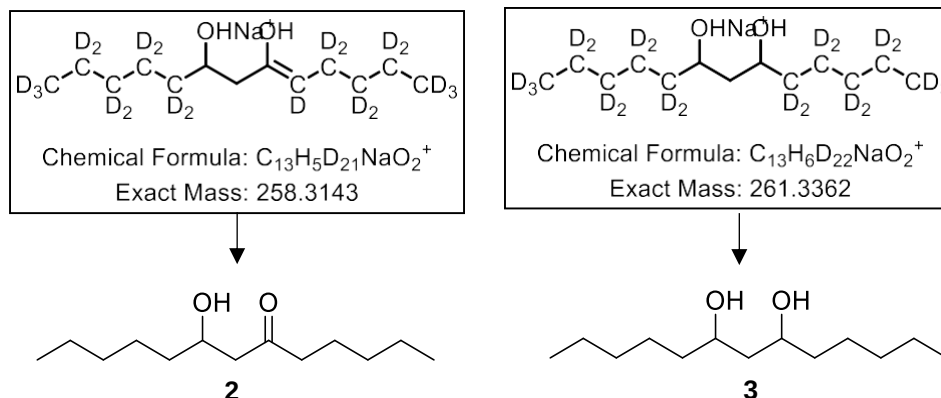
Plasmids	Antibiotic resistances
pRL278	Km ^r , Nm ^r
pRL278::nocE (cargo)	Km ^r , Nm ^r
pRL623 (helper)	Cm ^r
RP4 and pRL443 (conjugative)	Amp ^r

Table SII.12 - Primers used for the DiPaC strategy (PmlI restriction site in red, Gibson overhangs underlined)

Primer name	Primer Sequence	Description	Size (bp)	Reference
C-GFP for 1	CATGGTTAGCAAAGGTGAAG	Amplification of the pET28b-ptetO::gfpv2 vector backbone	6 552	12
spec-ptet-R	GGTCGATCCTCTTCTCTATC			
06071_v1_F1	<u>TCAGTGATAGAGAAGAGGATCGACCATGCC</u> <u>AAAATCTAATCAAGAAAAG</u>	Amplification of the <i>nocGHJK</i> with the Gibson overhangs	5678	This work
06071_v1_R1	<u>TTCACCTTTGCTAACCATG</u> <u>CACGTG</u> <u>TTAATA</u> <u>GATAAATGAATAGG</u>			
06071_v1.2_F2	<u>CCTATTCATTTATCTATTAAC</u> <u>CAC</u> <u>T</u> <u>T</u> <u>TAATGTTGT</u> <u>TGCAACTTTTCAGGTTTATTGG</u>	Amplification of the <i>nocABDEF</i> with the Gibson overhangs	6687	This work
06071_v1.2_R2	<u>TTCACCTTTGCTAACCATG</u> <u>CACGTG</u> <u>TTTGCTT</u> <u>ACTCTACGAGAAGAGC</u>			
06071_v1.2_NT_F1	ATGCCAAAATCTAATCAAGAAAAG	Amplification of the <i>nocGHJK</i> without Gibson overhangs	5634	This work
06071_v1.2_NT_R1	TTAATAGATAAATGAATAGG			
06071_v1.2_NT_F2	TTAATGTTGTTGCAACTTTTCAGGTTTATTGG	Amplification of the <i>nocABDEF</i> without Gibson overhangs	6633	This work
06071_v1.2_NT_R2	TTTGCTTACTCTACGAGAAGAGC			
Screen_ptetF1	TCCGACCTCATTAAGCAGC	Colony screening of pET-28:: <i>nocK</i> region	695	12
06071_colony_R1	CAGTACGAATCAGAGAATTCC			This work
06071_colony_F3	ATCAGCATCTAATTTGCAGCC	Colony screening of pET-28:: <i>nocA</i> region	574	This work
screen_GFP_R	TTACCGTTGGTCGCATCACC			12
DIPAC-F1-F	<u>TTATCTATTAAATGAACTT</u> <u>ACTAAATCTTCC</u>	Amplification of fragment 1 for ΔNocF DIPAC mutant with the Gibson overhangs	8136	This work
DIPAC-F1-R	<u>CTGTCTTCTTTATCATGCAACT</u> <u>CGTAGG</u>			
DIPAC-F2-F	<u>TTGCATGATAAAGAAGACAG</u> <u>TCATAAGTGC</u>	Amplification of fragment 2 for ΔNocF DIPAC mutant with the Gibson overhangs	9101	This work
DIPAC-F2-R	<u>TAAGTTCATTTAATAGATAAAT</u> <u>GAATAGGCGC</u>			
DIPAC-E1-F	<u>ACTTTAAATGATGATTTA</u> <u>TAAACCAATTTTGA</u> <u>GC</u>	Amplification of fragment 1 for ΔNocE DIPAC mutant with the Gibson overhangs	8123	This work
DIPAC-E1-R	<u>GAATTGAGATCTCGATCCTCTACG</u>			
DIPAC-E2-F	<u>GTAGAGGATCGAGATCTCAATT</u> <u>CG</u>	Amplification of fragment 2 for ΔNocE DIPAC mutant with the Gibson overhangs	10178	This work
DIPAC-E2-R	<u>TAAATCATCATTTAAAGT</u> <u>GTAAGAAACCATC</u> <u>TTC</u>			

a

supplementation experiment	m/z $[M+Na]^+$	Molecular formula	unlabeled m/z	labeling
d_{11} -hexanoic acid	258.3143	$C_{13}H_5D_{21}NaO_2^+$	214.1933	d_{21}
	261.3362	$C_{13}H_6D_{22}NaO_2^+$	216.2089	d_{22}
d_{15} -octanoic acid	252.2797	$C_{13}H_{15}D_{13}NaO_2^+$	216.2089	d_{13}



b

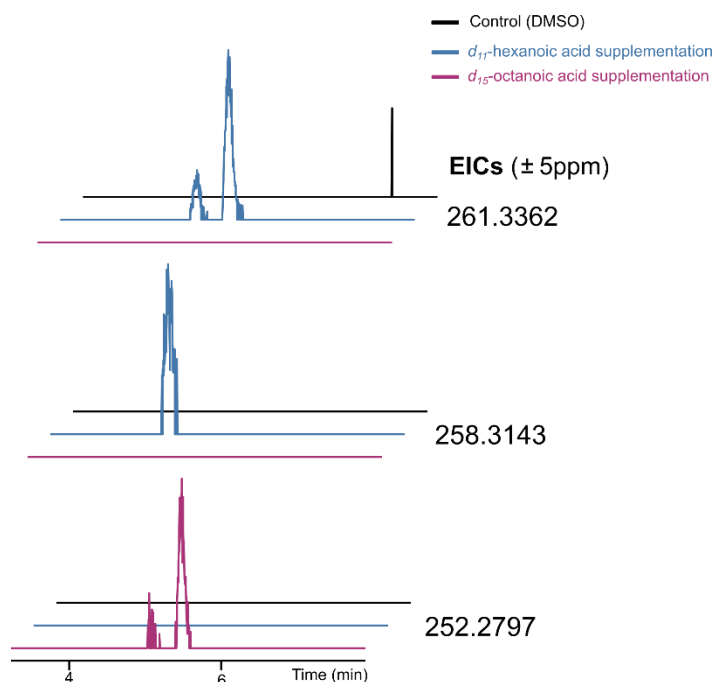


Figure SII.1 - Putative intermediates found in *Nodularia* sp. LEGE06071 supplementation experiments. (a) m/z values, chemical formulas, and putative structures of **2** and **3** intermediates found upon d_{11} -hexanoic acid and d_{15} -octanoic acid supplementation experiments in *Nodularia* sp. LEGE 06071 (b). LC-HRMS-derived EICs of d_{21} -**2** (m/z $[M+Na]^+$ 258.3143) and d_{22} -**3** (m/z $[M+Na]^+$ 261.3362) and d_{22} -**3** (m/z $[M+Na]^+$ 252.2797) in upon d_{11} -hexanoic acid and d_{15} -octanoic acid supplementation experiments in *Nodularia* sp. LEGE 06071.

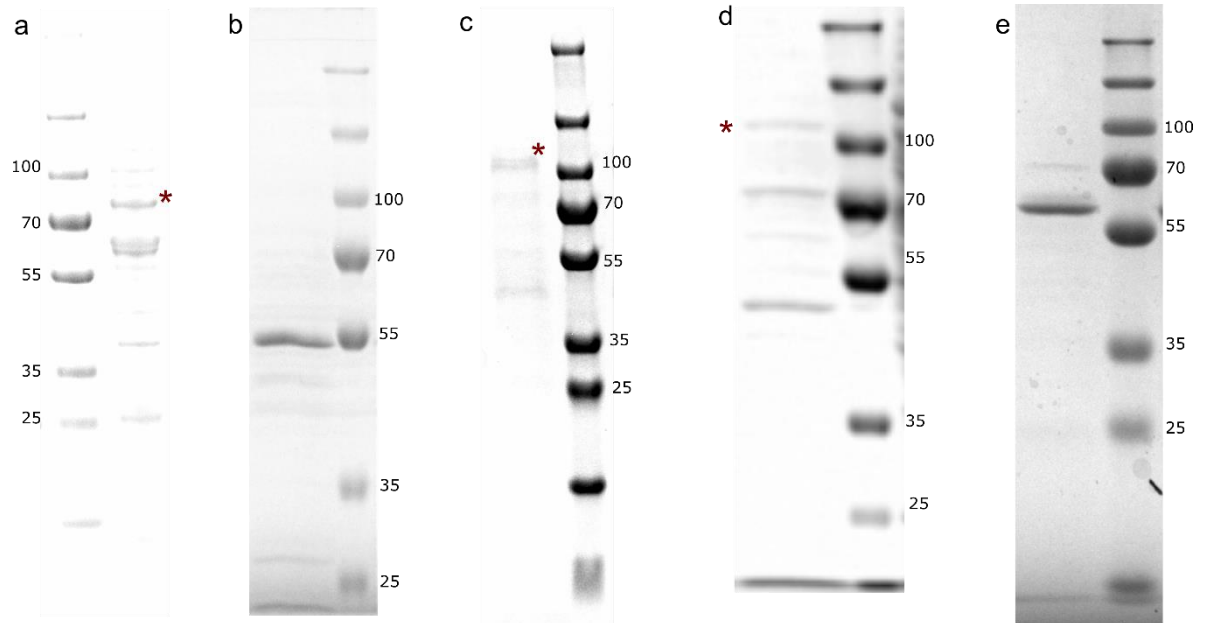


Figure SII.2 - SDS-Page of N-His₆-SUMO-NocE (a), N-Strep-SUMO-NocJ-His₆ (b), N-Strep-SUMO-pelB-NocF165-His₆ (c), N-His₆-SUMO-pelB-NocF (d) and N-Strep-SUMO-NocK-His₆ (e) purifications. All proteins were purified by Ni-IDA packed columns (Macherey Nagel) except for NocF purifications were performed with cobalt resin. Target proteins highlighted with *.

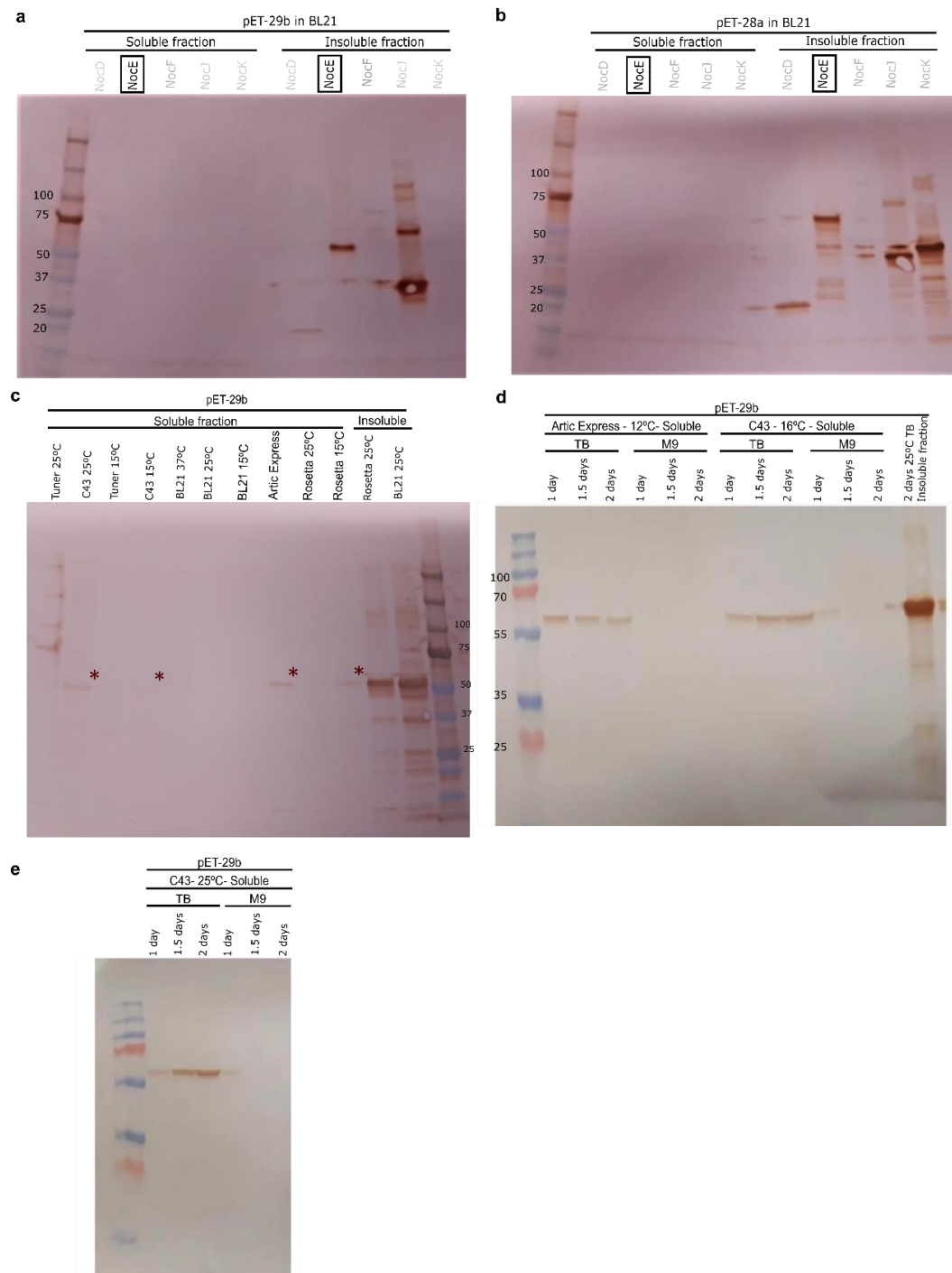


Figure SII.3 - Western blots of NocE expression tests using pET-29b and pET-28a, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation (unless indicated otherwise). (a) pET-29b::*nocE* and expressed in BL21(DE3) was unable to express any soluble NocE. (b) pET-28a::*nocE* and expressed in BL21(DE3) was unable to express any soluble NocE. (c) pET-29b::*nocE* and expressed in C43(DE3) at 25 °C and 15 °C, Rosetta(DE3) at 15 °C and Artic Express at 12 °C were able to produce soluble NocE. (d) pET-29b::*nocE* and expressed at 12 °C in Artic Express and 16 °C in C43 were able to produce soluble NocE. This production occurred in all expression times tests – 1, 1.5, 2 days. The expression was only observed in cultures grown in terrific broth and not in M9, indicating that the culture medium has a determinant role on the NocE expression. (e) pET-29b::*nocE* and expressed at 25 °C in C43 was able to produce soluble NocE using TB medium in all time points.

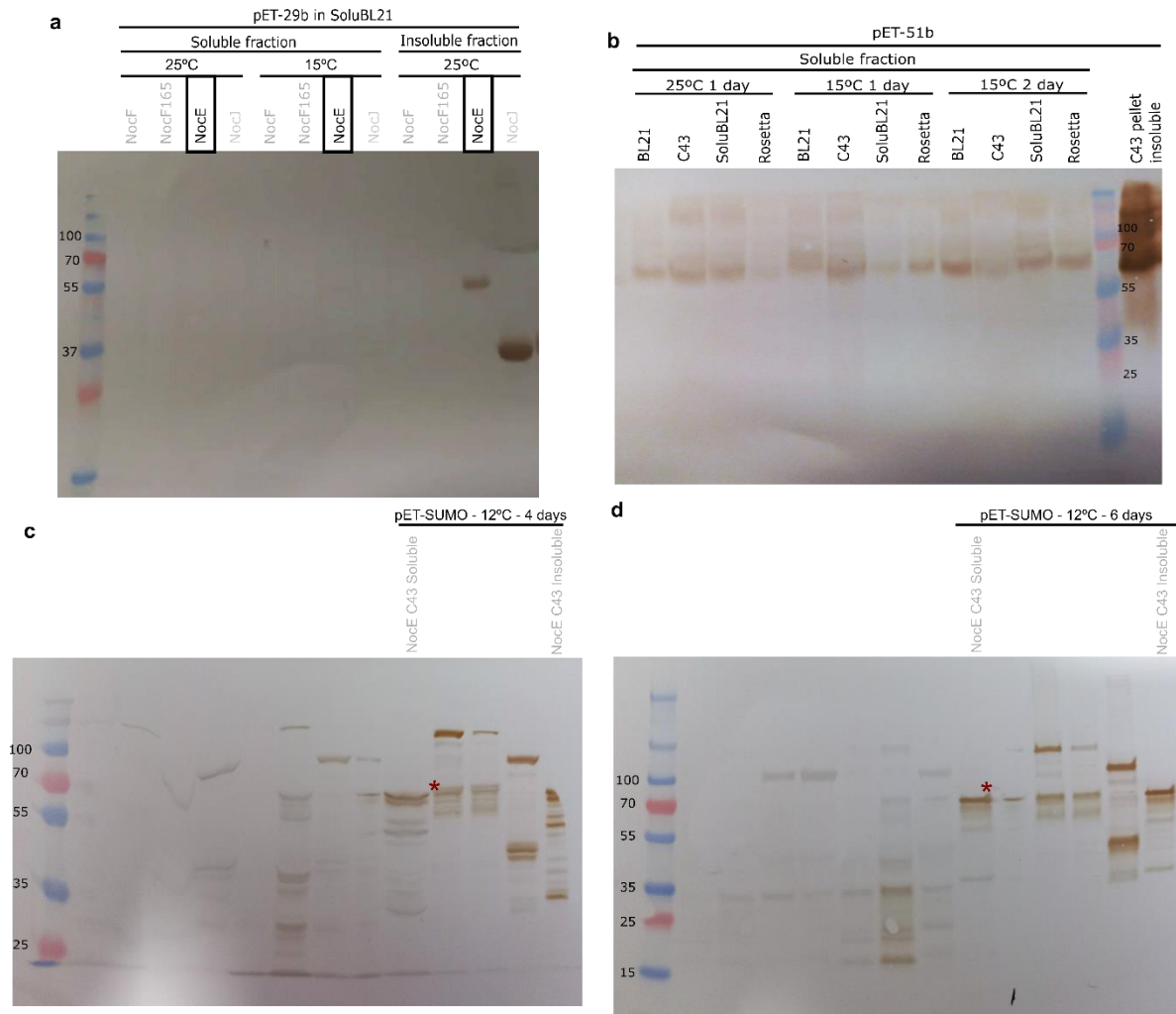


Figure SII.4 - Western blots of NocE expression tests using pET-29b, pET-51b, pET-His₆-SUMO-TEV, terrific broth (TB), 0.5 mM IPTG, 25 °C overnight (unless indicated otherwise). (a) pET-29b::*nocE* and expressed in SoluBL21(DE3) was unable to express any soluble NocE. (b) pET-51b::*nocE* and expressed in BL21(DE3), C43(DE3), Rosetta(DE3) and SoluBL21(DE3) was able to express soluble NocE in all tested conditions. (c) pET-His₆-SUMO::*nocE* and expressed at 12 °C for 4 days was able to produce soluble NocE. (d) pET-His₆-SUMO::*nocE* and expressed at 12 °C for 6 days was able to produce soluble NocE

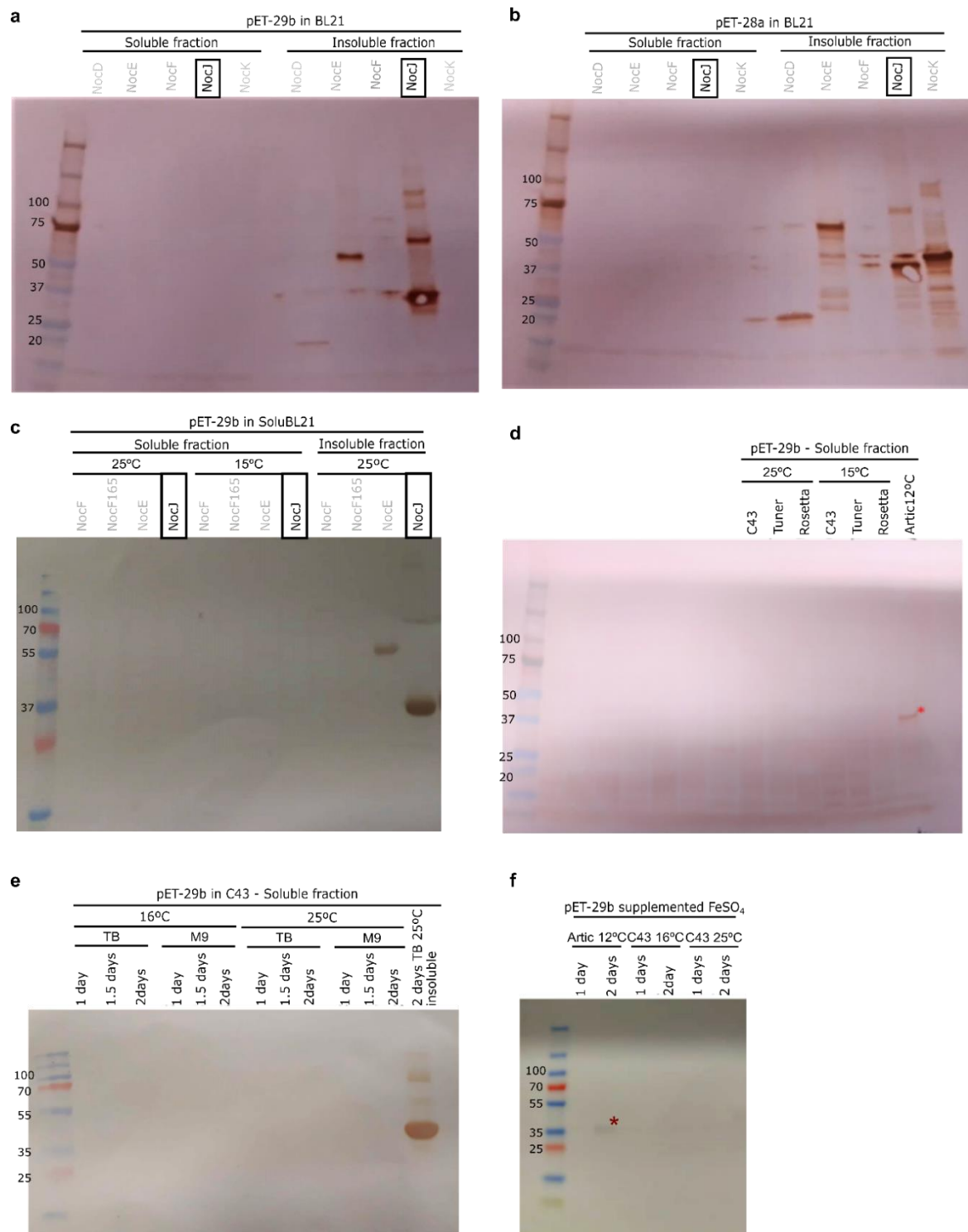


Figure SII.5 - Western blots of NocJ expression tests using pET-29b and pET-28a, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation (unless indicated otherwise). (a) pET-29b::*nocJ* and expressed in BL21(DE3) was unable to express any soluble NocJ. (b) pET-28a::*nocJ* and expressed in BL21(DE3) was unable to express any soluble NocJ. (c) pET-29b::*nocJ* and expressed in SoluBL21(DE3) was unable to express any soluble NocJ. (d) pET-29b::*nocJ* and expressed in Artic Express at 12 °C was able to produce a small amount of soluble NocJ. (e) pET-29b::*nocJ* and expressed C43(DE3) and expressed in Artic Express was able to produce a small amount of soluble NocJ after 2 days of expression in TB supplemented with 0.2 mM FeSO₄.

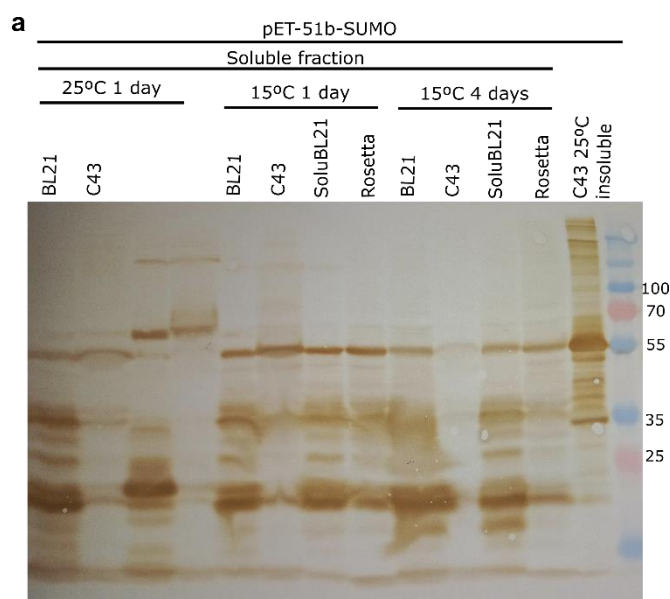


Figure SII.6 - Western blots of NocJ expression tests using pET-51b-SUMO, terrific broth (TB), 0.5 mM IPTG, 25 °C or 15 °C and 1 and 4 days incubation (unless indicated otherwise). pET-51b-SUMO::*nocJ* and expressed in BL21(DE3), C43(DE3), SoluBL21(DE3) and Rosetta(DE3) was able to produce significant amounts of soluble NocJ in all tested conditions. Incubation for 1 day at 15 °C provided the best expression conditions. The background on the western blot was caused by the use of an old antibody.

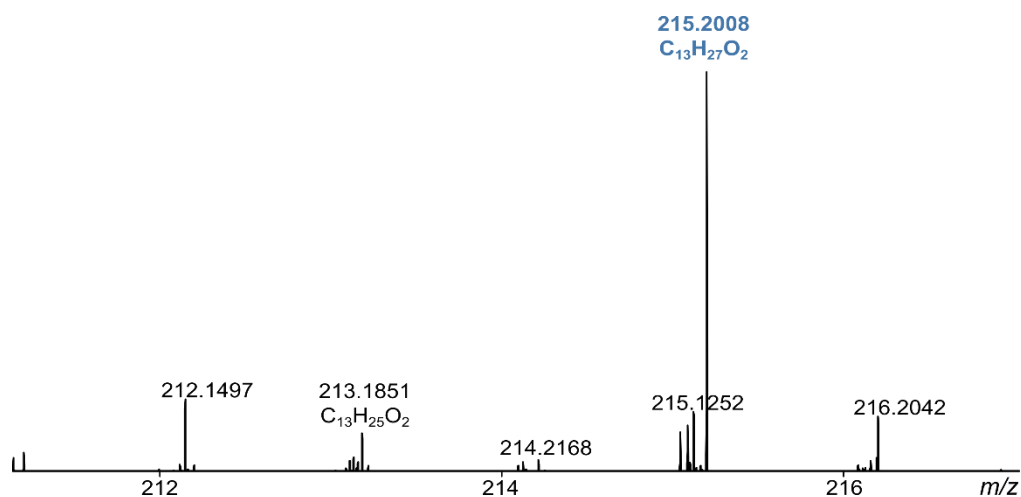


Figure SII.7 - HR-MS/MS spectrum of **2** (in blue) from NocE enzymatic assay with 6-tridecanone.

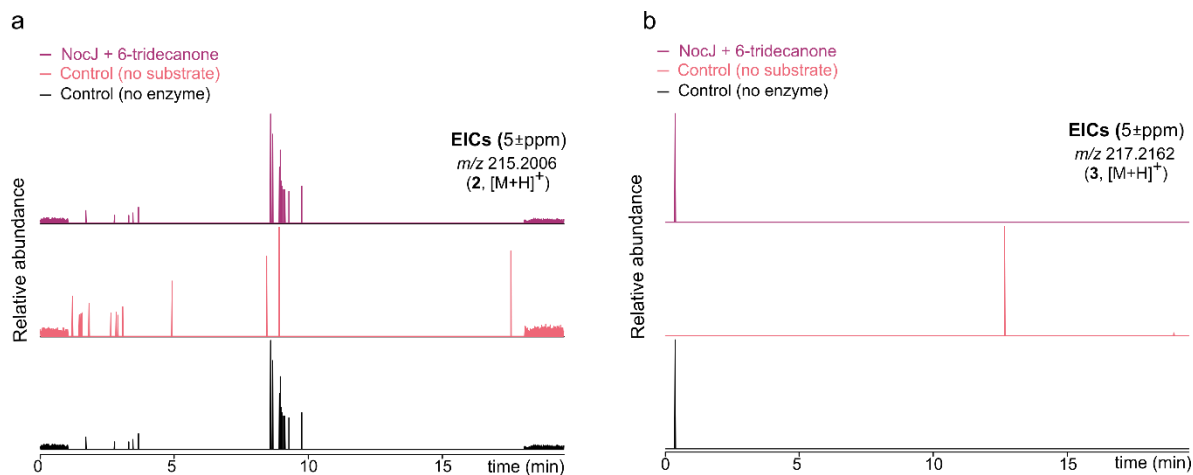


Figure SII.8 - LC-HRMS-derived EICs of **2** (a) and **3** (b) in NocJ enzymatic assays with 6-tridecanone. NocJ was not able to generate **2** or **3**.

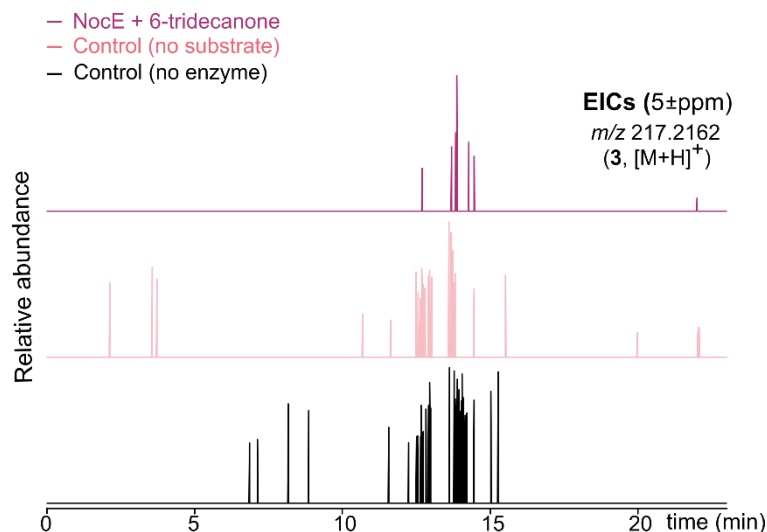


Figure SII.9 - LC-HRMS-derived EICs of **3** in NocE enzymatic assays with 6-tridecanone. NocE was not able to generate **3**.

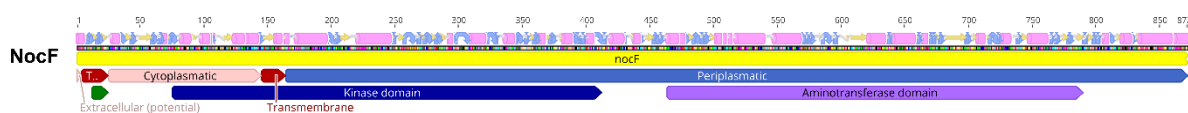


Figure SII.10 – Predicted domains of NocF and cell location. The kinase domain is located on the N-terminal region of the protein (74-412 aa) and the aminotransferase is located on the C-terminal (426-790 aa).

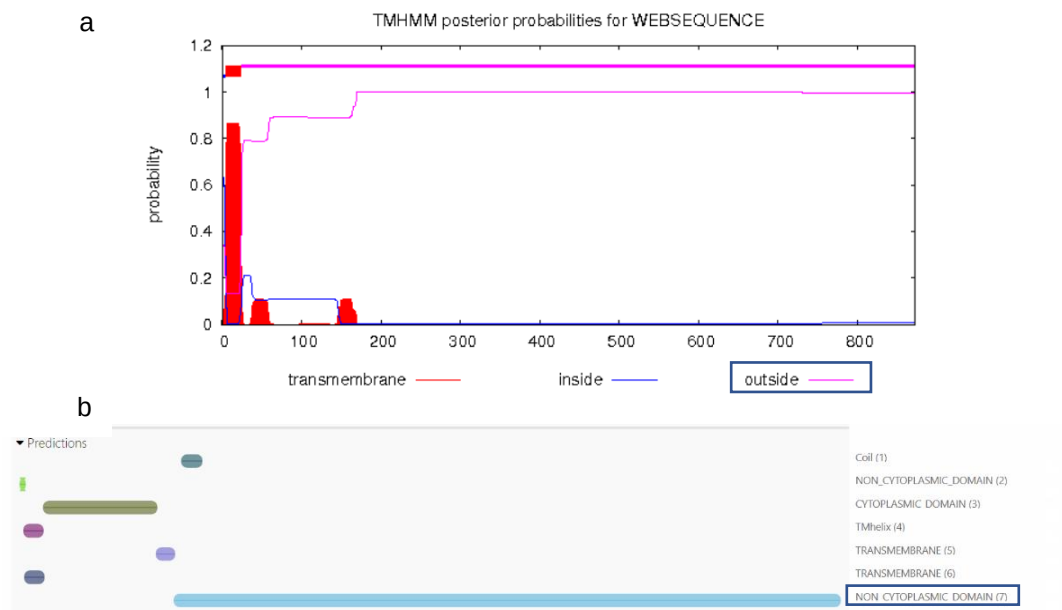


Figure SII.11 – Swiss model (a) and InterPro (b) prediction of NocF location in the cell. The N-terminal region of NocF is cytoplasmatic alternated with small transmembrane domains. After amino acid 165, the protein is predicted to be entirely periplasmatic.

GGATCCGATGCTGCTGCAGCTGTTTCGCTTTATTGGTCTGACCTTTATCTATCTGTGTGGTTATTTCTGACCACCATTTGGCAAAAATAAGGAAGAACGT
 ATTCTGAAACAGGGTCAGTGGATCTATCGTTATTTCTGCAGATGGGCCCGGTTTATATTAAGATTGGTCAGATTCTGGCCACCCGCGAGCATCTGATTCCGGAAA
 ATTGGGTGCAGATTCTGCGTGAACTGCAGGATAATGTTCCGAGCATGAGTGAAGCAGATACCCGTAATAATTATACCAAGATCTGCCGCTGGCATTGAAAAATGT
 TTTAGCGAATTTCTTCCAGCCGCTGGCAAGTGGCAGTATTGCCAGGTTTCATAGCGCAACCCCTGCTGACCGGTGCAAGATGGCAGTGAAGATTGTTAAAAA
 GAATGTGCGTAATCAGCTGAAACAGAATCTGGATATTATTCAGTATTTTCGTTGCCATTCTGGATTTTCTGGTGACCAAGTTCGCGCAACTGGGCCGCGCAACGT
 CTGCAGGAACCTGCGCAATCTGCTGATTATTCAGGCCGATATGAACAGGAAGCCAGAAACAGGAAGTATTAATAATTTCAAGAACCCCGATGTTAAAG
 TGCCGGCCATTATTAAGGATCTGAGTACCAGCAATCTGCTGTTTATGGAATTCATGGAAGGCATTCCGGGTAAAGATGTTCAAAAGTTGAACTGAAACGCAATCT
 GTTAGCACAGCGCTTTCAGGATACCATCTATACCATGCTGTATATGCATGGCCTGTGCCACGGTGACCCGATCCGGGTAAATTTTCTTTACCAAAGATGGTAGC
 ATCATTCTGCTGGATTATGGCATTACCGTGCAGCTGAGCGAAGATGAAAAATGGGGTCTGAGCAGTTTTTATTATGCATGCACCCGCAAGAAATGGGATGTGGCA
 GTGGAACGCTTTACCCAGCATTTTGTGACCGATAAAGATTATATTAGCCAGAATCTGAGCCAGTATGCCGAAGAAATTAGAGTGTGCTGCAGTATCATTTTGATA
 TTAGCACCCAGTGGAGTACCGTGAAGTATTTTTCAGGATGTTAGTAAATTTCTGCGTAAATATCGCGCCCGCTATACCCCAATTTTACCAAAGTGAAGTGGT
 TTTTCTGAGCTGCGAAGTTTTGCAAGCCAGATTGATCCGAATATTGATATTTGGGAAAAACGCCGTAATTTCACTGATCGTTATAGCCCGTATATGAATCCGGAA
 GTGAAAGAAACCTTTGATGCCTATTTTCAAGAAACCGCCCGAGTAGCCTGGCCATGCGTGATCGTGCAAAATAGTCTGGTGGCCCGACCCATATTAATCGC
 TATTTCTTTCCGAGCACCTATCCGGTTTTTGTGAAAAAGCCGATAAAAGTAAGATCGAAGATCTGGATGGCAATGTGTTTGTGGATATTAGTTGCGGTTATGGTCC
 GCATCTGCTGGGCTATGCACATCCGGTTATTAATCAGGCAATTAGCGAAGCCATTGCCAATGGCTTTGTTAATGCAATTGGTCATGAAGCCGAAGTACCCCTGGC
 CGAAATCTGGTGGAAGCACTGCCGGGCGCAGAAAAAGCCATTCTGAGTAATAGTGGTACCGAAAGCATTCTGCAGGCAATGCGTCTGTGTGCTGCCTATCGTA
 AAAAAAGATCGTGTGGCAAAATTTGAAGGCCATTATCATGGCTTTAGCGATCAGGGCATGGTTAGCAGTGGTTTTCGCTTTACCCGGTAAAAATTTGATCCGCAAGC
 CGATTGCCGCTACCCAGGCGACCGATAGTGCCACCGTGAAAAATACCCGTTTCTGCAGTATGTTTATTTGAAGGTCTGGAACGTCTGCGCAGTGAAGCAAGC
 GAACTGGCCTGTGTGATTTGCGAACCAGTCCGAGCCCTGCTGGCAAGCTATGATGTTGATTTTCTGACCAAACTGCGCGCAATTTGTACCGAACTGGATATTCG
 CTGATTTTATGATGAAGTTGTTAGTGGTTTTGCTGTGGCCTATGGCGGTGTGCAGAACTGCGCAGGCATTTATCCGGATCTGACCTGCCTGGGTAAAGTTATTGGC
 GGTGGCCTGCCGTGTGGTCCGTTGTGGGTAAAAAGAACTGATTGATATTGGTAAAGCAGTCAGGACCCCTTTTGCATTATGAAAAAGAAAGCATTGTGGGT
 GGTACCATGAGTGGCAATAGCATTACCTGTCTGGCCGGTACCGCAGCACTGACCTATCTGAAAGAACATCCGAAATCTATGTGAACTGGATAATAATACCAAC
 TGGCTGGGTGAGAACTGCTGGAATTTGGCCGAGTCATGGTGTTCGATTTCAGGTTAAAGCAATCGTAGCATTTTATGCTGACCTTTAGCCATAAAAGCGCC
 ACCTATTTTCTGAAAAACAGAGCGGTAGTAATTTAAAGTGAATCTGGCACTGGCCTATTATATGCGCAAAACATGGTATCTATATGCCGAACTGCATACCCGTAT
 GCTGAGTGCAGCATACCCAGGAAGATCTGGAAGTATTGACCCGATTCATCTGAGCCTGCAGGAATGCTGGAAGATGGCTTTTACCCCTGTAACTCGA
 G

GGATCC – BamHI
 CTCGAG – XhoI

Figure SII.12 – NocF sequence after codon-optimization for *E. coli*.

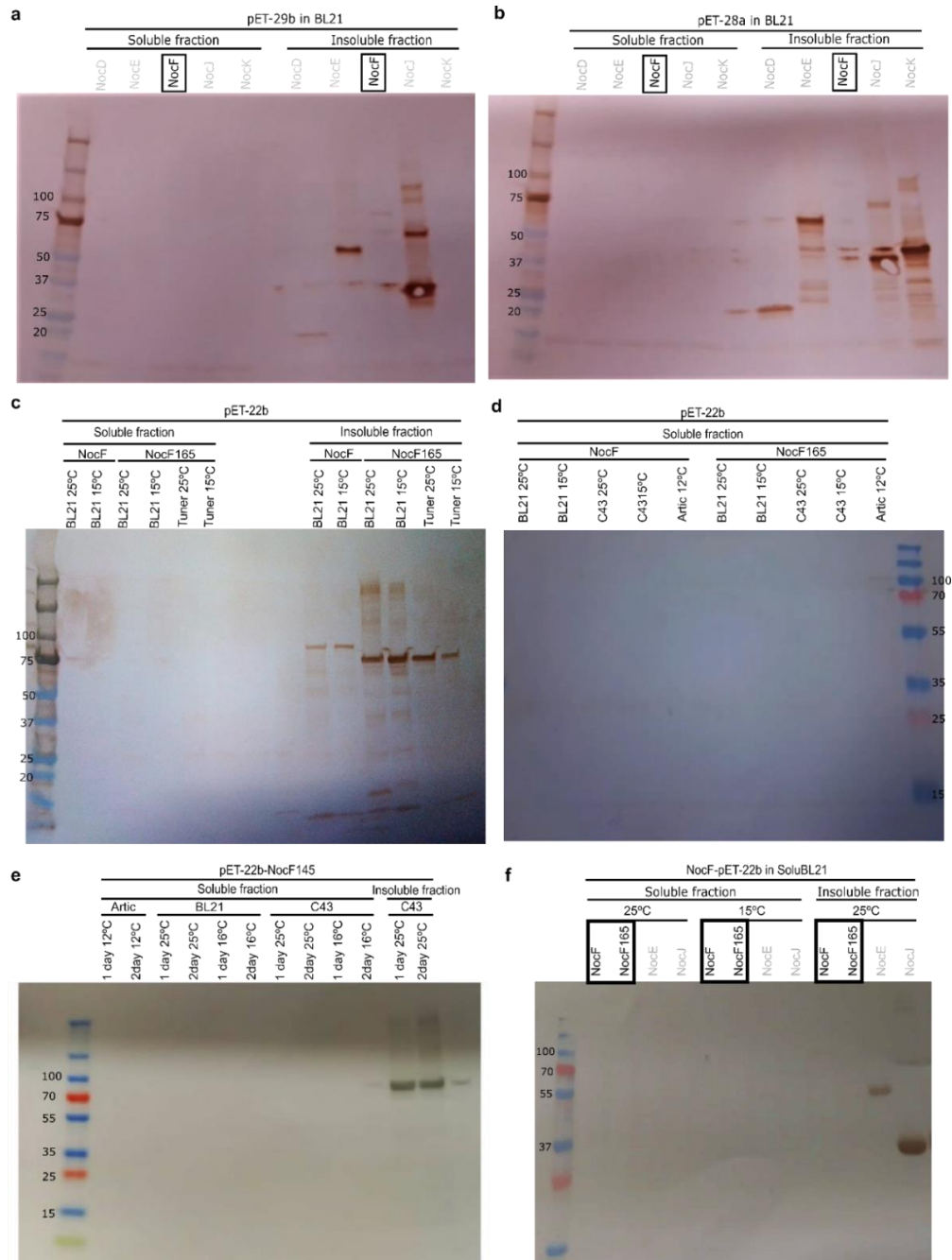


Figure SII.13 – Western blots of NocF (codon optimized) expression tests using pET-29b, pET-28a and pET-22b, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation (unless indicated otherwise). (a) pET-29b::*nocF* (original gene) and expressed in BL21(DE3) was unable to express any soluble and insoluble NocF. (b) pET-28a::*nocF* (original gene) and expressed in BL21(DE3) was unable to express any soluble and insoluble NocF. (c) pET-22b::*nocF165* (in BL21(DE3) and Tuner(DE3)) was able to express insoluble NocF, particularly with the truncated version, indicating that the *pelB* signal sequence is essential for NocF expression which is improved by deletion of the transmembrane domain (d) pET-22b::*nocF* and pET-22b::*nocF165* expressed in BL21(DE3), C43(DE3) and Artic Express was unable to generate soluble protein. (e) pET-22b::*nocF* and pET-22b::*nocF165* expressed in BL21(DE3), C43(DE3) and Artic Express was unable to generate soluble protein for either 1 day or 2 days expression and higher or lower temperatures (25 °C or 12 °C). (f) pET-22b::*nocF* and pET-22b::*nocF165* and expressed in SoluBL21(DE3) was unable to generate soluble protein.

