Exploiting cyanobacterial metabolism to unveil new natural products and inquire about beta-oxidation

Bárbara Rêgo e Pires

Master in Functional Biology and Biotechnology of Plants

Biology Department

2021/2022

Supervisor

Dr. Sandra Figueiredo, Junior Researcher, CIIMAR

Co-Supervisor

Dr. Pedro Leão, Principal Investigator, CIIMAR

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.
O Presidente do Júri,

Porto, _____/_____/_____

Declaração de Honra

Eu, Bárbara Rêgo e Pires, inscrita no Mestrado em Biologia Funcional e Biotecnologia de Plantas da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Bárbara Rêgo e Pires

Porto, 14/12/2022

Acknowledgements

I would like to thank, first of all, my supervisor, Dr. Sandra Figueiredo for all that she has taught me and for all the guidance and support throughout the last year. I truly appreciate the opportunity that you gave me and the chance you took on me to work on a field so different from my background and to allow me to be a part of these two exciting projects.

Thank you, also, to my co-supervisor and team leader, Dr. Pedro Leão for all the help and advice you gave me and for allowing me to be part of such an incredible team. To the Cyanobacterial Natural Products (CNP) team, I cannot thank you enough for all the support, the guidance, and patience in both the best and worst times. I would not have been able to have the amazing experience I had at CNP without each one of you. I would like to thank you all for your friendship, your teachings and for all the fun we had together, working with you is never boring. I must specially thank Sara, Dr. Lígia, Dr. Marco and Dr. Amaranta for also helping me during my work when I needed.

To my friends and, specially, to my family, thank you for supporting me during this time, as you always have, and thank you for being my greatest source of encouragement.

I must also thank the University of Porto and the Faculty of Sciences for my education.

The first project was developed under the European Research Council through a Starting Grant (759840) and the European Union's Horizon 2020 programme (WIDESPREAD, Grant Agreement 952374). The second project was supported by national funds through FCT – Fundação para a Ciência e Tecnologia, I.P. under the framework of the project with reference "EXPL/BIA- BQM/0649/2021" and through grants UIDB/04423/2020 and UIDP/04423/2020.

Resumo

As cianobactérias são um grupo de procariotas que utilizam oxigénio e realizam fotossíntese, sendo um dos grupos morfologicamente mais diversos. Estes organismos produzem vários compostos de interesse e com diversas aplicações biotecnológicas.

Durante muito tempo, a descoberta e isolamento de produtos naturais de cianobactérias foram, essencialmente, guiados por bioensaios. Atualmente, existem outras possibilidades para explorar a descoberta destes produtos, como estratégias de incorporação de precursores marcados com isótopos. Uma dessas estratégias utiliza ácidos gordos marcados com deutério, juntamente com técnicas de espectrometria de massa para detetar novos produtos naturais. Durante este trabalho, tivemos a oportunidade de aplicar essa estratégia para descobrir novas estruturas químicas em diferentes estirpes de cianobactérias.

Essa abordagem foi baseada em algumas características metabólicas específicas das cianobactérias; uma delas é a sua aparente falta de β-oxidação funcional. Algumas observações interessantes sobre a degradação de ácidos gordos foram feitas durante esse trabalho, as quais apoiam a ideia de que afinal poderá existir uma via catabólica de ácidos gordos nesses organismos. Essas observações levaram a novas investigações sobre a β-oxidação em cianobactérias, permitindo a formulação de novas hipóteses sobre estas vias metabólicas. Embora algumas dessas hipóteses tenham sido rejeitadas, conseguimos obter novas evidências sobre esses mecanismos nunca antes descritos em cianobactérias.

No geral, durante este projeto, novos compostos foram descobertos e quatro novos produtos naturais produzidos por cianobactérias ficaram próximos de ser isolados. Avanços importantes foram também feitos na área do metabolismo de ácidos gordos em cianobactérias.

Palavras-Chave: cianobactérias, produtos naturais, ácidos gordos, β-oxidação

Abstract

Cyanobacteria are an ancient group of oxygenic phototrophic procaryotes, being one of the most morphologically diverse groups. They can produce valuable compounds and have many interesting biotechnological applications.

For a long time, cyanobacterial natural product (NP) discovery and isolation was mostly guided by bioassays. Nowadays, there are other alternatives to explore cyanobacterial NPs discovery. One of those strategies uses fatty acids (FA) labeled with deuterium coupled to mass spectrometry (MS) techniques to detect new lipid related NPs. During this work, we had the opportunity to apply such strategy to find new compounds in different cyanobacterial strains.

This approach was based on some specific metabolic characteristics of cyanobacteria; one of those being an apparent lack of a functional β-oxidation pathway. However, some interesting observations were made during this previous work that support the idea that a catabolic pathway might exist in these organisms. Those findings led to further inquiries about β-oxidation in cyanobacteria, allowing the formulation of new hypotheses about the metabolic pathway for FA catabolism. Although some of those hypotheses were rejected, we were able to gain new insight into these mechanisms that have never been described before in cyanobacteria.

Overall, during this project, new compounds have been discovered; four new NPs produced by cyanobacteria were close to being isolated, and important progress has been made in the field of cyanobacteria metabolism, more specifically their mechanism for catabolizing FA.

Key words: cyanobacteria, natural products, fatty-acids, β-oxidation

Index

Figure Index

Table Index

Equation Index

List of abbreviations and chemical formulas

Aar – Acyl-ACP reductase

Aas – Acyl-acyl carrier protein

ACP – Acyl-carrier protein

ADO – Aldehyde deformylating oxygenase

ALDH – Aldehyde dehydrogenase

BGC – Biosynthetic gene cluster

C_{16}:0 - Palmitic acid

C_6:0 - Hexanoic acid

C_8:0 - Octanoic acid

CAL – CoA ligase

CIIMAR – Interdisciplinary center for marine and environmental research

CO_2 – Carbon dioxide

d_{11}-C6 – Perdeuterated hexanoic acid

d_{15}-C8 - Perdeuterated octanoic acid

d_{31}-C16 – Perdeuterated palmitic acid

DCM – Dichloromethane

EtOAc - Ethyl acetate

FA – Fatty acid

FAAL – Fatty acyl-AMP ligases

FAD – Fatty acid degradation

FC – Flash chromatography

FFA – Free fatty acids

GNPS – Global Natural Product Social Molecular Networking

H_2O – Water

Hex - Hexane

HPLC – High performance liquid chromatography

IPA – Isopropanol

Kan – Kanamycin

LB – Luria broth

LC-HRESIMS – Liquid Chromatography – High Resolution Electrospray Ionization Mass Spectrometry

LEGEcc – Blue Biotechnology and Ecotoxicology culture collection

MCS – Multiple cloning site

MeOH – Methanol

min – Minute

MS – Mass spectrometry

NIES – National Institute of Environmental Studies

NMR – Nuclear magnetic resonance

NP – Natural product

NRPS – Non-ribosomal Peptide Synthetases

PCC – Pasteur Culture Collection

PKS – Polyketide Synthases

RP – Ribosomal peptides

rpm – Rotations per minute

RP-SPE – Reverse phase – Solid phase extraction

SPE – Solid phase extraction

TLC – Thin-layer chromatography

$U^{13}C_{16}$ – Palmitic acid uniformly-labelled with carbon 13

$U^{13}C_{8}$ – Octanoic acid uniformly-labelled with carbon 13

VLC – Vacuum liquid chromatography

Introduction

Cyanobacteria are an ancient group of oxygenic phototrophic prokaryotes, morphologically very diverse and with a capacity to survive and thrive even in the most severe environments[1,2]. Additionally, these organisms are known to produce interesting NPs, which are a group of diverse small molecules with a wide range of biological properties[3,4].

In the past, cyanobacterial NPs were mainly discovered using bioassay-guided isolation strategies, which often lead to rediscovery of already known compounds.

In recent years, with the development of DNA sequencing, it became clear that there are yet several biosynthetic gene clusters (BGCs), which encode the genetic information associated to NPs.[5]

Chapter I of this dissertation builds on a previous approach that allows for both genome dependent and independent discovery of new NPs[6]. This work involved supplementation of deuterium-labelled FAs that were incorporated in lipid compounds produced by cyanobacteria. These species can then be detected by liquid chromatography – mass spectrometry (LC-MS) and, subsequently, isolated by a MS-guided isolation strategy.

On the present work, one particular strain, *Roholtiella* sp. LEGE 12411, showed promising results due to the discovery of a potential new NP. Another strain, *Kamptonema formosum* PCC 6407, from the Pasteur culture collection, was used to try to isolate several compounds, following their mass in each step of the isolation process.