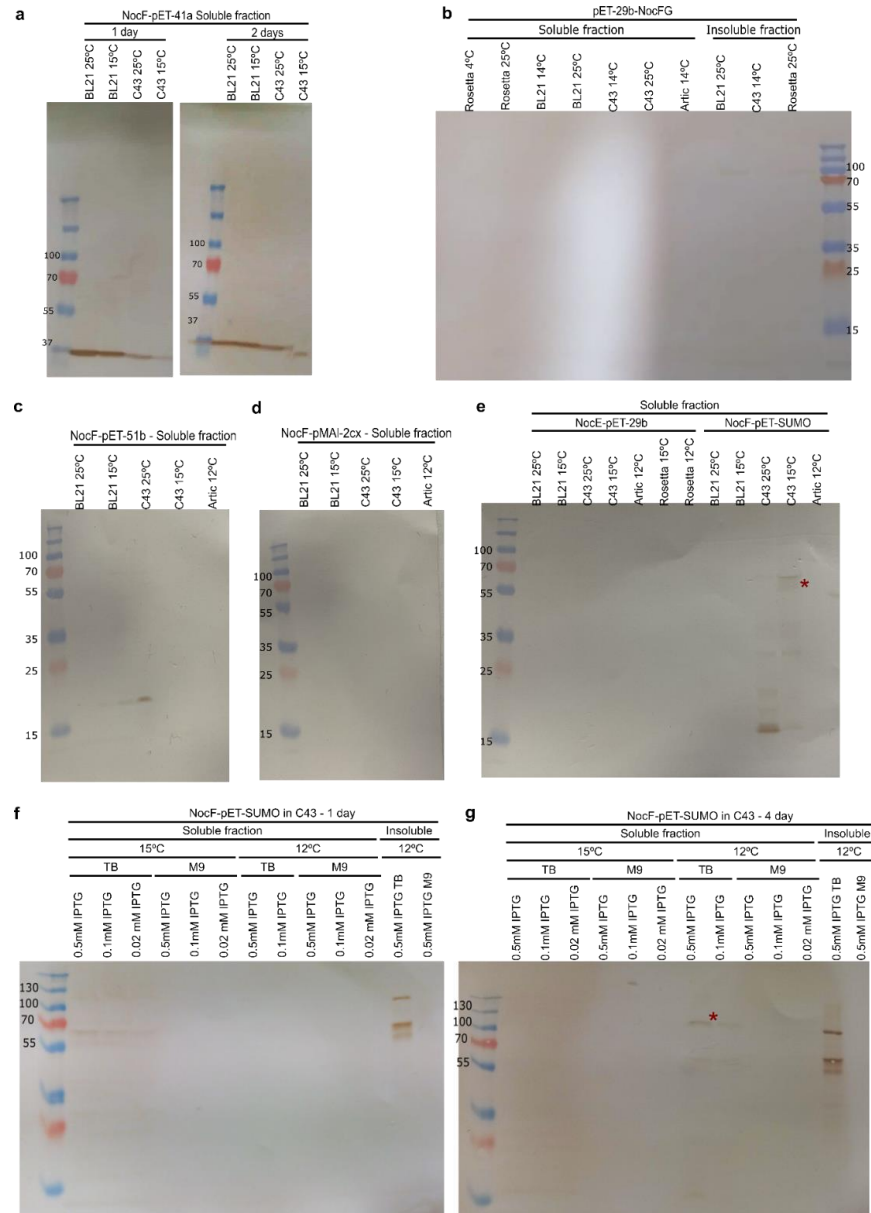


Figure S11.14 – Western blots of NocF (codon optimized) expression tests using pET-41a, pET-29b, pET-51b or pET-SUMO, terrific broth (TB), 0.5 mM IPTG at 25 °C and overnight incubation (unless indicated otherwise). (a) pET-41a::pelB-nocF and expressed in BL21(DE3) and C43(DE3) was unable to express any soluble NocF. (b) pET-29b::nocF (co-expression assay of the original genes) and expressed in BL21(DE3), C43(DE3), Rosetta(DE3) and Artic Express was unable to express any soluble NocF. (c) pET-51b::pelB-nocF in BL21(DE3), C43(DE3) and Artic Express was unable to express soluble NocF (d) pMAL-2cx harboring *nocF* in BL21(DE3), C43(DE3) and Artic Express was unable to express soluble NocF (e) pET-His₆-SUMO::pelB-nocF was able to provide a small amount of soluble protein in C43(DE3) grown at 15°C. This is the first condition that were able to provide any soluble NocF, indicating that the presence of the SUMO tag is essential for NocF solubility. (f) pET-His₆-SUMO::pelB-nocF in C43(DE3) was able to produce a very small amount of soluble protein in terrific broth media at 15°C after 1 day of expression. (g) pET-His₆-SUMO::pelB-nocF in C43(DE3) was able to produce soluble NocF when using terrific broth and 12°C for 4 days, indicating that lower temperatures and longer expression times are important to improve NocF solubility. The different IPTG concentration did not have a significant effect on NocF solubility, although 0.5mM IPTG provided slightly more protein.

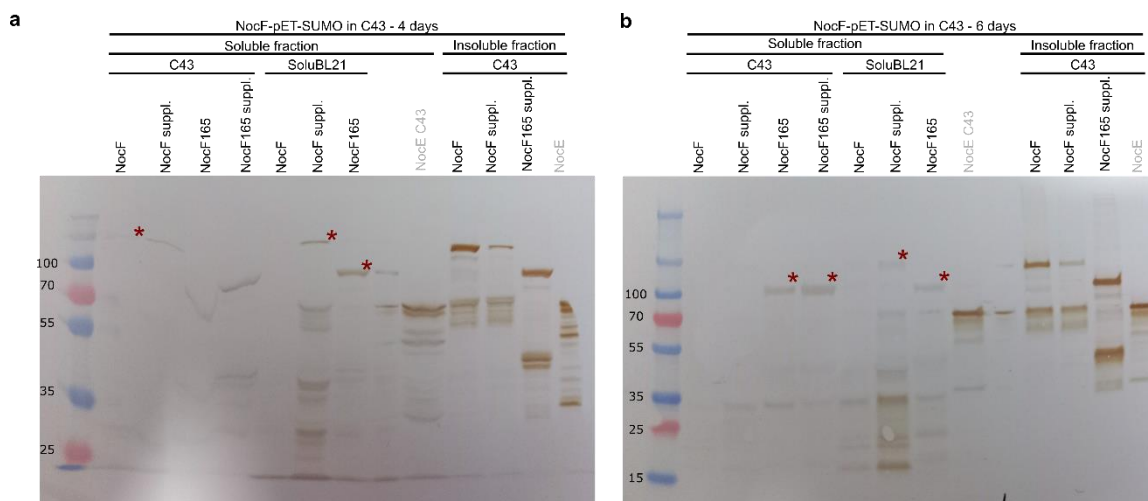


Figure SII.15 – Western blots of NocF (codon optimized) expression tests using pET-His6-SUMO-TEV, terrific broth (TB), 0.5 mM IPTG at 12 °C and several days of incubation (unless indicated otherwise). pET-His6-SUMO::*pelB-nocF* and pET-His6-SUMO::*pelB-nocF165* in C43(DE3) and SoluBL21(DE3) was able to produce significant amounts of soluble protein under several conditions after 4 days of expression (a) and 6 days (b) with or without supplementation with the predicted co-factor and substrate, PLP and 6-tricadenone, respectively. Despite the unspecific bands that showed up in the western blot likely caused using an old antibody, we can conclude that:

1. N-His₆-SUMO-*pelB*-NocF in C43(DE3) is soluble only after 4 days (not 6 days) and the supplementation improved the protein solubility;
2. N-His₆-SUMO-*pelB*-NocF165 in C43(DE3) is the most soluble version of NocF (especially after 6 days) and the supplementation is not influencing the solubility significantly;
3. N-His₆-SUMO-*pelB*-NocF in SoluBL21(DE3) is soluble especially after 6 days with supplementation;
4. N-His₆-SUMO-*pelB*-NocF165 in SoluBL21(DE3) is soluble with no supplementation;
5. N-His₆-SUMO-NocE in C43(DE3) is soluble, especially after 6 days.

In general, these results indicate the temperature plays a major effect on NocF solubility. Lower temperatures, which naturally require longer expression times, the deletion of the transmembrane N-terminal of NocF significantly increased the proteins solubility. Supplementation with the predicted co-factor and substrate also provided a positive effect but depending on the construct in use. To notice that these western blots and previous ones showed a second band with lower molecular weight whose origin it is not clear but might indicate NocF proteolysis or the expression of a truncated NocF-SUMO version.

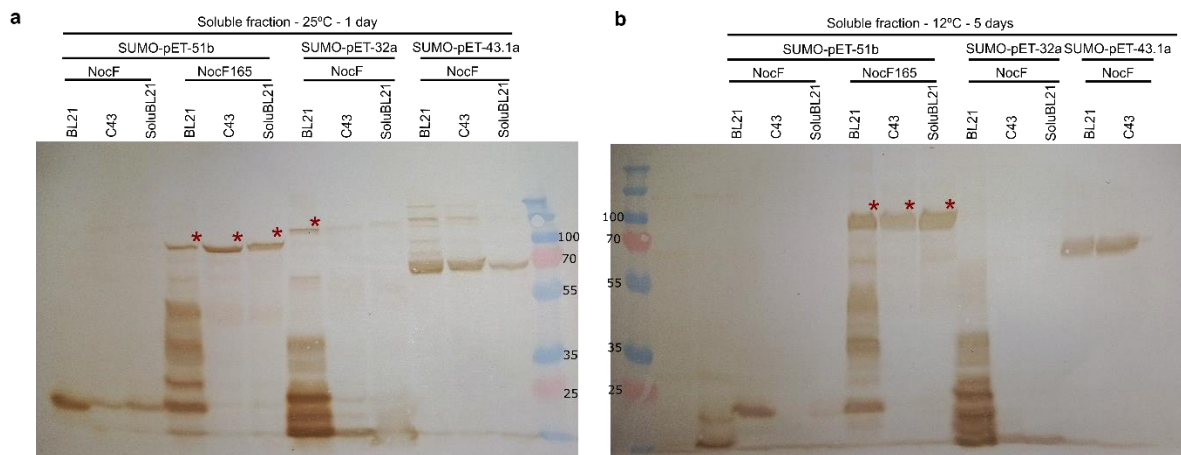


Figure SII.16 - Western blots of NocF expression tests using pET-32a, pET-43.1a, and N-His₆-SUMO-pelB-NocF using pET-51b in terrific broth (TB), 0.5 mM IPTG at 25 °C for 1 day or 12 °C or 5 days at 12 °C (unless indicated otherwise) with no co-factor/substrate supplementation. pET-51b harboring N-His₆-SUMO-pelB-NocF165 in BL21(DE3), C43(DE3) or SoluBL21(DE3) was able to produce high amounts of soluble protein after 1 day of expression (a) and 5 days (b). pET-43.1a::nocF unable to produce soluble NocF in any of the tested conditions and pET-32a::nocF was only able to express soluble NocF in BL21(DE3) at 25 °C for 1 day. The tripled tagged version of the protein - N-Strep-SUMO-pelB-NocF165-His₆-provided the best expression condition found for the protein when expressed at 12 °C for 5 days in either BL21(DE3), C43(DE3) and SoluBL21(DE3). These results indicate once again the positive effects of the transmembrane domain deletion, the use of the SUMO tag, the use of longer expression periods at lower temperatures and also the co-existence of the Strep and SUMO tags on NocF solubility.

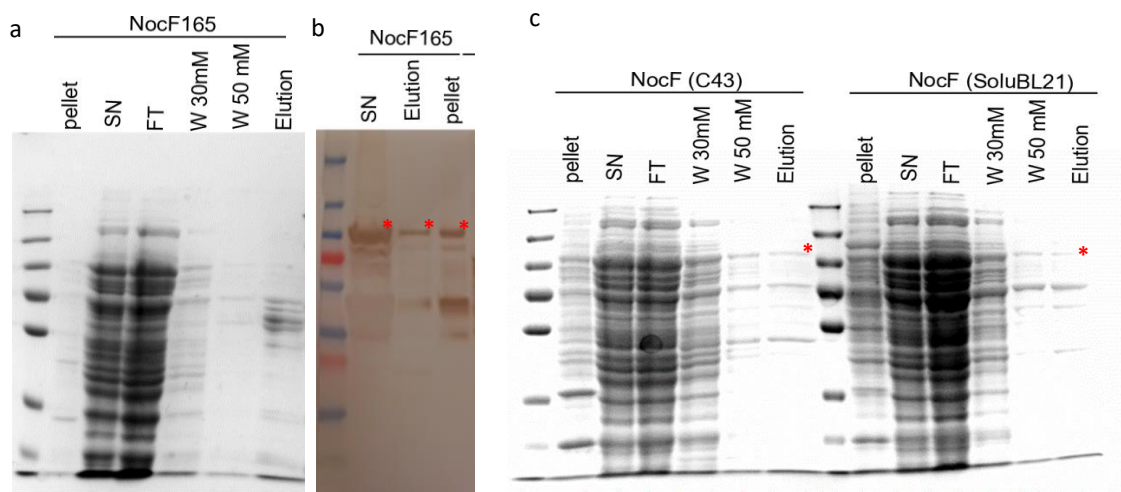


Figure SII.17 - NocF purification attempts. (a) Coomassie blue gel of N-His₆-SUMO-pelB-NocF165 purification attempt from C43(DE3) cells using nickel column. (b) Western blot of N-His₆-SUMO-pelB-NocF165 supernatant, elution and pellet fractions. (c) Coomassie blue gel of N-His₆-SUMO-pelB-NocF purification attempt from C43(DE3) cells (cobalt resin) and N-His₆-SUMO-pelB-NocF from SoluBL21(DE3) cells (nickel column). N-His₆-SUMO-pelB-NocF165 purification retrieved the protein, as confirmed by the western blot, but with several contaminant proteins. Additionally, the western blot showed that the supernatant fraction contained a large amount of soluble NocF which is not proportional to the amount of protein recovered by the elution, suggesting interaction issues between the column and the histidine tag. Regarding N-His₆-SUMO-NocF purification, we observed similar problems, recovering contaminated elution fractions.

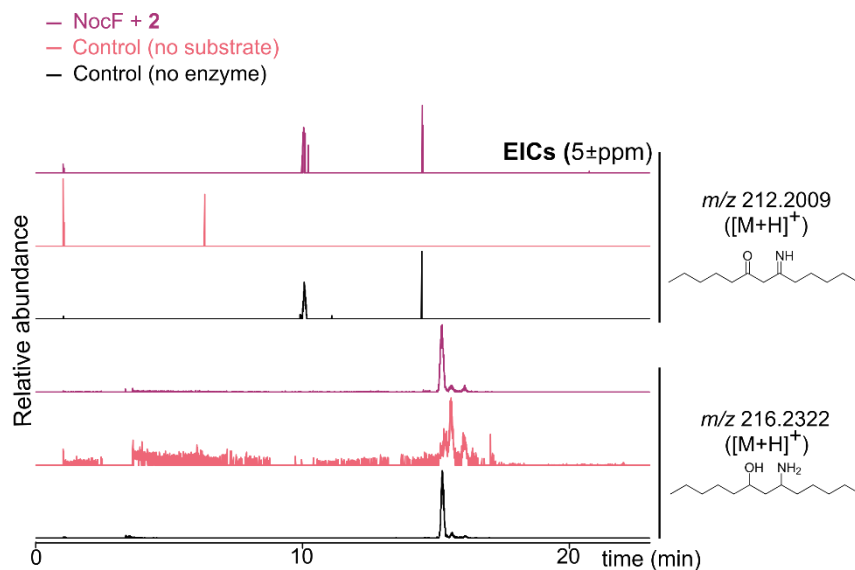


Figure SII.18 - LC-HRMS-derived EICs of the putative amine and imine products generated by NocF (except for 8-iminotrican-6-ol – Fig 3 of the main manuscript) in enzymatic assays with **2**. NocF was only able to generate **4**.

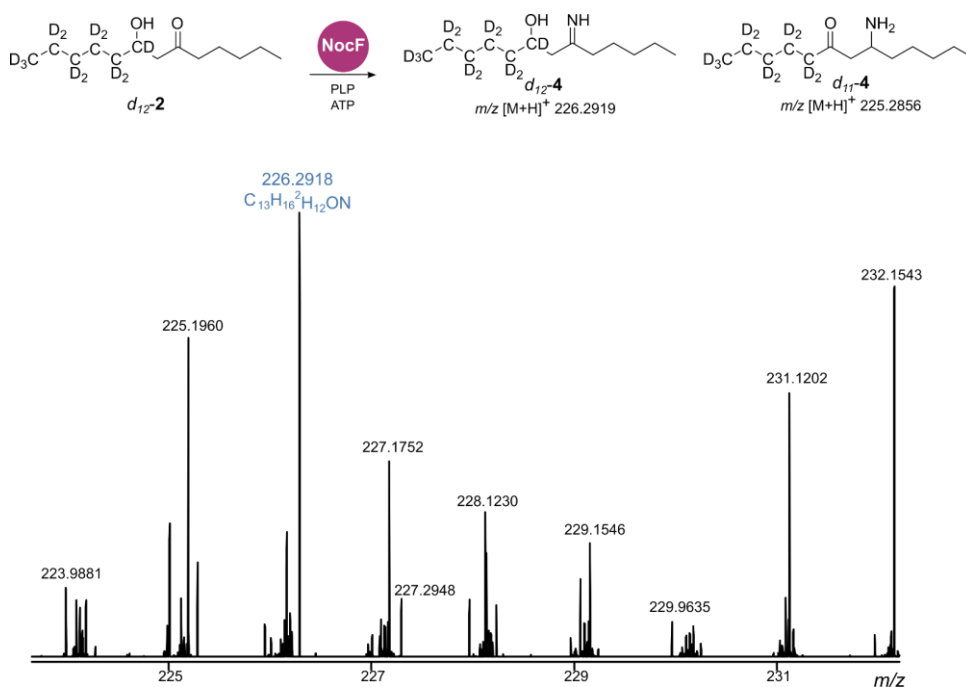


Figure SII.19 – Spectrum of m/z feature 226.2919 (d_{12} -**4**) detected upon NocF assay with d_{12} -**2**. Considering that C-4 position (directly connected to the oxygen) in **1** keeps 1 deuterium when fed with d_{15} -C8 and that no deuterons are lost when d_{12} -**2** is incubated with NocF, originating d_{12} -**4** and not d_{11} -**4**, this indicates that the ketone of **2** is converted into an imine, forming 8-iminotrican-6-ol, and not the alcohol is converted into an amine, which would form 8-aminotridecan-6-one.

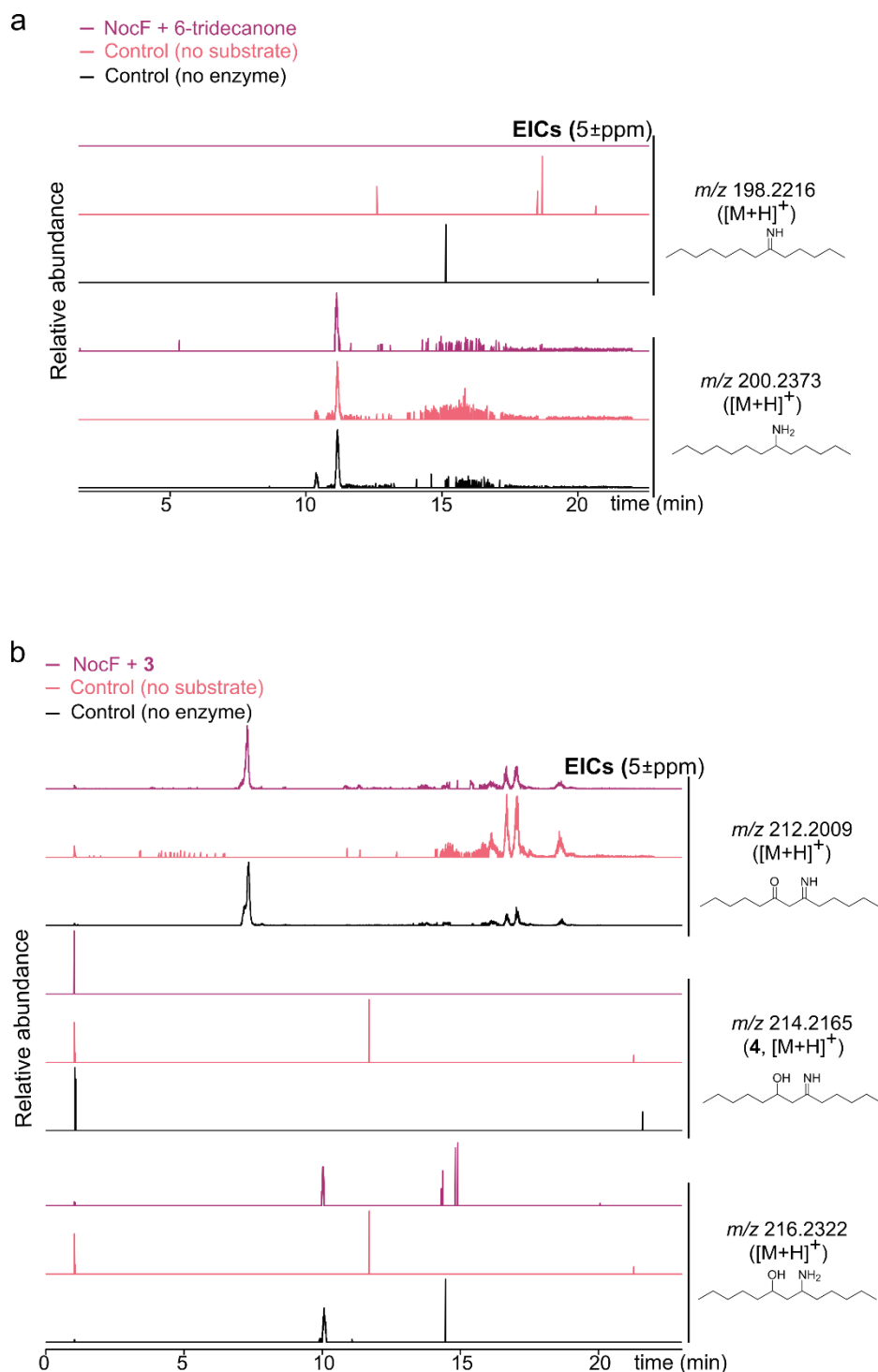


Figure SII.20 - LC-HRMS-derived EICs of the putative amine and imine product generated by NocF in enzymatic assays with 6-tridecanone (a) and **3** (b). NocF was not able to produce any of the putative products indicating that **2** is the true NocF substrate.

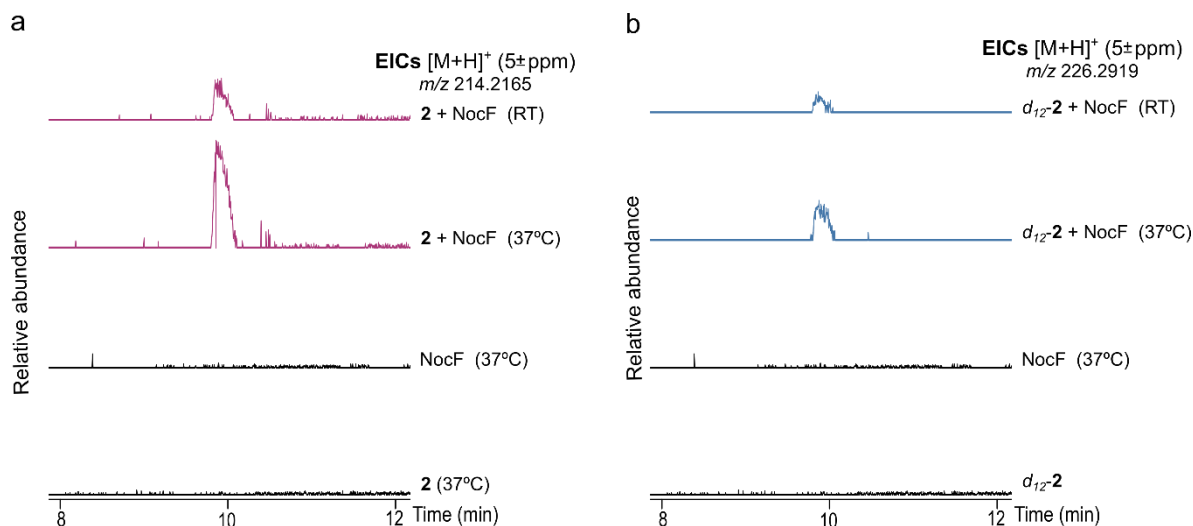


Figure SII.21 - LC-HRMS-derived EICs of **4** (a) and *d*₁₂-**4** (b) generated by enzymatic assays of N-His₆-SUMO-peIB-NocF at room temperature and 37°C overnight, with **2** and **3**, respectively.

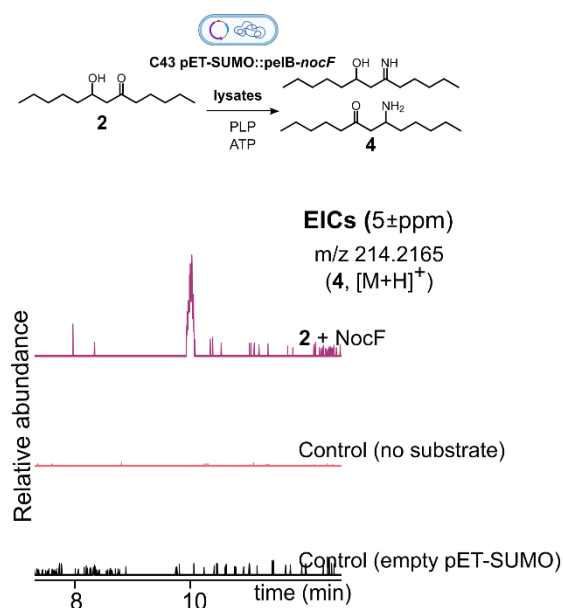


Figure SII.22 - LC-HRMS-derived EICs of **4** generated by *in vivo* assays of C43 harboring pET-SUMO::peIB-NocF supplemented with **2**.

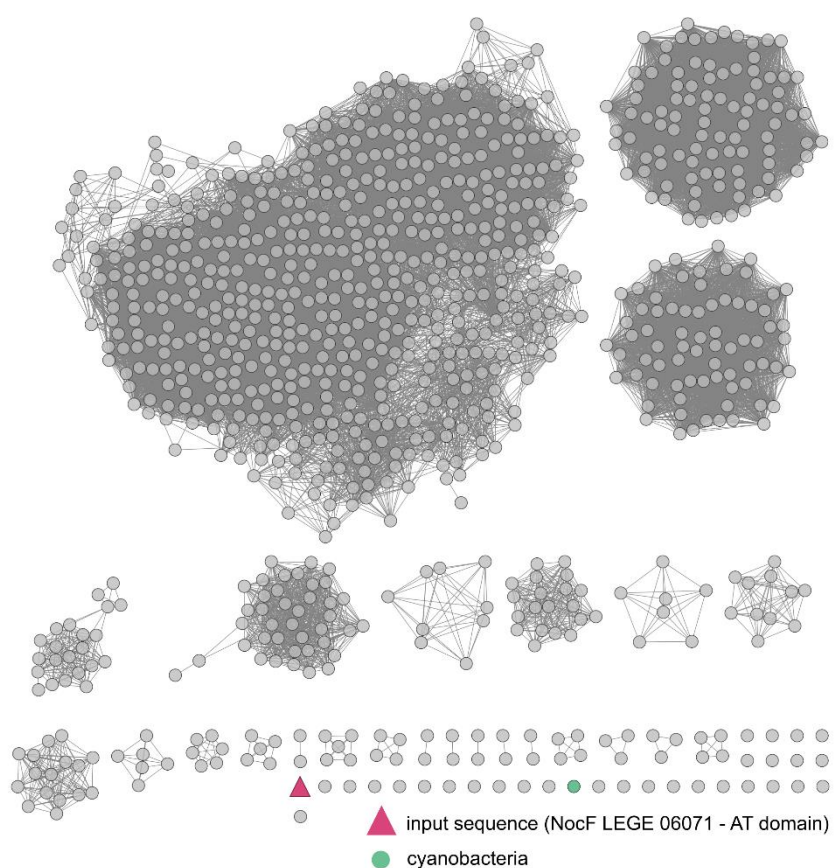


Figure SII.23 - Sequence similarity network (SSN) of NocF and the 1000 closest homologues indicating a separation of NocF and other homologues from the remaining sequences (%ID 42).

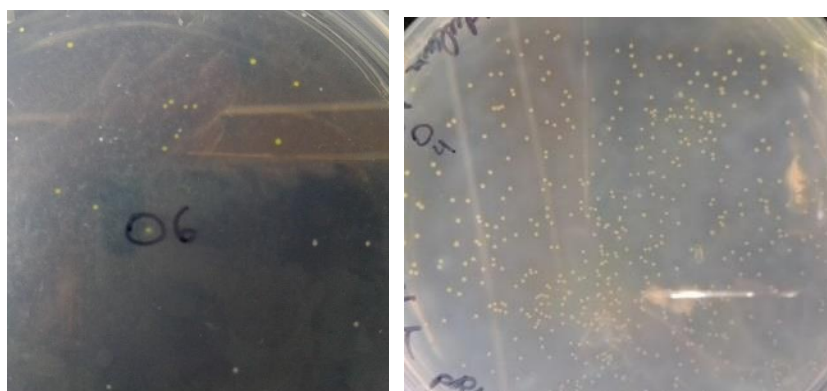


Figure SII.24 – BG11 neomycin plates after 1 month of conjugation with pRL278::*nocE* showed no cyanobacterial colonies. The plates showed only *Microbacterium* sp. and *Sphingomonas* sp. colonies (confirmed by 16S sequencing), which likely represent other bacterial strains that were present in *Nodularia* sp. LEGE06071 non-axenic culture.

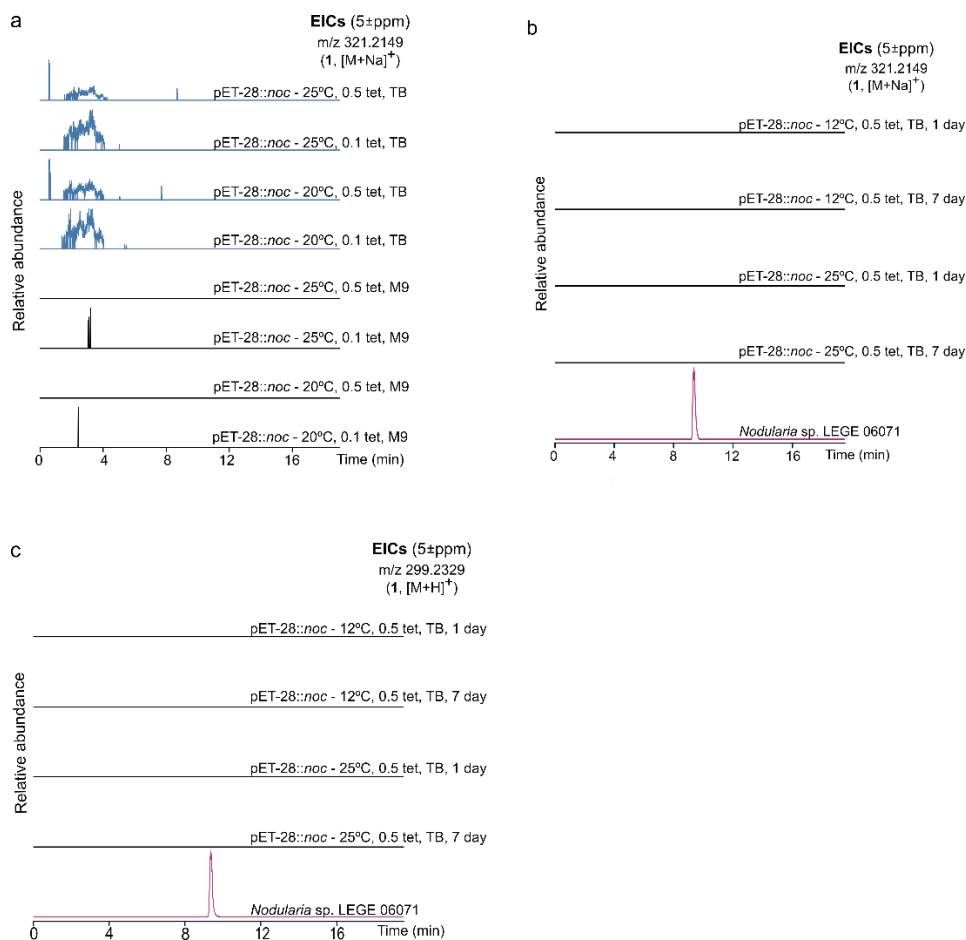


Figure SII.25 - LC-HRMS-derived EICs of **1** in several growth conditions of BAP1 pET-28::noc transformants. (a) LC-HRMS-derived EICs of **1** ([M+H]⁺) in BAP1 pET-28::noc transformants grown at 25 °C or 20 °C in terrific broth or M9 minimal media, induced with 0.5 µg mL⁻¹ of tetracycline for 4 days. (b) LC-HRMS-derived EICs of **1** ([M+Na]⁺) in BAP1 pET-28::noc transformants grown at 25 °C or 12 °C in terrific broth or M9 minimal media, induced with 0.5 µg mL⁻¹ of tetracycline for 1 and 7 days. (b) LC-HRMS-derived EICs of **1** ([M+H]⁺) in BAP1 pET-28::noc transformants grown at 25 °C or 12 °C in terrific broth or M9 minimal media, induced with 0.5 µg mL⁻¹ of tetracycline for 1 and 7 days.

supplementation experiment	m/z $[M+H]^+$	Molecular formula	Unlabeled m/z $[M+H]^+$	Molecular formula	labeling
d_{11} -hexanoic acid	225.2356	$C_{13}H_{17}D_{11}ON^+$	214.2165	$C_{13}H_{28}ON$	d_{11}
	236.3546	$C_{13}H_6D_{22}ON^+$			d_{22}
d_8 -L-methionine	N/A	N/A			N/A
d_{11} -hexanoic acid	283.3274	$C_{16}H_{23}D_{11}O_2N^+$	272.2584	$C_{16}H_{34}O_2N$	d_{11}
	294.3965	$C_{16}H_{12}D_{22}O_2N^+$			d_{22}
d_8 -L-methionine	276.2835	$C_{16}H_{30}D_4O_2N^+$			d_4

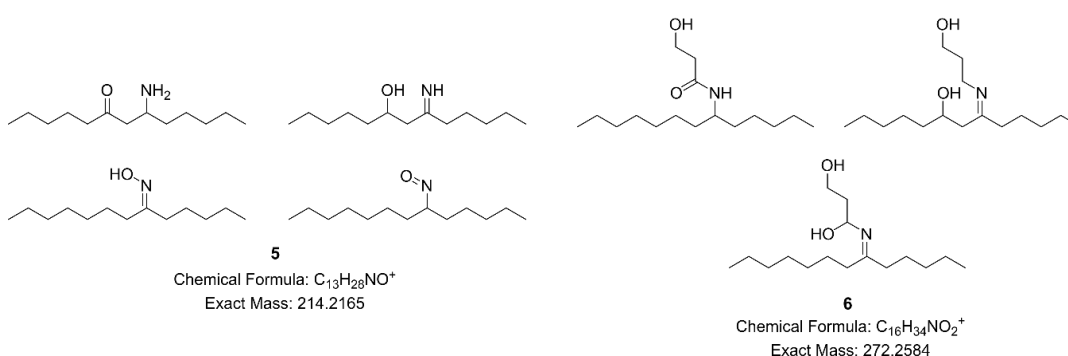


Figure SII.26 – Putative intermediates of the *noc* biosynthetic pathway identified in BAP1 *E. coli* harboring pET28::*noc*. (a) Putative *noc* intermediates, m/z values, predicted molecular formulas and labelling upon d_{11} -hexanoic acid and d_8 -L-methionine supplementation. (b) Hypothetical structures of compounds **5** (m/z 214.2165) and **6** (m/z 272.2582).

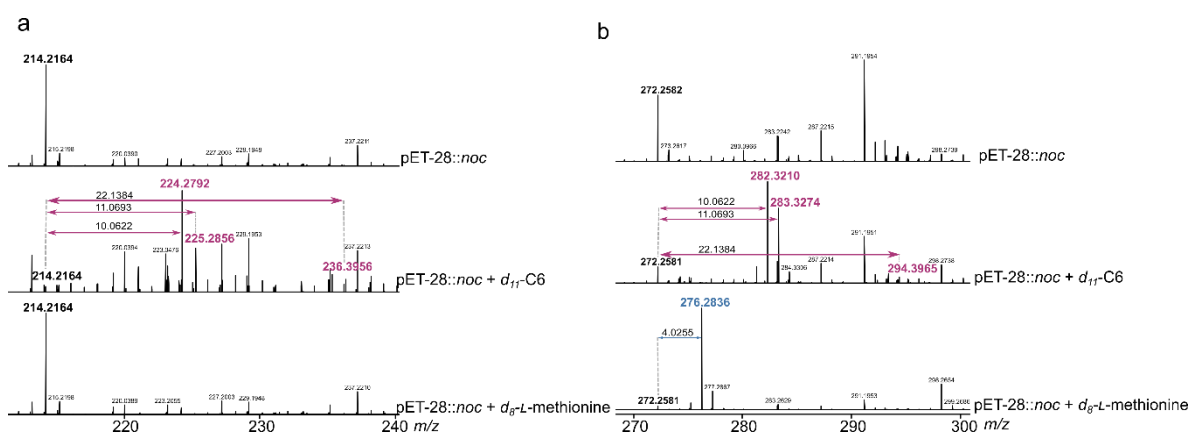


Figure SII.27 - HR-MS/MS spectrum of **5** (a) and **6** (b) under upon d_{11} -hexanoic acid and d_8 -L-methionine supplementation.

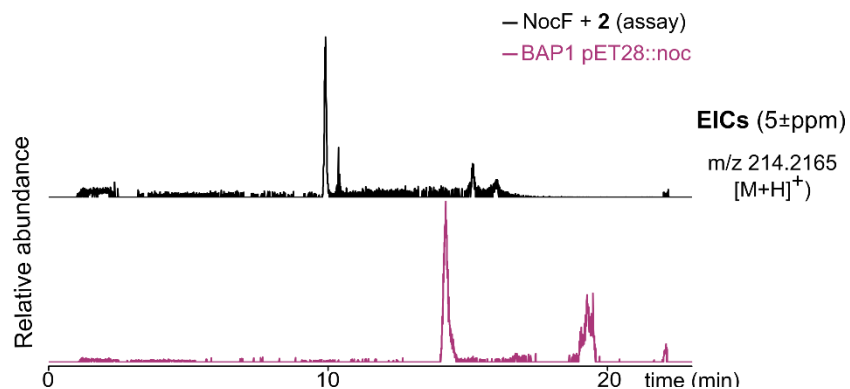


Figure SII.28 - LC-HRMS-derived EICs of m/z 214.2165 (**4** and **5**) in enzymatic assays of NocF with **2** and in the crude extracts of BAP1 pET-28::*noc*. The different retention time of the mass feature indicates that **4** and **5** are structural isomers and do not represent the same metabolite.

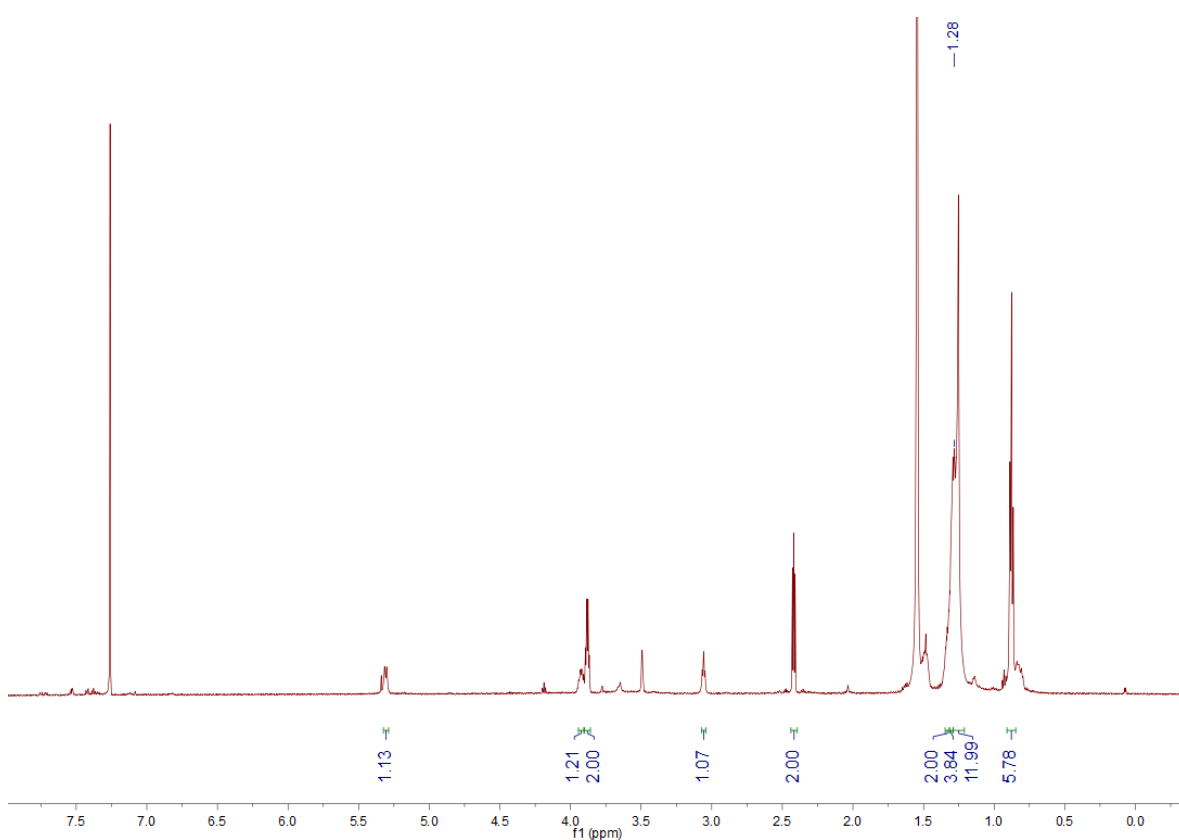


Figure SII.29 - ^1H NMR spectrum (600 MHz) of **6** in CD_3Cl .