As for Chapter II, some cyanobacterial FA metabolic specificities that were considered during the development of the strategy used in Chapter I, were further explored. In particular, a more comprehensive study of the pathway for FA catabolism in cyanobacteria was done by following different routes. To this end, supplementation assays with isotopic labelled precursors were performed in cyanobacterial strains that possess distinctive genomic characteristics, in order to analyse how this could affect the labelling patterns. Also, some molecular biology work was done to try to clone some genes of interest for further assays that would elucidate some questions about the β-oxidation pathway in cyanobacteria.

CHAPTER I – Discovery of cyanobacterial natural products

1. Introduction

1.1. Natural Products

NPs represent a large family of diverse chemical entities with various biological activities[7]. These compounds are small molecules produced naturally by any organism from both the primary metabolism – compounds involved in growth, development, and reproduction of the organism, being indispensable for its survival - and secondary metabolism – compounds involved in ecological interactions that confer an advantage to the organism but are not necessary for its survival[8].

It is more common for NPs to be associated with secondary, specialized metabolites. which facilitate a wide variety of mechanisms for chemical warfare, communication, nutrient acquisition, or stress protection[5].

Historically, NPs were considered the source of, virtually, all medicinal preparations. In fact, before high-throughput screening and the post-genomic era, more than 80 % of approved drugs were either NPs or inspired by NPs[9]. Nowadays, they are still taking part in clinical trials for drugs that can act as anticancer, antimicrobial, anti-infective, antidiabetic agents, among others[9,10].

Scientists have been fascinated by the chemical diversity and complexity of NPs. They have a large range of molecular weights, an interesting stereo-complexity, a large biodiversity, and a wide variety of pharmaceutical uses[7]. NPs encompass many chemical classes such as polyketides, nonribosomal peptides, saccharides, alkaloids, and terpenoids, which together represent an enormous diversity of chemical scaffolds[11]. Although most of the NPs produced by microorganisms or plants were not produced to specifically bind to human proteins, they have evolved for optimal interactions with biological macromolecules and are, therefore, the richest source of novel compound classes for biological studies and drug discovery[12].

Marine organisms are a source of bioactive NPs. Since the ocean covers most of the Earth’s surface, it is plausible that it also encompasses a huge species diversity, and as a result, a huge chemical diversity[13]. Also, marine organisms have gone through the longest evolutionary period and so, over time, that contributed to the diversity of molecules with unique structures, functional features, and bioactivities. It is also worth noting that, due to symbiotic relations, the real origin of most marine NPs is not marine invertebrates, as previously thought, but the microorganisms that live in symbiosis with them. In fact, almost all marine NPs approved as drugs have a bacterial origin[13].

1.2. Discovery of NPs

The study of NPs involves various areas of research, from microbiology to cell biology; chemical engineering; genetics; organic chemistry, among other. All these scientific fields working together have been able to create an area of investigation focused on NPs research[7].

Traditionally, most of the NP discovery has been achieved through bioactivity-guided fractionation of chemical extracts from individual microbes and plants. However, the associated high rate of rediscovery makes this approach less effective[5].

With the surge of DNA sequencing, it became clear that the genomes of NPs producing-organisms encoded BGCs that produce many metabolites that had never been observed in the laboratory. These BGCs encode several different molecules such as: enzymes; regulatory proteins; and transporters, all of which are necessary to produce, process and export specialized metabolites These observations led to the emergence of genome mining as a technology to identify the biosynthetic pathways for both known and unknown metabolites[11].

Currently there is an effort in NP research to uncover the compounds encoded in orphan BGCs. These are gene clusters for which the correspondent metabolite is still undiscovered[14]. Initially, several orphan BGCs had been identified by gene activation/deletion and metabolomic profiling but several novel strategies that do not require gene manipulation have since been introduced[15].

The traditional approach of genome-guided NP discovery is very dependent on the recognition of the BGCs in the genome and the prediction of its function, so it is very challenging to predict the product of a BGC that is not closely related to gene clusters of already known compounds[5]. Due to that difficulty, an effort was put into developing genome-independent strategies.

For that, various advances were made in the field of bioinformatics and analytical chemistry, especially in terms of MS, which allowed for an increase of the speed of the NP discovery process[16]. These new strategies have allowed researchers to try to reveal any kind of metabolite in a biological sample.

The development of a wide range of analytical platforms, such as gas chromatography, high-performance liquid chromatography (HPLC) and ultra-performance liquid chromatography coupled to MS, and nuclear magnetic resonance (NMR) spectroscopy allows for the detection, separation, characterization and even quantification of these metabolites[17].

Metabolomics is an important tool for metabolite profiling. Attempting to analyse several metabolites in the same experiment is referred to as untargeted metabolomics, which coupled with LC-MS techniques, is a commonly applied tool[18]. Results from data

preprocessing can be analyzed and mapped to metabolic pathways to give a biological context to untargeted metabolomics data. Computational algorithms extract raw LC-MS or GC-MS data and process them in a way that, connecting with databases search, makes sense of the information in a way that allows for detection of chemical species, assignment of metabolites to such species and integration of the metabolites in an integrated framework that leads to a better understanding of biological systems[19].

Also, community efforts like the Global Natural Product Social Molecular Networking (GNPS) enables dereplication of known compounds which is a critical step for annotation of MS/MS data, but also allows for the detection of analogs and, very importantly, it is an important tool in assisting with the discovery of new compounds[20].

1.3. Cyanobacteria

When talking about marine NPs, an important group of organisms that should be mentioned are cyanobacteria. These are an ancient group of oxygenic phototrophic prokaryotes from which chloroplasts have evolved. They are gram-negative bacteria with an extensive morphological diversity which goes from single-celled forms to differentiated multicellular organisms[21]. The variability observed in cyanobacteria led to the development of a section system to connect cyanobacteria with similar characteristics (Figure 1)[21].

Cyanobacteria are present is various environments, such as marine, freshwater, or terrestrial habitats[1], as well as more extreme marine and environmental ecosystems. The optimal temperature for most cyanobacterial species is higher than that of eukaryotic algae, which allows them to thrive in warmer climates. Another characteristic about cyanobacteria is their widespread tolerance to water stress and they are also among the most successful organisms in saline environments[22].

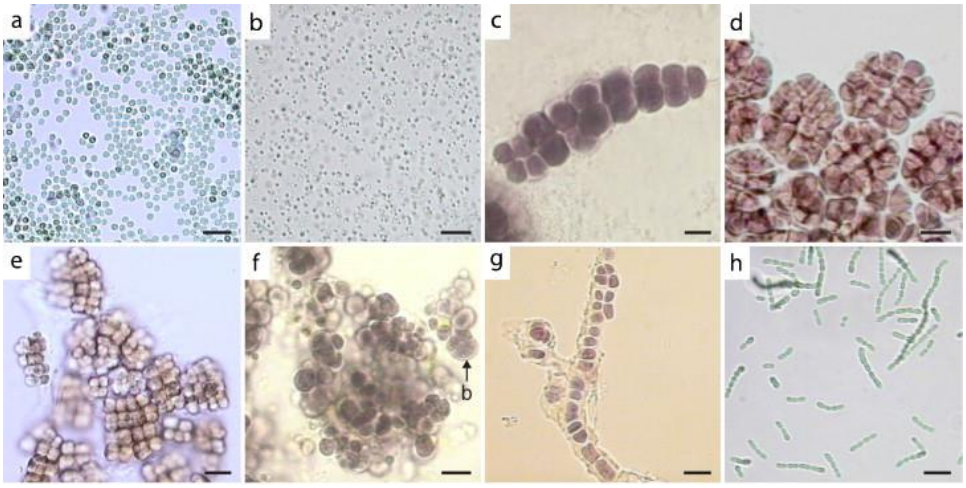


FIGURE 1 - REPRESENTATION OF CYANOBACTERIAL DIVERSITY. A – *SYNECHOCYSTIS SALINA* LEGE 06099, B - *CYANOBIUM* SP. LEGE 06184, C - *CHROOCOCCOPSIS* SP. LEGE 07168, D - *CHROOCOCCOPSIS* SP. LEGE 07187, E - *MYXOSARCINA* SP. LEGE 06146, F - *CHROOCOCCOPSIS* SP. LEGE 07161, G - *HYELLA* SP. LEGE 07179, H - *ROMERIA* SP. LEGE 06013 (ADAPTED FROM BRITO ET AL., 2012([23]))

Cyanobacteria carry out oxygenic photosynthesis, but, if necessary, some can also switch to an anoxygenic photosynthesis. Some filamentous cyanobacteria have heterocysts, which gives them the ability to fix atmospheric nitrogen[24].

It has been concluded that there were two routes of genome development in cyanobacteria. One route was an expansion strategy that was based in gene-family enlargement and allowed for a better adaptative potential. The other route focused on a simplification of the genome, which led to adaptations to highly specific niches[22].

Cyanobacteria are known to have several possible uses in human activities, such as in mariculture, food, fuel, fertilizers, colorants, and the production of various secondary metabolites[24]. These secondary metabolites include several bioactive NPs. The research of these molecules in marine cyanobacteria has revealed an impressive biosynthetic capacity for producing novel secondary metabolites[4].

An important field of research within cyanobacterial NPs is the study of BGCs. These are groups of at least two genes, clustered together, that encode pathways for the production of secondary metabolites[25].

There are different structural classes of BGCs, such as the nonribosomal peptide synthetase (NRPS) or the polyketide synthase (PKS) enzyme systems[26,27]. Some polyketide synthases can also form NRPS-PKS hybrids.

NRPSs are a type of BGC divided in different domains: the highly conserved adenylation domain, the peptidyl carrier domain and the condensation domain, which is important for the peptide elongation since it catalyzes the amide bond formation[28]. In PKSs, similarly to the condensation domain of NRPSs, there is a ketosynthase domain that catalyzes the same reaction[29]. Aside from these core domains, these enzymes can incorporate other types of domains that endorse the NP diversity[30].

Other major classes of secondary metabolites include ribosomal peptides (RPs), ribosomally-synthesized and post-translationally modified peptides (RiPPs), which possess a chemical diversity that is uncommon to ribosomal peptides, and terpenes known for, among other things, their odorific characteristics[27,31].

The complexity of the secondary metabolism of these organisms is, in fact, so big, that only a very small fraction of the compounds synthesized by cyanobacteria have been characterized[31].

1.4. Fatty-acid labelling for natural product discovery

Cyanobacteria grow slowly and there is a difficulty in genetically manipulating these organisms. These challenges stimulate advances in new strategies of discovery of cyanobacterial NPs[32].

To that end, new approaches like stable isotope-labelled precursor incorporation strategies have been applied to achieve this goal[6].

An interesting characteristic in cyanobacterial NPs is the common presence of FA-derived moieties. These can be incorporated into secondary metabolites through fatty acyl-AMP ligases (FAALs) that load fatty acyl units into the respective acyl carrier protein (ACP). Also, FAs can be recruited from the primary metabolism and incorporated into the NPs through the action of several different enzymes[6].

However, the size and type of the acyl group carried by these enzymes is variable and hard to predict, which means it is complicated to obtain structural predictions that are very precise for the compounds that result from the actions of such enzymes[6].

Another peculiarity about the FA metabolism of cyanobacteria is an apparent lack of a functional β-oxidation pathway[33].

These cyanobacterial features, alongside with the ability that these organisms possess to incorporate and elongate exogenous FAs, due to the acyl-acyl carrier protein synthetase (Aas) activity[34], have led to studies about how to take advantages of these properties to create new techniques for the discovery of NPs.

To that end, Figueiredo et al. ([6]) used lipidome labeling to discover new cyanobacterial NPs. With targeting of FA-incorporating BGCs, they created a method that allowed for the metabolites to be detected by MS techniques, more specifically, LC-HRESIMS. What was seen was that the metabolites that had incorporated the FA had an *m/z* increment that could be easily detected (Figure 2).

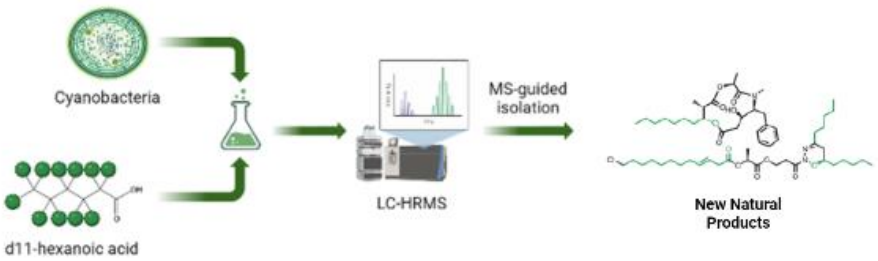


FIGURE 2 - DISCOVERY OF NEW CYANOBACTERIAL NPS USING FA LABELLING (CREATED WITH BIORENDER.COM)

This strategy allows for, not just targeted, but also untargeted discovery of metabolites. For targeted discovery, the BGCs present in the genome are searched for the coding region of the enzymes that incorporate the FA (FAAL enzymes). In the case of untargeted discovery, any cyanobacterial strain can be tested to uncover new compounds without having their genome.

The authors have shown, through the discovery of several compounds, that this method of supplementation of cyanobacterial strains with deuterated FAs is an example of the possibilities for the future of NP discovery[6].

2. Objectives

The main goal of this work was to use the MS-guided isolation strategy to isolate a novel compound from the cyanobacterial strain *Kamptonema formosum* PCC 6407, previously discovered using the FA labelling approach developed by Figueiredo et al. (2021), and to fully characterize it.

Another part of the project involved applying the supplementation assays developed by Figueiredo et al. (2021) to cyanobacterial strains present at the Blue Biotechnology and Ecotoxicology culture collection (LEGEcc), housed at the Interdisciplinary Center of Marine and Environmental Research (CIIMAR).

For this work, the goal was to screen several strains and discover new NPs that would then be isolated and characterized.

Both parts of this project are based on the FA labeling method for discovering new cyanobacterial NPs and the MS-guided isolation technique. With the use of these strategies, we wanted to have a straight forward approach that would allow us to fulfil our goals.

3. Materials and Methods

3.1. Culture of cyanobacterial strains

Most of the strains used for this work were part of the LEGEcc. This is a collection of cyanobacteria and microalgae created at CIIMAR which encompasses hundreds of strains with information relating to their origin, culture conditions, taxonomic-related data, among others.

The cyanobacterial strains were maintained at 25° C +/- 2° C with a photoperiod of light:dark of 14:10 (14 hours of light and 10 hours of darkness) and a light intensity of 10 – 30 µmol photons s^{-1} m^{-2}. They grew in Z8 medium or, in case of marine strains, Z8 + 25 g L^{-1} of tropic marine salt. Both media were prepared as described in Kotai[35].

The initial cultures were of 40 mL (Figure 3 - A) and, afterwards, those that were scaled-up were re-inoculated to higher volumes up to 20 L (Figure 3 - B).

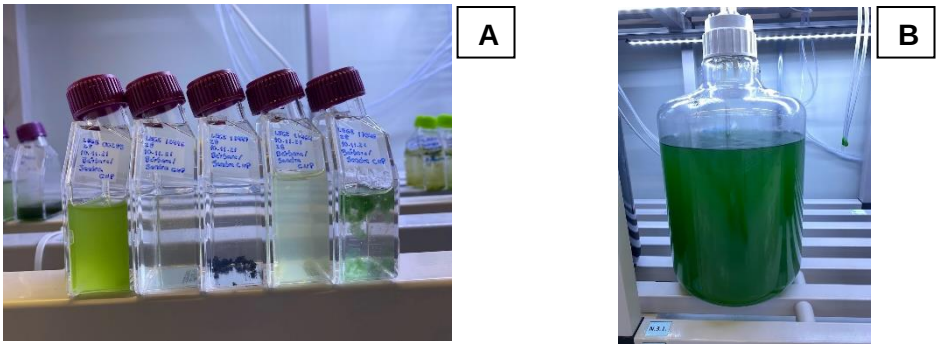


FIGURE 3 - – CULTURES OF CYANOBACTERIA. A – 40 ML INITIAL CULTURES OF DIFFERENT STRAINS OF CYANOBACTERIA. B – 20 L CULTURE OF *CUSPIDOTHRIX ISSATSCHENKOI* LEGE 03285

For this project, ten different cyanobacterial strains were chosen for stable isotope labelled FA incorporation experiments: *Nostocales* sp. LEGE 15495; *Gloeothece* sp. LEGE 16572; *Roholtiella* sp. LEGE 12411; *Plectonema cf. radiosum* LEGE 06105; *Nostocales* sp. LEGE 12452; *Hyella patelloides* LEGE 07179; *Nostocales* sp. LEGE 17548; *Desmonostoc muscorum* LEGE 12446; *Microcystis aeruginosa* LEGE 11464 and *Microcystis aeruginosa* LEGE 00239.