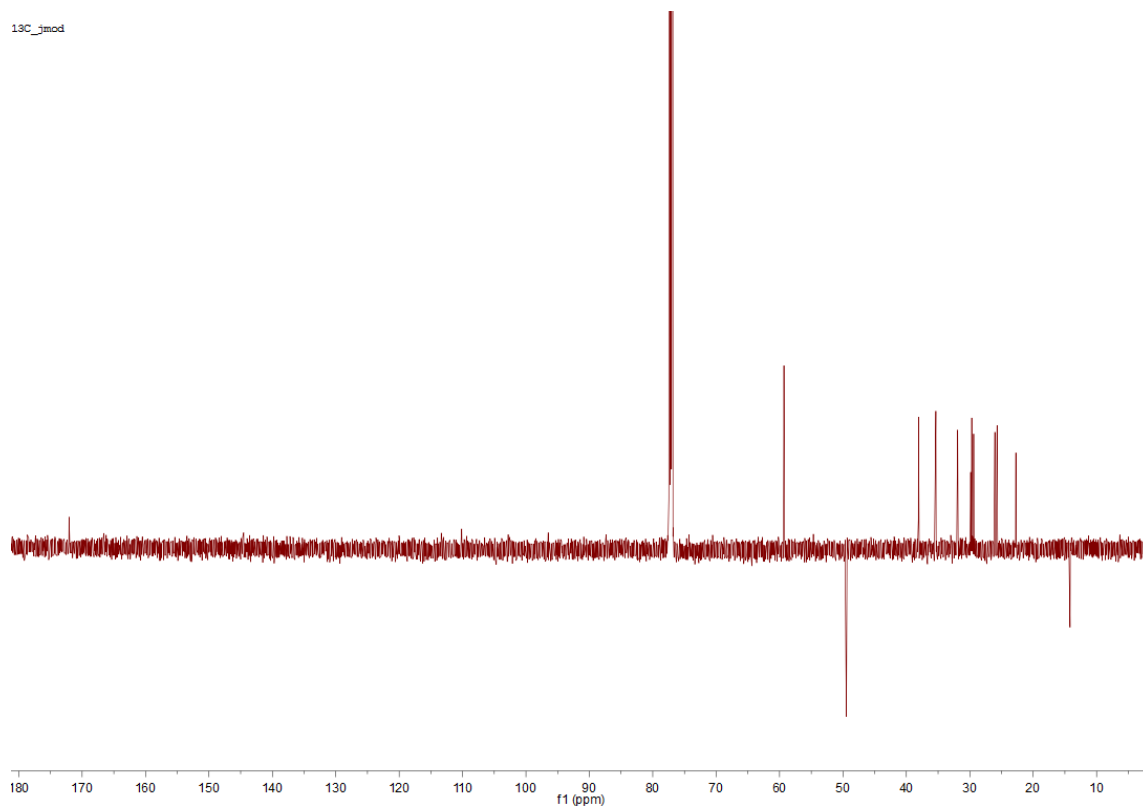


Figure SII.30 – ¹³C APT NMR spectrum (600 MHz) of **6** in CD₃Cl.

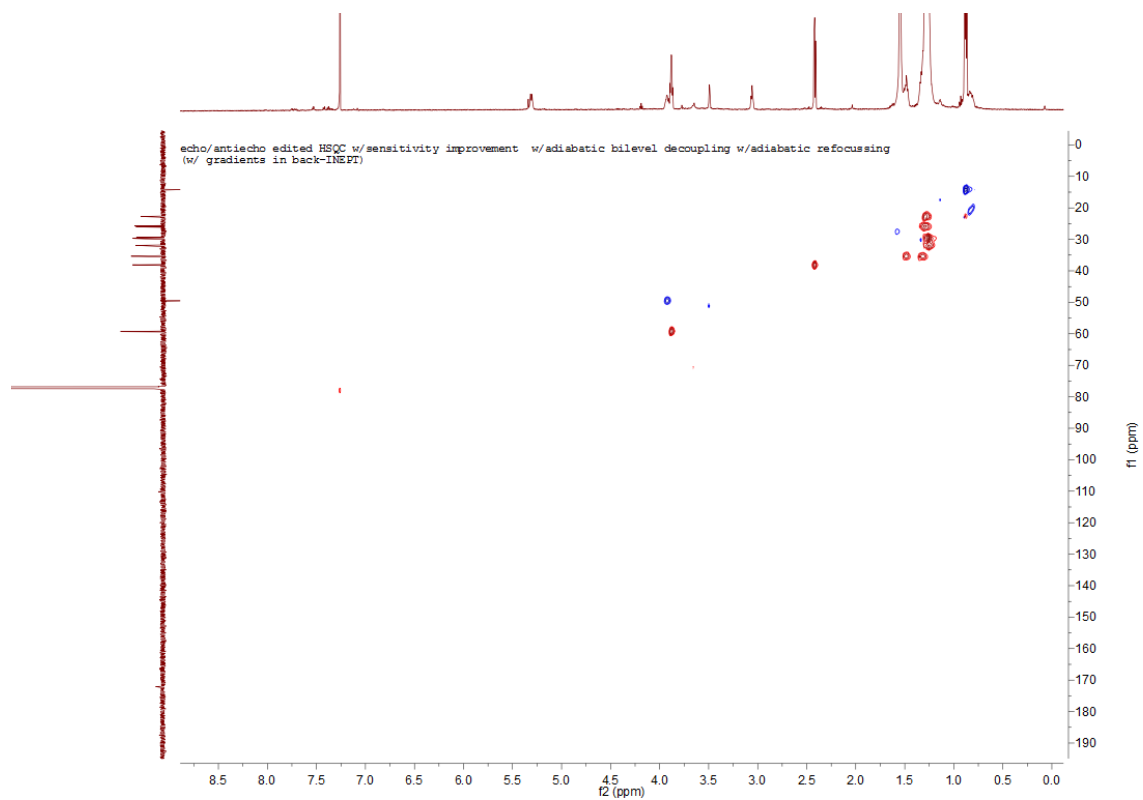


Figure SII.31 - HSQC spectrum (600 MHz) of **6** in CD₃Cl.

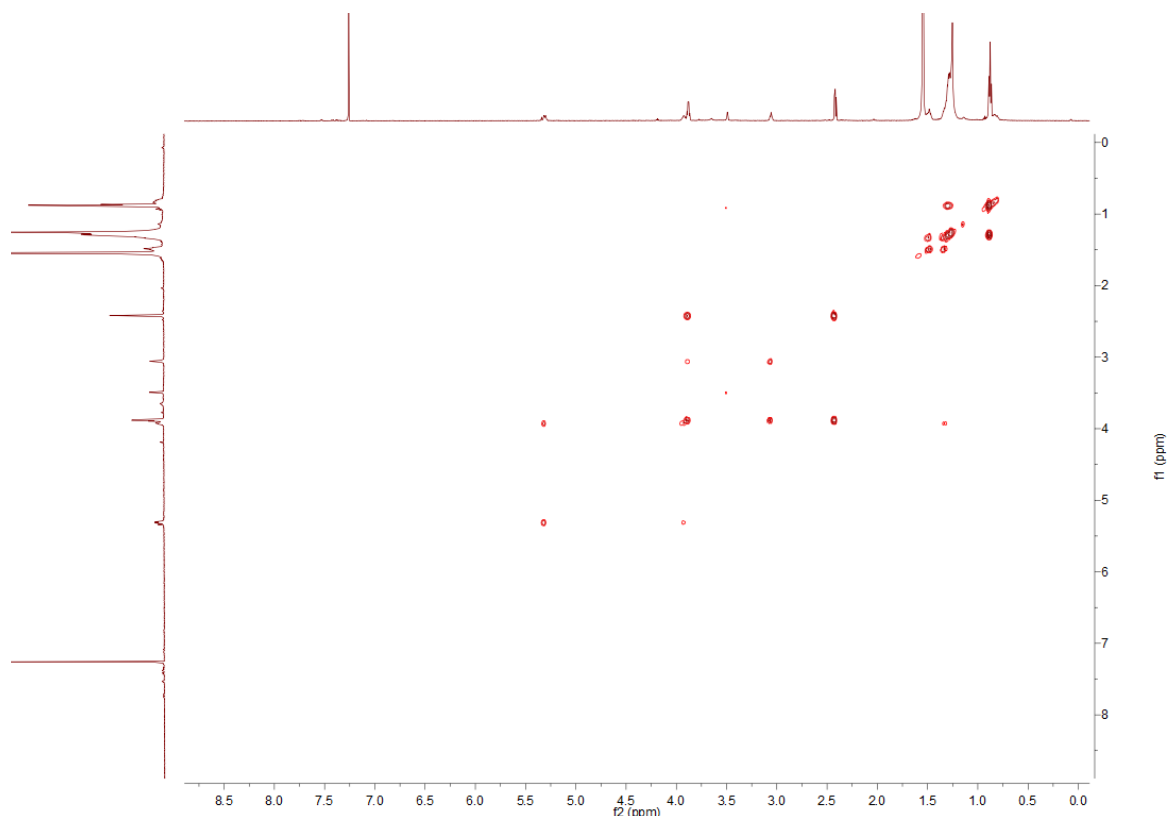


Figure SII.32 - COSY NMR spectrum (600 MHz) of **6** in CD₃Cl.

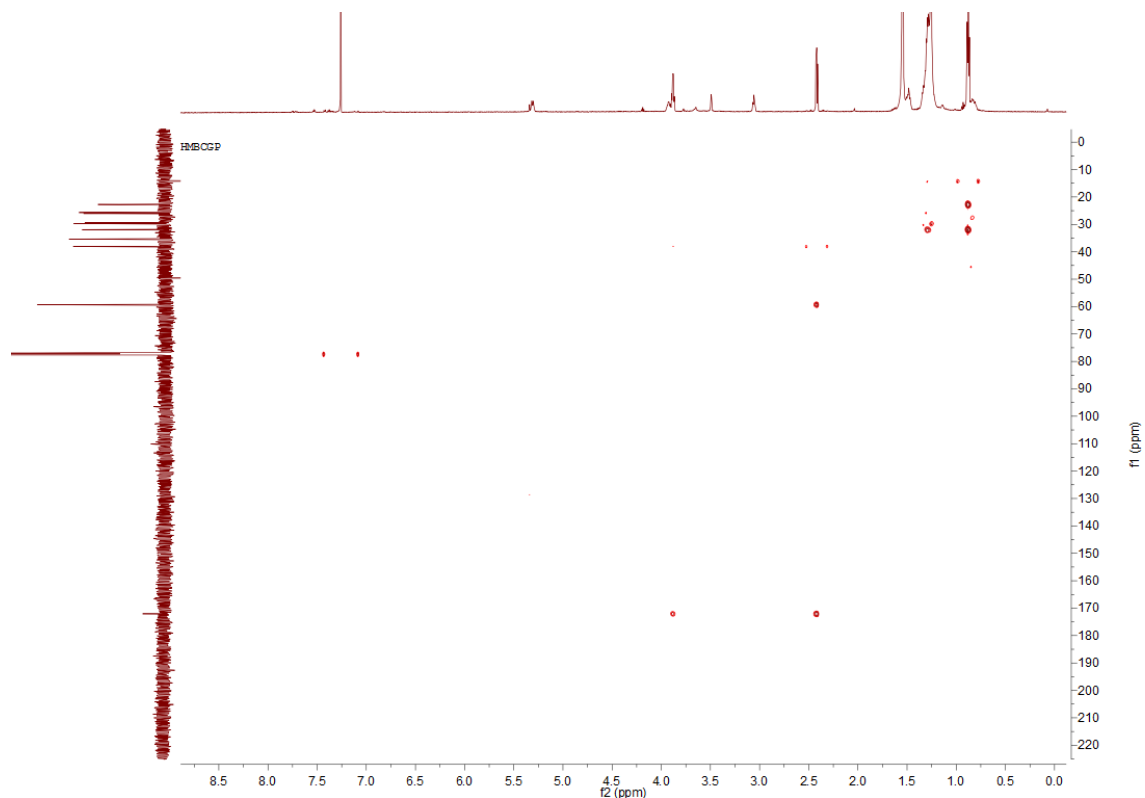


Figure SII.33 - HSQC spectrum (600 MHz) of **6** in CD₃Cl.

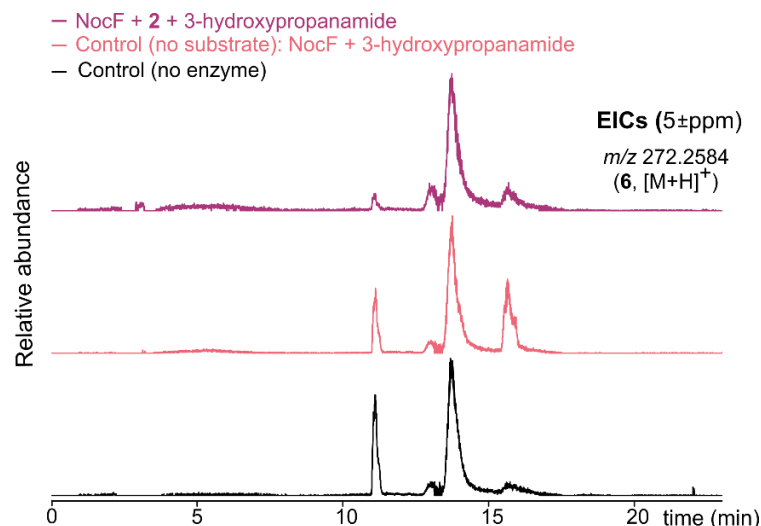


Figure SII.34 - LC-HRMS-derived EICs of **6** in NocF enzymatic assays incubated with **2** and 3-hydroxypropanamide as putative substrate. NocF was not able to generate **6** confirming its role as the ketone-imine enzyme responsible for the formation of **4**.

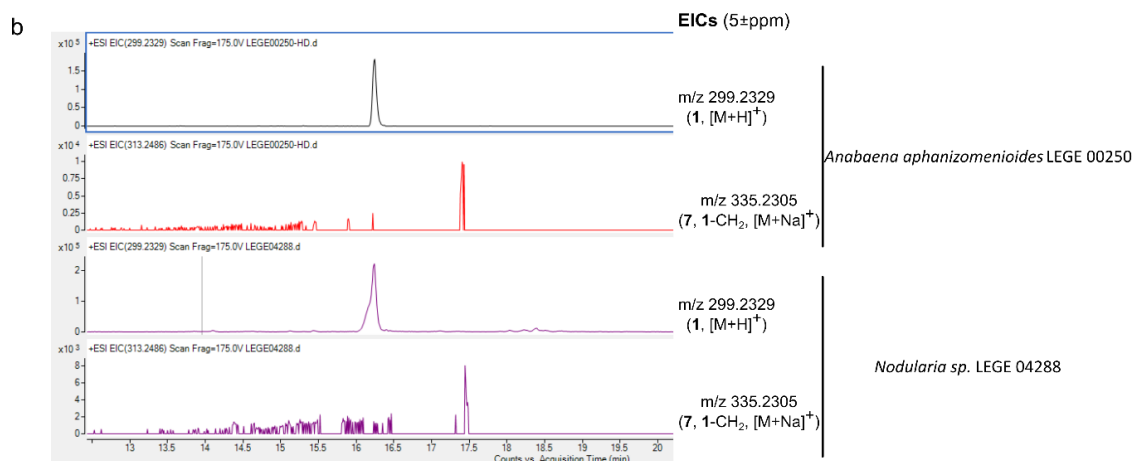
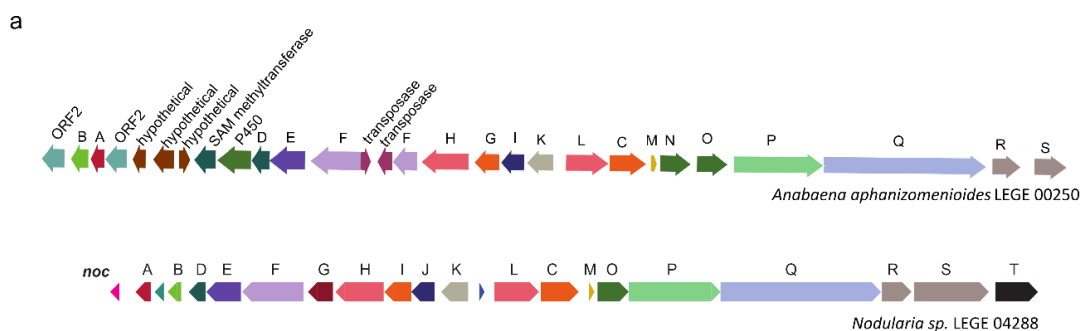


Figure SII.35 – Nocuelin A and B are produced by *Anabaena aphanizomenioides* LEGE 00250 and *Nodularia* sp. LEGE 04288. (a) *noc* BGC in *Anabaena aphanizomenioides* LEGE 00250 and *Nodularia* sp. LEGE 04288. LEGE 00250 *noc* BGC encodes several additional proteins that are not part of the the remaining homologues BGCs. This includes a SAM methyltransferase, a cytochrome P450 and several transporters and transposases. (b) EIC of **1** and **7** in the crude extracts of LEGE 00250 and LEGE 04288, both harboring the *nocD* gene in their BGCs.

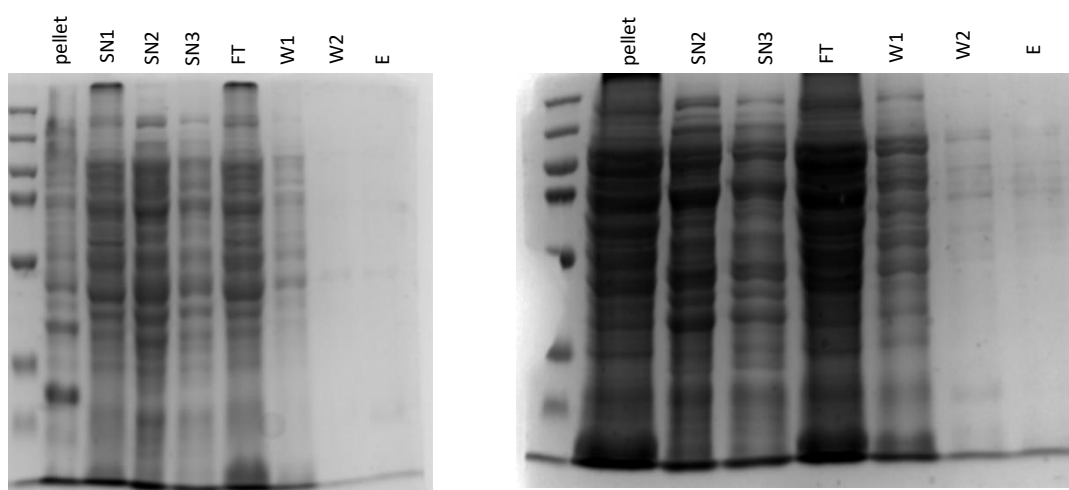


Figure SII.36 – Coomassie blue gels of the purification attempts of N-His₆-NocD(PCC) (a) and C-His₆-NocD(LEGE) (b) from BL21(DE3) cells (SN1 – first supernatant after cell disruption, SN2 – supernatant after the first ultracentrifugation, SN3 – supernatant after the second ultracentrifugation, FT – flow through, W1 – washing step 1 (30 mM imidazole), W2 – washing step 2 (50 mM imidazole), E – elution (200 mM imidazole)). Red star highlights the putative band corresponding to NocD.

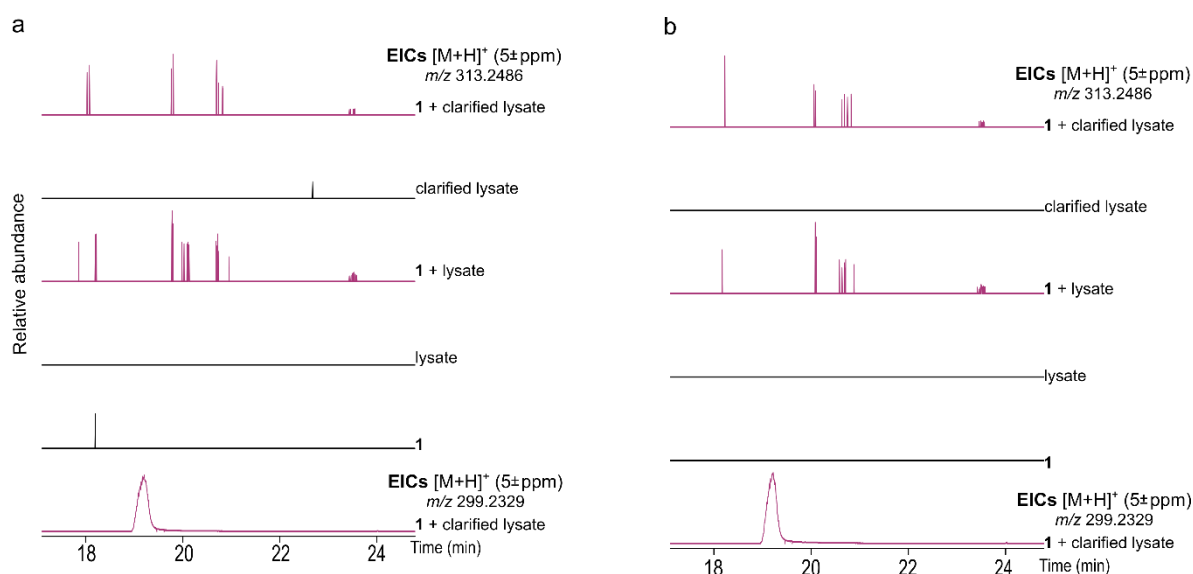


Figure SII.37 - LC-HRMS-derived EICs of **7** in *in vitro* assays of **1** with NocD lysates and clarified lysates. (a) LC-HRMS-derived EICs of **7** in *in vitro* assays with BL21(DE3) pET-29b::*nocD*(LEGE) lysate and clarified lysate incubated with **1**. (b) LC-HRMS-derived EICs of **7** in *in vitro* assays with BL21(DE3) pET-28a::*nocD*(PCC) lysate and clarified lysate incubated with **1**. Compound **7** was not detected in the assays.

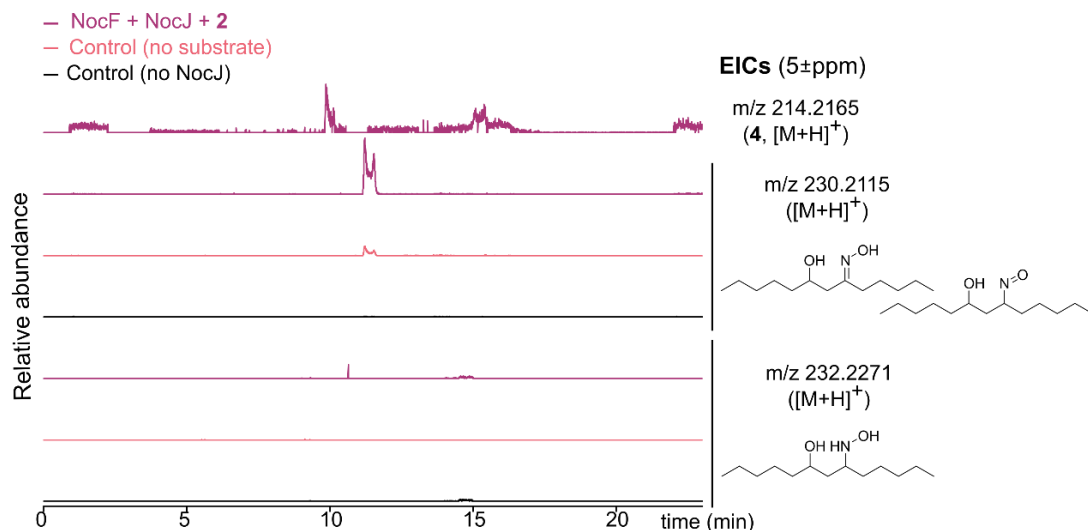


Figure SII.38 - LC-HRMS-derived EICs of putative products of NocJ oxygenation of **4**. A mass feature of m/z 230.2115 was identified in the coupled NocF/NocJ assay with **4** but the same feature is present in the control assay with no substrate, invalidating the origin of this feature as result of NocJ activity.

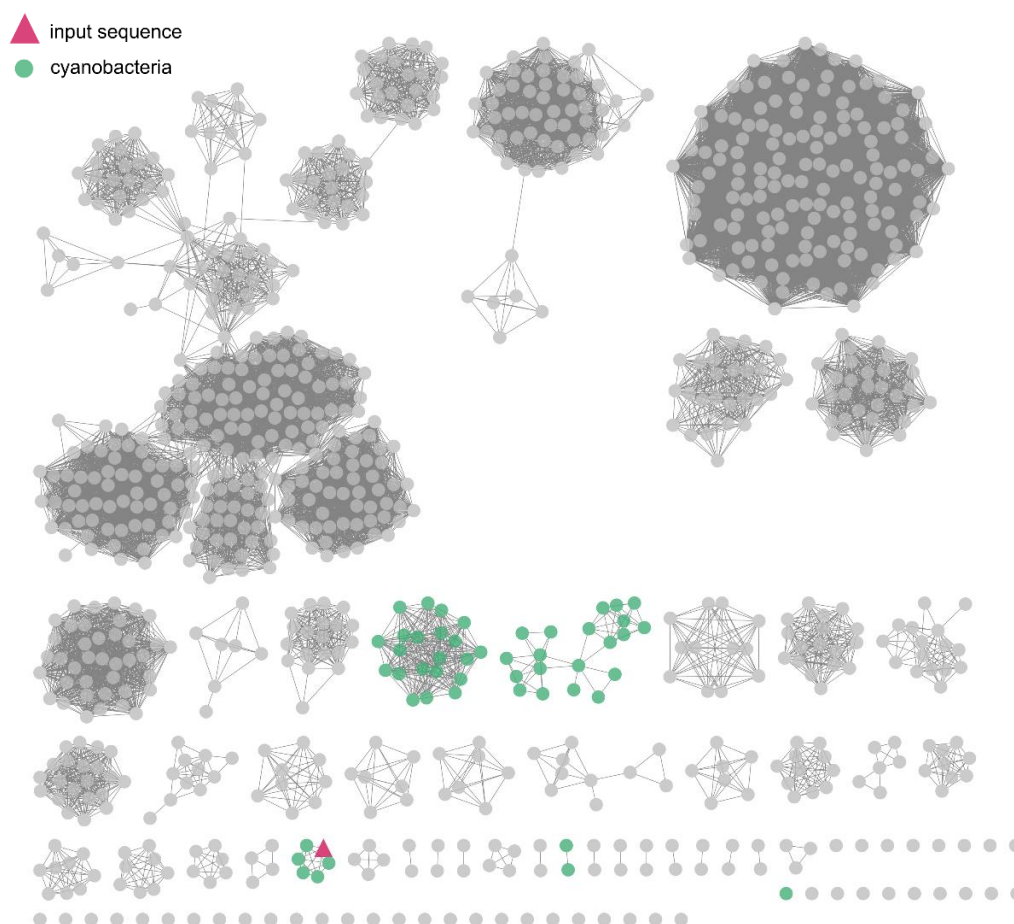


Figure SII.39 - Sequence similarity network (SSN) of NocJ and the 1000 closest homologues indicating a separation of NocJ and other homologues from the remaining sequences (%ID 49). NocJ from *Nodularia* sp. LEGE 06071 clades exclusively with other cyanobacterial strains.

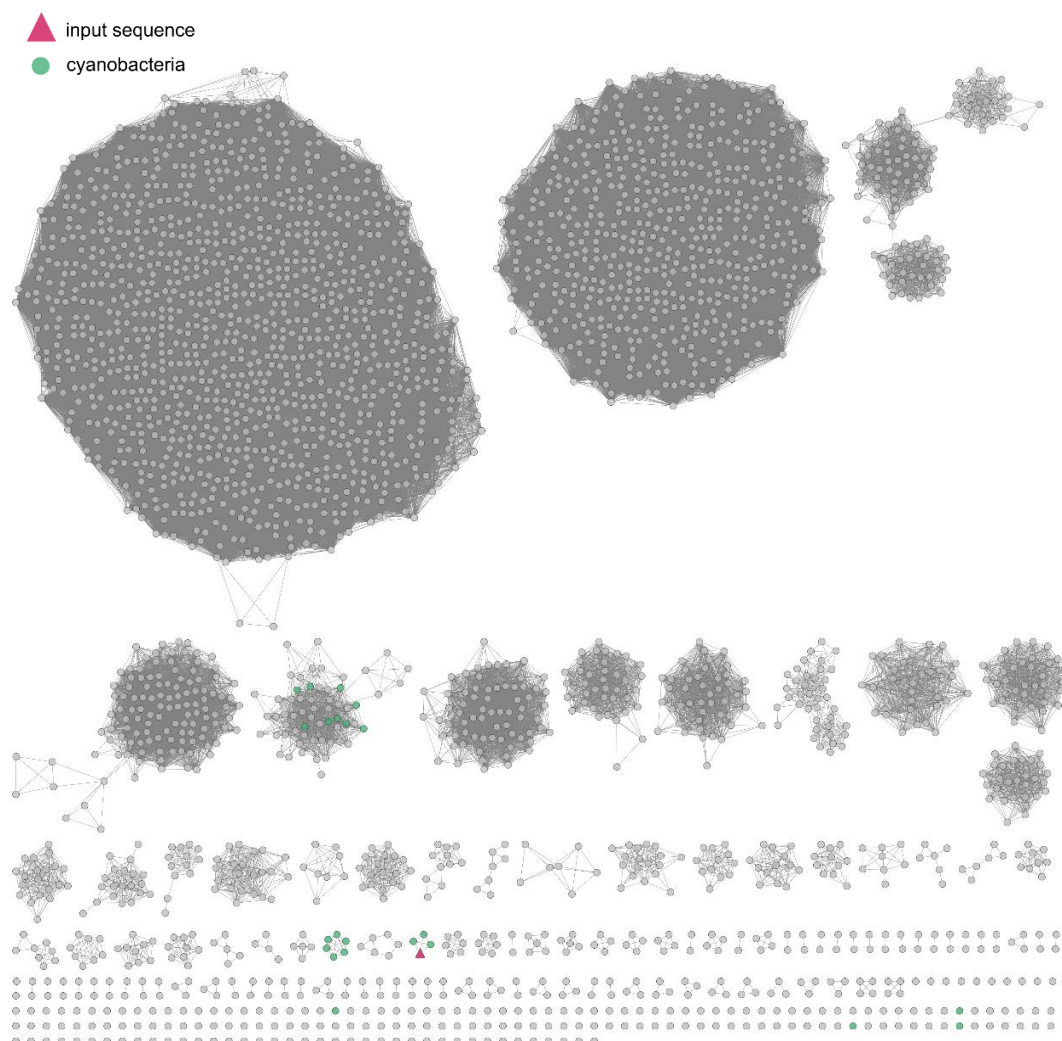


Figure SII.40 - Sequence similarity network (SSN) of NockK, the 1000 closest homologues and the IPR021829 InterPro family, indicating a separation of NockK and other homologues from the remaining sequences (%ID 55). NockK from *Nodularia* sp. LEGE 06071 clades exclusively with other cyanobacterial strains.

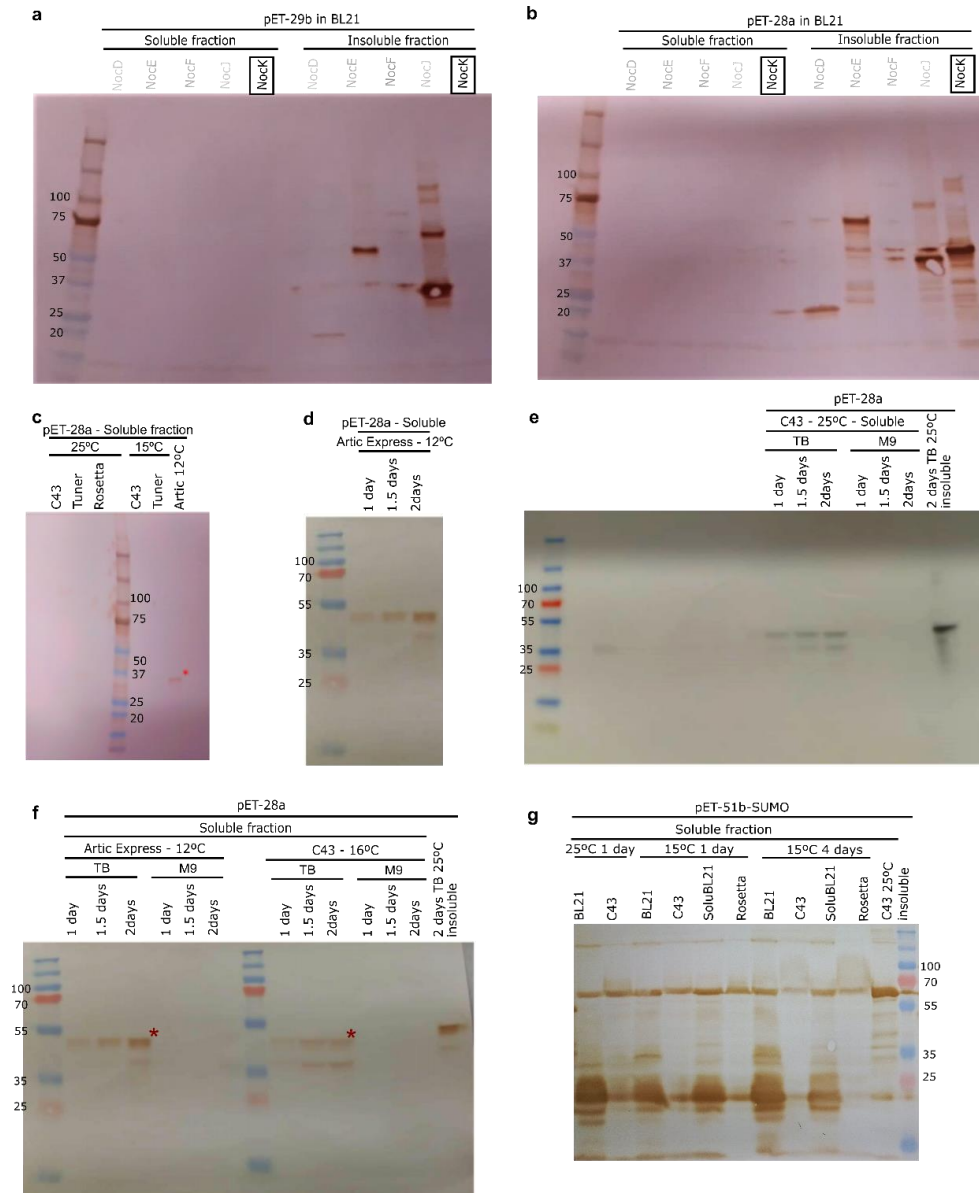


Figure SII.41 - Western blots of NockK expression tests using pET-28a and pET-51b-SUMO, terrific broth (TB), 0.5 mM IPTG, 25 °C and overnight incubation (unless indicated otherwise). (a) pET-29b::*nock* and expressed in BL21(DE3) was unable to express any soluble or insoluble NockK. (b) pET-28a::*nock* and expressed in BL21(DE3) was unable to express any soluble NockK but produced insoluble protein, indicating that the presence of the histidine tag on the N-terminal of the protein is important for its expression. (c) pET-28a::*nock* and expressed in Artic Express at 12 °C was able to produce a small amount of soluble NockK. (d) pET-28a::*nock* and expressed in Artic Express at 12 °C was able to produce significant amounts of soluble NockK after 1, 1.5 and 2 days of expression. (e) pET-28a::*nock* and expressed in Artic Express at 12 °C was able to produce soluble NockK after 1, 1.5 and 2 days of expression in Terrific broth but not in M9 minimal medium. (f) pET-28a::*nock* and expressed in Artic Express at 12 °C and C43(DE3) at 16 °C was able to produce significant amounts of soluble NockK after 1, 1.5 and 2 days of expression in terrific broth but not in M9, indicating the culture medium choice is crucial for NockK expression and solubility. (g) pET-51b-SUMO::*nock* and expressed in BL21(DE3), C43(DE3). SoluBL21(DE3) and Rosetta(DE3) was able to express significant amounts of soluble NockK in all tested conditions. This indicates, together with the NocF and NocJ expression tests highlights the potential of pET-51b-SUMO for protein expression.

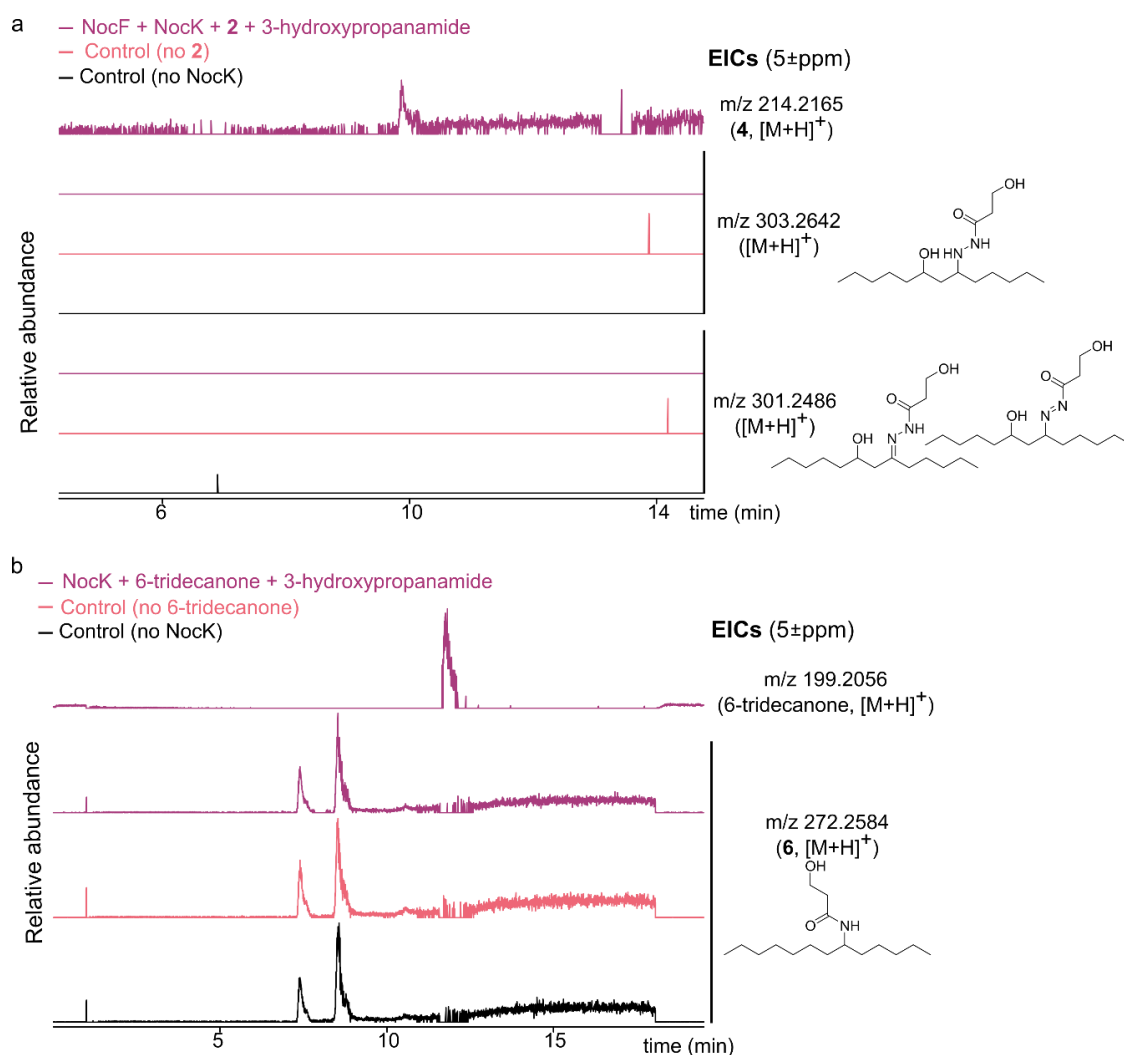


Figure SII.42 - LC-HRMS-derived EICs of putative products of NocK products generated from **4** and 6-tridecanone with 3-hydroxypropanamide. (a) LC-HRMS-derived EICs of putative products of NocK products generated from **4** (coupled assay with NocF and **2**) with 3-hydroxypropanamide. (b) LC-HRMS-derived EICs of **7** in NocK enzymatic assays with 6-tridecanone. None of the putative m/z features, absent in the controls, were detected in the assay mixtures.

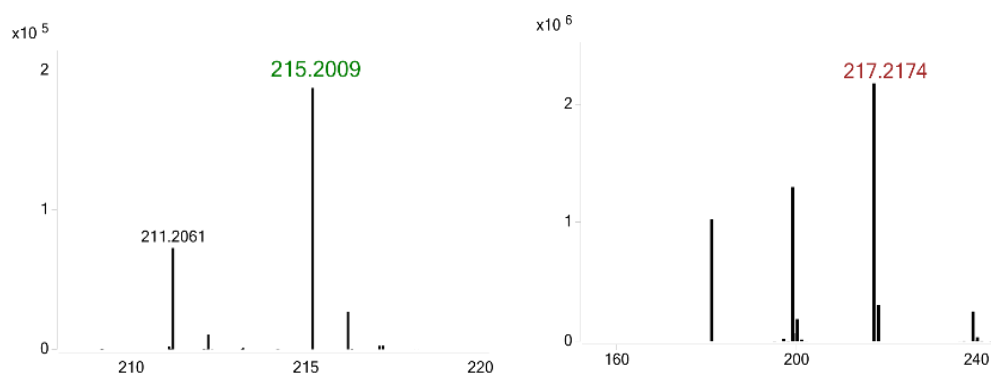


Figure SII.43 - HRMS-derived EICs of the synthetic compounds **2** (a) and **3** (b), confirming the expected masses.

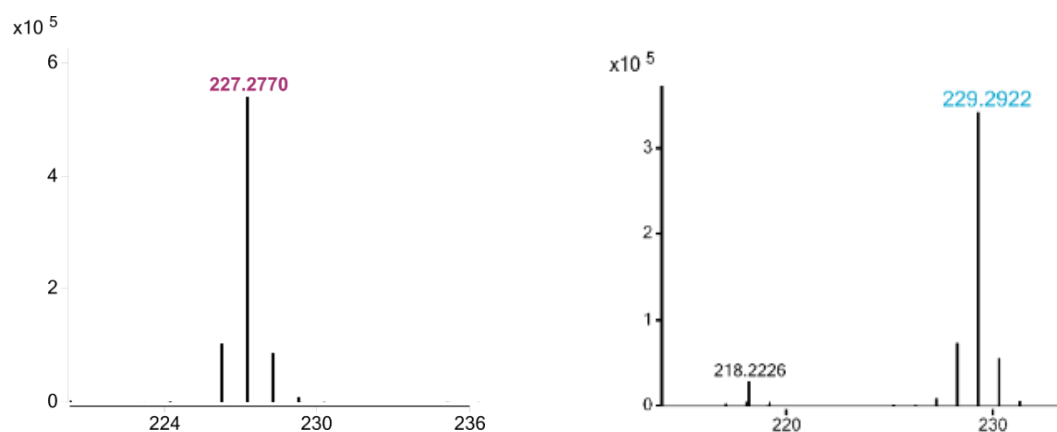


Figure SII.44 - HR-MS spectrum of synthetic compounds of d_{12} -2 (a) and d_{12} -3 (b) confirming the expected masses.

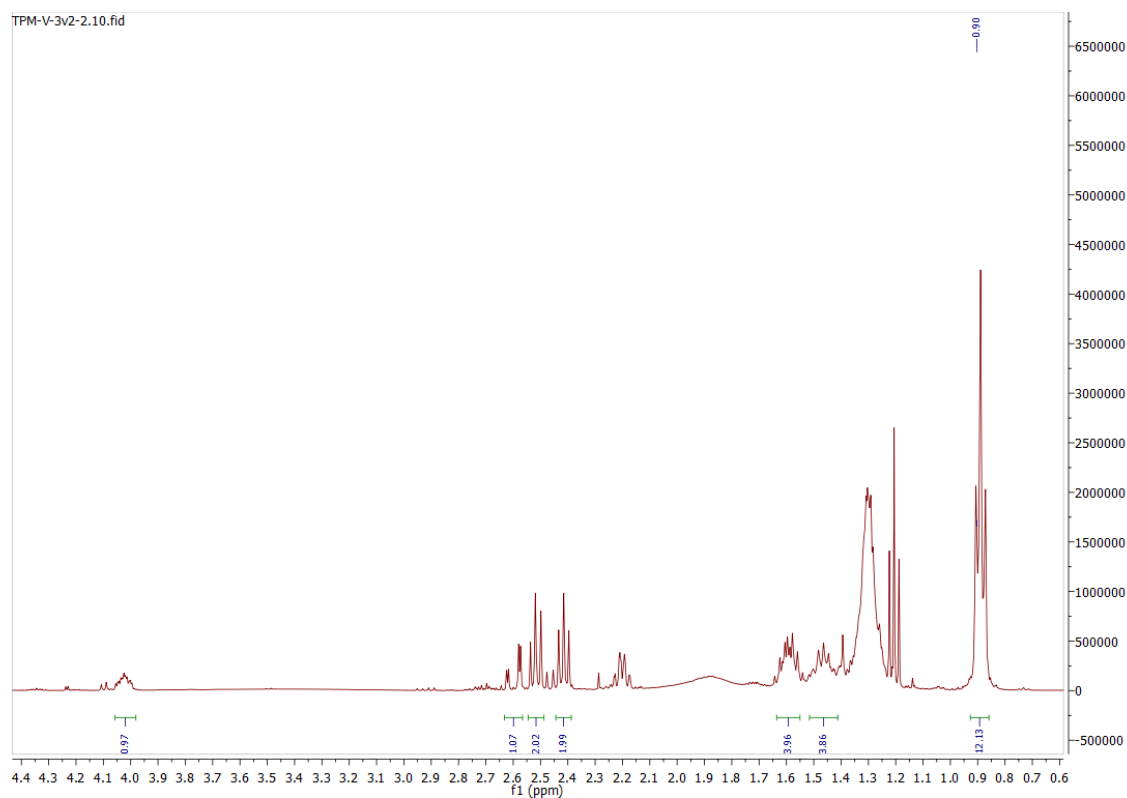


Figure SII.45 - ^1H NMR (400 MHz) of **2** in CD_3Cl .

TPM-V-4-15-37.10.fid

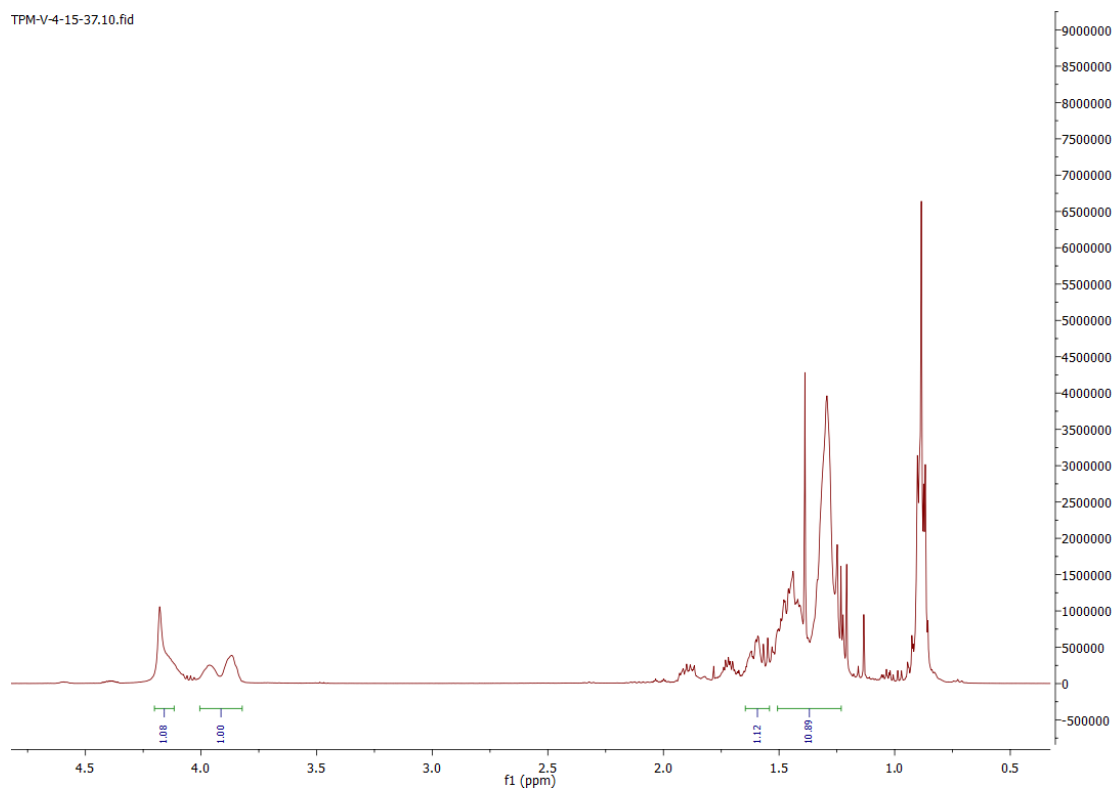


Figure SII.46 - ^1H NMR (400 MHz) of **3** in CD_3Cl .

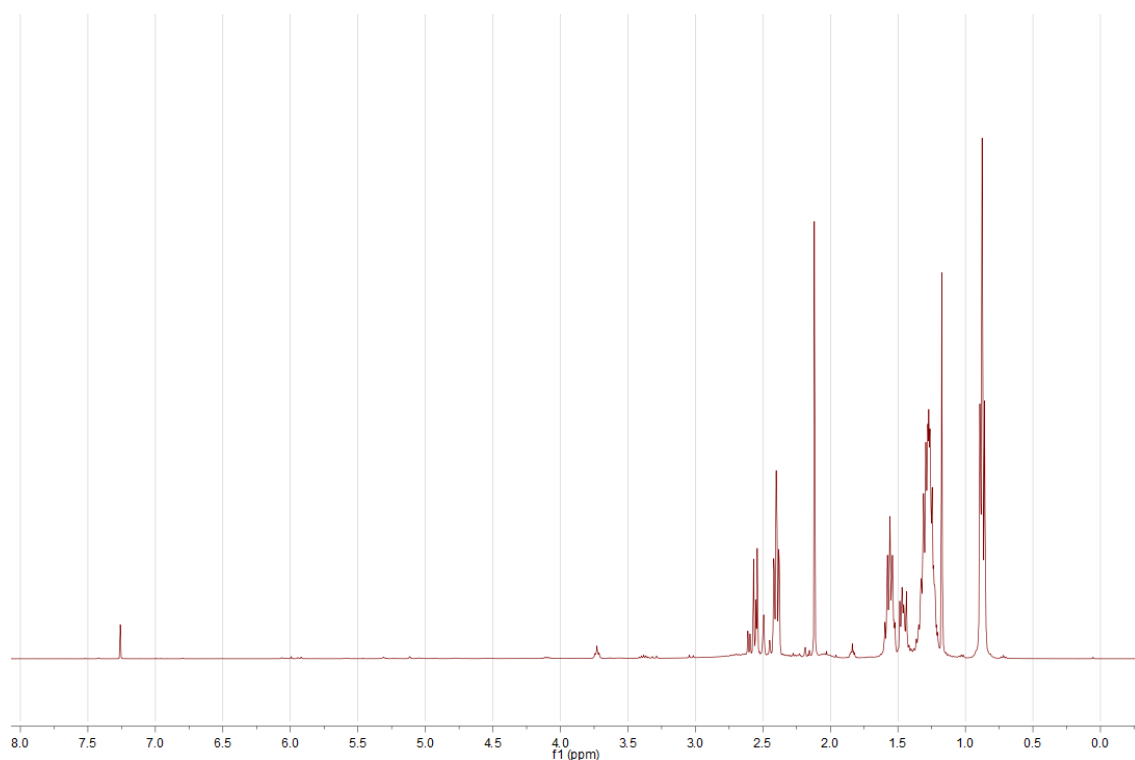


Figure SII.47 - ^1H NMR (400 MHz) of d_{12} -**2** in CD_3Cl .

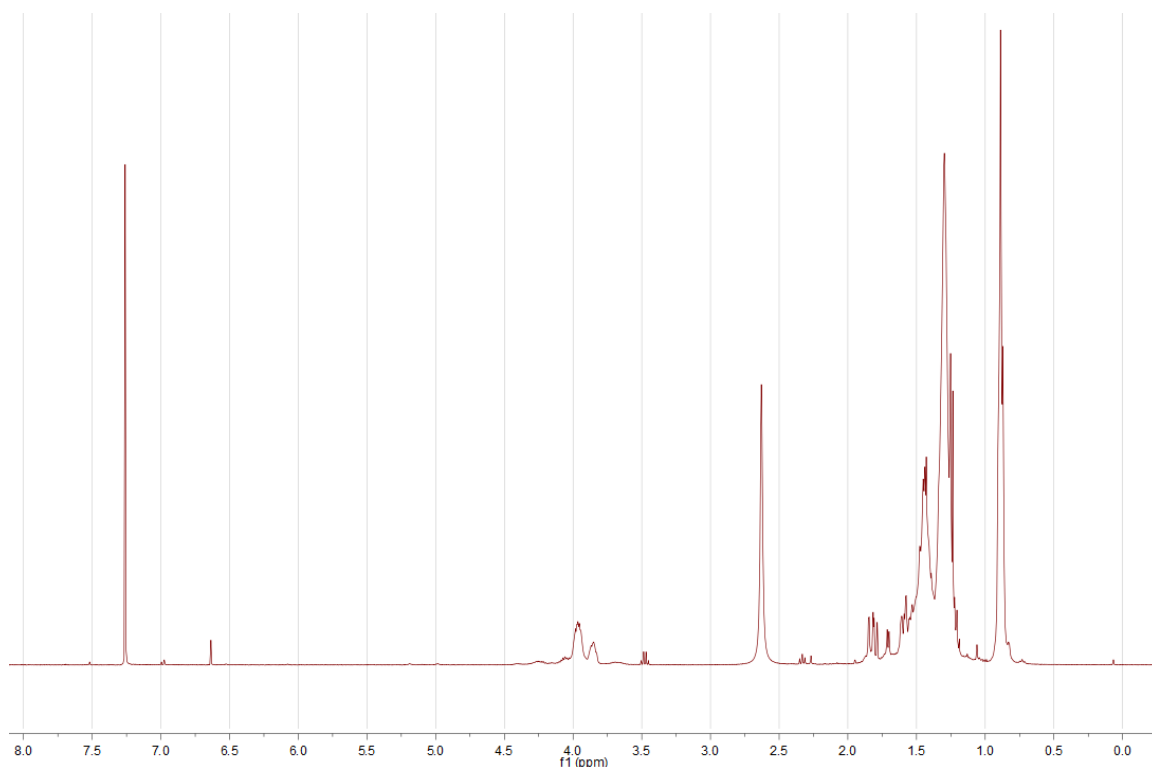


Figure SII.48 - ^1H NMR (400 MHz) of d_{12} -**3** in CD_3Cl

References

1. Kotai, J., *Instructions for Preparation of Modified Nutrient Solution Z8 for Algae*. Oslo: Norwegian Institute for Water Research: 1972.
2. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010**, *11*, 395-406.
3. 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.
4. Zallot, R.; Oberg, N.; Gerlt, J. A., The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58* (41), 4169-4182.
5. Black, T. A.; Cai, Y.; Wolk, C. P., Spatial expression and autoregulation of hetR, a gene involved in the control of heterocyst development in *Anabaena*. *Mol Microbiol* **1993**, *9* (1), 77-84.
6. Heidorn, T.; Camsund, D.; Huang, H. H.; Lindberg, P.; Oliveira, P.; Stensjo, K.; Lindblad, P., Synthetic biology in cyanobacteria engineering and analyzing novel functions. *Methods Enzymol.* **2011**, *497*, 539-579.
7. Jumper, J.; Evans, R.; Pritzel, A.; Green, T., et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589.
8. Van Draanen, N. A.; Arseniyadis, S.; Crimmins, M. T.; Heathcock, C. H., *J. Org. Chem.* **1991**, (56), 2499-2506.

9. Menche, D.; Hassfeld, J.; Li, J.; Rudolph, S., Total synthesis of archazolid A. *Journal of the American Chemical Society* **2007**, *129* (19), 6100-6101.
10. Martins, T. P.; Glasser, N. R.; Kountz, D. J.; Oliveira, P.; Balskus, E. P.; Leao, P. N., Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A. *ACS Chem. Biol.* **2022**, 2528-2537.
11. Voráčová, K.; Hájek, J.; Mareš, J.; Urajová, P., et al., The cyanobacterial metabolite nocuolin a is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLoS One* **2017**, *12* (3), e0172850.
12. Eusebio, N.; Castelo-Branco, R.; Sousa, D.; Preto, M.; D'Agostino, P.; Guider, T. A. M.; Leao, P. N., Discovery and Heterologous Expression of Microginins from *Microcystis aeruginosa* LEGE 91341. *ACS Synth. Biol.* **2022**, *11* (10), 3493–3503.

Chapter III

Manuscript

**NocC is a β,γ acyl-ACP dehydrogenase
involved in nocuolactylate biosynthesis**

NocC is a β,γ acyl-ACP dehydrogenase involved in nocuolactylate biosynthesis

Teresa P. Martins^{1,2}, Nate R. Glasser³, Emily P. Balskus^{3,4*}, Pedro N. Leão^{1*}

¹Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n 4450-208 Matosinhos, Portugal

²School of Medicine and Biomedical Sciences, University of Porto, Rua Jorge de Viterbo Ferreira 228, 4050-313, Porto, Portugal

³Department of Chemistry & Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, 02138 MA, USA

⁴Howard Hughes Medical Institute, Harvard University, Cambridge, MA USA

* corresponding authors

Abstract

Acyl-CoA dehydrogenases (ACADs) are a large family of flavin-dependent oxidoreductases that typically desaturate at the α,β position of various fatty acyl-CoA substrates. Some ACADs have been shown to perform unusual desaturations at less electrophilic sites like γ,δ . However, no ACAD that desaturates at the β,γ position has been characterized to date. Previous studies suggest the ACAD NocC, potentially involved in the biosynthesis of nocuolactylates, cyanobacterial natural products with a β,γ unsaturated C_{12} chain, could have such an activity. Here, we report that NocC is an acyl-ACP β,γ dehydrogenase. Using *in vitro* enzymatic assays and *in vivo* supplementation experiments, we confirmed that NocC catalyzes the formation of β,γ -acyl thioesters. This reaction likely involves initial abstraction of the α proton to generate an α,β unsaturated intermediate that is later shifted to the β,γ position. NocC shows promiscuity towards acyl chain length (C8 to C16), branching and halogenation. With this findings, we have characterized the first acyl-ACP dehydrogenase known to generate a β,γ unsaturated product.

Introduction

The presence of alkene groups in natural products is particularly interesting in terms of reactivity as they can be easily derivatized to introduce a wide range of other desirable functional groups.^{1,2} Alkenes are often generated during polyketide biosynthesis as a result of the sequential action of ketoreductase and dehydratase domains.³ However, they can also be formed oxidatively by desaturases or acyl-CoA dehydrogenases (ACADs). ACADs are a large family of flavin-dependent oxidoreductases that typically desaturate at the α,β

positions of various fatty acid acyl-CoA conjugates.⁴ These enzymes are involved in amino acid and in fatty-acid metabolism, catalyzing the first oxidative step of the β -oxidation cycle. Because of their central role in metabolism, deficiency in these enzymes is the cause of several metabolic diseases like medium-chain ACAD disease (MCADD).⁵⁻⁷ In secondary metabolism, these enzymes have been associated with a restricted number of biosynthetic pathways. However, ACADs are responsible for the introduction of double bonds in NPs like tacrolimus and tambjamines (Fig. III.1a).^{8,9}

Despite large differences in specificity, all ACADs use flavin adenine dinucleotide (FAD) as a cofactor, are structurally similar and share the same general mechanism for α,β -dehydrogenation.¹⁰ Unlike the majority of other dehydrogenases, where at least one hydrogen is connected to a heteroatom and is kinetically labile, acyl-CoA dehydrogenases break two kinetically stable C-H bonds. The mechanism of this α,β -dehydrogenation involves α C-H ($pK_a \approx 20$)¹⁰ deprotonation by a catalytic base, usually Glutamate, and concertedly the β C-H hydrogen is transferred as a hydride to FAD, forming FADH₂.¹⁰ Apart from α,β -dehydrogenation, some ACADs, such as TcsD⁸ and the homologous HliR,¹¹ involved in the biosynthesis of the clinical immunosuppressant FK506 also known as tacrolimus,¹² and the antibiotic haliangicin,¹¹ respectively, are able to perform unusual reactions like γ,δ dehydrogenation⁸ (Fig. III.1a). Furthermore, the biosynthesis of the tambjamines YP1, prodiginine related metabolites with antifungal, antifouling, and antimicrobial activities, the ACAD TamT has been hypothesized as a β,γ ACAD, but no characterization of the enzyme was performed.⁹ Thus, to our knowledge a β,γ fatty acyl-ACP dehydrogenase has not been reported to date.

In the polyketide biosynthetic pathways that produce rhizoxin D, ansamitocin P3 and mushroom polyenes, β,γ -unsaturation is generated by shifting the typical α,β -unsaturation.¹³⁻¹⁵ For example, in rhizoxin biosynthesis, the α,β unsaturated moiety is generated by RhiD module 7's dehydrogenase (DH) domain and subsequently shifted by a second modified DH domain that is part of RhiD module 8.¹³

Apart from polyketides, β,γ unsaturation is also found in other classes of natural products like the nocuolactylates¹⁶ (Fig. III.1b). These are hybrid molecules that result from the esterification of nocuolin A – a unique 1,2,3-oxadiazine¹⁷ – and chlorosphaerolactylates, esters of (poly)chlorinated dodecanoic acid and L-lactic acid.¹⁸ Each of these compounds was recently connected to the *noc* and *cly* BGCs,^{19,20} respectively; hence the nocuolactylates are likely produced by the entire *noc* locus, which comprises both *noc* and *cly* BGCs.¹⁶

The β,γ unsaturation in the nocuolactylates is part of a dodecanoic acid-derived acyl chain in the chlorosphaerolactylate substructure. While a PKS-encoding gene is present in the *cly* BGC, this fatty acyl chain is recruited and incorporated into the

chlorosphaerolactylates by a fatty acyl AMP-ligase (FAAL).¹⁹ Thus, the β,γ unsaturation in these molecules is not predicted to be of PKS origin. The *noc* locus also encodes two ACADs, NocC and NocI,²⁰ whose dehydrogenase activity was previously confirmed in our study²⁰ (Fig. III.1b, Table SIII.1). In this study, we show that NocC is a β,γ acyl-ACP dehydrogenase. This enzyme is responsible for the β,γ unsaturation of the nocuolactylates, which likely involves a multiple step mechanism that starts by the formation of an α,β unsaturation that later is shifted to the β,γ position. NocC shows high promiscuity, which could be of interest for biocatalysis.

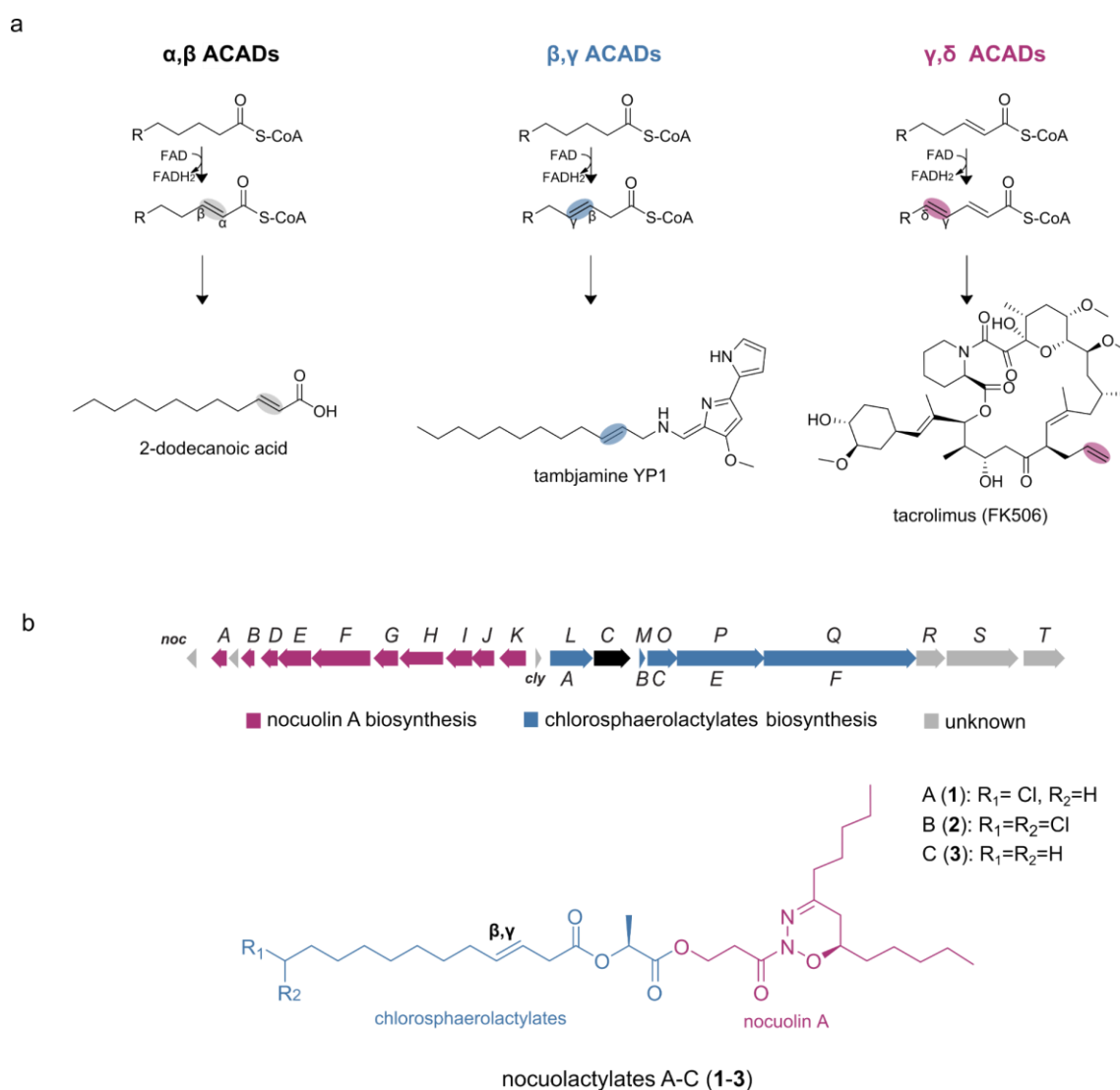


Figure III.1 – Acyl-CoA dehydrogenases and their unsaturated natural products. (a) Generic reactions performed by α,β , β,γ , γ,δ ACADs and unsaturated metabolites resulting from these reactions: 2-dodecanoic acid, tambjamine YP1 and tacromilus (FK506), respectively (for schematic simplicity, thioesters are represented as CoA thioesters, but ACADs can either use CoA or ACP thioesters). (b) The *noc* and *cly* biosynthetic gene clusters, found in the *noc* locus in *Nodularia* sp. LEGE06071 are involved in the biosynthesis of the nocuolactylates (1-3), which show a β,γ unsaturation in an acyl chain.

Results and Discussion

Identification of NocC as a putative β - γ acyl-ACP dehydrogenase

Nodularia sp. LEGE 06071 (hereafter *Nodularia*) and *Sphaerospermopsis* sp. LEGE 00249 (hereafter *Sphaerospermopsis*) harbor the *cly* BGC and, while *Sphaerospermopsis* was reported to produce only the chlorosphaerolactylates,^{18,19} *Nodularia* was reported to produce the hybrid nocuolactylates¹⁶ (Fig. III.1b). However, unlike the nocuolactylates found in *Nodularia*, the chlorosphaerolactylates from *Sphaerospermopsis* do not feature the β,γ unsaturation. We searched the crude extracts of both strains for the most abundant reported chlorosphaerolactylates (m/z [M-H]⁻ 339.1135, 305.1525) and their corresponding unsaturated versions (m/z [M-H]⁻ 337.0979, 303.1369) (Fig. III.2). *Sphaerospermopsis* produced fully saturated versions of the molecules (chlorosphaerolactylates A, **4** and B, **5**) and *Nodularia* produced the unsaturated versions, named chlorosphaerolactylates E (**6**) and F (**7**), in agreement with the structures of the corresponding nocuolactylates¹⁶ (**2** and **1**, respectively). Additionally, neither nocuolin A nor **1-3** nocuolactylates were found in *Sphaerospermopsis* suggesting that the *noc* BGC could be silent in this strain (Fig. SIII.1, Table SIII.2). Despite the high similarity between the *cly* BGCs of both strains, *Sphaerospermopsis* does not contain *nocC*. This gene was predicted to encode an acyl-CoA dehydrogenase and this activity was confirmed in our previous study²⁰ where NocC was tested in a coupled assay with ClyB-bound fatty acyl substrates. In available genomes

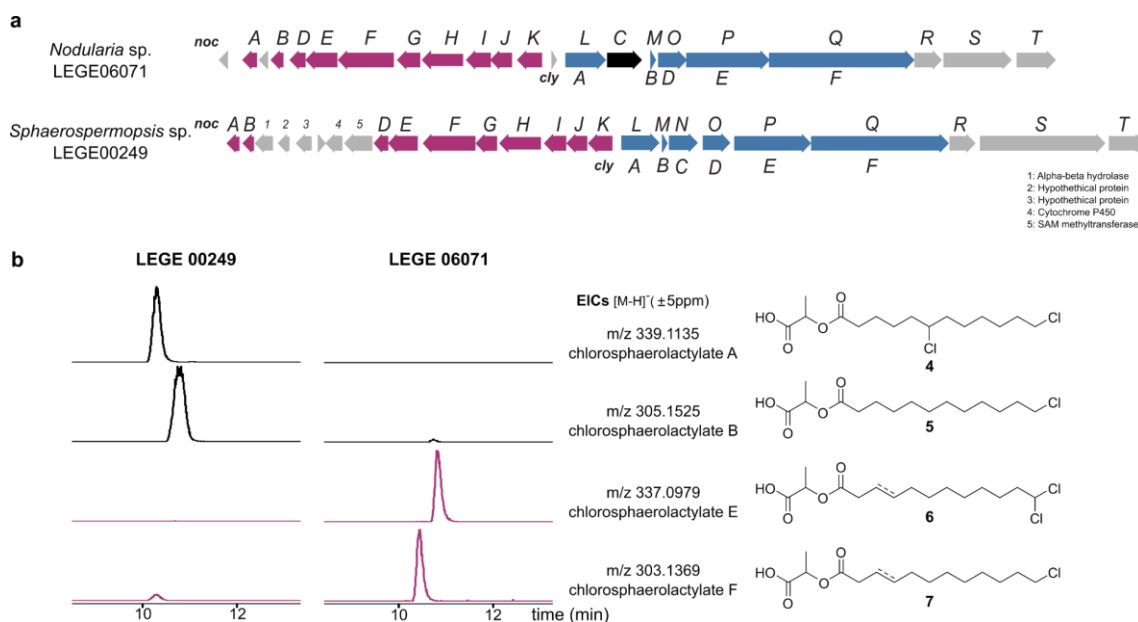


Figure III.2 – *nocC* is present in cyanobacteria producing unsaturated chlorosphaerolactylates like *Nodularia* and absent in cyanobacteria producing fully saturated chlorosphaerolactylates. (a) The *noc/cly* locus harbors the *nocC* gene in *Nodularia* which is absent in the same locus in *Sphaerospermopsis*. (b) EIC of saturated chlorosphaerolactylates A and B (**4**, **5**) and new unsaturated chlorosphaerolactylates E and F (**6**, **7**) (hypothetical structures) in both cyanobacterial extracts

containing the *noc* locus, *nocC* is found either in the *noc* or *cly* subclusters (Fig. SIII.2); however we considered that overall, the available evidence suggested that NocC could be the β,γ acyl-ACP dehydrogenase responsible for the C₁₂ chain unsaturation in the chlorosphaerolactylates.

NocC is a promiscuous β,γ acyl-ACP dehydrogenase

Given our previous success in expressing and purifying recombinant NocC,²⁰ we considered performing *in vitro* assays to study the putative involvement of NocC in generating the β,γ unsaturated acyl moiety of the chlorosphaerolactylates/nocuolactylates. C-His₆-NocC was heterologously expressed in *Escherichia coli* BL21(DE3) and purified using a nickel resin (Table SIII.3, SIII.4, Fig. SIII.3). The purified enzyme had yellow color typical of FAD-dependent enzymes (Fig. SIII.4). The UV-vis profile of the enzyme and LC-HRMS data confirmed the presence of the FAD cofactor, as expected for an ACAD (Fig. SIII.4). To test the regioselectivity of NocC, we performed coupled assays with ClyA (FAAL) and ClyB (ACP) using octanoic acid (hereafter C8) and isotopically labeled *d*₁₅-C8. ClyA loads the supplied fatty acid onto ClyB to generate a fatty acyl-ClyB thioester substrate for NocC. Although the natural substrate of NocC is expected to be dodecanoyl-ClyB, we chose to perform initial assays with C8, because of the commercial availability of 2-octenoic acid (α,β unsaturated) and 3-octenoic acid (β,γ unsaturated), which can be loaded onto ClyB by ClyA to generate standards in order to confirm the identity of the reaction products of NocC. The previously described dehydrogenase activity of NocC²⁰ was again observed for the new protein preparation, using octanoyl-ClyB thioester as substrate, by detecting a *m/z* shift of 1.0068 consistent with the loss of two protons in $[M+2H]^{2+}$ ions. The product generated by NocC in this assay matched the retention time of the 3-octenoyl-ClyB standard and not that of the 2-octenoyl-ClyB standard, confirming the role of NocC as β,γ acyl-ACP dehydrogenase (Fig. III.3a, Fig. SIII.5, SIII.6). The second ACAD in the *noc* locus – NocI – was also expressed, purified, and tested in this assay, but no dehydrogenase activity was detected with C8- or C12-ClyB thioesters (Fig. III.3a, Fig. SIII.7).

In the assay with *d*₁₅-C8-ClyB, NocC was expected to either generate an α,β unsaturated product which would result in the loss of 2 deuterons or a β,γ unsaturated product which would result in the same outcome or in the loss of 3 deuterons and gain of 1 proton due to multiple reaction steps. In this assay, the unsaturated product of NocC showed an isotopic pattern which indicates the loss of three deuterons and the gain of one proton (Fig. III.3b). This suggests that NocC performs a multi-step reaction to generate the β,γ unsaturation, which we interpret as an initial α,β unsaturation (loss of 2 deuterons)

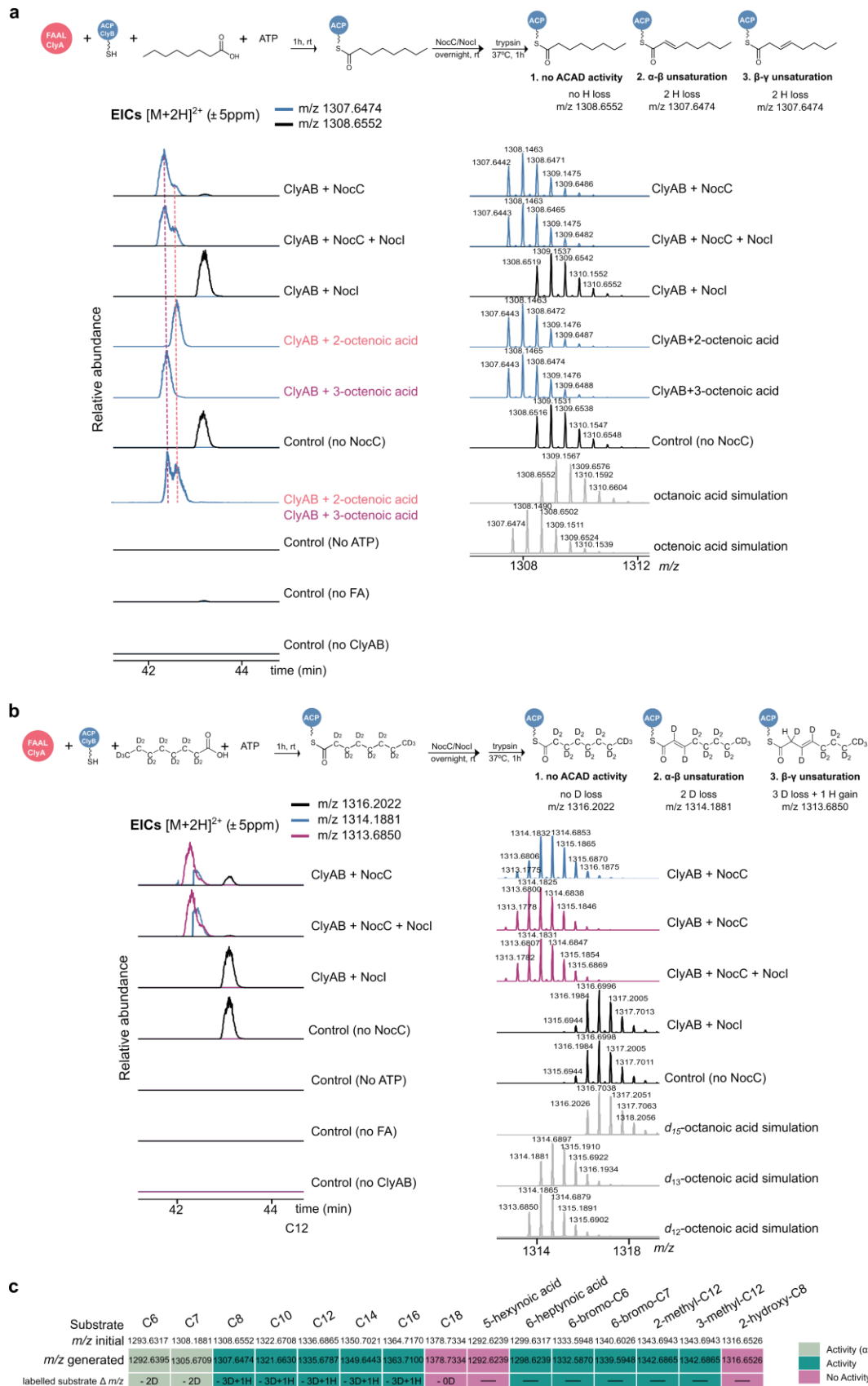


Figure III.3 – NocC is a β,γ acyl-ACP dehydrogenase. (a) LC-HRMS-derived EICs and spectrum of NocC enzymatic assay with C8 coupled with ClyAB showed the same retention time as the commercial standard *trans*-3-octenoic acids loaded into ClyA, proving the β,γ activity of this ACADs (b) LC-HRMS-derived EICs and spectrum of NocC enzymatic assays with *d*₁₅-C8 coupled with ClyAB showed the loss of 3 deuterons and gain of 1 proton, consistent with the initial formation of a α,β unsaturation followed by shifting for the final β,γ positions.(c) NocC is able to accept a large range of substrates showing promiscuity.

which is then shifted to the β,γ position (loss of a third deuteron and gain of one proton). This proposal matches the observed pattern in the EICs since the major *trans*-3-octenoic-ClyB thioester peak which is followed by a small *trans*-2-octenoic-ClyB thioester. To our knowledge, this is the first characterized acyl-ACP dehydrogenase able to perform β,γ unsaturation.

Since dodecanoic acid is predicted to be the main substrate of ClyAB²⁰, we sought to confirm the observed NocC activity with coupled enzymatic assays, generating d_{23} -dodecanoyl-, 2,2- d_2 -dodecanoyl- or dodecanoyl-ClyB substrates by supplying the respective fatty acids and ClyAB (Fig. SIII.7). In the 2,2- d_2 -dodecanoic acid assay, the isotopic pattern of the acyl-ClyB product generated by NocC indicates the loss of one deuteron and one proton, which is compatible with an α,β unsaturation or an α,β unsaturation followed by a β,γ unsaturation. For the d_{23} -dodecanoic acid assay, a loss of three deuterons and the gain of one proton was observed in the acyl-ClyB product, in a similar fashion to what was observed with the perdeuterated octanoic substrate. This outcome is also consistent with a two-step $\alpha,\beta/\beta,\gamma$ unsaturation shift.

We then tested the substrate scope of NocC using the previously described coupled assay and different fatty acids: C6, C7, C8, C10, C12, C14, C16 and C18 (Fig. III.3c, Figs. SIII.8, SIII.9). The enzyme proved to be promiscuous and showed dehydrogenase activity for acyl-ClyB thioesters ranging from C6 to C16. No activity was detected for the assay with C18, which could have been caused by an inability of ClyA to load this fatty acid onto ClyB. However, assays with the perdeuterated versions of the fatty acids indicated that the enzyme is only able to perform α,β unsaturation on the shorter-chain (C6 and C7) acyl-ClyB substrates, since only two deuterons were lost in the acyl moiety (Fig. SIII.9). For the remaining fatty acid substrates, the loss of three deuterons and the gain of a proton was again observed. NocC tolerated additional variability in the acyl chains, as it showed activity toward 6-bromo-heptanoyl-, 6-bromo-hexanoyl- and heptynoyl-ClyB and with branched fatty acyl-ClyB substrates 2-methyldodecanoyl- and 3-methyldodecanoyl-ClyB (Figs. SIII.8, SIII.9). Until recently, it was suggested that ACADs have strict substrate specificities^{21,22} and only a few exceptions have been reported, like the promiscuous long-chain acyl-CoA dehydrogenase FadE5.²³ The broad substrate scope of NocC is interesting from a biocatalysis perspective since it can generate diverse β,γ products, increasing the range of potential organic chemistry applications.²⁴ However, we were unable to fully reconstitute the β,γ unsaturation activity of NocC using d_{23} -dodecanoyl-SNAC substrate mimics; this yielded only the α,β unsaturated acyl thioester intermediate (Fig. SIII.10), and suggests that interaction of NocC with ClyB is important for the alkene shift.

Finally, to confirm the role of NocC *in vivo*, we supplemented *Nodularia* cultures with d_{23} -dodecanoic acid and investigated the isotope patterns of labelled **1-3**. As expected, and

taking into account deuterium loss due to halogenation, the deuterium-labelled versions of **1-3** indicated the loss of three deuterons due to the β,γ unsaturation (Fig. SIII.11). These data support that *in vivo* wild-type NocC performs the same two-step reaction observed in the *in vitro* assays with recombinant NocC.

NocC represents a new class of ACADs

To gain insight into how NocC compares to other ACADs, we searched for homologues in the NCBI non-redundant database. A BlastP search showed that after the first 12 hits, the identity percentage drops markedly to below 50% (Table SIII.5). We then performed a phylogenetic analysis of 308 different ACADs sequences from the main phylogenetic groups described so far. As expected, the resulting phylogenetic tree shows that NocC clades separate from all other characterized ACADs (Fig. III.4a, Table SIII.6). The NocC-containing clade comprises mostly sequences from Cyanobacteria, Proteobacteria and Myxobacteria. Apart from α,β unsaturation-generating ACADs, NocC also clades separately from γ,σ unsaturation-generating ACADs like HliR¹¹ and TcsD.⁸ A similar grouping can be observed in a sequence similarity network (SSN) and genomic neighborhood network (GNN) analyses (Fig. SIII.12, SIII.13). These analyses highlight the uniqueness of NocC and homologues among previously characterized ACADs.

To better understand how the primary structure of NocC compares to characterized ACADs that generate unsaturations at different positions, an alignment of the amino acid sequence of NocC with TcsD and α,β unsaturation-generating ACADs⁸ was carried out (Fig. SIII.14). Key residues were conserved among the enzymes; notably, Glu 373 is conserved and is predicted to localize close to the FAD co-factor and perform the deprotonation of the α -carbon acting as the catalytic base (Fig. III.4b).²⁵ This indicates that the first reaction step of NocC likely follows a typical ACAD concerted mechanism with α C-H hydrogen deprotonation by the catalytic base and transfer of the β C-H hydrogen as a hydride to FAD.¹⁰ To confirm the key role of this residue, we generated a E373A NocC mutant. The mutation caused loss of FAD binding as ascertained from a lack of yellow coloration (Fig. SIII.4) and loss of detectable activity in the *in vitro* assays using different substrates (Fig. III.4b, Fig. SIII.15).

In terms of tertiary structure, the AlphaFold 2.0^{26,27} model of NocC shows a typical ACAD structure with three subdomains: an N-terminal α -helix, a β -sheet domain and a C-terminal α -helix. FAD is predicted to bind in a pocket formed by the C-terminal α -helix domain, the β -sheet domain, and the C-terminal domain of the adjacent subunit,

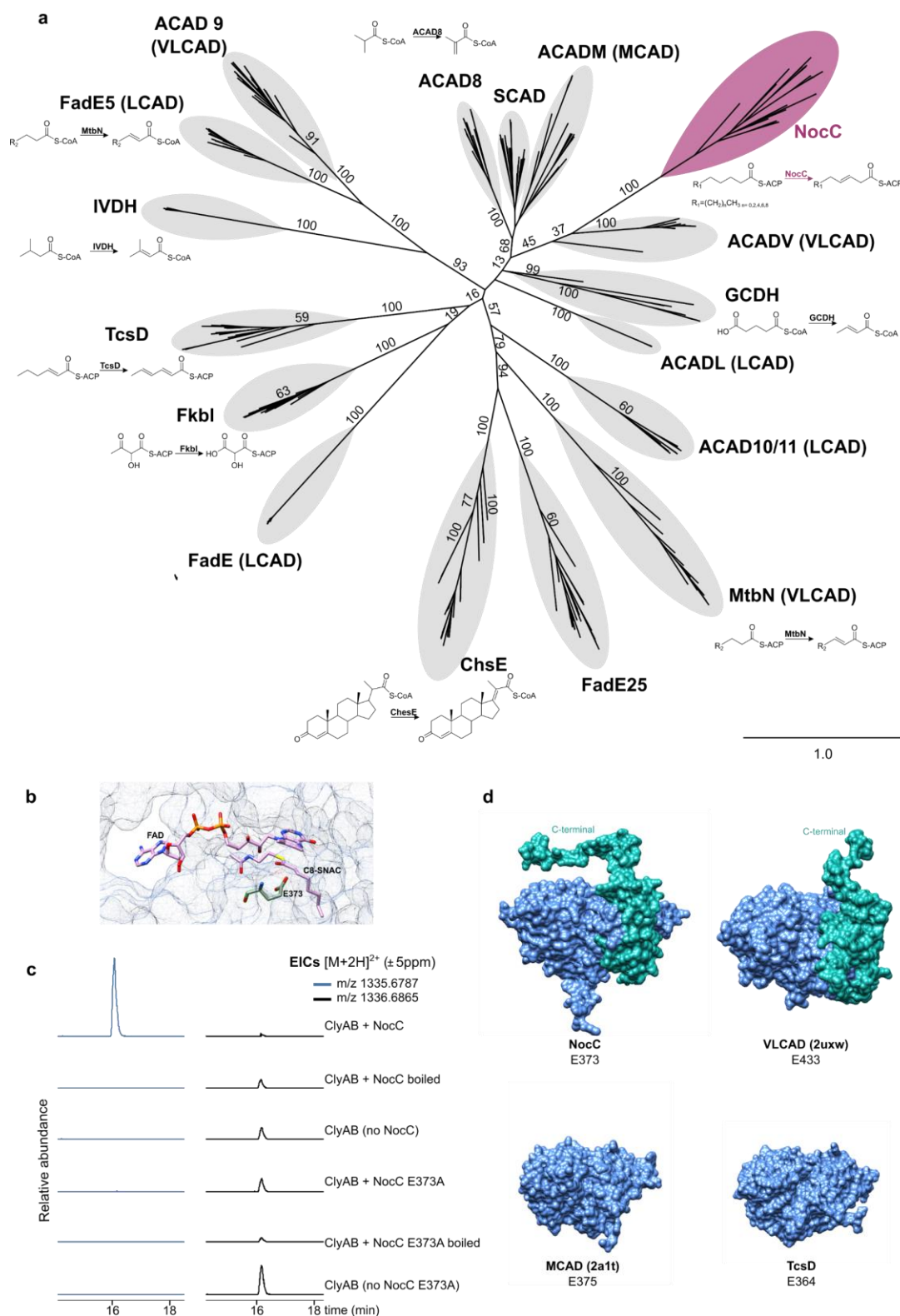


Figure III.4 – NocC represents the first characterized member of a unique group of ACADs. (a) Phylogenetic tree highlighting the relationship of NocC and its closest homologues with other ACADs (the scale bar indicates the degree of divergence as substitutions per site); characterized homologs and their reactivity is annotated for key clades. (b) AlphaFold model of NocC highlighting the catalytic E373, FAD and C8-SNAC (FAD and C8-SNAC were positioned based on NocC-2uxw crystal structure overlap) (c) LC-HRMS-derived EICs of NocC and NocC E373A enzymatic assay. (d) 3D structure of NocC AlphaFold model and comparison with VLCAD, MCAD and TcsD crystal structures.

underneath the Glu373 fatty acyl binding chamber (Fig. SIII.16). NocC contains a long C-terminal region that is absent in TcsD, and in short-chain and middle-chain ACADs (Fig. SII.14). Despite the low identity at the amino acid level, this long C-terminal region is exclusively present in long and very-long acyl-ACP dehydrogenases (VLCAD)²² (Fig. III.4c), being predicted to mediate interactions with membranes²⁸ (Fig. III.4c). Despite these similarities towards LCADs and VLCADs, the presence of the catalytic glutamate at position 373 is characteristic of medium-chain acyl-CoA dehydrogenases - MCADs (E376)²⁹ - and not VLCAD (E422).²² NocC thus presents a blend of structural characteristics of both MCADs and VLCADs.