These strains were maintained in cultures of up to 200 mL of medium for the supplementation assays, which will be explained next.

After the supplementation assays, *Roholtiella* sp. LEGE 12411 and *Plectonema cf. radiosum* LEGE 06105 appeared to produce new compounds and were scaled-up to 20 L.

With the strains growing in 20 L, we harvested the biomass, when necessary, and did a new inoculum, in order to continue to obtain biomass. The harvesting process was

done with two different strategies depending on the strain: through filtration with a nylon net (different nets with different pore sizes were used) or through centrifugation where the biomass remained in the pellet.

In each case, the biomass was then freeze-dried and kept at room temperature away from light.

3.2. Discovery of cyanobacterial natural products

The ten strains mentioned above were used for assays that consisted of doing a pulse supplementation with different FAs: triplicates of unlabelled samples, used as the control, and triplicates of deuterium labelled samples, which corresponded to the experimental condition. After the assay was done, the biomass was harvested for an organic extraction and the extracts were used for LC-HRESIMS. The data went through a comparative metabolomics analysis and dereplication process, to look for new compounds.

3.2.1. Genome mining

LEGEcc strains were used for genome mining.

For the selection of the ten cyanobacterial strains used in this work, the genomes of different strains were analysed to search the BGCs for the CoA ligase (CAL) domain, which is one of the NRPS/PKS domains and it encodes the FAAL enzymes that allow for the incorporation of FAs into the cells.

The antiSMASH[36] data for the genomes was already available and the selection of strains was done based on highest to lowest number of BGCs with the CAL domain in each genome.

3.2.2. Supplementation assays

The first step (day 01) was to grow the strain in 200 mL in the conditions mentioned above and with an agitation of 200 rotations per minute (rpm) for 72 h.

After this, on day 04 of the experiment, the culture was split into six Erlenmeyers with 25 mL each – three for the control conditions and three for the experimental conditions. Each of the control cultures was supplemented by three pulses of a 0.5 mM final concentration of hexanoic acid (C_6:0) and each of the experimental cultures was supplemented by three pulses of a 0.5 mM final concentration of deuterated hexanoic acid (d_{11}-C_6), from a 1000x concentrated solution of each FA in DMSO. The supplementation was repeated on days 06 and 08.

Simultaneously, on day 04 3 mL of the culture (1 mL/Eppendorf) were collected to do a chlorophyll *a* assay that was used as a measure of the growth of the strain, in case the assay had to be repeated. Next, each sample was centrifuged at 17.000 g for three minutes and a washing step was performed with the pellet. The chlorophyll *a* was extracted with a solution of 90 % methanol (MeOH) and 10 % water (H_2O). The mixture was left overnight in the dark with an agitation of 300 rpm. On day 05, another centrifugation was done in the conditions mentioned before and the supernatant was used to measure the absorbance of each triplicate at 663 nm, with a solution of 90 % MeOH and 10% H_2O as blank. The chlorophyll *a* concentration was measured using Equation 1.

EQUATION 1 - FORMULA USED TO CALCULATE THE CHLOROPHYLL *A* CONCENTRATION

$$\textbf{Chlorophyll } \boldsymbol{a} \textbf{ = Abs 663 nm x 12,7 (µg mL}^{-1}\textbf{)}$$

3.2.3. Harvesting biomass and organic extraction

At day 11 the biomass from each Erlenmeyer was collected through centrifugation at 7500 g for 10 minutes. Then the supernatant was discarded, and the pellet was washed and went through another centrifugation at 5000 g for 10 minutes. Afterwards the pellet was freeze dried for 24 to 48 hours.

After that period, the dry biomass was transferred to solvent resistant tubes and 15 mL of a mixture of Dichloromethane (DCM):MeOH (2:1 v/v) was added. The tubes were left for one hour in an orbital shaker and, after that, for 15 minutes in an ultrasonic bath.

The organic extract obtained was filtered through a Whatman 1 filter to a 100 mL round-bottom (RB) flask.

All the solvents were then dried in a rotavapor, and the extracts transferred to pre-weighted glass vials that were stored at -20° C until further analysis were performed.

3.2.4. LC-HRESIMS analysis

The LC-HRESIMS analysis of each sample was done using an Ultimate 3000 HPLC (Thermo Fisher Scientific), which is divided in the following modules: LPG-3400RS pump; WPS-3000RS autosampler; TCC-3X00RS column compartment; MWD-3000RS UV/Visible detector. This is coupled to a Q Exactive Focus Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Fisher Scientific). The equipment is controlled by the Q Exactive Focus (Exactive Series) 2.9 and Thermo Scientific Xcalibur 4.1.31.9 software.

For analysis, the crude extracts were resuspended in MeOH for a final concentration of 2 mg mL^{-1}. Then they were filtered with a 0.2 µm minisart filter and, approximately, 100 µL of the samples were transferred to LC-MS vials.

LC-HRESIMS data was obtained in Full Scan positive and negative modes, with a capillary voltage of HESI set to -3.8 kV and a capillary temperature of 300° C.

The analysis was done using an ACE UltraCore SuperC18 column, with a 2.5 µm particle size, 95 Å pore size and 75 x 2.1 mm dimensions. The column oven was set to 40° C.

The gradient program used two eluents: eluent A was H_2O:MeOH (1:1 v/v) with 0.1 % v/v formic acid); eluent B was isopropanol (IPA) with 0.1 % v/v formic acid

The program that was applied to process the samples and separate the compounds present in the organic extract had a elution at 0.4 mL min^{-1} with a gradient of 90 % eluent A to reach 65 % eluent B during 5 minutes, followed by an isocratic of 65 % eluent B for 12 minutes and a gradient of 25 % eluent A to reach 85 % eluent B for 9 minutes and, then, return to the initial conditions, followed by an isocratic of 90% eluent A for 2 minutes.

3.2.5. Comparative metabolomics

After the LC-HRESIMS run was completed, the data was used for a comparative metabolomics analysis. This consisted of using the MzMine 2.53 software to process, analyse and compare the MS data. Within the program, several parameters were defined for the analysis of both positive and negative modes, which led to a final list with the different compounds that were detected during the LC-HRESIMS run; their presence in each of the triplicates for both the control and experimental conditions; what type of deuterium incorporation they had; the retention time of the peak corresponding to the compound; and the peak area.

After, a list was created with all the peaks that corresponded with a deuterium labelling and that were in, at least, two different samples and not in the control. The next step was to confirm the presence of those peaks using the Xcalibur software to check for the $[M+H]^+$ or $[M-H]^-$ peak values and calculate the respective neutral mass.

3.2.6. Dereplication

Each neutral mass value obtained was introduced in the Dictionary of Natural Products 31.1 website[37] and the Natural Products Atlas website[38] for dereplication, i.e., to check if a compound with the same molecular mass value as the one found in the sample had already been reported. This avoided rediscovery of compounds.

Those same mass values were also introduced in the Lipid Maps website[39] to see if each value corresponded to a compound characterized as a lipidic primary metabolite, since the focus was only the secondary metabolites.

After each neutral mass value was introduced in all these different databases, the ones that did not provide any hits were selected to be looked at more in depth and the strains that were producing them were scale-up for a possible isolation of such compounds.

3.3. Isolation of new compounds from *Kamptonema formosum* PCC 6407

As mentioned before, the main goal of this project was the isolation of a new compound from *Kamptonema formosum* PCC 6407 (from here on referred to as PCC 6407), which, throughout the process, turned into the isolation of four different compounds.

3.3.1. Organic Extraction

The first step to isolate the compounds was an organic extraction with all the dry biomass that was previously collected and freeze dried.

The initial amount of dry biomass was 66.3 g. To that, a mixture of DCM:MeOH (2:1 v/v) was added, very similarly to what was described for the supplementation assays, but in a larger scale. It was necessary to use a total of around 2 L of the mixture in order to extract the organic compounds. The mixture would be frequently stirred to remove as much organic extract as possible. Sonication was occasionally used to also help with the removal of the organic extract.

The organic mixture was then filtered by vacuum, using a Whatman 41 paper filter and cloth, as shown in Figure 4, and the result of the filtration went into an RB flask, which was then evaporated under reduced pressure in a rotavapor to eliminate all the solvents.