Conclusion

This study shows that the β,γ unsaturation found in the nocuolactylates and chlorosphaerolactylates results from the activity of a β,γ acyl-ACP dehydrogenase, NocC. Despite the thermodynamically less stable β,γ -unsaturation when compared with α,β -unsaturation, these observations are compatible with the chemical structures of nocuolactylates, previously characterized by NMR. This is the first such enzyme characterized to date, and it accepts a wide range of acyl-ACP thioester substrates (from C8 to C16) including branched and halogenated versions. Isotope labelling experiments with perdeuterated fatty-acids are consistent with a multi-step mechanism, where an α,β unsaturation is formed and then shifted to yield a β,γ unsaturated thioester. NocC is predicted to have structural characteristics of both MCADs and VLCADs and we predict the reaction is initiated by a catalytic residue conserved in these ACADs (Glu373) but further studies are necessary to fully elucidate its catalytic mechanism. The ability of NocC to generate a wide range of β,γ unsaturated products, which can be easily derivatized to introduce desirable functionalities, might be particularly interesting for future biocatalytic applications, in particular interfacing with synthetic organic chemistry, where beta-gamma substrates are important building blocks.³⁰

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 and UIDP/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 J. Morais for assistance with cyanobacterial cultures.

Author contribution

T.P.M. carried out all experimentation. N.R.G. assisted with the experimentation. T.P.M., E.P.B. and P.N.L. designed research. T.P.M. and P.N.L. drafted the manuscript, with contributions from all co-authors.

References

1. Xu, J.; Fu, Y.; Luo, D. F.; Jiang, Y. Y.; Xiao, B.; Liu, Z. J.; Gong, T. J.; Liu, L., Copper-Catalyzed Trifluoromethylation of Terminal Alkenes through Allylic C-H Bond Activation. *Journal of the American Chemical Society* **2011**, *133* (39), 15300-15303.
2. Coombs, J. R.; Morken, J. P., Catalytic Enantioselective Functionalization of Unactivated Terminal Alkenes. *Angew. Chem.* **2016**, *55* (8), 2636-2649.
3. Keatinge-Clay, A. T., The structures of type I polyketide synthases. *Nat. Prod. Rep.* **2012**, *29* (10), 1050-1073.
4. Swigonova, Z.; Mohsen, A. W.; Vockley, J., Acyl-CoA Dehydrogenases: Dynamic History of Protein Family Evolution. *J. Mol. Evol.* **2009**, *69* (2), 176-193.
5. Merritt, J. L., 2nd; Chang, I. J., Medium-Chain Acyl-Coenzyme A Dehydrogenase Deficiency. In *GeneReviews((R))*, Adam, M. P.; Everman, D. B.; Mirzaa, G. M.; Pagon, R. A.; Wallace, S. E.; Bean, L. J. H.; Gripp, K. W.; Amemiya, A., Eds. Seattle (WA), 1993.
6. Lund, M.; Olsen, R. K. J.; Gregersen, N., A short introduction to acyl-CoA dehydrogenases; deficiencies and novel treatment strategies. *Expert Opinion on Orphan Drugs* **2015**, *3* (12), 1375-1386.
7. Mason, E.; Hindmarch, C. C. T.; Dunham-Snary, K. J., Medium-chain Acyl-CoA dehydrogenase deficiency: Pathogenesis, diagnosis, and treatment. *Endocrinol Diabetes Metab* **2023**, *6* (1), e385.
8. Blake-Hedges, J. M.; Pereira, J. H.; Cruz-Morales, P.; Thompson, M. G., et al., Structural Mechanism of Regioselectivity in an Unusual Bacterial Acyl-CoA Dehydrogenase. *J. Am. Chem. Soc.* **2020**, *142* (2), 835-846.
9. Marchetti, P. M.; Kelly, V.; Simpson, J. P.; Ward, M.; Campopiano, D. J., The carbon chain-selective adenylation enzyme TamA: the missing link between fatty acid and pyrrole natural product biosynthesis. *Org. Biomol. Chem.* **2018**, *16* (15), 2735-2740.
10. Ghisla, S.; Thorpe, C., Acyl-CoA dehydrogenases. A mechanistic overview. *Eur. J. Biochem.* **2004**, *271* (3), 494-508.
11. Sun, Y.; Feng, Z.; Tomura, T.; Suzuki, A., et al., Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Sci. Rep.* **2016**, *6*, 22091-22102.

12. Dumont, F. J., FK506, an immunosuppressant targeting calcineurin function. *Curr. Med. Chem.* **2000**, 7 (7), 731-748.
13. Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C., Functionally distinct modules operate two consecutive $\alpha,\beta \rightarrow \beta,\gamma$ double-bond shifts in the rhizoxin polyketide assembly line. *Angew Chem Int Ed Engl* **2010**, 49 (8), 1460-1464.
14. Brandt, P.; Garcia-Altare, M.; Nett, M.; Hertweck, C.; Hoffmeister, D., Induced Chemical Defense of a Mushroom by a Double-Bond-Shifting Polyene Synthase. *Angew Chem Int Ed Engl* **2017**, 56 (21), 5937-5941.
15. Taft, F.; Brunjes, M.; Knobloch, T.; Floss, H. G.; Kirschning, A., Timing of the $\Delta(10,12)$ - $\Delta(11,13)$ double bond migration during ansamitocin biosynthesis in *Actinosynnema pretiosum*. *J. Am. Chem. Soc.* **2009**, 131 (11), 3812-3813.
16. 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.
17. Voráčová, K.; Hájek, J.; Mareš, J.; Urajová, P., et al., The cyanobacterial metabolite nocuolin a is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLoS One* **2017**, 12 (3), e0172850.
18. Gutierrez-Del-Rio, I.; Brugerolle de Fraissinette, N.; Castelo-Branco, R.; Oliveira, F., et al., Chlorosphaerolactylates A-D: Natural Lactylates of Chlorinated Fatty Acids Isolated from the Cyanobacterium *Sphaerospermopsis* sp. LEGE 00249. *J. Nat. Prod.* **2020**, 83 (6), 1885-1890.
19. 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.
20. Martins, T. P.; Glasser, N. R.; Kountz, D. J.; Oliveira, P.; Balskus, E. P.; Leao, P. N., Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A. *ACS Chem. Biol.* **2022**, 2528-2537.
21. Ikeda, Y.; Okamura-Ikeda, K.; Tanaka, K., Purification and characterization of short-chain, medium-chain, and long-chain acyl-CoA dehydrogenases from rat liver mitochondria. Isolation of the holo- and apoenzymes and conversion of the apoenzyme to the holoenzyme. *J. Biol. Chem.* **1985**, 260 (2), 1311-1325.
22. McAndrew, R. P.; Wang, Y.; Mohsen, A. W.; He, M.; Vockley, J.; Kim, J. J., Structural basis for substrate fatty acyl chain specificity: crystal structure of human very-long-chain acyl-CoA dehydrogenase. *J. Biol. Chem.* **2008**, 283 (14), 9435-9443.
23. Chen, X.; Chen, J.; Yan, B.; Zhang, W.; Guddat, L. W.; Liu, X.; Rao, Z., Structural basis for the broad substrate specificity of two acyl-CoA dehydrogenases FadE5 from mycobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, 117 (28), 16324-16332.
24. Bornscheuer, U. T.; Kazlauskas, R. J., Catalytic promiscuity in biocatalysis: Using old enzymes to form new bonds and follow new pathways. *Angew. Chem.* **2004**, 43 (45), 6032-6040.
25. Thorpe, C.; Kim, J. J., Structure and mechanism of action of the acyl-CoA dehydrogenases. *FASEB J.* **1995**, 9 (9), 718-725.
26. Jumper, J.; Evans, R.; Pritzel, A.; Green, T., et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583-589.

27. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S., et al., AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **2022**, *50* (D1), D439-D444.
28. Goetzman, E. S.; Wang, Y.; He, M.; Mohsen, A. W.; Ninness, B. K.; Vockley, J., Expression and characterization of mutations in human very long-chain acyl-CoA dehydrogenase using a prokaryotic system. *Mol Genet Metab* **2007**, *91* (2), 138-147.
29. Kim, J. J.; Wang, M.; Paschke, R., Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (16), 7523-7527.
30. Thadathil, D. A.; Varghese, A.; Radhakrishnan, K. V., The Renaissance of Electro-Organic Synthesis for the Difunctionalization of Alkenes and Alkynes: A Sustainable Approach. *Asian Journal of Organic Chemistry* **2021**, *10* (11), 2820-2847.

Chapter III – Supplementary Information

Published in ACS ChemBio

<https://doi.org/10.1021/acscchembio.2c00464>.

**NocC is a β,γ acyl-ACP dehydrogenase
involved in nocuolactylate biosynthesis**

Methodology

1. General experimental procedures

Stable-isotope labeled (perdeuterated) fatty acids (d_{23} -dodecanoic acid) were obtained from CDN Isotopes. *Trans*-2-octenoic acid was obtained from Apollo scientific, *trans*-3-octenoic acid, 6-bromohexanoic acid, 5-hexynoic acid, 6-heptynoic acid, 7-bromoheptanoic acid from Fluorochem, 3-hydroxyoctanoic acid from Santa Cruz Biotechnology, 2-methyldodecanoic acid and 3-methyldodecanoic acid from BLD Pharmatech. All solvents purchased from VWR Chemicals, Thermo Fisher Scientific, Carlo Erba and Chem Lab NV were ACS grade, except for HPLC solvents (HPLC gradient grade) and LC-MS solvents (MS-grade). Deuterated solvents for NMR were purchased from Cambridge Isotope Laboratories. NMR spectra were acquired in Chloroform on a Bruker Avance III HD, 400 MHz, in the Materials Center of the University of Porto (CEMUP) and the NMR data were analyzed in MestReNova 12.0.3 (MestrelabResearch). LC-HRMS analyses were performed on Thermo Scientific UltiMate 3000 HPLC system, which consists of the following components: LPG-3400RS pump, WPS-3000RS autosampler, TCC-3x00RS column compartment, and MWD-3000RS UV-vis detector coupled to a Q Exactive Focus Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Fisher Scientific), controlled by Q Exactive Focus (Exactive Series) 2.9 and Thermo Scientific Xcalibur 4.1.31.9 software. LC-HRMS/MS analyses were carried out on the latter system.

Luria-Bertani Lennox (LB) medium, used for NocC expressions and Terrific Broth (TB) medium, used for ClyAB expressions, were purchased from Thermo Fisher Scientific. IPTG was acquired from NZYTech. Oligonucleotides were synthesized by StabVida and Sigma-Aldrich. PCR reactions were carried in Veriti 96-well thermal cycler (Applied Biosystems). DNA polymerases, restriction enzymes and T4 DNA ligase were purchased from Thermo Fisher Scientific and New England Biolabs. DNA fragments' extractions from gel and clean-up of restriction reactions were performed using NZYGelpure kit (NZYTech). Plasmid purifications were performed using NZYMiniprep (NZYTech). DNA Sanger sequencing was performed by Eurofins Scientific.

Extracted DNA and PCR reactions were separated in 1 to 1.5% (w/v) agarose gels containing 1× Tris/Acetic acid/EDTA (TAE) buffer (Bio-Rad), stained with GreenSafe Premium (NZYTech) and visualized and photographed on a gel documentation system (Gel Doc EZ Imager, Bio-Rad). NZYDNA Ladder III (NZYtech) and Precision plus all standards blue/Precision Plus (Biorad) were used as DNA and protein molecular weight markers, respectively. Proteins were separated in Mini-PROTEAN® TGX™ Precast Polyacrylamide gels – 4-15% (Biorad) containing 1x Tris/Glycine/SDS (TGS) buffer (Fisher Scientific) and stained with InstantBlue (Expedeon) or Bio-Safe Coomassie G-250 (Biorad). DNA and protein concentrations were determined using a DS-11 FX Spectrophotometer/Fluorometer

(DeNovix) and NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc.). Optical densities (600 nm) of *E. coli* cultures were determined using an Ultrospec 10 Spectrophotometer (GE Healthcare).

2. Plasmids, Strains and culture conditions

Nodularia sp. LEGE 06071 and *Sphaerospermopsis* sp. LEGE 00249 were acquired from the Blue Biotechnology and Ecotoxicology Culture Collection (LEGEcc) and grown in Z8 medium¹ at 25 °C, under a 16 h light (25 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$)/8 h dark cycle and continuous shaking (190 rpm). To assess the functional role of selected proteins encoded by genes in the *noc* gene cluster, the respective coding sequences were cloned in specific expression plasmids, and the proteins purified. To that end, pET-29b and pCDFDuet-1 were acquired from Merck and Novagen. *E. coli* TOP10 (Life Technologies) was used for cloning and *E. coli* BL21(DE3) (Nzytech) and BAP1 (Kerafast) were used for recombinant protein expression.

3. Supplementation experiments

One hundred milliliters of *Nodularia* culture were inoculated in 900 mL of Z8 medium and allowed to grow at 25 °C. After 4 days, 25 mL of culture were distributed evenly through 50 mL Erlenmeyer flasks (duplicates were performed for each condition). A pulse feeding protocol was used: right after inoculation (day 0) and at days 2 and 4, the cultures were supplemented with 0.17 mM of perdeuterated fatty acids: *d*₂₃-dodecanoic acid or DMSO (control) and the biomass harvested after 7 days of growth. After harvesting, fresh biomass from each experimental condition was extracted with DCM:MeOH (2:1) with vigorous vortexing at room temperature. The resulting extracts were filtered with a Whatman No. 1 filter paper, concentrated using a rotary evaporator (R-210, Buchi) and transferred to pre-weighed vials. Extracts from the feeding experiments were dissolved in MS grade methanol to a final concentration of 2 mg mL⁻¹ and analyzed by LC-HRMS and LC-HRMS/MS. The same organic extraction and LC-HRMS and LC-HRMS/MS was performed with *Nodularia* and *Sphaerospermopsis* cultures grown under standard conditions.

4. LC-HRMS and LC-HRMS/MS analysis

HRMS data were obtained in Full Scan positive and negative mode with a scan range of *m/z* 1502000, with a capillary voltage of H set to -3.5 kV and the capillary temperature to 263 °C. Sheath gas flow rate was set to 50 units. Cyanobacterial extracts were separated in an ACE UltraCore™ SuperC18, 2.5 μm particle size, 95 Å pore size and dimensions 75 × 2.1 mm. The column oven was set to 40 °C. The samples were eluted at 0.35 mL/min with a mobile phase of 0.1% formic acid in H₂O/MeOH 95:5 (v/v) (eluent A) and in 0.1%

formic acid in IPA/MeOH 95:5 (eluent B). The gradient program was as follows: linear gradient from 0.5% solution B to reach 10% solution B over 0.5 min, followed by an increase to 60% solution B in 8 min and then to 90% in 1 min; these conditions were kept for 6 min before returning to the initial conditions.

Quenched enzymatic assays with octanoic acid and d_{15} -octanoic acid (standards comparison) were separated in a Aeris peptide XB-C18 column (3.6 μ m particle size, 150 $\times$ 4.6 mm). The column oven was set to 40 °C. The flow was set to 0.5 mL/min with min with a mobile phase of 0.1% formic acid in H₂O (eluent A) and in 0.1 % formic acid in ACN (eluent B). The gradient program was as follows: isocratic 100 % A for 2 min, gradient from 100 % A to 55 % B over 43 min, held at 55 % B for 2 min, before returning to the initial conditions. The remaining quenched enzymatic assays were separated in the same column at 40 °C. The flow was set to 0.5 mL/min with min with a mobile phase of 0.1% formic acid in H₂O (eluent A) and in 0.1% formic acid in ACN (eluent B). The gradient program was as follows: isocratic 100 % A for 2 min, gradient from 100 % A to 100 % B over 16 min, held at 100 % B for 5 min, before returning to the initial conditions.

MS/MS parameters for the LC-HRMS/MS analysis were: resolution of 17500, with a 3 m/z isolation window, a loop count of 3, AGC target of 5×10^4 and collision energy of 30 eV.

5. Cloning, expression and purification

5.1. Cloning, expression and purification of C-His₆-ClyAB and C-His₆-NocI

Cloning and large-scale growth of C-His₆-ClyAB (C-His₆-NocML) and C-His₆-NocI were performed as described in Martins et al.² The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 $\times g$ and the clarified supernatant loaded into 1 mL Ni-IDA packed columns (Macherey Nagel). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50mM imidazole and the protein was eluted in 5 mL of elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction was dialyzed and concentrated in the Amicon ultra-4 4ml 3K (Merck Millipore) for 30 min at 7000 $\times g$ to reach a final concentration of 177.31 μ M (C-His₆-ClyAB) and 47.14 μ M (C-His₆-NocI) (Fig. SIII.3). The proteins were added dropwise to liquid nitrogen and stored at -80 °C.

5.2. Cloning, expression and purification of C-His₆-NocC

The *nocC* gene was amplified from *Nodularia* sp. LEGE 06071 gDNA with primers containing restriction sites for XhoI and NdeI (Table SIII.4).

The reaction contained 1 × Phusion HF Buffer, 400 μM dNTPs, 0.5 μM of each primer, 0.4 U Phusion polymerase (Thermo Scientific) and was performed under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 40 cycles of a denaturation step of 98 °C for 10 seconds, annealing at 58 °C for 30 seconds, extension at 72 °C for 1 min and the final extension at 72 °C for 7 min. PCR products were gel purified, ligated with T4 ligase (NewEngland Biolabs) and transformed in *E. coli* TOP10. The obtained constructs were confirmed by Sanger sequencing and then transformed into chemically competent *E. coli* BL21 (DE3) using the heat shock method. For the large-scale purification, transformants were grown in 100 mL LB medium supplemented with 50 μg mL⁻¹ kanamycin, overnight at 37 °C. Twenty-millilitres of this overnight grown culture were then inoculated into 2 × 2 L of TB medium supplemented with 50 μg mL⁻¹ kanamycin and 5 mM of MgSO₄ at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 30 °C, induced with 0.5 mM of IPTG and allowed to grow for 2 h. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 minutes at 5000 ×g and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl₂, 5 mM CaCl₂, 30 mM imidazole, 10 % (v/v) glycerol, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 ×g and the clarified supernatant loaded into 1 mL Ni-IDA packed columns (Macherey Nagel). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM imidazole and the protein was eluted in 5 mL elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed and concentrated the using the Amicon Ultra Centrifugal Filters (Merck) with 3.5 mL of dialysis buffer (20 mM HEPES, 50 mM NaCl, 1 mM DTT, 10% (v/v) glycerol) for 30 min at 7000 ×g to reach a final concentration of 97.82 μM (Fig. SIII.3). The protein was added dropwise to liquid nitrogen and stored as pellets at -80 °C.

5.3. Cloning, expression and purification of C-His₆-NocC E373A

C-His₆-NocC E373A was constructed using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's instruction.

PCR reaction was performed on pET-29b-NocC with 1× Q5 Hot Start High-Fidelity, 10 μM of each primer (Table SIII.4), and was carried out under the following conditions: initial denaturation at 98 °C for 30 seconds, followed by 35 cycles of a denaturation step of

98 °C for 10 seconds, annealing at 55 °C for 30 seconds, extension at 72 °C for 4 min and the final extension at 72 °C for 7 min. PCR products were gel purified, ligated with KLD treatment (NewEngland Biolabs), and transformed in *E. coli* TOP10. The obtained constructs were confirmed by Sanger sequencing and then transformed into chemically competent *E. coli* BL21 (DE3) using the heat shock method. For the large-scale purification, cultures containing 100 mL LB medium supplemented with 50 µg mL⁻¹ kanamycin, were grown overnight at 37 °C. Twenty-millilitres of this overnight grown culture were then inoculated into 2 × 2 L of TB medium supplemented with 50 µg mL⁻¹ kanamycin and 5 mM of MgSO₄ at 37 °C, 175 rpm. The cultures were grown to an OD of 0.5-0.6 and then placed at 30 °C, induced with 0.5 mM of IPTG and allowed to grow for 2 h. The cell pellets were recovered by centrifugation (10 min, 3000 ×g), then transferred to 50 mL centrifuge tubes, re-centrifuged for 10 minutes at 5000 ×g and flash frozen in liquid nitrogen. The pellets were resuspended in approximately 30 mL of chilled lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM MgCl₂, 5 mM CaCl₂, 30mM imidazole, 10 % (v/v) glycerol, pH 8.0) supplemented with Pierce Protease Inhibitor Mini Tablets EDTA-free (Thermo Scientific) and the cells homogenized with a pipette. The cell suspensions were lysed in a cell disruptor (CF Range, Constant Systems Ltd) by passing the suspension twice through the high-pressure chamber (20 000 psi), while all the lines were kept chilled. The cell pellets were centrifuged for 30 min at 19 500 ×g and the clarified supernatant loaded into 1mL Ni-IDA packed columns (Macherey Nagel). The column was washed with 5 mL lysis buffer and 5 mL of lysis buffer supplemented with 50 mM imidazole and the protein was eluted in 5mL elution buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol (v/v), 200 mM imidazole pH 8.0). The collected fractions were analyzed by SDS-PAGE and the elution fraction dialyzed and concentrated the using the Amicon Ultra Centrifugal Filters (Merck) with 3.5 mL of dialysis buffer (20 mM HEPES, 50 mM NaCl, 1 mM DTT, 10% (v/v) glycerol) for 30 min at 7000 ×g to reach a final concentration of 119.75 µM (Fig. SIII.3). The protein was added dropwise to liquid nitrogen and stored as pellets at -80 °C.

6. Enzymatic assays

Briefly, holo-ClyB was prepared by incubating 100 µM holo-ClyB co-expressed with ClyA, 1 mM fatty acid (hexanoic acid, *d*₁₁-hexanoic acid, octanoic acid, heptanoic acid, *d*₁₅-octanoic acid, decanoic acid, *d*₁₉-decanoic acid, dodecanoic acid, *d*₂₃-dodecanoic acid, tetradecanoic acid, *d*₂₇-tetradecanoic acid, hexadecanoic acid, *d*₃₁-hexadecanoic acid, octadecanoic, *d*₃₅-octadecanoic acid, 5-hexynoic acid, 6-heptynoic acid, 2-methyldodecanoic acid, 3-methyldodecanoic acid, 7-bromoheptanoic acid, 6-bromohexanoic acid, or 3-hydroxyoctanoic acid), 1 mM ATP, 2 mM MgCl₂ in 50 mM HEPES pH 8.0 for 1 h at room temperature. The sample procedure was applied when preparing the

trans-2-octenoyl-ClyB and *trans*-3-octenoyl-ClyB standards using *trans*-2-octenoic and *trans*-3-octenoic acids.

The mixture was then split into two conditions: a control with no enzyme addition (instead, we used 50 mM HEPES pH 8.0), and a test condition with the addition of 20 μ M NocC and/or NocI. The reactions were incubated in Eppendorf tubes overnight at room temperature, digested with trypsin protease MS grade (Thermo Scientific) and quenched with an equal volume of 25% formic acid (aq). The reactions were centrifuged at 16 000 $\times g$ for 5 min and analyzed by LC-HRMS. The necessary controls with no ATP, no ClyAB and no substrate were prepared simultaneously. For the NocC and NocC E373A boiled controls, the enzymes were denatured at 90 °C for 15 min before adding to the reaction.

The same procedure was applied when testing the mutant NocC E373A.

7. Phylogenetic analysis

BlastP amino acid sequences of the NocC homologues with lowest e-value were retrieved from the NCBI database (34 homologues retrieved). Their amino acidic sequences were aligned, together with different groups of characterized ACAD and their homologues (274) based on previous literature data³ using MUSCLE alignment, from within Geneious R11 (Biomatters). The resulting alignment (with a total of 308 sequences) was trimmed to its core region and contained 590 positions. RaxML-HPC2 on XSEDE was ran in CIPRES⁴ to compute the maximum-likelihood phylogenetic tree, using pseudocounts and 1000 rate categories of sites.

8. Sequence similarity networks and Genomic Neighborhood Diagrams

The sequence similarity network (SSN) was generated using the EFI-EST sever, following a “Sequence BLAST (Basic Local Alignment Search Tool)” of NocC (WP_194003091.1) as input⁵ using negative log e-values of 5 for UniProt BLAST sequence retrieval. The 5000 retrieved sequences and the query sequence were then used to generate the SSNs with an alignment score threshold of 140 and a minimum length of 200. The network was visualized in Cytoscape (v3.90). The full SSN obtained in the previous step was used to generate Genomic Neighborhood Diagrams (GNDs) using the EFI-GNT tool.⁵ A Neighborhood Size of 10 was used and the Lower Limit for Co-occurrence was 20 %. The resulting GNDs were visualized in Cytoscape.

HR-MS: calcd. for $C_{16}H_{23}NO_2S$ - $([M+H]^+)$ 325.3592, found 325.3588

^1H NMR (500 MHz, $CDCl_3$): δ 3.44 (q, J = 6.1 Hz, 2H), 3.02 (t, J = 6.5 Hz, 2H), 2.01 (s, 3H)

10. NocC structural model

NocC sequences were submitted to Alphafold⁷ and the protein models generated following the standard parameters. 3D images of the proteins were obtained in Chimera.

Table SIII.1 - *noc* and *cly* annotation from LEGE 06071 and LEGE 00249

LEGE 06071		LEGE 00249	
NocA	Methylthioribulose-1-phosphate dehydratase (MTA salvage pathway)	NocA	Methylthioribulose-1-phosphate dehydratase (MTA salvage pathway)
NocB	Acireductone Dioxygenase (MTA salvage pathway)	NocB	Acireductone Dioxygenase (MTA salvage pathway)
NocD	Methyltransferase (putative)	NocD	Methyltransferase (putative)
NocE	Monooxygenase (putative)	NocE	Monooxygenase (putative)
NocF	Aminotransferase (putative)	NocF	Aminotransferase (putative)
NocG	Ketosynthase	NocG	Ketosynthase
NocH	Fatty-acyl AMP ligase (FAAL)	NocH	Fatty-acyl AMP ligase (FAAL)
NocI	Acyl-ACP dehydrogenase	NocI	Acyl-ACP dehydrogenase
NocJ	Rieske oxygenase (putative)	NocJ	Rieske oxygenase (putative)
NocK	Hypothetical protein	NocK	Hypothetical protein
ClyA	Fatty-acyl AMP ligase (FAAL)	ClyA	Fatty-acyl AMP ligase (FAAL)
ClyB	ACP	ClyB	ACP
NocC	Acyl-ACP dehydrogenase	ClyC	Halogenase
ClyD	Halogenase	ClyD	Halogenase
ClyE	Polyketide synthase (PKS)	ClyE	Polyketide synthase (PKS)
ClyF	Non-ribosomal peptide synthase (NRPS)	ClyF	Non-ribosomal peptide synthase (NRPS)

Table SIII.2 – Molecular formula and calculated mass of nocuolin A, chlorosphaerolactylates and nocuolactylates ions mentioned in this work

Compound	Molecular formula	ions	Calculated mass
Nocuolin A	C ₁₆ H ₃₀ N ₂ O ₃	[M+H] ⁺	299.2329
		[M+Na] ⁺	321.2149
Nocuolactylate A	C ₃₁ H ₅₃ N ₂ O ₆ Cl	[M+H] ⁺	585.3665
		[M+Na] ⁺	607.3484
d ₁₉ -Nocuolactylate A	C ₃₁ H ₃₄ D ₁₉ N ₂ O ₆ Cl	[M+Na] ⁺	626.46769
Nocuolactylate B	C ₃₁ H ₅₂ N ₂ O ₆ Cl ₂	[M+H] ⁺	619.3275
		[M+Na] ⁺	641.3095
d ₁₈ -Nocuolactylate B	C ₃₁ H ₃₄ D ₁₈ N ₂ O ₆ Cl ₂	[M+Na] ⁺	659.42244
Nocuolactylate C	C ₃₁ H ₅₄ N ₂ O ₆	[M+H] ⁺	551.4055
		[M+Na] ⁺	573.3874
d ₂₀ -Nocuolactylate C	C ₃₁ H ₃₄ D ₂₀ N ₂ O ₆	[M+Na] ⁺	593.51294
Chlorosphaerolactylate A	C ₁₅ H ₂₆ Cl ₂ O ₄	[M-H] ⁻	339.1135
Chlorosphaerolactylate B	C ₁₅ H ₂₇ ClO ₄	[M-H] ⁻	305.1525
Chlorosphaerolactylate E	C ₁₅ H ₂₄ Cl ₂ O ₄	[M-H] ⁻	337.0979
Chlorosphaerolactylate F	C ₁₅ H ₂₅ ClO ₄	[M-H] ⁻	303.1369

Table SIII.3 - Expression conditions used for all proteins purified in this work

Protein	Predicted MW (kDa)	Expression condition					
		plasmid	<i>E. coli</i> strain	Media	Temp.	IPTG/ mM	time
NocC	60.52	pET-29b(+)	BL21(DE3)	LB	30 °C	0.5	2h
NocI	42.18	pET-29b(+)	BL21(DE3)	TB	25 °C	0.5	2h
NocML	61.64	pCDF-Duet1	BAP1	TB	25 °C	0.5	4h
	11.54						
NocC E373A	60.45	pET-29b(+)	BL21(DE3)	LB	30 °C	0.5	2h

Table SIII.4 – Primers used in this work. Green: NdeI restriction site; Red: XhoI restriction site

Gene/ Vector	primer	Sequence	Reference
<i>nocC</i>	Fw	CTTG CATATG CACTTCAATAGTCAAAGC	2
	Rv	TCCG CTCGAG GGCATTATTACAAAG	2
<i>nocI</i>	Fw	GCG CATATG CATTTTTTACACAAAG	2
	Rv	AA CTCGAG AGGCATCATTAAAGG	2
<i>nocML</i>	NocM-Fw	CAGCCATCACCATCATCACCACGGCAGCACGCAACTTTTTGAAGACA AC	2
	NocM-Rv	CAAGCTTGTGACCTGCAGTTATTAGTTTTCTCAGCCACATATTC	2
	NocL-Fw	GTATATTAGTTAAGTATAAGAAGGAGATATACCATGAATGTTTCGTTTG ACCG	2
	NocL-Rv	GTAGACGAGTCCATGTGCTTATTATTGAAGTGCATTAGCTTTTTCAC	2
pCDF Duet-	MCS1-Rv	CTGCAGGTCGACAAGCTTG	2
	MCS1-Fw	GTGGTGATGATGGTGTATGGCTG	2
	MCS2-Rv	GGTATATCTCCTTCTTATACTTAATAATACTAAGATGG	2
NocC E373A	Fw	CCTACTGCAACTTTACAAATGC	This study
	Rv	CCCTTCAAAGATCCGAAAC	This study

Table SIII.5 - BlastP results of NocC against nr database of NCBI (20/12/2022)

Name	Strain	Query cover	E value	Per. Identity	Accession number
Acyl-CoA/Acyl-ACP dehydrogenase	Unclassified <i>Nodularia</i>	100	0	100	WP_194003091.1
NocC	<i>Nodularia</i> sp. HBU26	100	0	97.20	AQX77691.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Thricormus variabilis</i>	100	0	94.58	WP_127052825.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Thricormus</i> sp. NMC-1	100	0	94.21	WP_071187597.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Anabaena lutea</i>	99	0	94.01	WP_190711361.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Sphaerospermopsis aphanizomenoides</i>	100	0	92.15	WP_200322371.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Sphaerospermopsis aphanizomenoides</i>	99	0	88.72	WP_193942045.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Nostoc</i> sp. 232	100	0	82.24	WP_094339716.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Nostoc</i> sp. NOS(2021)	100	0	82.62	WP_265273523.1
Acyl-CoA/Acyl-ACP dehydrogenase	<i>Moorea</i> sp. SIOASIH	97	0	71.46	MBN3895521.1
Acyl-CoA dehydrogenase family protein	<i>Calotrix</i> sp. HK-06	96	0	61.82	WP_073619120.1
Acyl-CoA dehydrogenase family protein	<i>Symplocastrum</i> sp. BBK-W-15	96	2e-180	52.22	WP_254011703.1
TPA AMP-dependent synthetase	Cyanobacteria bacterium UBA11162	95	1e-171	52.44	HBL14898.1
Acyl-CoA dehydrogenase family protein	<i>Sphaerospermopsis aphanizomenoides</i>	97	4e-169	49.33	WP_200322352.1

Table SIII.6 - Accession numbers of all sequences used in the ACADs phylogenetic tree

MBH4410084.1	WP_159247370.1	KZC13622.1	WP_163050860.1
WP_126490718.1	WP_193854954.1	XP_030431709.1	WP_025601786.1
TfB24742.1	WP_073619120.1	tr G1SS41	WP_215828288.1
WP_214554278.1	MBN3895521.1	sp P50544	WP_027895834.1
WP_224259149.1	WP_071187597.1	EH804463.1	WP_202893249.1
WP_224233864.1	AQX77691.1	XP_032471267.1	WP_190149693.1
WP_139541170.1	NocC_LEGE06071	XP_032136424.1	WP_242709836.1
WP_161617674.1	MCB0167207.1	VLCAD_2UXW_1	NEY33908.1
WP_249022280.1	WP_006530792.1	XP_047291887.1	GHD95847.1
WP_258990510.1	NEO71021.1	sp P49748	WP_020732115.1
WP_085448671.1	WP_189200288.1	XP_023369739.1	WP_068164596.1
WP_148048035.1	WP_064410479.1	tr W5PLW6	WP_208821713.1
EGE2519660.1	WP_078360753.1	sp P48818	WP_171200000.1
WP_000973082.1	sp Q3L887	XP_053218192.1	WP_201360692.1
WP_228595108.1	WP_163672684.1	XP_051755842.1	HGB84841.1
EGF8088302.1	WP_065021725.1	XP_041947733.1	AMM72019.1
sp Q47146	sp O53666	KAA0716552.1	WP_012826362.1
EGD4879485.1	WP_012836049.1	tr A0A1Q5L3H9	WP_138665572.1
EAA9194633.1	WP_084986644.1	tr A0A209C2S4	WP_219609722.1
EBQ0990076.1	WP_067673656.1	tr A0A6G3T9N0	WP_154786914.1
WP_156016613.1	WP_150400863.1	tr A0A4S2UT61	WP_067647886.1
EFR4790376.1	WP_195091618.1	tr A0A2P7ZBC5	MBF6467992.1
WP_200074235.1	WP_073446575.1	tr Q9X7Y2	WP_232052297.1
WP_100271470.1	WP_090770938.1	tr A0A7U8UTM1	WP_074336954.1
MBJ3560293.1	WP_092929872.1	tr A0A837GR78	WP_090354976.1
WP_200041904.1	WP_224387550.1	tr A0A6M9XUG7	WP_064856084.1
NMO00952.1	WP_073460335.1	sp Q9UKU7	WP_083071051.1
QKT06249.1	tr A0A383SOQ2	sp Q9D7B6	sp P96855
WP_068970485.1	tr A0A8J3A0J7	tr M1F902	WP_053270253.1
WP_079679238.1	tr A0A4D4JG90	RCL77382.1	WP_020730832.1
WP_085261519.1	tr A0A480AKB9	MBL6767158.1	WP_068276898.1
WP_031673781.1	tr A0A3G9IK33	tr J9DHR3	WP_224098974.1
sp P71858	tr A0A6F8TEU9	VXC53169.1	WP_135124168.1
WP_085234819.1	tr A0A3T1AMK6	WP_020593915.1	WP_135360336.1
WP_036406186.1	tr A0A4Y3QUC5	MBT5157793.1	WP_091510040.1
WP_122444899.1	tr A0A4Y3N6V4	WP_035535899.1	WP_089245281.1
WP_023368411.1	tr A0A4Y3WWN1	MBU2606069.1	WP_024102869.1
WP_135124079.1	tr A0A511DPS5	MBK8198780.1	WP_025431766.1
MCB0942061.1	tr A0A7I7P0U2	WP_035598668.1	WP_049697714.1
WP_232996274.1	tr A0A6N4WF38	tr Q0C3R6	WP_105941326.1
WP_040773330.1	tr A0A511MQ36	VLCAD_2Z1Q_1	WP_251734250.1
WP_067858519.1	tr A0A7I7X2D0	NP_001120800.1	MY586322.1
WP_011206979.1	tr A0A2V0MR89	sp P11310	WP_075928180.1
WP_182357913.1	tr A0A7I7YWG5	MCAD_2A1T_	WP_134714591.1
WP_143772136.1	tr A0A7I9VD55	XP_037857788.1	MB03676547.1
WP_229579872.1	tr A0A2C9TOX9	XP_025240877.1	WP_006350831.1
WP_084326786.1	tr A0A6N4VPK1	XP_005610547.1	AAF86388.1
WP_050657066.1	tr A0A7I7ZIN3	MCAD_1UDY_1	WP_141576713.1
WP_095038805.1	tr A0A810PAD6	XP_045333841.1	WP_093460655.1
WP_048705265.1	MBT2388387.1	XP_032726017.1	WP_095582811.1
WP_048633805.1	WP_229828999.1	WP_193332862.1	WP_114933623.1
WP_206716735.1	SOD43653.1	WP_135228557.1	WP_102934018.1
WP_124721996.1	WP_067044802.1	MCB2185809.1	WP_229236435.1
WP_078632281.1	MBJ8338078.1	SCAD_Pp	sp P33224
WP_176423102.1	WP_024802972.1	MCD5411076.1	MB59312656.1
WP_235685702.1	WP_254750991.1	sp P52042	EAQ9964153.1
QKW94310.1	WP_201876556.1	NLU08464.1	MBJ8681536.1
WP_121280682.1	WP_197987664.1	WP_182088470.1	WP_187259215.1
MBA3772887.1	WP_132431031.1	SCAD_Me	WP_049848606.1
WP_128729722.1	MBA3249200.1	pdb 1BUC A	WP_202770987.1
WP_125250937.1	WP_055551085.1	WP_014017064.1	WP_153531614.1
WP_172872510.1	TcsD	sp Q06319	WP_063061438.1
WP_141404599.1	WP_062330499.1	MCH4178860.1	WP_161925866.1
WP_122835417.1	WP_157251707.1	NLH45428.1_1	WP_163823652.1
WP_076912742.1	MBN6053817.1	WP_144349141.1	WP_153338778.1
WP_060292829.1	WP_019815635.1	NNF45702.1	OBG40251.1
WP_042584692.1	WP_128462586.1	KAB2891741.1	WP_085134792.1
WP_132020641.1	tr U5E8C8	RMF44656.1	WP_174494180.1
WP_175805856.1	tr A0A1B1KD02	WP_121473086.1	PRC51732.1
POM15303.1	sp Q92947	WP_126428940.1	sp Q73ZP8
WP_059678272.1	sp Q2KHZ9	WP_093194752.1	sp P63432
WP_233089864.1	tr A0A7I8BL66	WP_100009860.1	sp P9WQF8
NCQ83073.1	tr G8PY94	WP_050350944.1	WP_053855957.1
WP_185237859.1	tr A0A653B395	tr A0A1Y6JKA6	WP_229477903.1
sp Q8K370	tr A0A250VTK1	tr H8FSA4	sp Q1BBA3
sp Q6JQN1	XP_045477856.1	tr A0A061QDQ5	sp Q5ZHT1
sp B3DMA2	sp Q709F0	tr W5PGK3	sp Q80XL6

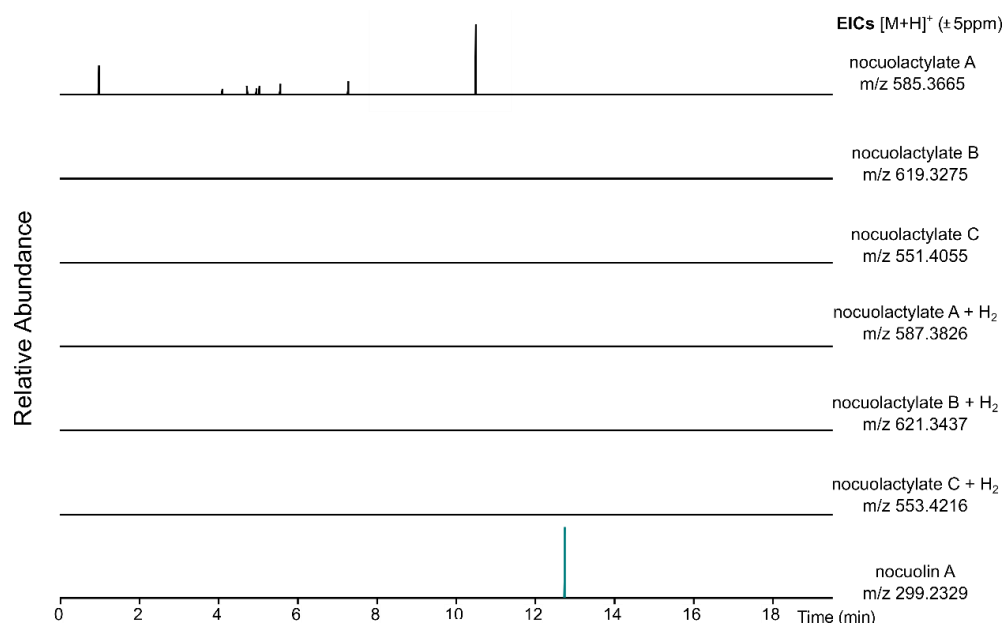


Figure SIII.1- EICs of nocuolin A, nocuolactylates A, B, C and putative saturated versions in the crude extract of *Sphaerospermopsis* LEGE 00249.

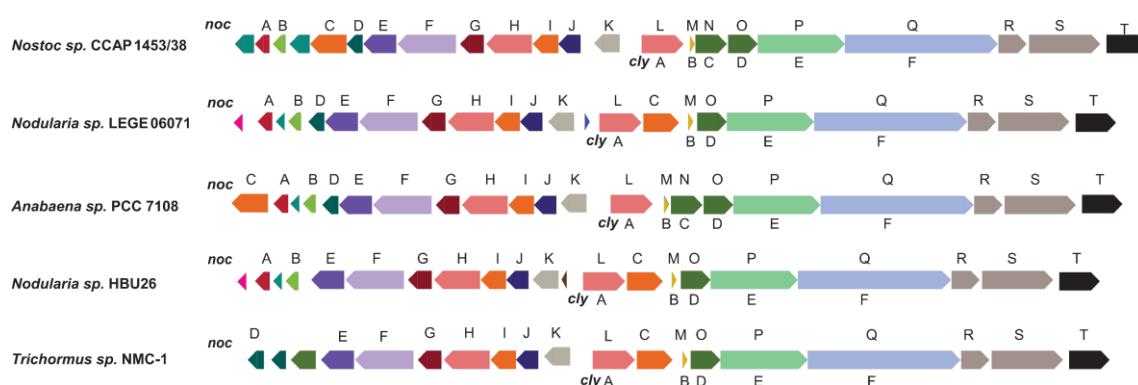


Figure SIII.2 - Schematic representation of the *noc* and *cly* BGCs from five cyanobacterial genomes. NocC (in orange) is the only mobile gene of the cluster, being positioned either on the *noc* or *cly*.

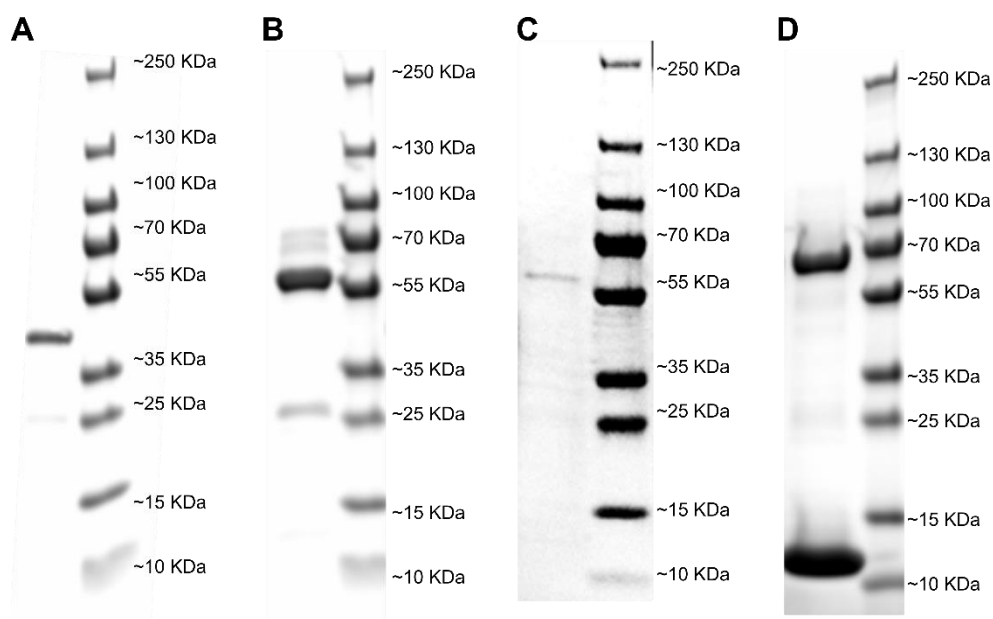


Figure SIII.3— SDS-Page of C-His₆-NocI (A), C-His₆-NocC (B), C-His₆-NocC E373A (C), C-His₆-ClyAB (D) purifications. All proteins were purified by Ni-IDA packed columns (Macherey Nagel).

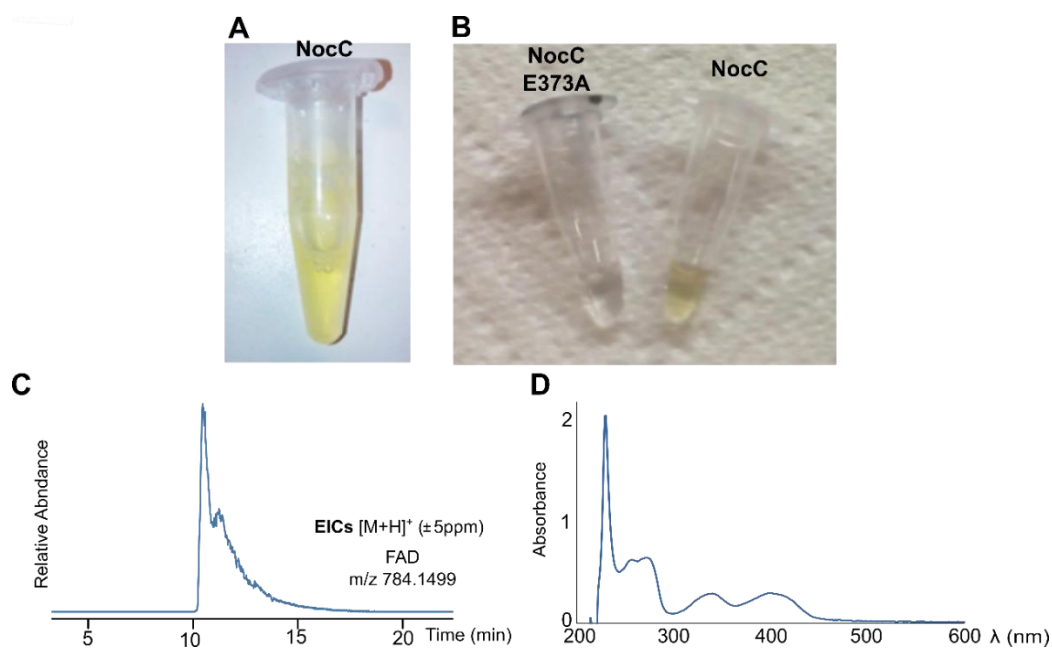


Figure SIII.4 – FAD binding of NocC causes the yellow coloration of the enzyme. (a) C-His₆-NocC after purification. (b) NocC E373A lost its FAD binding ability, losing the typical yellow color of NocC. (c) EICs of FAD in C-His₆-NocC trypsinated sample. (d) UV-Vis spectrum of C-His₆-NocC with a typical FAD-absorbance profile.

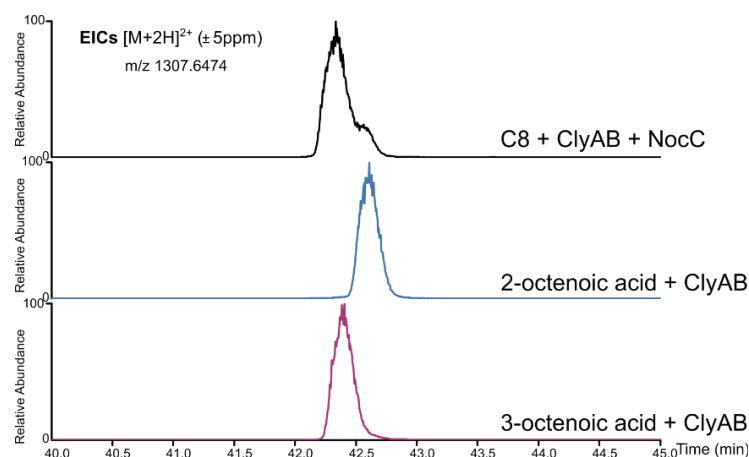


Figure SIII.5 –NocC enzymatic assay peaks compared with *trans*-2-octenoyl-ClyB and *trans* 3-octenoyl-ClyB generated by incubation of the *trans*-2-octenoic and *trans*-3-octenoic acids with ClyAB (zoom in).

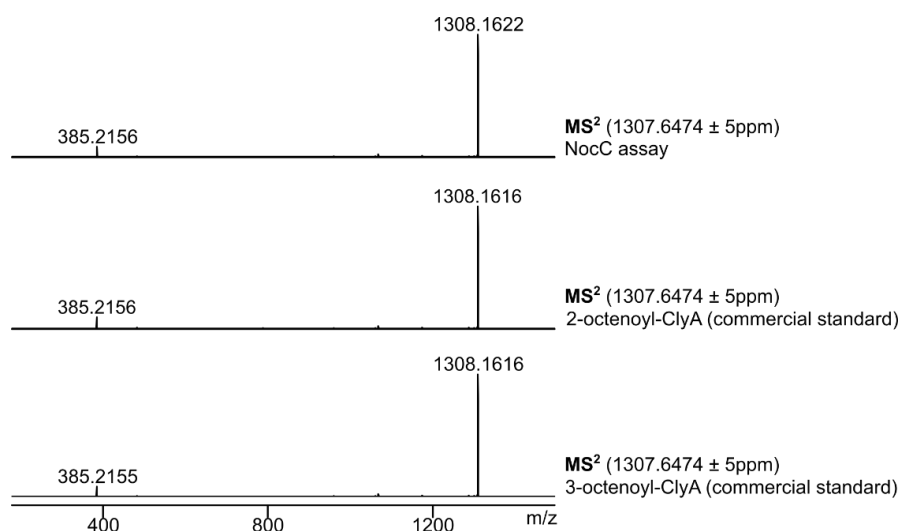


Figure SIII.6 – HR-LCMS² of NocC generated C8-ClyB thioester (*m/z* 1307.6474) and comparison with the *trans*-2-octenoyl-ClyB and *trans*-3-octenoyl-ClyB.

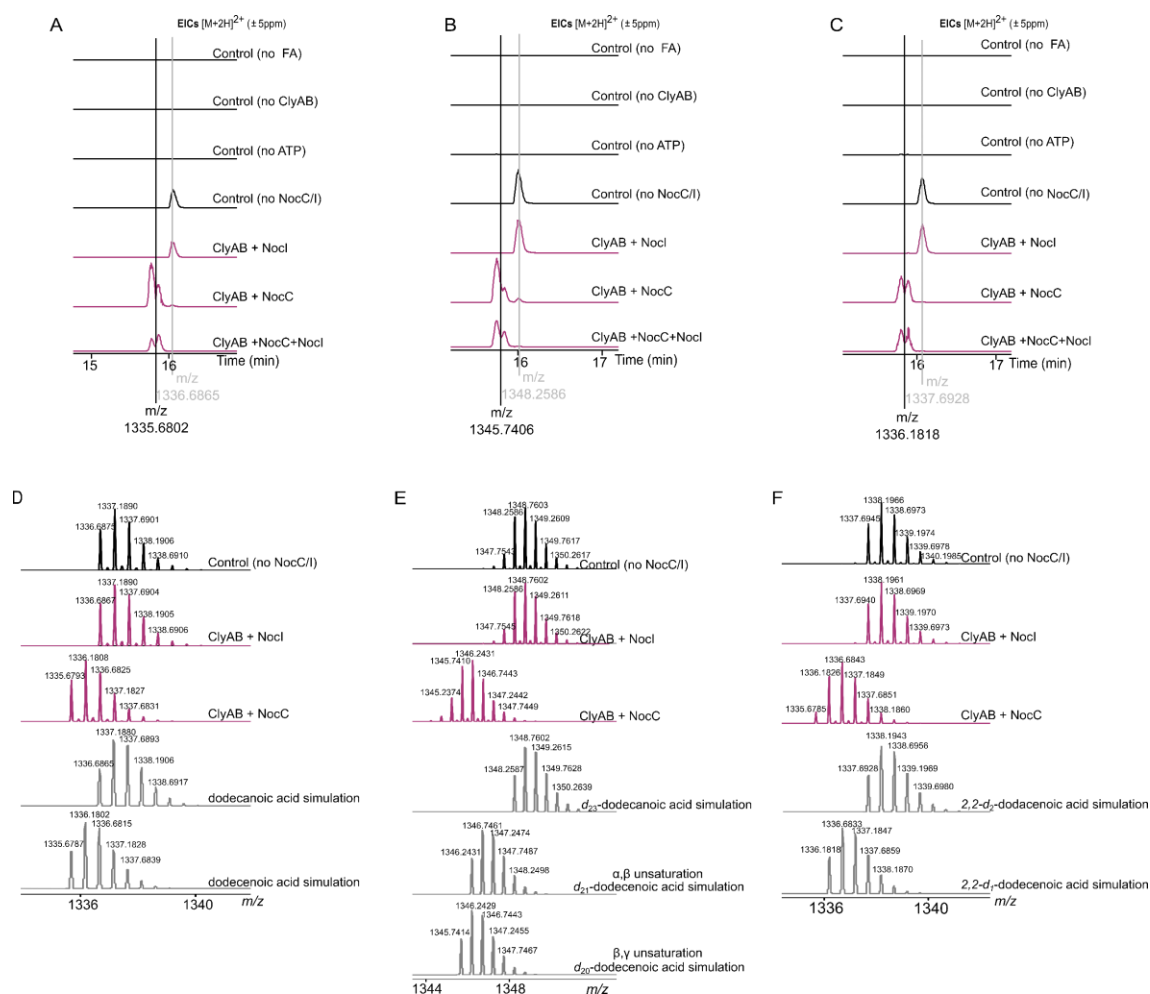


Figure SIII.7 – NocC and NocI enzymatic assays with C12, 2,2- d_2 -C12 and d_{23} -C12. EICs and spectra for C12 assay (A, D), d_{23} -C12 (B, E) and 2,2- d_2 -C12 (C, F).

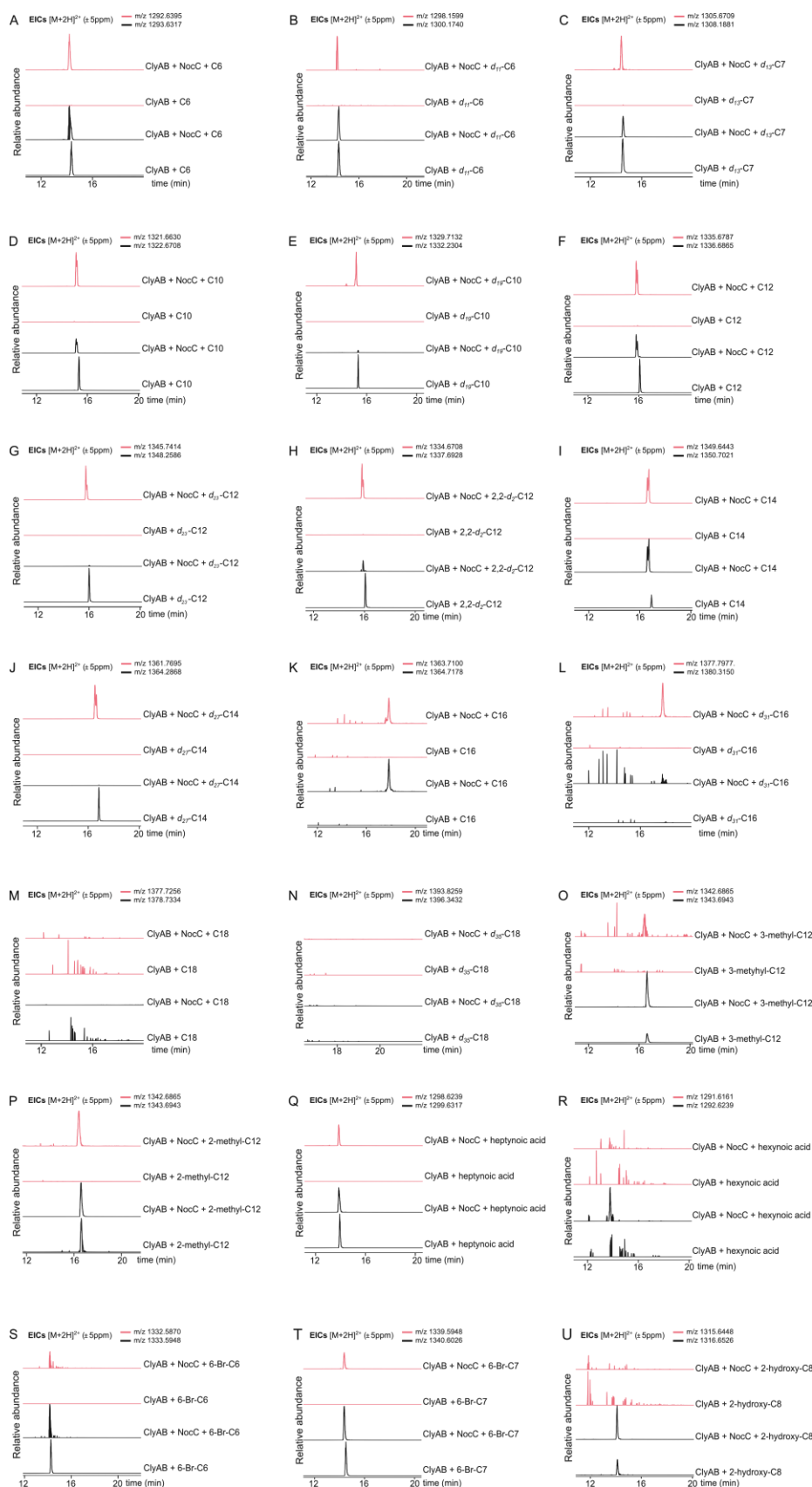
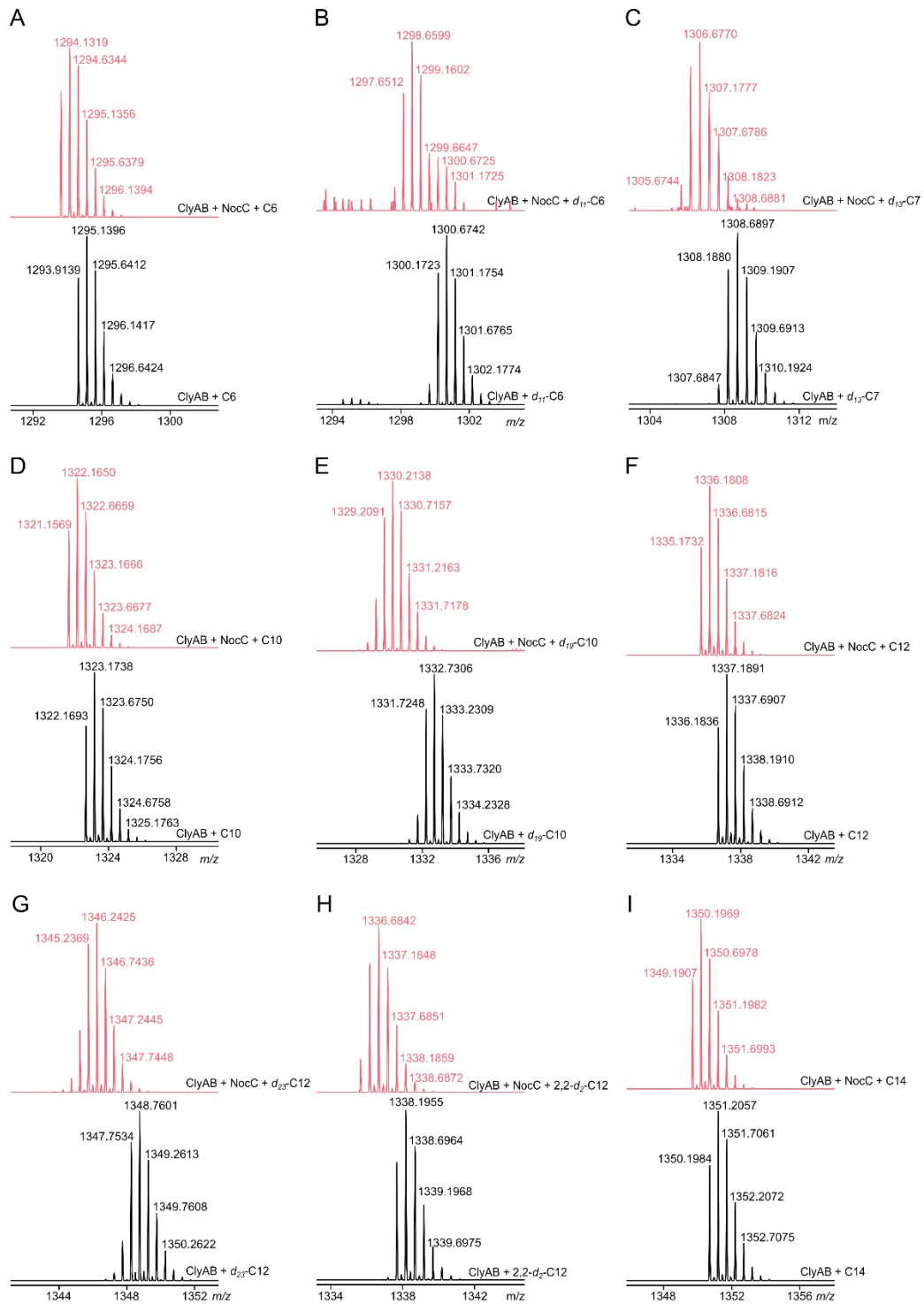


Figure SIII.8 – NocC substrate scope. EICs of the NocC assay with C6 (A), d₁₁-C6 (B), d₁₃-C7 (C), C10 (D), d₁₉-C10 (E), C12 (F), d₂₃-C12 (G), 2,2-d₂-C12 (H), C14 (I), d₂₇-C14 (J), C16 (K), d₃₁-C16 (L), C18 (M), d₃₅-C18 (N), 3-methyl-C12 (O), 2-methyl-C12 (P), 6-heptynoic acid (Q), 5-hexynoic acid (R), 6-bromo-C6 (S), 6-bromo-C7 (T), 2-hydroxy-C8 (U).



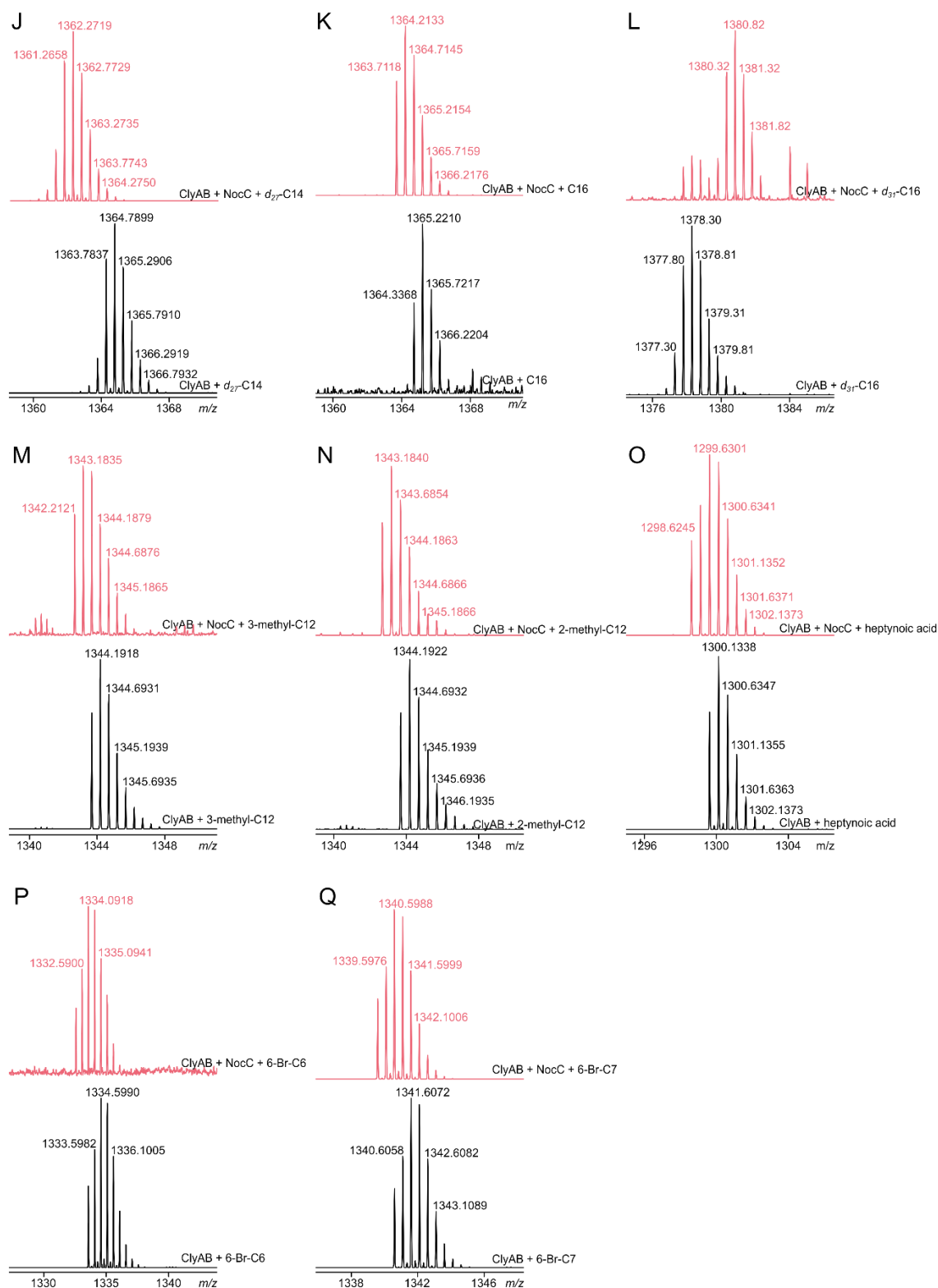


Figure SIII.9 – NocC substrate scope. MS Spectra of NocC assay with C6 (A), d_{11} -C6 (B), d_{13} -C7 (C), C10 (D), d_{19} -C10 (E), C12 (F), d_{23} -C12 (G), 2,2- d_2 -C12 (H), C14 (I), d_{27} -C14 (J), C16 (K), d_{31} -C16 (L), C18 (M), d_{35} -C18 (N), 3-methyl-C12 (O), 2-methyl-C12 (P), 6-heptynoic acid (Q), 5-hexynoic acid (R), 6-bromo-C6 (S), 6-bromo-C7 (T), 2-hydroxy-C8 (U).

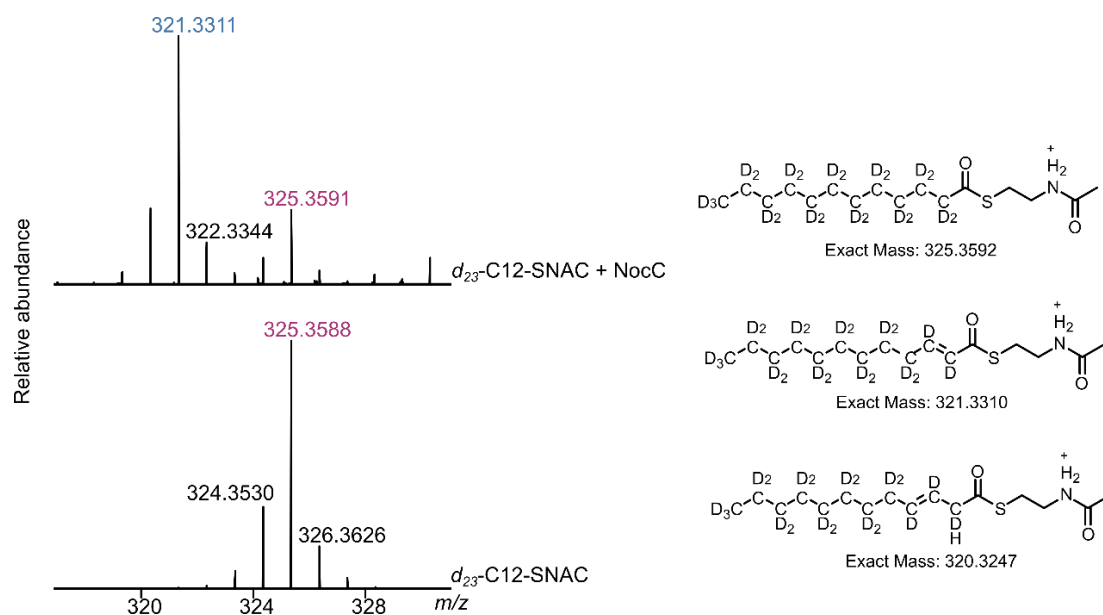


Figure SIII.10 – NocC enzymatic assays with C12-SNAC (A) and d_{23} -C12-SNACs (B). The loss of only 2 deuterium in the d_{23} -C12-SNACs indicated that the enzyme is unable to perform the second reaction step (α,β shifting to β,γ) indicating that the ACP (ClyA) is crucial for the reactivity of the enzyme.

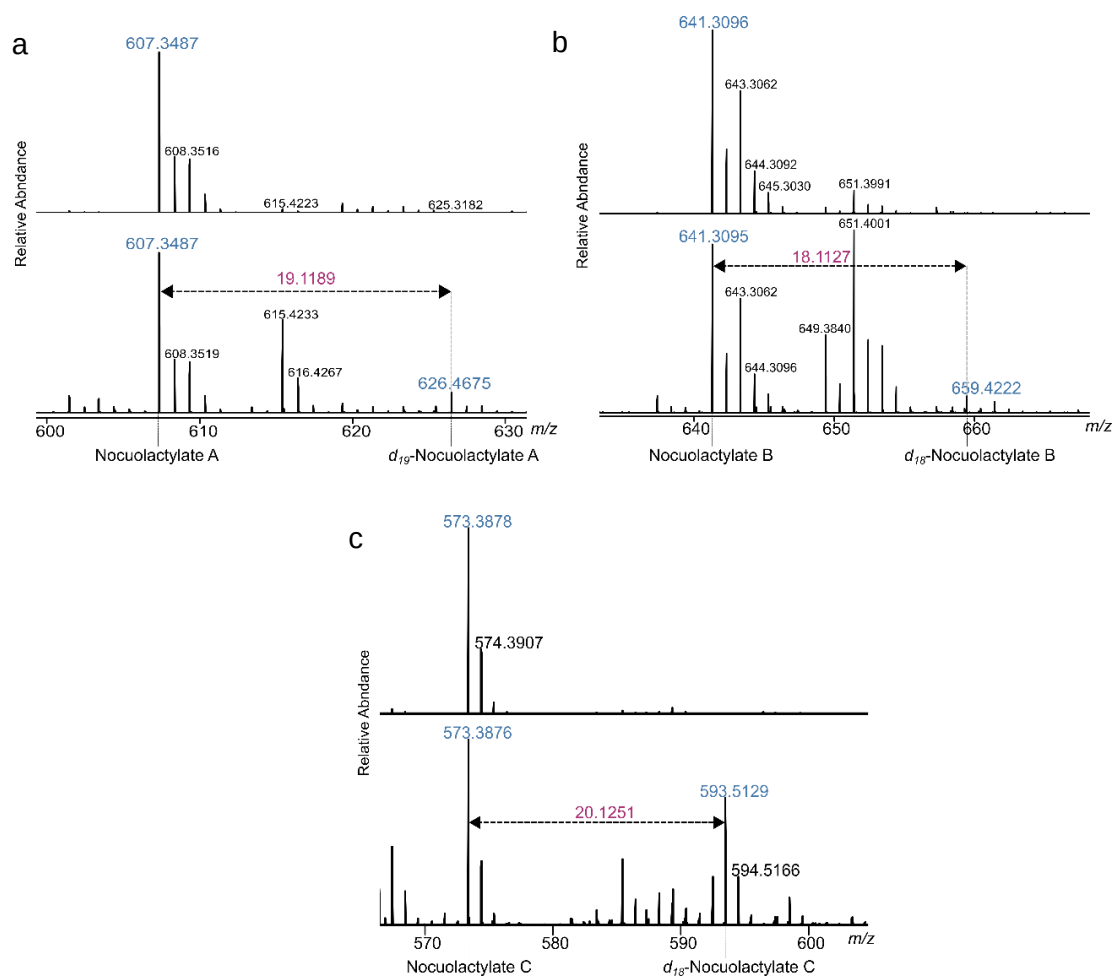


Figure SIII.11 – *Nodularia* sp. LEGE 06071 supplementation experiments with d_{23} -C12, highlighting the labeled peaks of noculactylate A with 19 deuterons (A), noculactylate B with 18 deuterons (B) and noculactylate C with 20 deuterons (C).

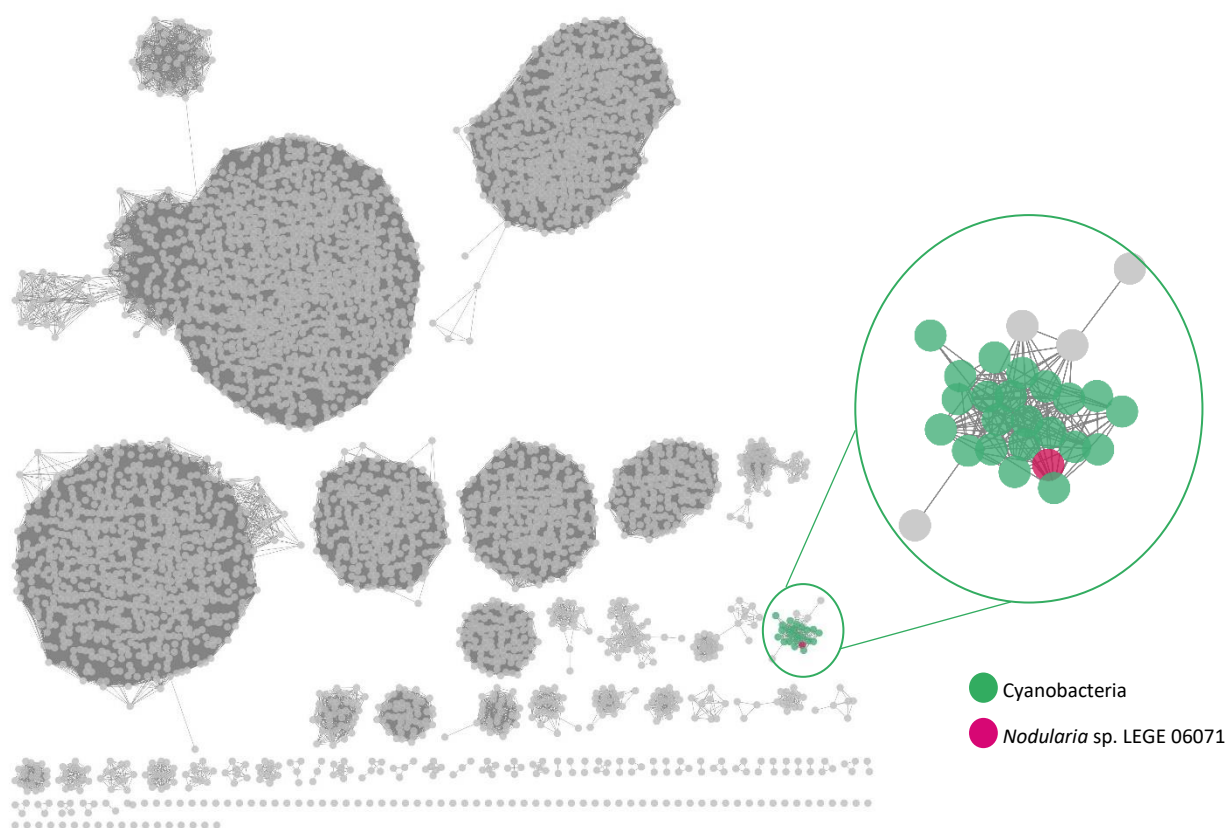


Figure SIII.12 – Sequence similarity network of NocC with the top 5000 BLAST hits (alignment score 140). NocC clusters apart from other ACADs, in a node dominated by cyanobacteria.

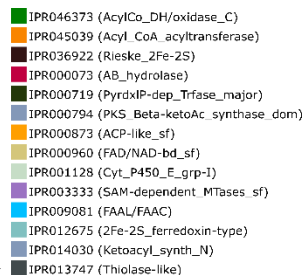
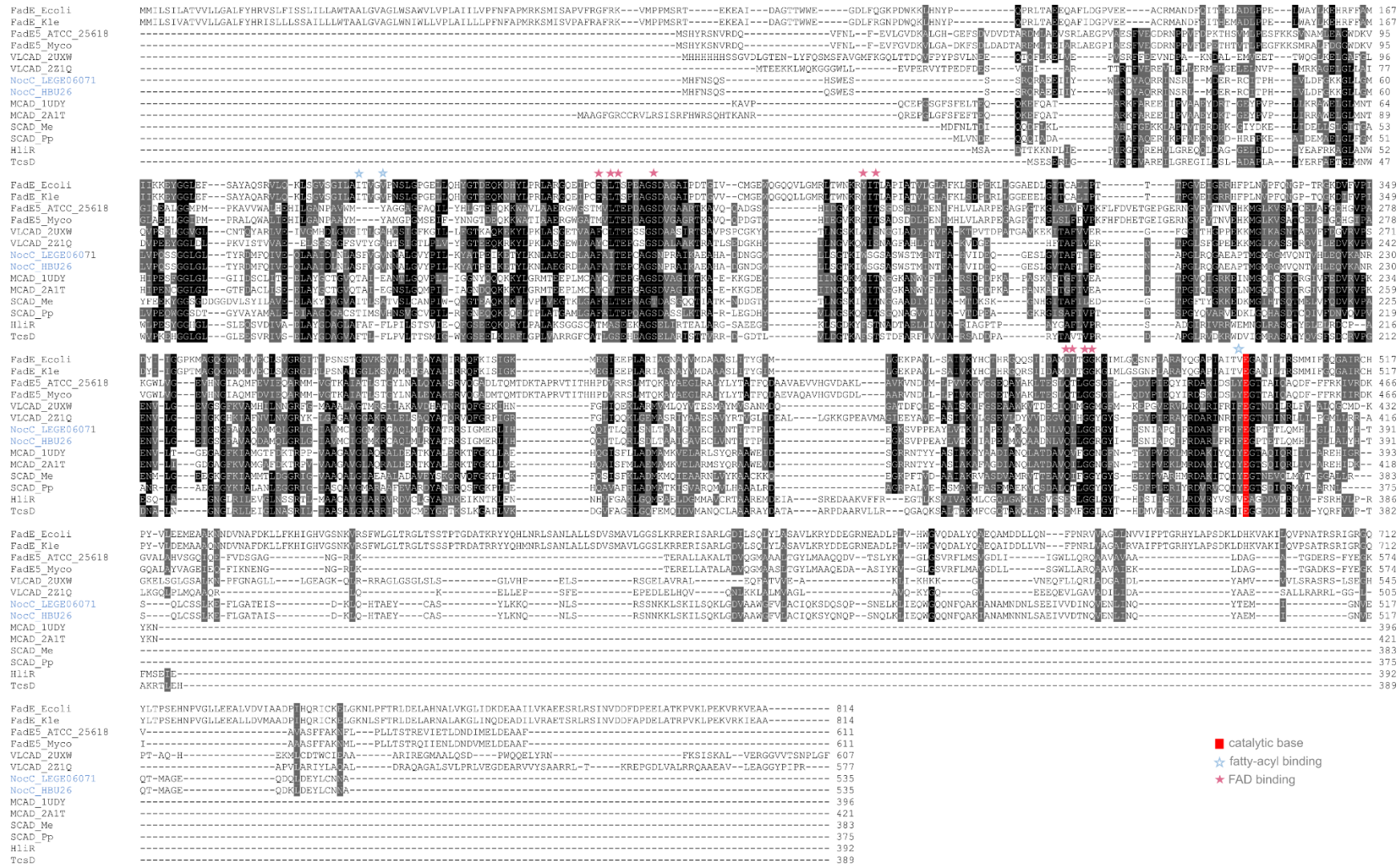


Figure SIII.13 - Diversity and genomic context of NocC-like enzymes (in red) BGCs. Genomic Neighborhood Diagrams (GNDs) showing examples of the different BGC architectures found among the SSN cluster encoding NocC homologues. *noc*-like genes and other biosynthetic genes are highlighted (see legend).



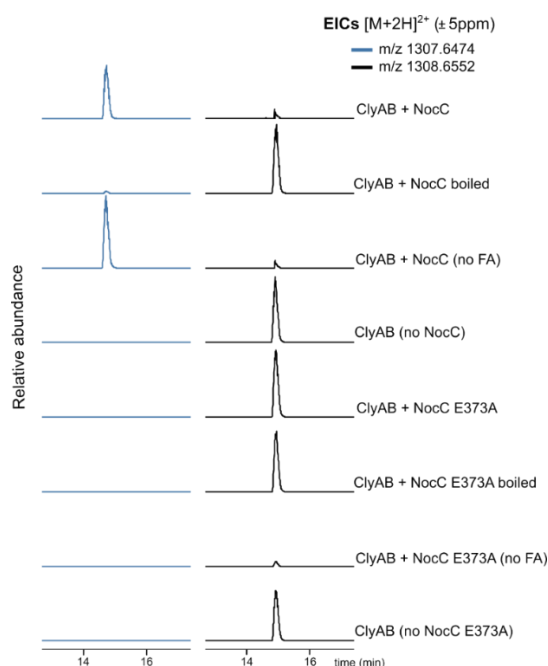


Figure SIII.15 – EICs of NocC and NocC E373A enzymatic assays with octanoic acid. The presence of loaded ClyAB-C8 in the assay lacking C8 is caused by *in vivo* loading of the substrate since ClyAB was expressed in BAP1 *E. coli* which introduce the PPant arm and consequently can load FA directly *in vivo*.

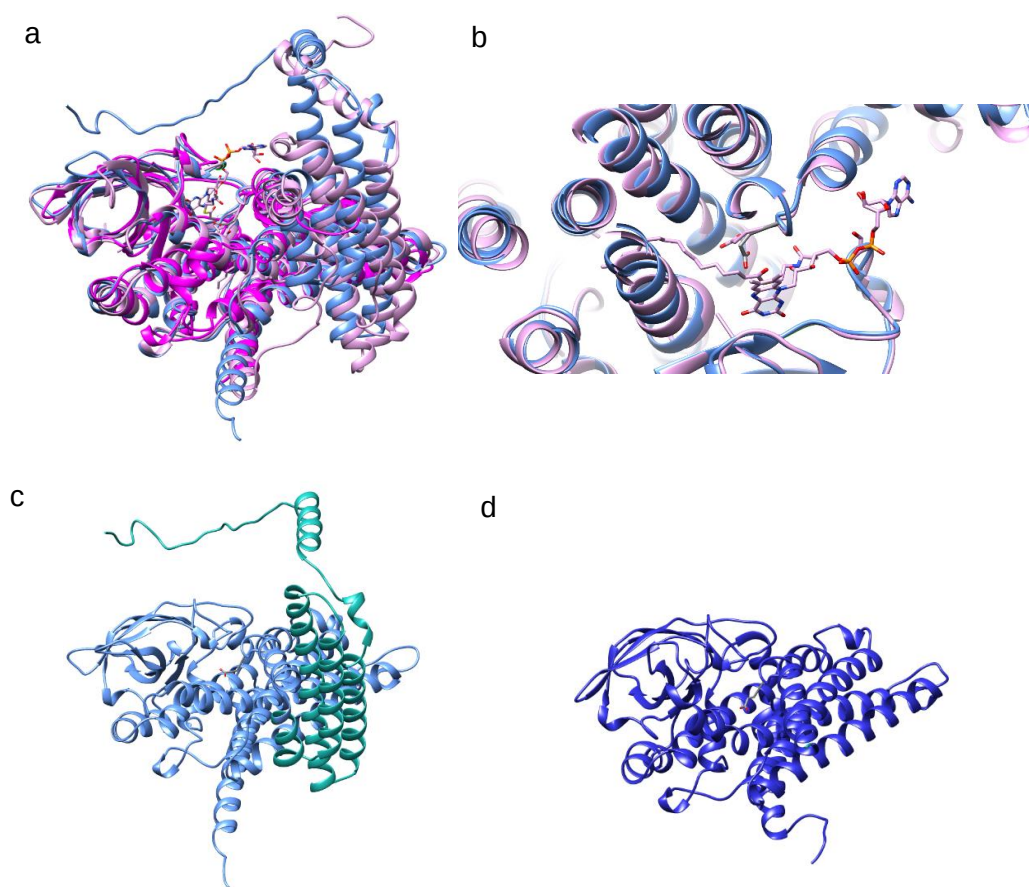


Figure SIII.16 – Structure comparison of TcsD (magenta), VLCAD 2uwX (light pink) and NocC (blue) (A) and detailed active site highlighting the catalytic glutamate (B). AlphaFold model of NocC highlighting the longer C-terminal (green). (C) Crystal structure of a MCAD – 2a1t.