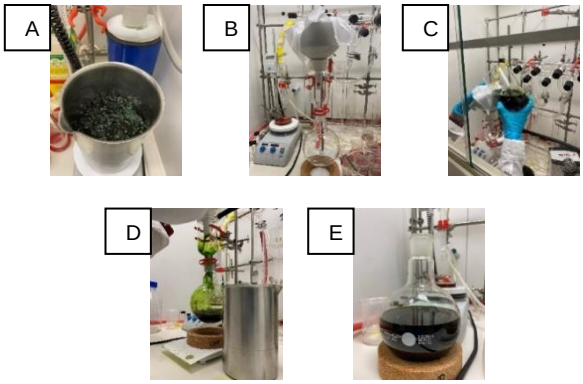


FIGURE 4 - ORGANIC EXTRACTION OF PCC 6407 SHOWING THE DRY BIOMASS WITH THE SOLVENTS (A); THE APPARATUS USED FOR THE FILTERING OF THE EXTRACT IN THE (B) AND THE FILTERING PROCESS (C-E)

These steps were repeated multiple times to extract as much as possible. The extract was then transferred to a pre-weighted glass vial.

3.3.2. Vacuum Liquid Chromatography (VLC)

The organic extract was fractionated by VLC. First, silica gel 60 (0.040-0.063 mm) was added in the same amount as the crude extract (8.4 g) and the mixture was dried in the rotavapor until we obtained a loose powder for the dry loading. The VLC apparatus (Figure 5) was filled with sea sand, silica gel 60, the dry loading of the sample, more sand and, finally, cotton to protect the sample from the addition of the eluents. The VLC column was then connected to a vacuum pump.

FIGURE 5 - VLC APPARATUS FOR THE VLC OF THE CRUDE EXTRACT OF PCC 6407

The elution gradient was done starting from a non-polar solvent and moving towards a polar solvent, for what is also called a normal phase chromatography. In this case, an increasing polarity gradient from hexane (Hex) and ethyl acetate (EtOAc) to MeOH was used for elution, beginning with EtOAc:Hex (1:9 v/v) and moving towards 100 % MeOH.

The collected fractions were pooled by volume (around 250 mL each fraction) into RB flasks. The fractions were then dried and transferred to previously weighted glass vials

to be analysed by LC-MS to verify the presence of the compounds of interest, as was done for all the following fractionation steps.

3.3.3. Flash Chromatography (FC)

The next chromatographic step was an automated FC. The fractions from the VLC where the compounds of interest were present, were fractionated by FC using a Pure Chromatography system (Büchi) with a reverse phase C18 silica gel cartridge, 230-400 mesh 40-63 µg (Silicycle) and a flow rate of 25 mL min^{-1} (Figure 6).

The fractionation program began with a polar mixture of solvents and moved towards a non-polar solvent. In this case, a decreasing polarity gradient from H_2O:MeOH (1:1 v/v) to IPA was used for elution, beginning with 100 % H_2O:MeOH (1:1 v/v) and moving towards 95 % IPA and 5 % H_2O:MeOH. (1:1 v/v).

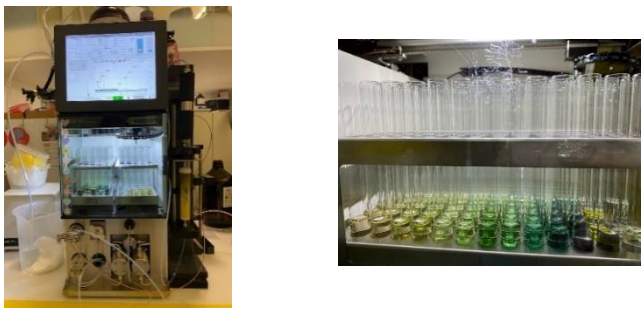


FIGURE 6 - FC APPARATUS FOR THE FRACTIONATION OF PCC 6407

Throughout the isolation process, other FCs had to be done, however, for those fractionations a normal phase elution was used.

For this second FC a silica gel cartridge 230-400 mesh 40-63 µm (Silicycle) was used with a flow rate of 14 mL min^{-1}. An increasing polarity gradient from EtOAc:Hex (3:7 v/v) to 100 % MeOH was used for elution.

The fractions were pooled aided by their Thin Layer Chromatography (TLC) (aluminum silica gel F254 TLS plates from Merck) analysis profiles.

A third FC was performed with the same cartridge as before and a flow rate of 10 mL min^{-1}. An increasing polarity gradient from 100 % EtOAc to DCM:MeOH (4:1 v/v) was used for elution and the fractions were polled in the same manner as for the second FC.

3.3.4. Solid Phase Extraction (SPE)

The fractions of interest from the first FC were combined and used for a reverse phase SPE (RP-SPE). For this fractionation a Visiprep™ SPE Vacuum Manifold, standard, 12-

port model with silica gel 0.8 g, 230-400 mesh 40-63 µg (Silicycle) was used. A gradient of H_2O:MeOH (3:1 v/v) to 100 % IPA was chosen for elution and a cleaning step with DCM and acetone was added in the end. A Strata C18-E column, 55 µm particle size, 70 Å pore size and dimensions 500 m^2 g^{-1} was used (Figure 7).

After this step, another RP-SPE was done using a Strata C18-E column, 55 µm particle size, 70 Å pore size and dimensions 500 m^2 g^{-1}. A similar gradient was used, that also went from H_2O:MeOH (3:1 v/v) to 100 % IPA.

FIGURE 7 - SPE APPARATUS FOR THE FRACTIONATION OF PCC 6407

3.3.5. High Performance Liquid Chromatography (HPLC)

HPLC was used for the final purification steps. For this, two different equipments were used: the Jasco P-2000 polarimeter controlled by SpectraManager 2.14.02 software with the UV spectra obtained on a UV-1600PC spectrometer controlled by MWAVE 1.0.20 software (VWR), which was used for HPLC 01 and HPLC 03; and the Dionex Ultimate 3000 controlled by the Chromeleon 7.2.8 software, which was used for all other HPLCs.

Also, three different types of columns were used: an ACE-C18 semipreparative column with a 5 µm particle size and dimensions 250/10.0 mm, for HPLC 01 with a flow of 3 mL min^{-1}; a Phenomenex Synergi 4u Fusion - RP semipreparative column with a 4 µm particle size and dimensions 250/10.0 mm for HPLC 02, also with a flow of 3 mL min^{-1}; and a Surf C18 100 A column with a 5 µm particle size and dimensions 250/4.6 mm with a flow of 1 mL min^{-1} for all other HPLCs.

The samples were always prepared in the minimal amount of solvent (MeOH) necessary to completely dissolve it, taking into consideration the capacity of the different columns.

3.3.5.1. HPLC 01

To process this sample, the elution program used three eluents, which were the same for all HPLCs. Eluent A was H_2O, eluent B was MeOH, and eluent C was IPA.

This program began with 95 % eluent B and 5 % eluent A, and it went up to 100 % eluent B in 15 minutes.

After that, from 15 to 30 minutes it went from 100 % eluent B to 54% eluent B and 46 % eluent C and from 30 to 33 minutes it changed to 70 % eluent C and 30 % eluent B. It stayed that way from 33 to 38 minutes and then, from 38 to 42 minutes, it came back to the initial conditions, and it remained with those until the end of the run at 45 minutes.

3.3.5.2. HPLC 02

For this sample, the elution program began with 95 % eluent B and 5 % eluent A, and it went up to 100 % eluent B in 15 minutes.

After that, from 15 to 30 minutes it went from 100 % eluent B to 54% eluent A and 46 % eluent C and from 30 to 33 minutes it changed to 70 % eluent C and 30 % eluent B. It stayed that way from 33 to 40 minutes and then, from 40 to 44 minutes, it came back the initial conditions and it remained with those until the end of the run at 50 minutes.

3.3.5.3. HPLC 03

In this HPLC, the elution program began with 60 % eluent B and 40 % eluent A for 10 minutes and it went up to 100 % eluent B from 10 to 15 minutes.

After that, from 15 to 25 minutes it stayed at 100 % eluent B and then, from 25 to 30 minutes it went to 50 % eluent B and 50 % eluent C, and that mixture remained from 30 to 40 minutes. From 40 to 42 minutes, it came back the initial conditions and it remained with those until the end of the run at 50 minutes.