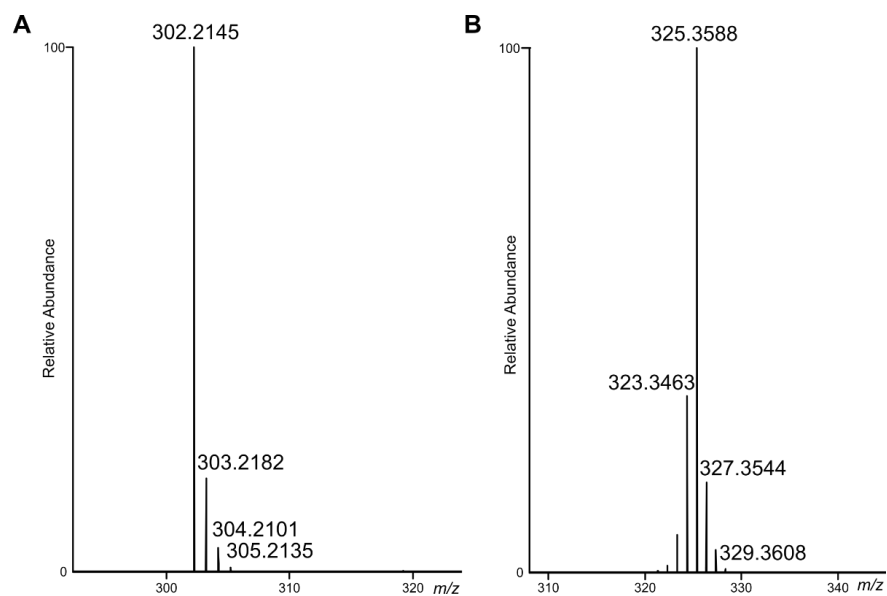


Figure SIII.17 – MS spectrum of the synthesized C12-SNAC (A) and d_{23} -C12-SNAC (B).

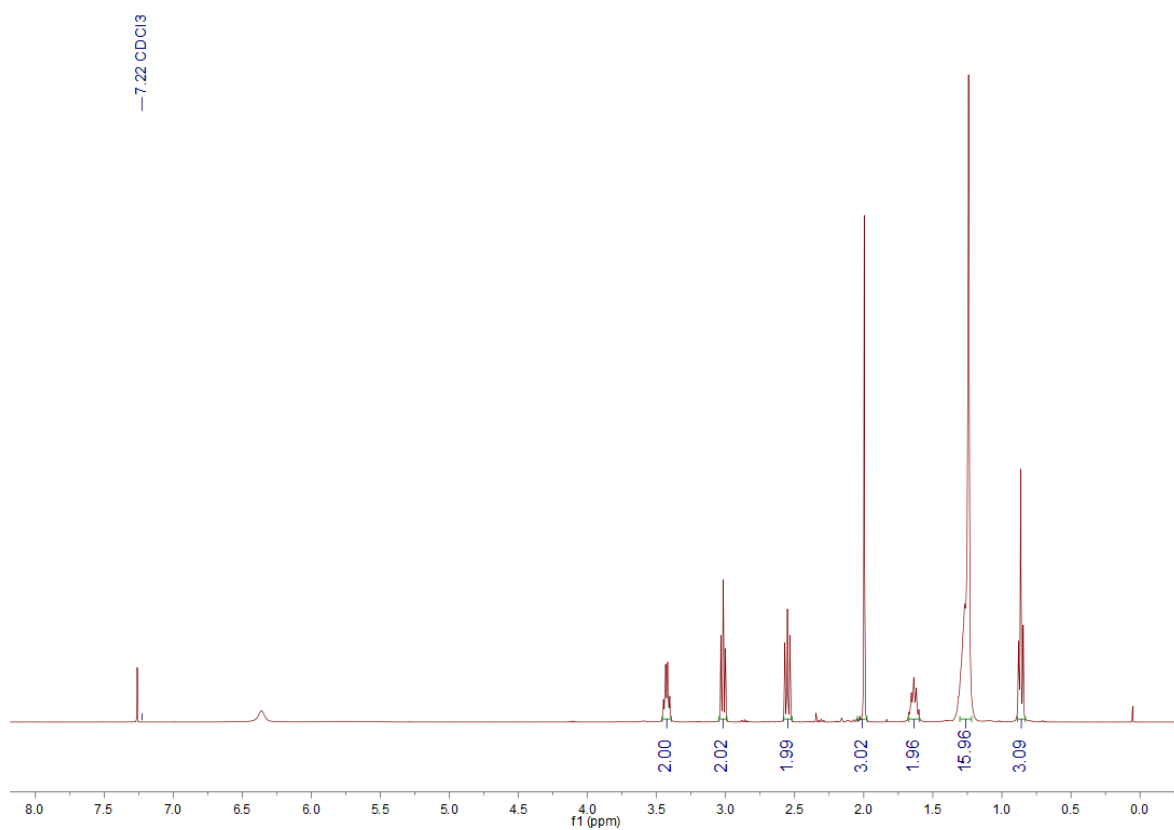


Figure SIII.18 - 400 MHz ^1H spectrum of C12-SNAC (CDCl_3).

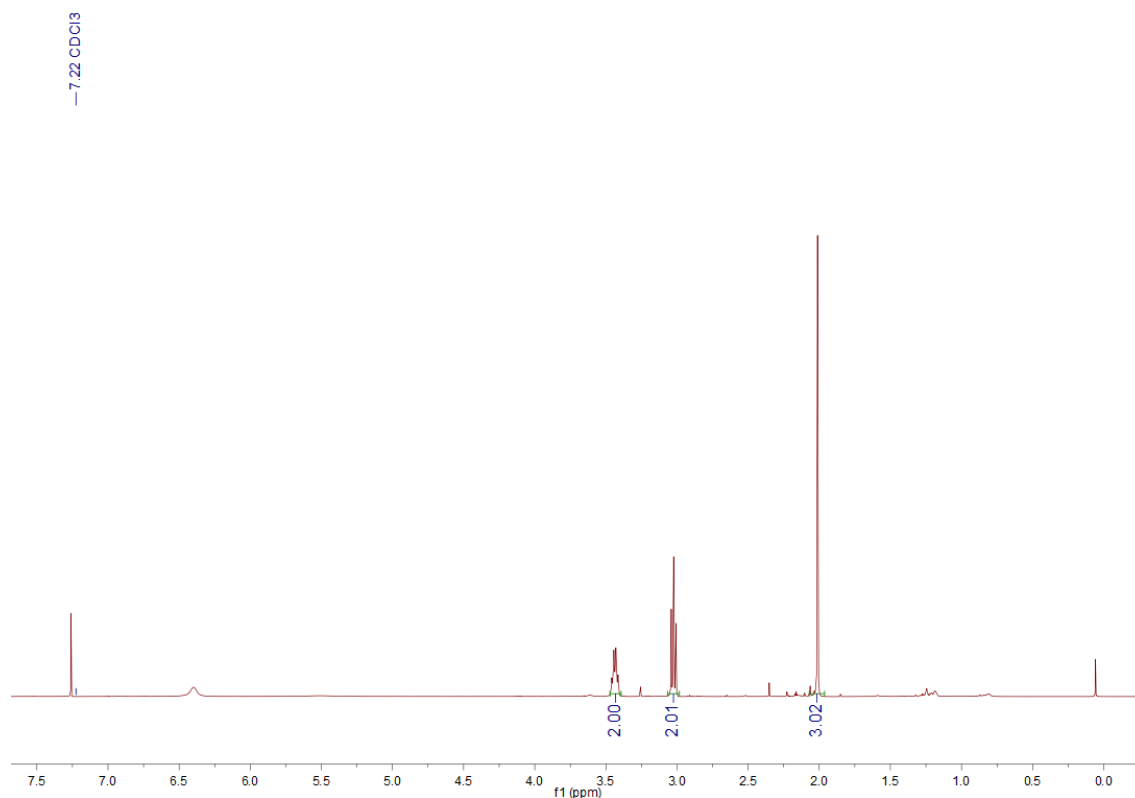


Figure SIII.19 - 400 MHz ^1H spectrum of $\text{d}_{23}\text{-C}_{12}\text{-SNAC}$ (CDCl_3).

References

1. Kotai, J., *Instructions for Preparation of Modified Nutrient Solution Z8 for Algae*. Oslo: Norwegian Institute for Water Research: 1972.
2. Martins, T. P.; Glasser, N. R.; Kountz, D. J.; Oliveira, P.; Balskus, E. P.; Leao, P. N., Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A. *ACS Chem. Biol.* **2022**, 2528-2537.
3. 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.
4. Miller, M. A.; Pfeiffer, W.; Schwartz, T., Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 2010; pp 1-8.
5. Zallot, R.; Oberg, N.; Gerlt, J. A., The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, 58 (41), 4169-4182.
6. Sun, Y.; Feng, Z.; Tomura, T.; Suzuki, A., et al., Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Sci. Rep.* **2016**, 6, 22091-22102.
7. Jumper, J.; Evans, R.; Pritzel, A.; Green, T., et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583-589.

General Discussion and Conclusions

General Discussion and Conclusions

In this work we focused on the characterization of the initial biosynthetic steps in nocuolin A (**1**) biosynthesis and on the study of the origin of an unusual unsaturation during chlorosphaerolactylates biosynthesis.

In Chapter 1, using supplementation experiments with stable isotopes, we identified the three building blocks involved in the biosynthesis of **1**: hexanoic acid, octanoic acid and L-methionine (Fig.1). We proved that both FAs are activated and loaded by NocH, a FAAL/ACP pair, in similar proportions, which contrasts with canonical FAALs that typically show preference for a single free fatty-acid substrate.^{188,189} We showed that the resulting hexanoyl-NocH and octanoyl-NocH thioesters are condensed by NocG, a stand-alone KS, with remarkable selectivity, forming a new C-C bond between carbons 5 and 6 of **1** (Fig. B.1). 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 enzyme. Additionally, despite the ability of NocH to load both substrates and the typical promiscuity of KSs,^{200,281} C₁₁ or C₁₅ products were not detectable in NocG activity assays. NocG performs a non-decarboxylative Claisen condensation reaction, originating a C₁₃ β -ketoacid, which is predicted to spontaneously decarboxylate to 6-tridecanone (Fig. B.1). We show that NocG is part of a phylogenetically well-supported clade whose members are found mainly in cyanobacteria and actinobacteria, with no other characterized members. NocG was thus considered to typify a new class of acylation KSs that use saturated acyl-ACP thioesters as substrates. The phylogenetic clade of NocG was previously considered part of a distant branch of the PpyS clade; the latter enzyme, however, performs a non-decarboxylative Claisen condensation followed by lactonization with high substrate promiscuity, to generate photopyrones.²⁸¹ We hypothesize that NocG catalyzes a canonical KS reaction that is initiated by the deprotonation of the α -carbon of octanoyl-NocH thioester by Glu100, forming a hydrogen bond with the α -carbon of the substrate and creating a nucleophilic species that subsequently attacks the carbonyl carbon of hexanoyl-NocH thioester to form the new C-C bond.²⁸¹

In the same chapter, we proved that L-methionine is involved in the biosynthesis of the 3-hydroxypropanoyl moiety of **1** through an as-yet-unclear mechanism that incorporates C-2 to C-4 and N-1 of this amino acid into **1**, involving a bond cleavage between C-1 and C-2. This suggests that L-methionine, and consequently SAM, transfer the aminopropyl moiety²³¹ (C₄N) to **1** (Fig. B.1). Similar SAM transferences were recently reported by Van Lanen *et al.*²⁸² and Barra *et al.*²⁸³ in the biosynthesis studies of muraymycin and altemicidin, respectively, which highlights the versatile biosynthetic role of this prosthetic group. Still,

none of these reactions can explain the observed incorporation pattern for stable isotope-labeled methionine substrates into **1**, indicating that generation of the 3-hydroxypropanoyl scaffold in **1** involves unprecedented biochemical transformations. Furthermore, NocA was biochemically characterized and NocB was bioinformatically predicted to be a methylthioribulose-1-phosphate dehydratase and acireductone dioxygenase, respectively, both methionine salvage pathway enzymes. This indicates that NocA and NocB play an accessory role in the biosynthesis of **1**, by recycling L-methionine.

Still in chapter 1, we proved that octanoic acid was incorporated into the western side of **1** and retaining one deuterium in position C-4, which could only be reconciled with functionalization through direct C-H bond activation. This was proved in Chapter 2, when we reconstituted two biosynthetic steps following 6-tridecanone formation in the biosynthesis of **1** (Fig. B.1).

In Chapter 2, enzymatic assays showed that 6-tridecanone, the product of NocG activity, was converted to 8-hydroxytridecan-6-one by direct C-H bond activation performed by the flavin-dependent enzyme, NocE (Fig. B.1). NocE performs a C-hydroxylation in an unactivated carbon with no direct connection with nucleophilic groups, to our knowledge unprecedented among FMOs,^{210,284} and lacks close characterized homologues. Furthermore, the predictable hydride transfer catalyzed by FMOs is compatible with the single deuterium loss in the C-4 position of **1**. Following the formation of 8-hydroxytridecan-6-one, we showed that NocF – an aminotransferase/kinase hybrid enzyme – appears to perform an unusual transamination reaction, converting the ketone of 8-hydroxytridecan-6-one into what likely is an imine group, generating 8-iminotrican-6-ol (Fig. 1B). The role of NocF as an aminotransferase was further confirmed by generating *nocF*-null mutant versions of the *noc* pathway in *E. coli*, which were unable to generate nitrogen-containing intermediates. If the structure of NocF's resulting product is confirmed to be an imine (something we were unable to demonstrate in this work), then this enzyme contrasts with the canonical transaminases, which convert ketones into amines.²⁸⁵ Furthermore, it is unclear if the kinase domain of NocF plays a catalytic role in this reaction, or a non-catalytic role related with protein folding²⁸⁶⁻²⁸⁸ (chaperonin-like) or potentially membrane transport.^{289,290} Like NocE, the uniqueness of the aminotransferase domain of NocF was supported by it generating a singleton in the SSN, separately from all the remaining protein family homologues. Despite this, further studies are necessary to confirm and fully characterize NocE and NocF reactivities and reaction mechanisms, namely the role of the kinase domain and the amino donor in NocF.

In the same chapter, we also identified the first analogue of **1**, nocuolin B. This methylated analogue is around 200-400 fold less abundant than **1** and it is produced in several *noc* harboring cyanobacterial strains. We have shown that NocD, a putative

methyltransferase, is likely involved in the methylation of **1** to generate nocuolin B, which goes in line with the absence of *nocD* in some nocuolin A-producing strains (Fig. B.1). However, further studies are required to firmly establish the structure of nocuolin B and the potential role of NocD in its biosynthesis.

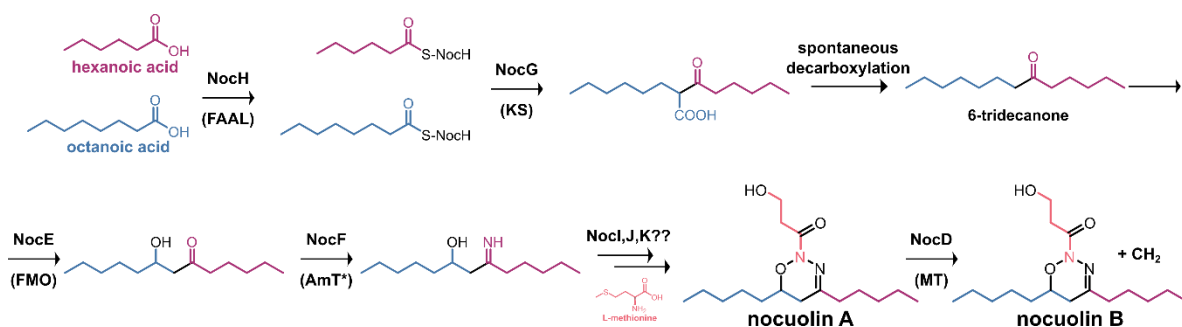


Figure B.1. - Partial biosynthesis of nocuolin A and nocuolin B based on the work of chapters 1 and 2 (AmT*-aminotransferase/kinase domains). The formation of the imine intermediate by NocF will be confirmed in future studies.

Overall, in chapters I and II, we have established the initial biosynthetic steps of nocuolin A biosynthesis, from the initial fatty-acid precursors (C6 and C8) to the putative formation of 8-iminotrican-6-ol (Fig. B.1). We provided the first biochemical evidence connecting the putative *noc* pathway to **1** and revealed several instances of unusual biochemistry leading to a unique molecular scaffold.

We have successfully characterized a new class of KSs (NocG), an unusual FAAL-ACP pair (NocH), a FMO (NocE) and an AmT (NocF). NocG and NocE characterization revoked its previously annotation as acyl-ACP reductase and N-monooxygenase,¹⁴⁷ respectively. Furthermore, we have provided initial proof of the unique reactivities of NocE and NocF and insights into NocD putative activity as a methyltransferase. We have also proved the activities of NocA and NocB as methionine salvage pathway enzymes and NocI as an acyl-ACP dehydrogenase with a currently unknown role. In total, we have expressed and purified 10/10 Noc enzymes and characterized 7/10. Currently, only NocI (acyl-ACP dehydrogenase), NocJ (putative Rieske oxygenase) and NocK (hypothetical protein) have unknown roles in the biosynthesis of **1**. Naturally, we predict that these enzymes might be involved in the formation of the 1,2,3-oxadiazine ring. However, NocI showed acyl-ACP dehydrogenase activity in *in vitro* assays¹⁷⁴ which is hardly reconcilable with the ring formation. Furthermore, due to the restricted number of uncharacterized Noc enzymes and the identification of an *E. coli* *noc* intermediate with a 3-hydroxypropanamide moiety, we hypothesize that the 3-hydroxypropanoyl moiety might be formed in **1** as result of the direct incorporation of 3-hydroxypropanamide. Both NocJ and NocK present sequences that only clade with cyanobacterial homologues (uncharacterized) in SSNs. The closest structural

homologues to NocK are N-methyltransferases and SAM-dependent methyltransferases, indicating a putative role regarding SAM and potentially the formation/incorporation of the 3-hydroxypropanoyl moiety in **1**.

From our studies, the biosynthesis of **1** seems to be performed as follows: first, NocG and NocH perform the C₁₃ alkyl chain formation, then NocE and NocF diversify the alkyl chain with introduction of oxygen and nitrogen atoms and potentially NocI, NocJ and NocK could cooperate in the later biosynthetic steps regarding the formation of the 1,2,3-oxadiazine ring. Additionally, NocA and NocB play an accessory role as methionine salvage pathway enzymes, recycling methionine for further catalytic cycles. In any case, future investigations on the biosynthesis of **1**, namely of the transformations leading to the formation of the 1,2,3-oxadiazine moiety, are now facilitated by the identification of these intermediates and biosynthetic steps.

Finally, in Chapter 3 we focused on an unusual structural motif found in the nocuolin A-related chlorosphaerolactylates and nocuolactylates. Using commercial standards and enzymatic assays with perdeuterated FAs, we showed that the C₁₂ chain β,γ unsaturation found in the nocuolactylates¹⁴⁵ and chlorosphaerolactylates (E and F) results from the activity of a β,γ acyl-ACP dehydrogenase, NocC (Fig. B.2). In chlorosphaerolactylate-producing strains, unsaturated versions are exclusively found when the *cly/noc* BGC contains a *nocC* homologue. NocC is the first enzyme characterized to date to perform this transformation and presents promiscuous activity, accepting a wide range of acyl-ACP thioester substrates (from C8 to C16) including branched and halogenated versions. NocC activity was also confirmed with *in vivo* assays using supplementation experiments of *Nodularia* sp. LEGE 06071 with perdeuterated fatty acids. Several lines of evidence, but in particular the loss of three deuterons in *in vivo* as well as *in vitro* assays, suggested that NocC might first generate an α,β unsaturation that is then shifted to yield a β,γ unsaturated thioester. Alkene shifting from α,β -unsaturation to β,γ -unsaturation has been reported in polyketides like rhizoxin D, ansamitocin P3 and mushroom polyenes,²⁹¹⁻²⁹³ but not on ACADs. NocC acts on acyl-ClyB substrates, previously activated by the FAAL ClyA. *In vitro* assays with *d*₂₃-dodecanoyl-SNAC substrate mimics were unable to fully reconstitute the β,γ unsaturation activity, yielding only the α,β unsaturated acyl thioester intermediate, which indicates a crucial interaction of NocC with ClyB for the putative alkene shift.

A phylogenetic analysis showed that NocC clades mostly with Cyanobacteria, Proteobacteria and Myxobacteria homologues, separately from all other characterized ACADs. Furthermore, NocC presents mixed structural features of both MCADs and VLCADs, namely the catalytic glutamate in position 373 (or similar) from MCADs²⁹⁴ and a

longer C-terminal from VLCADs.²⁹⁵ Regarding the reaction mechanism, presence of the catalytic glutamate, whose role was confirmed by mutagenesis, indicates that the first reaction step of NocC likely follows a typical ACAD concerted mechanism. This is initiated by catalytic glutamate deprotonation of the α C-H hydrogen and transfer of the β C-H hydrogen as a hydride to FAD, but further studies are necessary to fully elucidate this catalytic mechanism, particularly the predicted second reaction step that shifts the unsaturation. The ability of NocC to generate a wide range of β,γ unsaturated products, which can be part of biotechnological relevant compounds such as tambjamines (bipyrrole NPs with antimicrobial, antimalarial and cytotoxic properties), or can easily be derivatized to introduce desirable functionalities, might be particularly interesting for future biocatalytic applications, in particular interfacing with synthetic organic chemistry, where beta-gamma substrates are important building blocks.²⁹⁶

Overall, in chapter III we characterized NocC as the first β,γ acyl-ACP dehydrogenase known to date. NocC introduced β,γ unsaturation in the chlorosphaerolactylates E and F and nocuolactylates and showed considerable promiscuity, which might be promising for biotechnological applications.

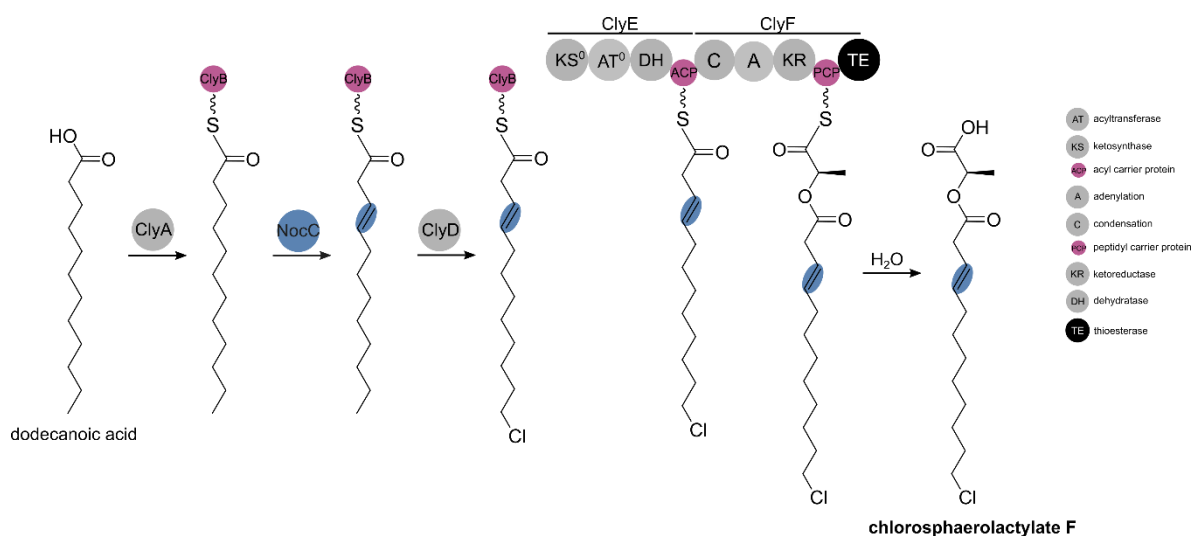


Figure B.2 - Proposed biosynthetic pathway of the β,γ unsaturated chlorosphaerolactylate F in *Nodularia* sp. LEGE 06071, highlighting NocC and the β,γ unsaturation introduced by its activity – in blue (adapted from ²³⁵).

The characterization of new enzymes in this work, particularly new classes of enzymes, represents not only an important addition to the scientific body of knowledge on biochemical transformations, but also a promising starting point for the future development of new biocatalysts.

General References

1. Ramos, V.; Morais, J.; Castelo-Branco, R.; Pinheiro, A., et al., Cyanobacterial diversity held in microbial biological resource centers as a biotechnological asset: the case study of the newly established LEGE culture collection. *J. Appl. Phycol.* **2018**, *30* (3), 1437-1451.
2. Hunter, P., Harnessing Nature's wisdom. Turning to Nature for inspiration and avoiding her follies. *EMBO Rep* **2008**, *9* (9), 838-840.
3. Demain, A. L.; Fang, A., The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* **2000**, *69*, 1-39.
4. Cutler, H. G.; Hill, R. A.; Ward, B. G.; Rohitha, B. H.; Stewart, A., Antimicrobial, insecticidal, and medicinal properties of natural product flavors and fragrances. *Biotechnology for Improved Foods and Flavors* **1996**, *637*, 51-66.
5. Dias, D. A.; Urban, S.; Roessner, U., A historical overview of natural products in drug discovery. *Metabolites* **2012**, *2* (2), 303-336.
6. Cragg, G. M.; Newman, D. J., Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta* **2013**, *1830* (6), 3670-3695.
7. Mann, J., *Murder, magic, and medicine*. Oxford University Press: Oxford, 1992.
8. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629-661.
9. Kinghorn, A. D.; Chin, Y. W.; Swanson, S. M., Discovery of natural product anticancer agents from biodiverse organisms. *Curr Opin Drug Discov Devel* **2009**, *12* (2), 189-196.
10. Fleming, A., On the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference to Their Use in the Isolation of *B. Influenzae*. *Br. J. Exp. Pathol.* **1929**, *10* (3), 226-236.
11. El-ries, M. A.; Abu-el-wafa, S. M.; Aly, F. A.; El-behairy, M. A., Structural Elucidation of the Cu(II)- Streptomycin Complex. Indirect Estimation of Streptomycin by Atomic Absorption Spectroscopy. *Anal. Lett.* **1985**, *18* (15), 1905-1913.
12. Carter, O. L.; Mcphail, A. T.; Sim, G. A., Optically Active Organometallic Compounds .I. Absolute Configuration of (-)-1,1'-Dimethylferrocene-3-Carboxylic Acid by X-Ray Analysis of Its Quinidine Salt. *J Chem Soc A* **1967**, (3), 365-373.
13. Pettit, G. R.; Kamano, Y.; Herald, C. L.; Tuinman, A. A., et al., Antineoplastic Agents .136. The Isolation and Structure of a Remarkable Marine Animal Antineoplastic Constituent - Dolastatin 10. *Journal of the American Chemical Society* **1987**, *109* (22), 6883-6885.
14. Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T., Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **1971**, *93* (9), 2325-2327.
15. Butora, G.; Hudlicky, T., *The Story of Morphine Structure Elucidation: 100 Years of Deductive Reasoning*. JAI: Stamford, CT, 1998; Vol. 4.
16. Dutcher, J. D., The discovery and development of amphotericin B. *Dis. Chest* **1968**, *54*, 296-298.
17. Sehgal, S. N.; Baker, H.; Vezina, C., Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization. *J. Antibiot. (Tokyo)* **1975**, *28* (10), 727-732.
18. Berdy, J., Bioactive microbial metabolites. *J. Antibiot. (Tokyo)* **2005**, *58* (1), 1-26.

19. Walsh, C. T.; Fischbach, M. A., Natural Products Version 2.0: Connecting Genes to Molecules. *Journal of the American Chemical Society* **2010**, *132* (8), 2469-2493.
20. Medema, M. H. e. a., Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015**, (11), 625–631.
21. Clardy, J.; Walsh, C., Lessons from natural molecules. *Nature* **2004**, *432* (7019), 829-837.
22. Parenti, F.; Ciabatti, R.; Cavalleri, B.; Kettenring, J., Ramoplanin: a review of its discovery and its chemistry. *Drugs Exp Clin Res* **1990**, *16* (9), 451-455.
23. Murata, M.; Naoki, H.; Matsunaga, S.; Satake, M.; Yasumoto, T., Structure and Partial Stereochemical Assignments for Maitotoxin, the Most Toxic and Largest Natural Non-Biopolymer. *Journal of the American Chemical Society* **1994**, *116* (16), 7098-7107.
24. Butler, M. S.; Newman, D. J., Mother Nature's gifts to diseases of man: the impact of natural products on anti-infective, anticholesteremics and anticancer drug discovery. *Prog. Drug Res.* **2008**, *65* (1), 3-44.
25. Scott, T. A.; Piel, J., The hidden enzymology of bacterial natural product biosynthesis. *Nat Rev Chem* **2019**, *3* (7), 404-425.
26. Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G., Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (22), 5601-5606.
27. Scherlach, K.; Hertweck, C., Mining and unearthing hidden biosynthetic potential. *Nature Communications* **2021**, *12* (1), 3864-3876.
28. Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M., et al., Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015**, *11* (9), 625-631.
29. Podust, L. M.; Sherman, D. H., Diversity of P450 enzymes in the biosynthesis of natural products. *Nat Prod Rep* **2012**, *29* (10), 1251-1266.
30. Perry, C.; de Los Santos, E. L. C.; Alkhalaf, L. M.; Challis, G. L., Rieske non-heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. *Nat Prod Rep* **2018**, *35* (7), 622-632.
31. Chopra, T.; Banerjee, S.; Gupta, S.; Yadav, G.; Anand, S.; Surolia, A.; Roy, R. P.; Mohanty, D.; Gokhale, R. S., Novel intermolecular iterative mechanism for biosynthesis of mycoketide catalyzed by a bimodular polyketide synthase. *PLoS Biol.* **2008**, *6* (7), e163.
32. Staunton, J.; Wilkinson, B., Biosynthesis of Erythromycin and Rapamycin. *Chem Rev* **1997**, *97* (7), 2611-2630.
33. Della Sala, G.; Mangoni, A.; Costantino, V.; Teta, R., Identification of the Biosynthetic Gene Cluster of Thermoactinoamides and Discovery of New Congeners by Integrated Genome Mining and MS-Based Molecular Networking. *Front Chem* **2020**, *8*, 397-411.
34. Nivina, A.; Yuet, K. P.; Hsu, J.; Khosla, C., Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **2019**, *119* (24), 12524-12547.
35. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of complex polyketides in a metabolically engineered strain of E-coli. *Science* **2001**, *291* (5509), 1790-1792.

36. Maithani, D.; Sharma, A.; Gangola, S.; Chaudhary, P.; Bhatt, P., Insights into applications and strategies for discovery of microbial bioactive metabolites. *Microbiol. Res.* **2022**, 261-279, 127053.
37. Hertweck, C., The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem.* **2009**, 48 (26), 4688-4716.
38. Helfrich, E. J. N.; Piel, J., Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **2016**, 33 (2), 231-316.
39. Yu, D. Y.; Xu, F. C.; Zeng, J.; Zhan, J. X., Type III polyketide synthases in natural product biosynthesis. *lubmb Life* **2012**, 64 (4), 285-295.
40. Eusebio, N.; Rego, A.; Glasser, N. R.; Castelo-Branco, R.; Balskus, E. P.; Leao, P. N., Distribution and diversity of dimetal-carboxylate halogenases in cyanobacteria. *BMC Genomics* **2021**, 22 (1), 633-648.
41. Olano, C.; Mendez, C.; Salas, J. A., Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat. Prod. Rep.* **2010**, 27 (4), 571-616.
42. Sherman, D. H.; Rath, C. M.; Mortison, J.; Scaglione, J. B.; Kittendorf, J. D., Biosynthetic Principles in Marine Natural Product Systems. In *Handbook of Marine Natural Products*, Fattorusso, E.; Gerwick, W. H.; Tagliatela-Scafati, O., Eds. Springer Netherlands: Dordrecht, 2012; pp 947-976.
43. Sussmuth, R. D.; Mainz, A., Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem.* **2017**, 56 (14), 3770-3821.
44. Walsh, C. T.; Chen, H. W.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L. S.; Marshall, C. G.; Miller, D. A.; Patel, H. M., Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **2001**, 5 (5), 525-534.
45. Fisch, K. M., Biosynthesis of natural products by microbial iterative hybrid PKS-NRPS. *Rsc Advances* **2013**, 3 (40), 18228-18247.
46. Roberts, M. F.; Wink, M., Introduction. In *Alkaloids: Biochemistry, Ecology, and Medicinal Applications*, Roberts, M. F.; Wink, M., Eds. Springer US: Boston, MA, 1998; pp 1-7.
47. O'Connor, S., Alkaloid Biosynthesis. *Wiley Encyclopedia of Chem Biol* **2009**, 1 (17), 1-16.
48. Moghaddam, J. A.; Jautzus, T.; Alanjary, M.; Beemelmans, C., Recent highlights of biosynthetic studies on marine natural products. *Org. Biomol. Chem.* **2021**, 19 (1), 123-140.
49. Lichman, B. R., The scaffold-forming steps of plant alkaloid biosynthesis. *Nat. Prod. Rep.* **2021**, 38 (1), 103-129.
50. Katoh, A.; Ohki, H.; Inai, K.; Hashimoto, T., Molecular regulation of nicotine biosynthesis. *Plant Biotechnol.* **2005**, 22 (5), 389-392.
51. Ziegler, J.; Facchini, P. J., Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.* **2008**, 59, 735-769.
52. Rohmer, M., The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **1999**, 16 (5), 565-574.
53. Miziorko, H. M., Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Archives of Biochemistry and Biophysics* **2011**, 505 (2), 131-143.

54. Dickschat, J. S., Bacterial terpene cyclases. *Nat. Prod. Rep.* **2016**, 33 (1), 87-110.
55. Helfrich, E. J. N.; Lin, G. M.; Voigt, C. A.; Clardy, J., Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **2019**, 15, 2889-2906.
56. Zheng, R. R.; Zhu, Z. Y.; Wang, Y. L.; Hu, S. Y., et al., UGT85A84 Catalyzes the Glycosylation of Aromatic Monoterpenes in *Osmanthus fragrans* Lour. Flowers. *Frontiers in Plant Science* **2019**, 10.
57. Awakawa, T.; Zhang, L. H.; Wakimoto, T.; Hoshino, S.; Mori, T.; Ito, T.; Ishikawa, J.; Tanner, M. E.; Abe, I., A Methyltransferase Initiates Terpene Cyclization in Teleocidin B Biosynthesis. *Journal of the American Chemical Society* **2014**, 136 (28), 9910-9913.
58. Bergman, M. E.; Davis, B.; Phillips, M. A., Medically Useful Plant Terpenoids: Biosynthesis, Occurrence, and Mechanism of Action. *Molecules* **2019**, 24 (21), 3961-2984.
59. Nelson, D. R., Cytochrome P450 diversity in the tree of life. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2018**, 1866 (1), 141-154.
60. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A., et al., Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **2013**, 30 (1), 108-160.
61. Montalban-Lopez, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R., et al., New developments in RiPP discovery, enzymology and engineering. *Nat Prod Rep* **2021**, 38 (1), 130-239.
62. Yang, X.; van der Donk, W. A., Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. *Chemistry (Easton)* **2013**, 19 (24), 7662-7677.
63. Duan, Y.; Niu, W.; Pang, L.; Bian, X.; Zhang, Y.; Zhong, G., Unusual Post-Translational Modifications in the Biosynthesis of Lasso Peptides. *Int. J. Mol. Sci.* **2022**, 23 (13), 7231-7248.
64. Mordhorst, S.; Ruijne, F.; Vagstad, A. L.; Kuipers, O. P.; Piel, J., Emulating nonribosomal peptides with ribosomal biosynthetic strategies. *RSC Chem Biol* **2023**, 4 (1), 7-36.
65. Hudson, G. A.; Zhang, Z.; Tietz, J. I.; Mitchell, D. A.; van der Donk, W. A., In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2015**, 137 (51), 16012-16015.
66. Du, Y. L.; Alkhalaf, L. M.; Ryan, K. S., In vitro reconstitution of indolmycin biosynthesis reveals the molecular basis of oxazolinone assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112 (9), 2717-2722.
67. Jankowitsch, F.; Kuhm, C.; Kellner, R.; Kalinowski, J.; Pelzer, S.; Macheroux, P.; Mack, M., A novel N,N-8-amino-8-demethyl-D-riboflavin Dimethyltransferase (RosA) catalyzing the two terminal steps of roseoflavin biosynthesis in *Streptomyces davawensis*. *J. Biol. Chem.* **2011**, 286 (44), 38275-38285.
68. Jhulki, I.; Chanani, P. K.; Abdelwahed, S. H.; Begley, T. P., A Remarkable Oxidative Cascade That Replaces the Riboflavin C8 Methyl with an Amino Group during Roseoflavin Biosynthesis. *Journal of the American Chemical Society* **2016**, 138 (27), 8324-8327.
69. Yee, D. A.; Niwa, K.; Perlatti, B.; Chen, M.; Li, Y.; Tang, Y., Genome mining for unknown-unknown natural products. *Nat. Chem. Biol.* **2023**, 633-640.

70. Zhou, Y.; Wu, S.; Li, Z., Cascade Biocatalysis for Sustainable Asymmetric Synthesis: From Biobased L-Phenylalanine to High-Value Chiral Chemicals. *Angew Chem Int Ed Engl* **2016**, *55* (38), 11647-11650.
71. Wender, P. A.; Miller, B. L., Synthesis at the molecular frontier. *Nature* **2009**, *460* (7252), 197-201.
72. De Regil, R.; Sandoval, G., Biocatalysis for Biobased Chemicals. *Biomolecules* **2013**, *3* (4), 812-847.
73. Sheldon, R. A.; Woodley, J. M., Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2018**, *118* (2), 801-838.
74. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* **2012**, *485* (7397), 185-194.
75. Pyser, J. B.; Chakrabarty, S.; Romero, E. O.; Narayan, A. R. H., State-of-the-Art Biocatalysis. *Acs Central Science* **2021**, *7* (7), 1105-1116.
76. Zetsche, L. E.; Narayan, A. R. H., Broadening the scope of biocatalytic C-C bond formation. *Nature Reviews Chemistry* **2020**, *4* (7), 334-346.
77. Lowell, A. N.; DeMars, M. D., 2nd; Slocum, S. T.; Yu, F., et al., Chemoenzymatic Total Synthesis and Structural Diversification of Tylactone-Based Macrolide Antibiotics through Late-Stage Polyketide Assembly, Tailoring, and C-H Functionalization. *J. Am. Chem. Soc.* **2017**, *139* (23), 7913-7920.
78. Nakamura, H.; Schultz, E. E.; Balskus, E. P., A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **2017**, *13* (8), 916-921.
79. Mazzaferro, L. S.; Huttel, W.; Fries, A.; Muller, M., Cytochrome P450-catalyzed regio- and stereoselective phenol coupling of fungal natural products. *J. Am. Chem. Soc.* **2015**, *137* (38), 12289-12295.
80. Zwick, C. R., 3rd; Renata, H., Evolution of Biocatalytic and Chemocatalytic C-H Functionalization Strategy in the Synthesis of Manzacidin C. *J. Org. Chem.* **2018**, *83* (14), 7407-7415.
81. Neugebauer, M. E.; Sumida, K. H.; Pelton, J. G.; McMurry, J. L.; Marchand, J. A.; Chang, M. C. Y., A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **2019**, *15* (10), 1009-1016.
82. Ng, T. L.; Rohac, R.; Mitchell, A. J.; Boal, A. K.; Balskus, E. P., An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **2019**, *566* (7742), 94-99.
83. Krug, S.; Boch, M.; Daniel, H.; Nimphius, W.; Muller, D.; Michl, P.; Rinke, A.; Gress, T. M., Streptozocin-Based Chemotherapy in Patients with Advanced Neuroendocrine Neoplasms - Predictive and Prognostic Markers for Treatment Stratification. *PLoS One* **2015**, *10* (12), e0143822.
84. Waldman, A. J.; Balskus, E. P., Discovery of a Diazo-Forming Enzyme in Cremeomycin Biosynthesis. *J. Org. Chem.* **2018**, *83* (14), 7539-7546.
85. Young, R. J.; Flitsch, S. L.; Grigalunas, M.; Leeson, P. D.; Quinn, R. J.; Turner, N. J.; Waldmann, H., The Time and Place for Nature in Drug Discovery. *JACS Au* **2022**, *2* (11), 2400-2416.
86. Grigalunas, M.; Brakmann, S.; Waldmann, H., Chemical Evolution of Natural Product Structure. *J. Am. Chem. Soc.* **2022**, *144* (8), 3314-3329.

87. Bornscheuer, U. T.; Buchholz, K., Highlights in biocatalysis - Historical landmarks and current trends. *Eng. Life Sci.* **2005**, *5* (4), 309-323.
88. Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P., et al., Biocatalysis. *Nature Reviews Methods Primers* **2021**, *1* (1).
89. Chakrabarty, S.; Wang, Y.; Perkins, J. C.; Narayan, A. R. H., Scalable biocatalytic C-H oxyfunctionalization reactions. *Chem. Soc. Rev.* **2020**, *49* (22), 8137-8155.
90. Koszeleivski, D.; Lavandera, I.; Clay, D.; Guebitz, G. M.; Rozzell, D.; Kroutil, W., Formal Asymmetric Biocatalytic Reductive Amination. *Angew. Chem.* **2008**, *47* (48), 9337-9340.
91. Akoh, C. C.; Chang, S. W.; Lee, G. C.; Shaw, J. F., Biocatalysis for the Production of Industrial Products and Functional Foods from Rice and Other Agricultural Produce. *Journal of Agricultural and Food Chemistry* **2008**, *56* (22), 10445-10451.
92. Shin, J. S.; Kim, B. G.; Liese, A.; Wandrey, C., Kinetic resolution of chiral amines with omega-transaminase using an enzyme-membrane reactor. *Biotechnology and Bioengineering* **2001**, *73* (3), 179-187.
93. Birmingham, W. R.; Starbird, C. A.; Panosian, T. D.; Nannemann, D. P.; Iverson, T. M.; Bachmann, B. O., Bioretrosynthetic construction of a didanosine biosynthetic pathway. *Nat. Chem. Biol.* **2014**, *10* (5), 392-399.
94. Khan, U.; Selamoglu, Z., Use of Enzymes in Dairy Industry: A Review of Current Progress. *Arch Razi Inst* **2020**, *75* (1), 131-136.
95. Paul, P. E. V.; Sangeetha, V.; Deepika, R. G., Chapter 9 - Emerging Trends in the Industrial Production of Chemical Products by Microorganisms. In *Recent Developments in Applied Microbiology and Biochemistry*, Buddolla, V., Ed. Academic Press: 2019; pp 107-125.
96. Katsimpouras, C.; Stephanopoulos, G., Enzymes in biotechnology: Critical platform technologies for bioprocess development. *Curr. Opin. Biotechnol.* **2021**, *69*, 91-102.
97. Chekan, J. R.; McKinnie, S. M. K.; Moore, M. L.; Poplawski, S. G.; Michael, T. P.; Moore, B. S., Scalable Biosynthesis of the Seaweed Neurochemical, Kainic Acid. *Angew Chem Int Ed Engl* **2019**, *58* (25), 8454-8457.
98. Kino, K.; Furuya, T. Production of (-)-rotundone from α -guaiene catalyzed by cytochrome P450 enzymes. EP3255151A2, 13 December 2017, 2017.
99. Kelly, S. A.; Pohle, S.; Wharry, S.; Mix, S.; Allen, C. C. R.; Moody, T. S.; Gilmore, B. F., Application of omega-Transaminases in the Pharmaceutical Industry. *Chem Rev* **2018**, *118* (1), 349-367.
100. Maria, P. D.; Gonzalo, G.; Alcantara, A. R., Biocatalysis as Useful Tool in Asymmetric Synthesis: An Assessment of Recently Granted Patents (2014-2019). *Catalysts* **2019**, *9* (10), 802-844.
101. Debarge, S.; Erdman, D. T.; O'Neill, P. M.; Kumar, R.; Karmilowicz, M. J. Process and Intermediates for the Preparation of Pregabalin. WO2014155291A1, 2 October 2014, 2014.
102. Occhipinti, A.; Berlicki, L.; Giberti, S.; Dziedziola, G.; Kafarski, P.; Forlani, G., Effectiveness and mode of action of phosphonate inhibitors of plant glutamine synthetase. *Pest Manage. Sci.* **2010**, *66* (1), 51-58.