3.3.5.4. HPLC 04

This elution program began with 90 % eluent B and 10 % eluent A for 5 minutes and it changed to 69 % eluent B with 27 % eluent C and 4 % eluent A from 5 to 10 minutes and it stayed that way from 10 to 40 minutes. After that, from 40 to 43 minutes it changed to 70 % eluent C and 30 % eluent B and that remained from 43 to 50 minutes. From 50 to 52 minutes, it came back the initial conditions and it remained with those until the end of the run at 55 minutes.

3.3.5.5. HPLC 05

This sample was processed with an elution program that began with 90 % eluent B and 10 % eluent A for 5 minutes and it changed to 91 % eluent B with 5 % eluent C and 4 % eluent A from 5 to 10 minutes and it stayed that way from 10 to 60 minutes. After that, from 60 to 65 minutes it changed to 70 % eluent C and 30 % eluent B and that remained from 65 to 70 minutes. From 70 to 72 minutes, it came back the initial conditions and it remained with those until the end of the run at 75 minutes.

3.3.5.6. HPLC 06

For the HPLC, the elution program began with 90 % eluent B and 10 % eluent A for 5 minutes and it changed to 69 % eluent B with 27 % eluent C and 4 % eluent A from 5 to 10 minutes and it stayed that way from 10 to 40 minutes. After that, from 40 to 43 minutes it changed to 70 % eluent C and 30 % eluent B and that remained from 43 to 53 minutes. From 53 to 55 minutes, it came back the initial conditions and it remained with those until the end of the run at 60 minutes.

3.3.5.7. HPLC 07

This HPLC used an elution program that began with 90 % eluent B and 10 % eluent A for 5 minutes and it changed to 91 % eluent B with 5 % eluent C and 4 % eluent A from 5 to 10 minutes and it stayed that way from 10 to 60 minutes. After that, from 60 to 65 minutes it changed to 70 % eluent C and 30 % eluent B and that remained from 65 to 70 minutes. From 70 to 72 minutes, it came back the initial conditions and it remained with those until the end of the run at 75 minutes.

3.3.6. LC-HRESIMS analysis of the fractionation steps

Every fraction from each fractionation step was dried and transferred to pre-weighted glass vials. After, all the fractions were resuspended in MeOH (2 mg mL^{-1} for most of the fractionation steps and 1 mg mL^{-1} for the HPLC fractions) to be analysed by LC-HRESIMS. The program used, however, was not always the same, it went through two optimization steps to allow for the better visualization of our masses of interest.

For the analysis of the fractions resultant from the VLC, the first FC and the first SPE, the program used was the same as for the supplementation assays, which was previously described.

After the second SPE, this program was optimized and it began using three eluents: eluent A was H_2O with 0.1 % v/v formic acid, eluent B was MeOH with 0.1 % v/v formic acid, and eluent C was IPA with 0.1 % v/v formic acid.

The fractions were processed with an elution at 0.3 mL min^{-1} with a gradient of 95 % eluent A and 5 % of eluent B to reach 95 % eluent B for 5 minutes, followed by an isocratic of 95 % eluent B for 20 minutes and a gradient to 90 % eluent C, 5 % eluent A and 5 % eluent B for 5 minutes, followed by an isocratic for 7 minutes. Then, it returned to the initial conditions in 2 minutes, followed by an isocratic of 95 % eluent A for 4 minutes.

This program was used to analyse every fraction until HPLC 05. At that point, the program was optimized once again. The gradient program that was used to process the samples and separate the compounds present in the organic extract had a elution at 0.3 mL min^{-1} with a gradient of 95 % eluent A and 5 % of eluent B to reach 85 % eluent B during 5 minutes, followed by an isocratic of 85 % eluent B for 20 minutes and a gradient to 90 % eluent C, 5 % eluent A and 5 % eluent C for 3 minutes, followed by an isocratic for 5 minutes. Then, return to the initial conditions in 2 minutes, followed by an isocratic of 95% eluent A for 4 minutes. This program was used to analyse the results of all the HPLCs that followed.

3.3.7. Nuclear Magnetic Resonance (NMR)

After HPLC 05, some peaks observed in the chromatogram were able to be isolated. Of those, one peak that possibly corresponded to one of the masses that were being pursued, was chosen for NMR.

For that the sample was prepared in deuterated solvents (Cambridge Isotope Laboratories) and analysed in NMR 600 MHz. This analysis was performed at the Materials Center of the University of Porto (CEMUP) on a BRUKER AVANCE III HD 600 MHz 14.1 Tesla, in liquid samples on 5 mm tubes compatible with the frequency of the equipment. The resultant data were analysed in MestReNova 12.0.3 (MestrelabResearch).

4. Results

4.1. Discovery of cyanobacterial natural products

The first part of this project consisted of using the method developed by Figueiredo et al. (2021)[6] to discover new NPs from cyanobacteria using FA labelling. For this, 10 different cyanobacterial strains were supplemented. All of these strains were obtained from LEGEcc at CIIMAR.

These strains were selected through genome mining based on antiSMASH analysis. All had a maximum of four and minimum of two different CAL domains, the BGC domain associated with the FAAL enzymes for FA incorporation (Figure 8). In this way, the selected strains had a greater potential to produce secondary metabolites that incorporated FA residues.

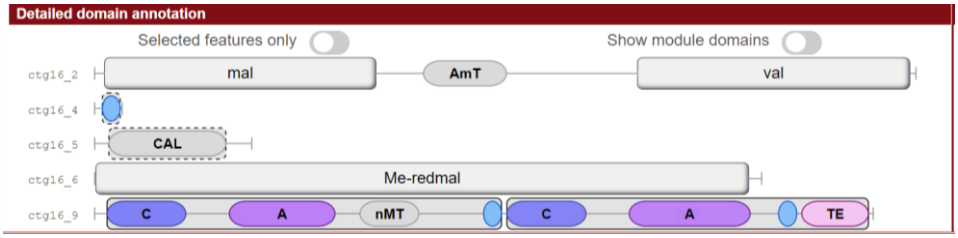


FIGURE 8 - EXAMPLE OF THE PRESENCE OF A CAL DOMAIN IN A BGC OF ONE OF THE SELECTED STRAINS (*NOSTOCALES CYANOBACTERIUM* LEGE 17548)

The 10 selected strains went through the process of supplementation; organic extraction; LC-HRESIMS; comparative metabolomics analysis and dereplication (Table 1).

TABLE 1 - STRAINS USED FOR THE SUPPLEMENTATION ASSAYS WITH THE MEAN OF THE CHLOROPHYL *A* ASSAY AND THE MEAN OF THE MASS OF EXTRACTS OBTAINED

Strain	Chlorophyl *a* (µg/mL)	Extract (mg) - mean
LEGE 15495	1.308	0.932 +/- 0.578
LEGE 16572	3.251	1.280 +/- 0.640
LEGE 12411	1.270	0.875 +/- 0.505
LEGE 06105	0.838	2.560 +/- 2.500
LEGE 12452	1.651	0.270 +/- 0.200
LEGE 07179	0.203	0.920 +/- 0.320
LEGE 17548	1.372	0.625 +/- 0.515
LEGE 12447	0.914	1.470 +/- 0.610
LEGE 11464	0.889	1.705 +/- 1.005
LEGE 00239	1.410	1.845 +/- 1.315

All these were studied to search for new NPs. Some strains only showed to be producing compounds that had already been characterized in other strains, or that were considered primary lipids. In the case of *Plectonema cf. radiosum* LEGE 06105, some MS features, which seemed to correspond to new compounds, were discovered. However, after further analysis, these were either primary lipids or just compounds

present in very low abundance. *Roholtiella* sp. LEGE 12411 was the most promising strain since, after the analysis was complete, it was concluded that the strain was producing what seemed to be a secondary metabolite that had not been described yet. This compound (Table 2), interestingly, was not found after the traditional organic extraction with DCM:MeOH (2:1 v/v). In fact, this strain was extracted first with a mixture of MeOH:H_2O (9:1 v/v) due to an experimental error, and the new MS features discovered resulted from this methanolic extraction.