103. Green, B. M.; Gradley, M. L. Two-step Enzymic Synthesis of L-glufosinate from Racemic Mixture. US20170253897A1, 7 September 2017, 2017.
104. Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B., Industrial biocatalysis today and tomorrow. *Nature* **2001**, 409 (6817), 258-268.
105. de Regil, R.; Sandoval, G., Biocatalysis for biobased chemicals. *Biomolecules* **2013**, 3 (4), 812-47.
106. Panke, S.; Wubbolts, M., Advances in biocatalytic synthesis of pharmaceutical intermediates. *Curr. Opin. Chem. Biol.* **2005**, 9 (2), 188-194.
107. Hama, S.; Noda, H.; Kondo, A., How lipase technology contributes to evolution of biodiesel production using multiple feedstocks. *Curr. Opin. Biotechnol.* **2018**, 50, 57-64.
108. Feng, S.; Hao Ngo, H.; Guo, W.; Woong Chang, S.; Duc Nguyen, D.; Cheng, D.; Varjani, S.; Lei, Z.; Liu, Y., Roles and applications of enzymes for resistant pollutants removal in wastewater treatment. *Bioresour. Technol.* **2021**, 335, 125278-125290.
109. Castenholz R.W., B. D. R., *The Archae and the Deeply Branching and Phototrophic Bacteria*. Bergey's Springer: New York, 2001.
110. Oren, A., *Cyanobacteria : an economic perspective*. 1st edition ed.; John Wiley&Sons, Ltd.: 2014.
111. Brocks, J. J.; Logan, G. A.; Buick, R.; Summons, R. E., Archean molecular fossils and the early rise of eukaryotes. *Science* **1999**, 285 (5430), 1033-1036.
112. Hoashi, M.; Bevacqua, D. C.; Otake, T.; Watanabe, Y.; Hickman, A. H.; Utsunomiya, S.; Ohmoto, H., Evidence of the early evolution of oxygenic photoautotrophs in 3.46 Billion-year-old haematite crystals from Pilbara, Western Australia. *Nature Geoscience* **2009**, 2, 301–306.
113. Schopf J. W.; Hayes J. M.; R., W. M., *Evolution of Earth's earliest ecosystems: recent progress and unsolved problems*. In *Earth's Earliest Biosphere: its Origin and Evolution*. Princeton University Press: 1983; p 361-384.
114. Falcon, L. I.; Magallon, S.; Castillo, A., Dating the cyanobacterial ancestor of the chloroplast. *ISME J* **2010**, 4 (6), 777-783.
115. Kasting, J. F.; Siefert, J. L., Life and the evolution of Earth's atmosphere. *Science* **2002**, 296 (5570), 1066-1068.
116. Fay, P.; Walsby, A. E., Metabolic activities of isolated heterocysts of the blue-green alga *Anabaena cylindrica*. *Nature* **1966**, 209 (5018), 94-95.
117. Livingstone, D.; Jaworski, G. H. M., The viability of akinetes of blue-green algae recovered from the sediments of Rostherne Mere. *British Phycology Journal* **1980**, 15, 357–364.
118. Rippka, R.; DeReuelles, J.; Waterbury, J. B.; Herdman, M.; Stanier, R. Y., Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Journal of General Microbiology* **1979**, 111, 1-61.
119. Dvořák, P.; Casamatta, D. A.; Hašler, P.; Jahodářová, E.; Norwich, A. R.; Pouličková, A., Diversity of the Cyanobacteria. In *Modern Topics in the Phototrophic Prokaryotes*, Hallenbeck, P. C., Ed. Cham: Springer: 2017; pp 3-46.

120. Saini, D. K.; Pabbi, S.; Shukla, P., Cyanobacterial pigments: Perspectives and biotechnological approaches. *Food Chem Toxicol* **2018**, *120*, 616-624.
121. Pointing, S. B.; Chan, Y.; Lacap, D. C.; Lau, M. C.; Jurgens, J. A.; Farrell, R. L., Highly specialized microbial diversity in hyper-arid polar desert. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (47), 19964-19969.
122. Friedmann, E. I.; Ocampo, R., Endolithic blue-green algae in the dry valleys: primary producers in the antarctic desert ecosystem. *Science* **1976**, *193* (4259), 1247-1249.
123. Oren, A., Cyanobacteria: biology, ecology and evolution. In *Cyanobacteria: An Economic Perspective*, LJS, N. K. S. A. a., Ed. John Wiley&Sons, Ltd.: 2014; pp 3-20.
124. 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.
125. Sivonen, K.; Leikoski, N.; Fewer, D. P.; Jokela, J., Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol* **2010**, *86* (5), 1213-1225.
126. Wiegand, C.; Pflugmacher, S., Ecotoxicological effects of selected cyanobacterial secondary metabolites: a short review. *Toxicol Appl Pharmacol* **2005**, *203* (3), 201-218.
127. Martins, T. P.; Arsin, S.; Fewer, D. P.; Leão, P. N., UV-protective secondary metabolites from cyanobacteria. In *The Pharmacological Potential of Cyanobacteria*, Lopes, G.; Silva, M.; Vasconcelos, V., Eds. 2022; pp 107-144.
128. Nozzi, N. E.; Oliver, J. W.; Atsumi, S., Cyanobacteria as a Platform for Biofuel Production. *Front Bioeng Biotechnol* **2013**, *1*, 7-13.
129. Karkos, P. D.; Leong, S. C.; Karkos, C. D.; Sivaji, N.; Assimakopoulos, D. A., Spirulina in clinical practice: evidence-based human applications. *Evid. Based Complement. Alternat. Med.* **2011**, *2011*, 531053-531057.
130. Agarwal, P.; Soni, R.; Kaur, P.; Madan, A.; Mishra, R.; Pandey, J.; Singh, S.; Singh, G., Cyanobacteria as a Promising Alternative for Sustainable Environment: Synthesis of Biofuel and Biodegradable Plastics. *Front. Microbiol.* **2022**, *13*, 939347-939372.
131. Jones, A. C.; Monroe, E. A.; Eisman, E. B.; Gerwick, L.; Sherman, D. H.; Gerwick, W. H., The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat. Prod. Rep.* **2010**, *27* (7), 1048-1065.
132. Dixit, R. B.; Suseela, M. R., Cyanobacteria: potential candidates for drug discovery. *Antonie Van Leeuwenhoek* **2013**, *103* (5), 947-961.
133. Tan, L. T., Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry* **2007**, *68* (7), 954-979.
134. Gerwick, W. H.; Moore, B. S., Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem Biol* **2012**, *19* (1), 85-98.
135. Luesch, H.; Moore, R. E.; Paul, V. J.; Mooberry, S. L.; Corbett, T. H., Isolation of dolastatin 10 from the marine cyanobacterium *Symploca* species VP642 and total stereochemistry and biological evaluation of its analogue symprostatin 1. *J. Nat. Prod.* **2001**, *64* (7), 907-910.

136. 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.
137. Jones, A. C.; Gu, L.; Sorrels, C. M.; Sherman, D. H.; Gerwick, W. H., New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. *Curr. Opin. Chem. Biol.* **2009**, *13* (2), 216-223.
138. Dahiya, R.; Dahiya, S.; Fuloria, N. K.; Jankie, S., et al., Natural Thiazoline-Based Cyclodepsipeptides from Marine Cyanobacteria: Chemistry, Bioefficiency and Clinical Aspects. *Curr. Med. Chem.* **2021**, *28* (38), 7887-7909.
139. Leao, P. N.; Nakamura, H.; Costa, M.; Pereira, A. R.; Martins, R.; Vasconcelos, V.; Gerwick, W. H.; Balskus, E. P., Biosynthesis-Assisted Structural Elucidation of the Bartolosides, Chlorinated Aromatic Glycolipids from Cyanobacteria. *Angew. Chem.* **2015**, *5* (48), 14895-14895.
140. Nagatsu, A.; Kajitani, H.; Sakakibara, J., Muscoride A: A new oxazole peptide alkaloid from freshwater cyanobacterium *Nostoc muscorum*. *Tetrahedron Lett.* **1995**, *36* (23), 4097-4100.
141. Martins, T. P.; Rouger, C.; Glasser, N. R.; Freitas, S.; de Fraissinette, N. B.; Balskus, E. P.; Tasdemir, D.; Leao, P. N., Chemistry, bioactivity and biosynthesis of cyanobacterial alkylresorcinols. *Nat Prod Rep* **2019**, *36* (10), 1437-1461.
142. Leao, P. N.; Martins, T. P.; Abt, K.; Reis, J. P. A.; Figueiredo, S.; Castelo-Branco, R.; Freitas, S., Incorporation and modification of fatty acids in cyanobacterial natural products biosynthesis. *Chem Commun (Camb)* **2023**.
143. Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E., Structures of cylindrocycphanes a-f. *Tetrahedron Lett.* **1992**, *48* (15), 3001-3006.
144. Stratmann, K.; Burgoyne, D. L.; Moore, R. E.; Patterson, G. M. L.; Smith, C. D., Hapalosin, a Cyanobacterial Cyclic Depsipeptide with Multidrug-Resistance Reversing Activity. *J. Org. Chem.* **1994**, *59* (24), 7219-7226.
145. 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.
146. Beresovsky, D.; Hadas, O.; Livne, A.; Sukenik, A.; Kaplan, A.; Carmeli, S., Toxins and biologically active secondary metabolites of *Microcystis* sp isolated from Lake Kinneret. *Isr. J. Chem.* **2006**, *46* (1), 79-87.
147. Voráčová, K.; Hájek, J.; Mareš, J.; Urajová, P., et al., The cyanobacterial metabolite nocuolin a is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLoS One* **2017**, *12* (3), e0172850.
148. Gutierrez-Del-Rio, I.; Brugerolle de Fraissinette, N.; Castelo-Branco, R.; Oliveira, F., et al., Chlorosphaerolactylates A-D: Natural Lactylates of Chlorinated Fatty Acids Isolated from the Cyanobacterium *Sphaerospermopsis* sp. LEGE 00249. *J. Nat. Prod.* **2020**, *83* (6), 1885-1890.
149. Sitachitta, N.; Gerwick, W. H., Grenadadiene and grenadamide, cyclopropyl-containing fatty acid metabolites from the marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **1998**, *61* (5), 681-684.

150. Moosmann, P.; Ueoka, R.; Gugger, M.; Piel, J., Aranazoles: Extensively Chlorinated Nonribosomal Peptide-Polyketide Hybrids from the Cyanobacterium *Fischerella* sp PCC 9339. *Org. Lett.* **2018**, *20* (17), 5238-5241.
151. Jaki, B.; Orjala, J.; Sticher, O., A novel extracellular diterpenoid with antibacterial activity from the cyanobacterium *Nostoc commune*. *J. Nat. Prod.* **1999**, *62* (3), 502-503.
152. Grindberg, R. V.; Ishoey, T.; Brinza, D.; Esquenazi, E., et al., Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS One* **2011**, *6* (4), e18565.
153. Soares, A. R.; Engene, N.; Gunasekera, S. P.; Sneed, J. M.; Paul, V. J., Carriebowlinol, an antimicrobial tetrahydroquinolinol from an assemblage of marine cyanobacteria containing a novel taxon. *J. Nat. Prod.* **2015**, *78* (3), 534-538.
154. Chang, Z.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J.; Sherman, D. H.; Gerwick, W. H., Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2004**, *67* (8), 1356-1367.
155. Schultz, E. E.; Braffman, N. R.; Luescher, M. U.; Hager, H. H.; Balskus, E. P., Biocatalytic Friedel-Crafts Alkylation Using a Promiscuous Biosynthetic Enzyme. *Angew Chem Int Ed Engl* **2019**, *58* (10), 3151-3155.
156. Hillwig, M. L.; Liu, X., A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **2014**, *10* (11), 921-923.
157. Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E. P., Cylindrocyclophane Biosynthesis Involves Functionalization of an Unactivated Carbon Center. *Journal of the American Chemical Society* **2012**, *134* (45), 18518-18521.
158. Morinaka, B. I.; Lakis, E.; Verest, M.; Helf, M. J., et al., Natural noncanonical protein splicing yields products with diverse beta-amino acid residues. *Science* **2018**, *359* (6377), 779-782.
159. Gu, L. C.; Wang, B.; Kulkarni, A.; Geders, T. W., et al., Metamorphic enzyme assembly in polyketide diversification. *Nature* **2009**, *459* (7247), 731-735.
160. Zhang, Y. e. a., LimF is a versatile prenyltransferase catalyzing histidine-C-geranylation on diverse nonnatural substrates. *Nature Catalysis* **2022**, (5), 682–693.
161. Le Goff, G.; Ouazzani, J., Natural hydrazine-containing compounds: Biosynthesis, isolation, biological activities and synthesis. *Bioorg Med Chem* **2014**, *22* (23), 6529-6544.
162. Carbone, M.; Li, Y.; Irace, C.; Mollo, E., et al., Structure and cytotoxicity of phidianidines A and B: first finding of 1,2,4-oxadiazole system in a marine natural product. *Org. Lett.* **2011**, *13* (10), 2516-2519.
163. Pasha, M. A.; S., M.; Panigrahi, N., Review of synthetic strategies in the development of oxadiazine scaffolds. *Mediterranean Journal of Chemistry* **2019**, *8* (5), 338-364.
164. Verma, G.; Marella, A.; Shaquiquzzaman, M.; Akhtar, M.; Ali, M. R.; Alam, M. M., A review exploring biological activities of hydrazones. *J. Pharm. Bioallied Sci.* **2014**, *6* (2), 69-80.

165. 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.
166. Mohareb, R. M.; Schatz, J., Anti-tumor and anti-leishmanial evaluations of 1,3,4-oxadiazine, pyran derivatives derived from cross-coupling reactions of beta-bromo-6H-1,3,4-oxadiazine derivatives. *Bioorg Med Chem* **2011**, *19* (8), 2707-2713.
167. 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*, 224-237.
168. Waldman, A. J.; Ng, T. L.; Wang, P.; Balskus, E. P., Heteroatom-Heteroatom Bond Formation in Natural Product Biosynthesis. *Chem Rev* **2017**, *117* (8), 5784-5863.
169. Yang, H.; Chen, S. L., Enzymatic N-N bond formation: Mechanism for the N-nitroso synthesis catalyzed by non-heme iron SznF enzyme. *J. Catal.* **2021**, *398*, 44-53.
170. 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.
171. Kawai, S.; Hagihara, R.; Shin-Ya, K.; Katsuyama, Y.; Ohnishi, Y., Bacterial Avenalumic Acid Biosynthesis Includes Substitution of an Aromatic Amino Group for Hydride by Nitrous Acid Dependent Diazotization. *Angew Chem Int Ed Engl* **2022**, e202211728.
172. Kawai, S.; Sugaya, Y.; Hagihara, R.; Tomita, H.; Katsuyama, Y.; Ohnishi, Y., Complete Biosynthetic Pathway of Alazopeptin, a Tripeptide Consisting of Two Molecules of 6-Diazo-5-oxo-l-norleucine and One Molecule of Alanine. *Angew Chem Int Ed Engl* **2021**, *60* (18), 10319-10325.
173. Zhao, G.; Peng, W.; Song, K.; Shi, J.; Lu, X.; Wang, B.; Du, Y. L., Molecular basis of enzymatic nitrogen-nitrogen formation by a family of zinc-binding cupin enzymes. *Nat Commun* **2021**, *12* (1), 7205-7215.
174. Martins, T. P.; Glasser, N. R.; Kountz, D. J.; Oliveira, P.; Balskus, E. P.; Leao, P. N., Biosynthesis of the Unusual Carbon Skeleton of Nocuolin A. *ACS Chem. Biol.* **2022**, 2528-2537.
175. Waldman, A. J.; Pechersky, Y.; Wang, P.; Wang, J. X.; Balskus, E. P., The Cremeomycin Biosynthetic Gene Cluster Encodes a Pathway for Diazo Formation. *ChemBioChem* **2015**, *16* (15), 2172-2175.
176. Barry, S. M.; Challis, G. L., Mechanism and Catalytic Diversity of Rieske Non-Heme Iron-Dependent Oxygenases. *ACS Catal* **2013**, *3* (10), 2362-2370.
177. Ferraro, D. J.; Gakhar, L.; Ramaswamy, S., Rieske business: Structure-function of Rieske non-heme oxygenases. *Biochemical and Biophysical Research Communications* **2005**, *338* (1), 175-190.
178. Senda, M.; Kishigami, S.; Kimura, S.; Fukuda, M.; Ishida, T.; Senda, T., Molecular mechanism of the redox-dependent interaction between NADH-dependent ferredoxin reductase and rieske-type [2Fe-2S] ferredoxin. *J. Mol. Biol.* **2007**, *373* (2), 382-400.
179. Swigonova, Z.; Mohsen, A. W.; Vockley, J., Acyl-CoA Dehydrogenases: Dynamic History of Protein Family Evolution. *J. Mol. Evol.* **2009**, *69* (2), 176-193.
180. Xiong, J. H., Fatty Acid Oxidation in Cell Fate Determination. *Trends in Biochemical Sciences* **2018**, *43* (11), 854-857.

181. Ghisla, S.; Thorpe, C., Acyl-CoA dehydrogenases. A mechanistic overview. *Eur. J. Biochem.* **2004**, *271* (3), 494-508.
182. Battaile, K. P.; Nguyen, T. V.; Vockley, J.; Kim, J. J. P., Structures of isobutyryl-CoA dehydrogenase and enzyme-product complex - Comparison with isovaleryl- and short-chain acyl-CoA dehydrogenases. *J. Biol. Chem.* **2004**, *279* (16), 16526-16534.
183. Sun, Y.; Feng, Z.; Tomura, T.; Suzuki, A., et al., Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Sci. Rep.* **2016**, *6*, 22091-22102.
184. Blake-Hedges, J. M.; Pereira, J. H.; Cruz-Morales, P.; Thompson, M. G., et al., Structural Mechanism of Regioselectivity in an Unusual Bacterial Acyl-CoA Dehydrogenase. *J. Am. Chem. Soc.* **2020**, *142* (2), 835-846.
185. Fujimori, D. G.; Hrvatin, S.; Neumann, C. S.; Strieker, M.; Marahiel, M. A.; Walsh, C. T., Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (42), 16498-16503.
186. Arora, P.; Goyal, A.; Natarajan, V. T.; Rajakumara, E., et al., Mechanistic and functional insights into fatty acid activation in Mycobacterium tuberculosis. *Nat. Chem. Biol.* **2009**, *5* (3), 166-173.
187. Trivedi, O. A.; Arora, P.; Sridharan, V.; Tickoo, R.; Mohanty, D.; Gokhale, R. S., Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature* **2004**, *428* (6981), 441-445.
188. 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), e00425.
189. 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.
190. Patil, G. S.; Kinatukara, P.; Mondal, S.; Shambhavi, S., et al., A universal pocket in FAALs ensures redirection of fatty acid pool away from CoA-based activation. *bioRxiv* **2021**, (10), e70067.
191. Leao, P. N.; Martins, T. P.; Abt, K.; Reis, J. P. A.; Figueiredo, S.; Castelo-Branco, R.; Freitas, S., Incorporation and modification of fatty acids in cyanobacterial natural products biosynthesis. *Chem. Commun.* **2023**, *59* (30), 4436-4446.
192. Freitas, S.; Castelo-Branco, R.; Wenzel-Storjohann, A.; Vasconcelos, V. M.; Tasdemir, D.; Leao, P. N., Structure and Biosynthesis of Desmamides A-C, Lipoglycopeptides from the Endophytic Cyanobacterium *Desmonostoc muscorum* LEGE 12446. *J. Nat. Prod.* **2022**, *85* (7), 1704-1714.
193. Eusebio, N.; Castelo-Branco, R.; Sousa, D.; Preto, M.; D'Agostino, P.; Guider, T. A. M.; Leao, P. N., Discovery and Heterologous Expression of Microginins from *Microcystis aeruginosa* LEGE 91341. *ACS Synth. Biol.* **2022**, *11* (10), 3493-3503.

194. Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H., Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **2004**, *11* (6), 817-833.
195. Ramaswamy, A. V.; Sorrels, C. M.; Gerwick, W. H., Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2007**, *70* (12), 1977-1986.
196. Micallef, M. L.; D'Agostino, P. M.; Sharma, D.; Viswanathan, R.; Moffitt, M. C., Genome mining for natural product biosynthetic gene clusters in the Subsection V cyanobacteria. *BMC Genomics* **2015**, *16* (1), 669-689.
197. Chen, A.; Jiang, Z.; Burkart, M. D., Enzymology of standalone elongating ketosynthases. *Chem. Sci.* **2022**, *13* (15), 4225-4238.
198. Olsen, J. G.; Kadziola, A.; von Wettstein-Knowles, P.; Siggaard-Andersen, M.; Larsen, S., Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **2001**, *9* (3), 233-243.
199. Mindrebo, J. T.; Chen, A.; Kim, W. E.; Re, R. N.; Davis, T. D.; Noel, J. P.; Burkart, M. D., Structure and Mechanistic Analyses of the Gating Mechanism of Elongating Ketosynthases. *Acs Catalysis* **2021**, *11* (12), 6787-6799.
200. Fuchs, S. W.; Bozhuyuk, K. A. J.; 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 Occuring Ketosynthase. *Angew. Chem.* **2013**, *52* (15), 4108-4112.
201. Proschak, A.; Zhou, Q. Q.; Schoner, T.; Thanwisai, A.; Kresovic, D.; Dowling, A.; Ffrench-Constant, R.; Proschak, E.; Bode, H. B., Biosynthesis of the Insecticidal Xenocycloins in *Xenorhabdus bovienii*. *ChemBioChem* **2014**, *15* (3), 369-372.
202. Goblirsch, B. R.; Frias, J. A.; Wackett, L. P.; Wilmot, C. M., Crystal Structures of *Xanthomonas campestris* OleA Reveal Features That Promote Head-to-Head Condensation of Two Long-Chain Fatty Acids. *Biochemistry* **2012**, *51* (20), 4138-4146.
203. Steffen-Munsberg, F.; Vickers, C.; Kohls, H.; Land, H., et al., Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. *Biotechnol. Adv.* **2015**, *33* (5), 566-604.
204. Voss, M.; Xiang, C.; Esque, J.; Nobili, A.; Menke, M. J.; Andre, I.; Hohne, M.; Bornscheuer, U. T., Creation of (R)-Amine Transaminase Activity within an alpha-Amino Acid Transaminase Scaffold. *ACS Chem. Biol.* **2020**, *15* (2), 416-424.
205. Liepman, A. H.; Olsen, L. I., Genomic analysis of aminotransferases in *Arabidopsis thaliana*. *Crit. Rev. Plant Sci.* **2004**, *23* (1), 73-89.
206. Guo, F.; Berglund, P., Transaminase biocatalysis: optimization and application. *Green Chem.* **2017**, *19* (2), 333-360.
207. Koszelewski, D.; Tauber, K.; Faber, K.; Kroutil, W., omega-Transaminases for the synthesis of non-racemic alpha-chiral primary amines. *Trends Biotechnol.* **2010**, *28* (6), 324-332.

208. Noguchi, T.; Isogai, S.; Terada, T.; Nishiyama, M.; Kuzuyama, T., Cryptic Oxidative Transamination of Hydroxynaphthoquinone in Natural Product Biosynthesis. *J. Am. Chem. Soc.* **2022**, *144* (12), 5435-5440.
209. Xie, L. A.; Zang, X.; Cheng, W.; Zhang, Z. A.; Zhou, J. H.; Chen, M. B.; Tang, Y., Harzianic Acid from *Trichoderma afroharzianum* Is a Natural Product Inhibitor of Acetohydroxyacid Synthase. *Journal of the American Chemical Society* **2021**, *143* (25), 9575-9584.
210. Furst, M. J. L. J.; Gran-Scheuch, A.; Aalbers, F. S.; Fraaije, M. W., Baeyer-Villiger Monooxygenases: Tunable Oxidative Biocatalysts. *Acs Catalysis* **2019**, *9* (12), 11207-11241.
211. Deng, Y. M.; Zhou, Q.; Wu, Y. Z.; Chen, X.; Zhong, F. R., Properties and Mechanisms of Flavin-Dependent Monooxygenases and Their Applications in Natural Product Synthesis. *Int. J. Mol. Sci.* **2022**, *23* (5), 2622-2647.
212. Badieyan, S.; Bach, R. D.; Sobrado, P., Mechanism of N-Hydroxylation Catalyzed by Flavin-Dependent Monooxygenases. *J. Org. Chem.* **2015**, *80* (4), 2139-2147.
213. Mugge, C.; Heine, T.; Baraibar, A. G.; van Berkel, W. J. H.; Paul, C. E.; Tischler, D., Flavin-dependent N-hydroxylating enzymes: distribution and application. *Applied Microbiology and Biotechnology* **2020**, *104* (15), 6481-6499.
214. Cashman, J. R., Flavin monooxygenases. In *Enzyme Systems that Metabolize Drugs and Other Xenobiotics*, C., I., Ed. John Wiley & Sons, Ltd: 2022; pp 67-93.
215. Phillips, I. R.; Shephard, E. A., Flavin-containing monooxygenases: new structures from old proteins. *Nat. Struct. Mol. Biol.* **2020**, *27* (1), 3-4.
216. Cashman, J. R.; Zhang, J., Human flavin-containing monooxygenases. *Annual Review of Pharmacology and Toxicology* **2006**, *46*, 65-100.
217. Eswaramoorthy, S.; Bonanno, J. B.; Burley, S. K.; Swaminathan, S., Mechanism of action of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (26), 9832-9837.
218. Taylor, K. L.; Ziegler, D. M., Studies on Substrate-Specificity of the Hog Liver Flavin-Containing Monooxygenase - Anionic Organic Sulfur-Compounds. *Biochem. Pharmacol.* **1987**, *36* (1), 141-146.
219. Thodberg, S.; Neilson, E. J. H., The "Green" FMOs: Diversity, Functionality and Application of Plant Flavoproteins. *Catalysts* **2020**, *10* (3), 329-350.
220. Abdelraheem, E.; Thair, B.; Varela, R. F.; Jockmann, E., et al., Methyltransferases: Functions and Applications. *ChemBioChem* **2022**, *23* (18), e202200212.
221. Liscombe, D. K.; Louie, G. V.; Noel, J. P., Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **2012**, *29* (10), 1238-1250.
222. Bruender, N. A.; Thoden, J. B.; Kaur, M.; Avey, M. K.; Holden, H. M., Molecular Architecture of a C-3 '-Methyltransferase Involved in the Biosynthesis of D-Tetronitrose. *Biochemistry* **2010**, *49* (28), 5891-5898.
223. Schubert, H. L.; Blumenthal, R. M.; Cheng, X. D., Many paths to methyltransfer: a chronicle of convergence. *Trends in Biochemical Sciences* **2003**, *28* (6), 329-335.
224. O'Hagan, D.; Schmidberger, J. W., Enzymes that catalyse S(N)2 reaction mechanisms. *Nat. Prod. Rep.* **2010**, *27* (6), 900-918.

225. Bauerle, M. R.; Schwalm, E. L.; Booker, S. J., Mechanistic Diversity of Radical S-Adenosylmethionine (SAM)-dependent Methylation. *J. Biol. Chem.* **2015**, *290* (7), 3995-4002.
226. Sekowska, A.; Ashida, H.; Danchin, A., Revisiting the methionine salvage pathway and its paralogues. *Microb. Biotechnol.* **2019**, *12* (1), 77-97.
227. Schneider, P.; Henssen, B.; Paschold, B.; Chapple, B. P.; Schatton, M.; Seebeck, F. P.; Classen, T.; Pietruszka, J., Biocatalytic C3-Indole Methylation-A Useful Tool for the Natural-Product-Inspired Stereoselective Synthesis of Pyrroloindoles. *Angew. Chem.* **2021**, *60* (43), 23412-23418.
228. Bennett, M. R.; Shepherd, S. A.; Cronin, V. A.; Micklefield, J., Recent advances in methyltransferase biocatalysis. *Curr. Opin. Chem. Biol.* **2017**, *37*, 97-106.
229. Nelson, J. T.; Lee, J.; Sims, J. W.; Schmidt, E. W., Characterization of SafC, a catechol 4-O-methyltransferase involved in saframycin biosynthesis. *Applied and Environmental Microbiology* **2007**, *73* (11), 3575-3580.
230. Albers, E., Metabolic Characteristics and Importance of the Universal Methionine Salvage Pathway Recycling Methionine from 5'-Methylthioadenosine. *Iubmb Life* **2009**, *61* (12), 1132-1142.
231. Sauter, M.; Moffatt, B.; Saechao, M. C.; Hell, R.; Wirtz, M., Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochemical Journal* **2013**, *451*, 145-154.
232. Sekowska, A.; Denervaud, V.; Ashida, H.; Michoud, K.; Haas, D.; Yokota, A.; Danchin, A., Bacterial variations on the methionine salvage pathway. *BMC Microbiol.* **2004**, *4*, 9-26.
233. Nakano, T.; Ohki, I.; Yokota, A.; Ashida, H., MtnBD Is a Multifunctional Fusion Enzyme in the Methionine Salvage Pathway of *Tetrahymena thermophila*. *PLoS One* **2013**, *8* (7), e67385.
234. North, J. A.; Sriram, J.; Chourey, K.; Ecker, C. D.; Sharma, R.; Wildenthal, J. A.; Hettich, R. L.; Tabita, F. R., Metabolic Regulation as a Consequence of Anaerobic 5-Methylthioadenosine Recycling in *Rhodospirillum rubrum*. *Mbio* **2016**, *7* (4), e00855.
235. 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.
236. Kleigrew, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B., et al., Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. *J. Nat. Prod.* **2015**, *78* (7), 1671-1682.
237. Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T., antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **2013**, *41* (Web Server issue), 204-212.
238. Boddy, C. N., Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J Ind Microbiol Biotechnol* **2014**, *41* (2), 443-450.
239. Skinnider, M. A.; Dejong, C. A.; Rees, P. N.; Johnston, C. W.; Li, H.; Webster, A. L.; Wyatt, M. A.; Magarvey, N. A., Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res.* **2015**, *43* (20), 9645-9662.
240. Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R., The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **2012**, *7* (3), e34064.

241. Tidgewell, K.; Clark, B. R.; Gerwick, W. H., The Natural Products Chemistry of Cyanobacteria. In *Comprehensive Natural Products II: Chemistry and Biology, Vol 2: Natural Products Structural Diversity-II: Secondary Metabolites: Sources, Structures and Chemical Biology*, 2010; pp 141–188.
242. Zallot, R.; Oberg, N.; Gerlt, J. A., The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, 58 (41), 4169-4182.
243. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **2015**, 1854 (8), 1019-1037.
244. Jumper, J.; Evans, R.; Pritzel, A.; Green, T., et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583-589.
245. Diana, P.; Cirrincione, G., Techniques for Biosynthesis. In *Biosynthesis of Heterocycles: From Isolation to Gene Cluster*, John Wiley & Sons, Inc: 2015; pp 51-72.
246. Kaczmarzyk, D.; Fulda, M., Fatty acid activation in cyanobacteria mediated by acyl-acyl carrier protein synthetase enables fatty acid recycling. *Plant Physiol.* **2010**, 152 (3), 1598-1610.
247. Covington, B. C.; Xu, F.; Seyedsayamdost, M. R., A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annual Review of Biochemistry, Vol 90, 2021* **2021**, 90, 763-788.
248. Ramey, C. J.; Baron-Sola, A.; Aucoin, H. R.; Boyle, N. R., Genome Engineering in Cyanobacteria: Where We Are and Where We Need To Go. *ACS Synth Biol* **2015**, 4 (11), 1186-1196.
249. Heidorn, T.; Camsund, D.; Huang, H. H.; Lindberg, P.; Oliveira, P.; Stensjo, K.; Lindblad, P., Synthetic biology in cyanobacteria engineering and analyzing novel functions. *Methods Enzymol.* **2011**, 497, 539-579.
250. Dhakal, D.; Chen, M. Y.; Luesch, H.; Ding, Y. S., Heterologous production of cyanobacterial compounds. *J. Ind. Microbiol. Biotechnol.* **2021**, 48 (3-4).
251. Griesse, M.; Lange, C.; Soppe, J., Ploidy in cyanobacteria. *FEMS Microbiol. Lett.* **2011**, 323 (2), 124-131.
252. Kufryk, G. I.; Sachet, M.; Schmetterer, G.; Vermaas, W. F., Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency. *FEMS Microbiol. Lett.* **2002**, 206 (2), 215-219.
253. Al-Haj, L.; Lui, Y. T.; Abed, R. M.; Gomaa, M. A.; Purton, S., Cyanobacteria as Chassis for Industrial Biotechnology: Progress and Prospects. *Life (Basel)* **2016**, 6 (4), 42-59.
254. Eaton-Rye, J. J., The construction of gene knockouts in the cyanobacterium *Synechocystis* sp. PCC 6803. *Methods Mol. Biol.* **2004**, 274, 309-324.
255. Xu, Y.; Alvey, R. M.; Byrne, P. O.; Graham, J. E.; Shen, G.; Bryant, D. A., Expression of genes in cyanobacteria: adaptation of endogenous plasmids as platforms for high-level gene expression in *Synechococcus* sp. PCC 7002. *Methods Mol. Biol.* **2011**, 684, 273-93.

256. Zhang, Y.; Pu, H.; Wang, Q.; Cheng, S.; Zhao, W.; Zhang, Y.; Zhao, J., PII is important in regulation of nitrogen metabolism but not required for heterocyst formation in the Cyanobacterium *Anabaena* sp. PCC 7120. *J. Biol. Chem.* **2007**, *282* (46), 33641-8.
257. Cong, L.; Ran, F. A.; Cox, D.; Lin, S., et al., Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, *339* (6121), 819-823.
258. Li, H.; Shen, C. R.; Huang, C. H.; Sung, L. Y.; Wu, M. Y.; Hu, Y. C., CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metab. Eng.* **2016**, *38*, 293-302.
259. Wendt, K. E.; Ungerer, J.; Cobb, R. E.; Zhao, H.; Pakrasi, H. B., CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Microb Cell Fact* **2016**, *15* (1), 115-123.
260. Ungerer, J.; Pakrasi, H. B., Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Sci. Rep.* **2016**, *6*, 39681-39690.
261. Yao, L.; Cengic, I.; Anfelt, J.; Hudson, E. P., Multiple Gene Repression in Cyanobacteria Using CRISPRi. *ACS Synth. Biol.* **2016**, *5* (3), 207-212.
262. Gao, X.; Wang, P.; Tang, Y., Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Applied Microbiology and Biotechnology* **2010**, *88* (6), 1233-1242.
263. Greunke, C.; Duell, E. R.; D'Agostino, P. M.; Glockle, A.; Lamm, K.; Gulder, T. A. M., Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **2018**, *47*, 334-345.
264. D'Agostino, P. M.; Gulder, T. A. M., Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth. Biol.* **2018**, *7* (7), 1702-1708.
265. Liu, Q. S.; Shen, Q. Y.; Bian, X. Y.; Chen, H. N., et al., Simple and rapid direct cloning and heterologous expression of natural product biosynthetic gene cluster in *Bacillus subtilis* via Red/ET recombineering. *Sci. Rep.* **2016**, *6*, 34623-34633.
266. Tippelt, A.; Nett, M., *Saccharomyces cerevisiae* as host for the recombinant production of polyketides and nonribosomal peptides. *Microbial Cell Factories* **2021**, *20* (1), 121-145.
267. Gomez-Escribano, J. P.; Bibb, M. J., *Streptomyces coelicolor* as an Expression Host for Heterologous Gene Clusters. *Natural Product Biosynthesis by Microorganisms and Plants, Pt C* **2012**, *517*, 279-300.
268. Chiang, Y. M.; Oakley, C. E.; Ahuja, M.; Entwistle, R.; Schultz, A.; Chang, S. L.; Sung, C. T.; Wang, C. C. C.; Oakley, B. R., An Efficient System for Heterologous Expression of Secondary Metabolite Genes in *Aspergillus nidulans*. *Journal of the American Chemical Society* **2013**, *135* (20), 7720-7731.
269. Taton, A.; Unglaub, F.; Wright, N. E.; Zeng, W. Y., et al., Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Res.* **2014**, *42* (17), e136.
270. Zhang, J. J.; Tang, X. Y.; Moore, B. S., Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.* **2019**, *36* (9), 1313-1332.

271. Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (5), 1957-1962.
272. Feng, Z. Y.; Kim, J. H.; Brady, S. F., Fluostatins Produced by the Heterologous Expression of a TAR Reassembled Environmental DNA Derived Type II PKS Gene Cluster. *Journal of the American Chemical Society* **2010**, *132* (34), 11902-11903.
273. Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6* (5), 343-345.
274. Duell, E. R.; D'Agostino, P. M.; Shapiro, N.; Woyke, T.; Fuchs, T. M.; Gulder, T. A. M., Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*. *Microbial Cell Factories* **2019**, *18*, 32-43.
275. Zerikly, M.; Challis, G. L., Strategies for the discovery of new natural products by genome mining. *ChemBioChem* **2009**, *10* (4), 625-633.
276. Fan, Y. F.; Zhu, S. X.; Hou, F. B.; Zhao, D. F.; Pan, Q. S.; Xiang, Y. W.; Qian, X. K.; Ge, G. B.; Wang, P., Spectrophotometric Assays for Sensing Tyrosinase Activity and Their Applications. *Biosensors-Basel* **2021**, *11* (8), 290-312.
277. Bothner, B.; Chavez, R.; Wei, J.; Strupp, C.; Phung, Q.; Schneemann, A.; Siuzdak, G., Monitoring enzyme catalysis with mass spectrometry. *J. Biol. Chem.* **2000**, *275* (18), 13455-13459.
278. Vang, J. Y.; Her, C.; Krishnan, V. V., NMR based real-time enzyme kinetics on estimating the inhibitory effect of sucralose in the enzymatic conversion of sucrose. *Biophys. Chem.* **2021**, *268*, 106495-106503.
279. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **2014**, *5*, 172-199.
280. Jana, S.; Deb, J. K., Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl Microbiol Biotechnol* **2005**, *67* (3), 289-298.
281. 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.
282. Cui, Z.; Overbay, J.; Wang, X. C.; Liu, X. D., et al., Pyridoxal-5'-phosphate-dependent alkyl transfer in nucleoside antibiotic biosynthesis. *Nat. Chem. Biol.* **2020**, *16* (8), 904-911.
283. Barra, L.; Awakawa, T.; Shirai, K.; Hu, Z. J.; Bashiri, G.; Abe, I., beta-NAD as a building block in natural product biosynthesis. *Nature* **2021**, *600* (7890), 754-758.
284. Landry, Z.; Swan, B. K.; Herndl, G. J.; Stepanauskas, R.; Giovannoni, S. J., SAR202 Genomes from the Dark Ocean Predict Pathways for the Oxidation of Recalcitrant Dissolved Organic Matter. *Mbio* **2017**, *8* (2), e00413.
285. Slabu, I.; Galman, J. L.; Lloyd, R. C.; Turner, N. J., Discovery, Engineering, and Synthetic Application of Transaminase Biocatalysts. *ACS Catalysis* **2017**, (7), 8263-8284.
286. Brasseur, G.; Tron, P.; Dujardin, G.; Slonimski, P. P.; BrivetChevillotte, P., The nuclear ABC1 gene is essential for the correct conformation and functioning of the cytochrome bc(1) complex and

the neighbouring complexes II and IV in the mitochondrial respiratory chain. *Eur. J. Biochem.* **1997**, *246* (1), 103-111.

287. Jasinski, M.; Sudre, D.; Schansker, G.; Schellenberg, M.; Constant, S.; Martinoia, E.; Bovet, L., AtOSA1, a Member of the Abc1-Like Family, as a New Factor in Cadmium and Oxidative Stress Response. *Plant Physiol.* **2008**, *145* (2), 719-731.

288. Bousquet, I.; Dujardin, G.; Slonimski, P. P., ABC1, a novel yeast nuclear gene has a dual function in mitochondria: it suppresses a cytochrome b mRNA translation defect and is essential for the electron transfer in the bc 1 complex. *EMBO J.* **1991**, *10* (8), 2023-2031.

289. Zhang, Q. Q.; Liu, H. J.; Liu, X.; Jiang, D. Q., et al., Discovery of the first macrolide antibiotic binding protein in Mycobacterium tuberculosis: a new antibiotic resistance drug target. *Protein & Cell* **2018**, *9* (11), 971-975.

290. Nunez-Samudio, V.; Chesneau, O., Functional interplay between the ATP binding cassette Msr(D) protein and the membrane facilitator superfamily Mef(E) transporter for macrolide resistance in Escherichia coli. *Res. Microbiol.* **2013**, *164* (3), 226-235.

291. Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C., Functionally distinct modules operate two consecutive alpha,beta-->beta,gamma double-bond shifts in the rhizoxin polyketide assembly line. *Angew Chem Int Ed Engl* **2010**, *49* (8), 1460-1464.

292. Brandt, P.; Garcia-Altares, M.; Nett, M.; Hertweck, C.; Hoffmeister, D., Induced Chemical Defense of a Mushroom by a Double-Bond-Shifting Polyene Synthase. *Angew Chem Int Ed Engl* **2017**, *56* (21), 5937-5941.

293. Taft, F.; Brunjes, M.; Knobloch, T.; Floss, H. G.; Kirschning, A., Timing of the Delta(10,12)-Delta(11,13) double bond migration during ansamitocin biosynthesis in *Actinosynnema pretiosum*. *J. Am. Chem. Soc.* **2009**, *131* (11), 3812-3813.

294. Kim, J. J.; Wang, M.; Paschke, R., Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (16), 7523-7527.

295. McAndrew, R. P.; Wang, Y.; Mohsen, A. W.; He, M.; Vockley, J.; Kim, J. J., Structural basis for substrate fatty acyl chain specificity: crystal structure of human very-long-chain acyl-CoA dehydrogenase. *J. Biol. Chem.* **2008**, *283* (14), 9435-9443.

296. Thadathil, D. A.; Varghese, A.; Radhakrishnan, K. V., The Renaissance of Electro-Organic Synthesis for the Difunctionalization of Alkenes and Alkynes: A Sustainable Approach. *Asian Journal of Organic Chemistry* **2021**, *10* (11), 2820-2847.

Appendix - Chapter I

Copyrights ACS ChemBio

American Chemical Society's Policy on Theses and Dissertations

This policy addresses permission to include your article(s) or portions of text from your article(s) in your thesis. Reuse/Republication of the Entire Work in Theses or Collections: Authors may reuse all or part of the Submitted, Accepted or Published Work in a thesis or dissertation that the author writes and is required to submit to satisfy the criteria of degree-granting institutions. Such reuse is permitted subject to the ACS' "Ethical Guidelines to Publication of Chemical Research".

More information: [Use of Materials from ACS Journals](#)