TABLE 2 – NEUTRAL MASS; [M+H]+ AND RETENTION TIME OF THE COMPOUND DISCOVERED

M	[M-H]-	Retention time
1006.41877	1005.41147	0.60 min

4.2. Isolation of new compounds from *Kamptonema formosum* PCC 6407

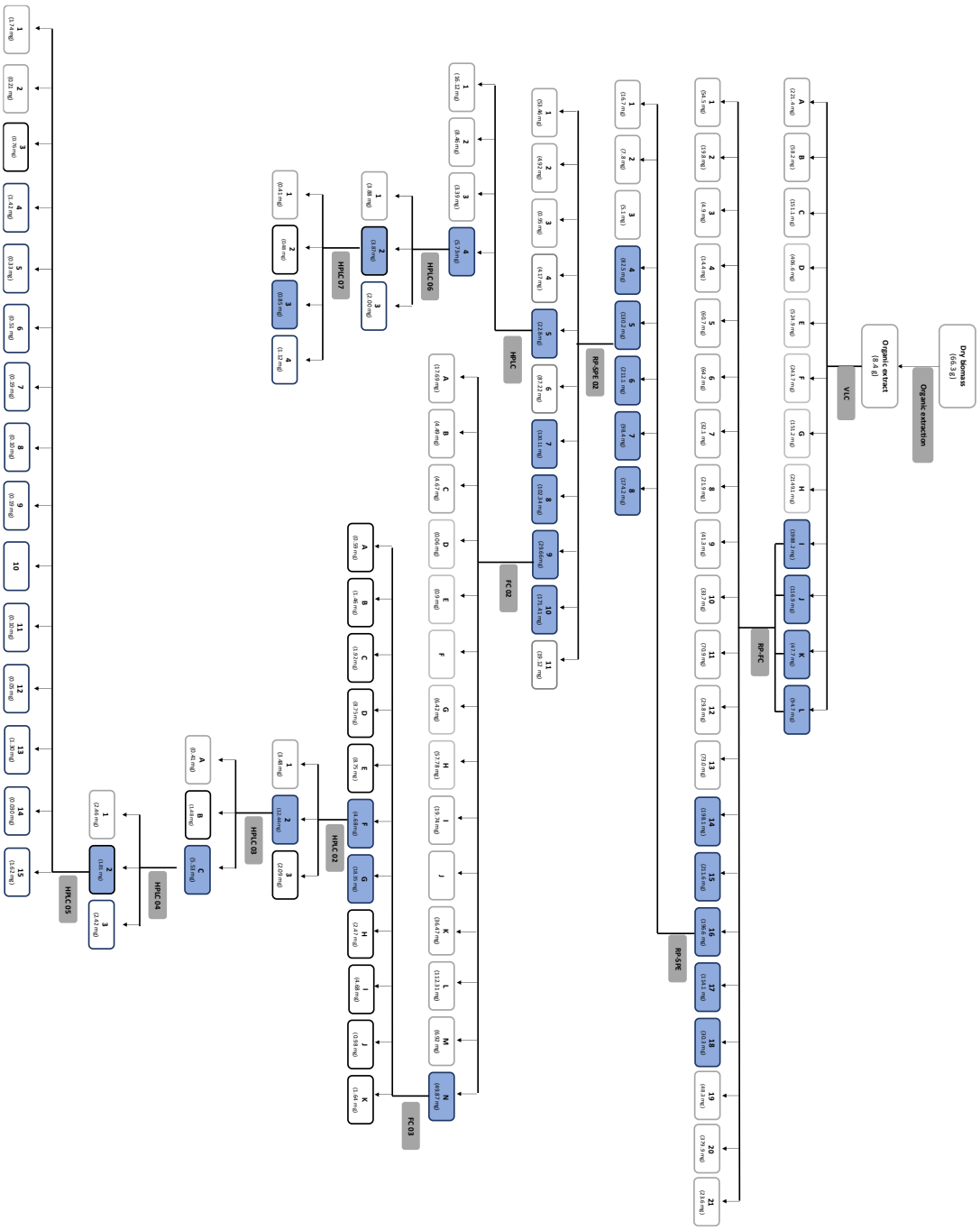


FIGURE 9 - ISOLATION SCHEME FOR PCC 6407

The isolation of compounds from PCC 6407 involved several fractionation steps that are represented in Figure 9 and that will be explained in detail in the next sections. At first only one compound was being pursued, which had an $[M+H]^+$ peak at *m/z*= 682.53660 (HRESIMS) and was previously found during the studies with the supplementation assays. After MS^2 analysis, the fragmentation pattern indicated that it might be a lipopeptide, due to some amino acid fragments found. After the second FC, three more compounds started being pursued for isolation, since it seemed like they were related to the previous one. These new compounds had an $[M+H]^+$ peak at *m/z*= 668.52090, *m/z* = 696.55236 and *m/z*= 682.53660 (Table 3). The first two peaks have a difference of a CH_2 radical in comparison with the initial compound (HRESIMS) and the one with *m/z*= 682.53660 is an isomer of the compound that began this work.

TABLE 3 - NEUTRAL MASS; $[M+H]^+$ AND RETENTION TIME OF THE COMPOUNDS FOR ISOLATION

M	$[M+H]^+$	Retention time (min)
681.52930	682.53660	16.67
667.51360	668.52090	15.21
681.52930	682.53660	17.36
695.54506	696.55236	19.35

4.2.1. Organic Extraction

The 66.3 g of dry biomass grown in Nalgene 20 L bottles was used for the organic extraction. From that biomass 8.4 g of organic extract were obtained to be used for the fractionation steps that followed.

4.2.2. VLC

The first fractionation step was the VLC that was done using the extract previously obtained.

The extract was divided in 12 different fractions, whose corresponding mobile phase, volume, and mass is described in Table 4.

The division of the sample in different fractions was done based on volume. However, at the end of the run, there was still some sample in the column, which was indicated by the presence of colour, and so, MeOH, DCM and acetone were added at the end to clean the VLC column.

TABLE 4 - VLC FRACTIONS WITH THE CORRESPONDING MOBILE PHASE, VOLUME AND MASS - THE FRACTIONS WHERE THE COMPOUND OF INTEREST IS PRESENT ARE HIGHLIGHTED IN BLUE

Fractions	Mobile Phase	Volume (mL)	m (g)
1	EtOAc:Hex (1:9 v/v)	500	0.22141
2	20 % EtOAc:Hex (1:4 v/v)	250	0.05821
3	30 % EtOAc:Hex (3:7 v/v)	250	0.15106
4	40 % EtOAc:Hex (2:3 v/v)	250	0.4066
5	EtOAc:Hex (3:2 v/v)	250	0.52487
6	EtOAc:Hex (4:1 v/v)	250	0.24366
7	EtOAc	250	0.15122

8	MeOH:EtOAc (1:3 v/v)	250	2.1491
9	MeOH	500	1.9882
10	MeOH + DCM + Acetone	750	0.11692
11	MeOH	500	0.04768
12	MeOH	500	0.09469

These fractions were analysed by LC-HRESIMS, using the program previously described in section **3.2.3.**, to look for the compound of interest, which was present in fractions 9 to 12, specially fraction 9.

After this fractionation the total mass of extract obtained was 6.1517 g for a total yield of 73 %.

4.2.3. FC

After the VLC, fractions 9 to 12 (2.2475 g), which were the fractions where the compound was present, were pooled together and used as the sample for the next fractionation step (FC). This resulted in 106 test tubes which were automatically obtained from the equipment based on ELSD and the UV chromatogram with UV channels of 254 nm; 220 nm; 280 nm and 320 nm. After, a manual analysis was done based on the ELSD and UV profiles and the test tubes were grouped in 21 fractions (Table 5). All the fractions were analysed by LC-HRESIMS, in the same conditions as mentioned before, to identify the fractions that contained the compound.

TABLE 5 - FC FRACTIONS WITH THE CORRESPONDING MOBILE PHASE, DIVISION BY TEST TUBES, MASS AND PEAK AREA – THE FRACTIONS WHERE THE COMPOUND OF INTEREST IS PRESENT ARE HIGHLIGHTED IN BLUE

Fractions	Mobile phase	Test tubes	Fraction (mg)	Peak area
01	IPA:(MeOH:H_2O 1:1 v/v) (0:1 – 1:4)	1-15	54.460	4.21E+07
02	IPA:(MeOH:H_2O 1:1 v/v) (1:4 v/v)	16-17	19.75	3.44E+08
03	IPA:(MeOH:H_2O 1:1 v/v) (3:7 v/v)	18-24	4.88	2.59E+08
04	IPA:(MeOH:H_2O 1:1 v/v) (2:3 v/v)	25-31	14.39	4.20E+07
05	IPA:(MeOH:H_2O 1:1 v/v) (1:1 v/v)	32-37	60.7	7.20E+07
06	IPA:(MeOH:H_2O 1:1 v/v) (1:1 v/v)	38-41	64.19	1.52E+08
07	IPA:(MeOH:H_2O 1:1 v/v) (1:1 v/v)	42-47	32.08	4.22E+08
08	IPA:(MeOH:H_2O 1:1 v/v) (1:1 v/v)	48-50	21.95	3.82E+08
09	IPA:(MeOH:H_2O 1:1 v/v) (1:1 v/v)	51-57	41.28	2.92E+08
10	IPA:(MeOH:H_2O 1:1 v/v) (3:2 v/v)	58-62	33.74	3.21E+08
11	IPA:(MeOH:H_2O 1:1 v/v) (3:2 v/v)	63-2’	70.91	3.89E+08
12	IPA:(MeOH:H_2O 1:1 v/v) (7:3 v/v)	3’-6’	29.8	2.33E+08
13	IPA:(MeOH:H_2O 1:1 v/v) (7:3 v/v)	7’-14’	73.03	2.62E+08
14	IPA:(MeOH:H_2O 1:1 v/v) (7:3 v/v)	15’-17’	198.14	7.92E+08
15	IPA (MeOH:H_2O) (4:1 v/v)	18’-20’	211.63	7.10E+08
16	IPA (MeOH:H_2O) (8:2 v/v)	21’-24’	196.63	1.52E+09
17	IPA (MeOH:H_2O) (9:1 v/v)	25’-28’	114.15	7.47E+08
18	IPA (MeOH:H_2O) (9:1 v/v)	29’-35’	30.34	1.16E+09
19	IPA	36’-41’	48.32	5.94E+08
20	IPA+DCM	IPA+DCM	379.94	1.70E+08
21	DCM	DCM	23.62	7.96E+08

After the VLC, the association between the compound and the eluents indicated that it would need strong eluents to be removed from the total extract and that is why IPA was added early in the program.

The presence of the compound in the last cleaning step is probably due to the use of a strong eluent, which took with it a small part of the compound that had remained in the column.

From this fractionation 1.7239 g of biomass were obtained for a yield of 77 %.

4.2.4. SPE

The fractions where the compound was present (14 to 18 and 21 - 774.51 mg) in the FC were used for a SPE. From that 8 fractions were obtained (Table 6). The fractions where the compound of interest was present were identified in the same way as mentioned for the FC.

TABLE 6 - SPE FRACTIONS WITH THE CORRESPONDING MOBILE PHASE, VOLUME, MASS AND PEAK AREA - THE FRACTIONS WHERE THE COMPOUND OF INTEREST IS PRESENT ARE HIGHLIGHTED IN BLUE

Fractions	Mobile phase	Volume (mL)	Fraction (mg)	Peak area
01	MeOH:H_2O (1:3 v/v)	100	16.73	1.14E+09
02	MeOH:H_2O (1:1 v/v)	100	7.76	3.75E+08
03	IPA:(MeOH:H_2O 1:1 v/v) (3:7 v/v)	100	5.15	1.92E+07
04	IPA:(MeOH:H_2O 1:1 v/v) (1:1 v/v)	250	84.47	8.01E+08
05	IPA:(MeOH:H_2O 1:1 v/v) (3:2 v/v)	100	130.2	1.17E+09
06	IPA:(MeOH:H_2O 1:1 v/v) (4:1 v/v)	100	211.14	9.74E+08
07	IPA	100	98.4	7.19E+08
08	DCM + Acetone	250	174.2	4.68E+08

After this SPE, the fractions where the compound was present corresponded to a mass of 698.41 for a yield of 90 %.

The fractions where the compound was present had a higher amount of biomass and were still very complex, so, before a more precise method of isolation was used, another SPE was done to further simplify the sample.

4.2.5. SPE 02

Fractions 4 to 8 from the previous isolation step (698.41 mg) were used for a second SPE with the goal of simplifying the fractions where the desired compound was present. This fractionation resulted in 11 fractions (Table 7) and the ones where the compound was present were identified by LC-HRESIMS, as mentioned before.

TABLE 7 - SPE FRACTIONS WITH THE CORRESPONDING MOBILE PHASE, VOLUME, MASS AND PEAK AREA - THE FRACTIONS WHERE THE COMPOUND OF INTEREST IS PRESENT ARE HIGHLIGHTED IN BLUE

Fractions	Mobile phase	Volume (mL)	Fraction (mg)	Peak area
01	MeOH:H_2O (1:3 v/v)	100	53.46	4.20E+08
02	MeOH:H_2O (1:1 v/v)	100	4.92	6.17E+08
03	MeOH:H_2O (3:1 v/v)	100	0.95	1.01E+08
04	MeOH:H_2O (17:3 v/v)	250	4.17	2.43E+08
05	MeOH:H_2O (19:1 v/v)	100	22.8	9.48E+08
06	MeOH:H_2O (19:1 v/v)	150	87.22	2.83E+08
07	MeOH	100	130.11	1.36E+09
08	MeOH:IPA (1:1 v/v)	250	102.34	9.22E+08
09	IPA	250	29.66	5.51E+08
10	DCM	210	171.41	6.19E+08
11	Acetone	200	19.12	1.92E+08

Unlike in the previous fractionations where a mixture of MeOH:H_2O (1:1 v/v) was used, for this SPE the eluents changed to different parts of MeOH and H_2O to see if the compound would react with that instead of the stronger eluents, mostly used for cleaning, and that bring a lot of complexity to the fractions because of removing a lot of compounds from the extract.

After this step the total extract was 626.16 mg for a yield of 90 %.

4.2.6. HPLC

After the previous fractionation, LC-HRESIMS results showed that the compound that was being followed was present in fractions 7 to 10. However, those were still complex fractions and, since the compound was also present in fraction 5 (22.8 mg), a simpler fraction and with less mass, that fraction was used to move forward with the HPLC to try to isolate the compound and also optimize the HPLC program for a future isolation of this compound.

After several programs were tested, the sample was processed using the program described in the Materials and Methods section, which resulted from looking at the previous results and understanding that IPA was needed, but not wanting to start with a high percentage, in order to not make all the compounds come out at the same time.
The program was divided in four sub-fractions based on the chromatographic profile shown in figure 10.

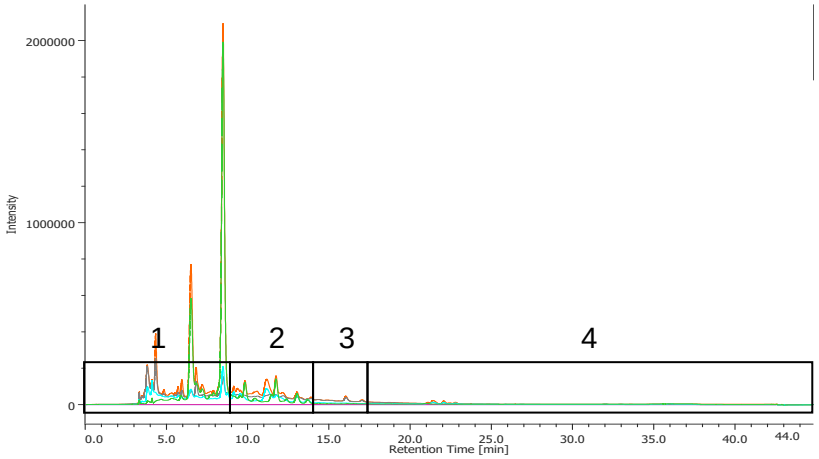


FIGURE 10 - CHROMATOGRAPHIC PROFILE OBTAINED FOR THE HPLC (ORANGE – 254 NM; GREY – 366 NM; BLUE – 475 NM; GREEN – 662 NM)

After the HPLC, a LC-HRESIMS was done to confirm the presence of the compound. The compound was only present in sub-fraction 4, that goes from the end of sub-fraction 3 until the end of the run, as is shown in Figure 10.
After this HPLC, 33.7 mg of extract were obtained for a yield of 148 %, probably due to an error during the weighing of the samples.

4.2.7. FC 02

With the remaining fractions obtained from the second SPE (fractions 7 to 10 – 433.52 mg), due to the high mass and complexity that they represented, they were pooled together for another FC.

The second FC, this time normal phase, resulted in 110 test tubes that were divided in 14 fractions using the same criteria as for the first FC, as well as what was observed after doing the TLC (Table 9).

For the design of this program, the most recent fractionations were analysed instead of the results from the VLC, which was the only previous fractionation step done in normal phase, and so, that is why it was not possible to properly fractionate the compounds in this FC.

TABLE 8 - FC FRACTIONS WITH THE CORRESPONDING MOBILE PHASE, DIVISION BY TEST TUBES, MASS AND PEAK AREA OF THE PEAK WITH *M/Z* = 682.53660 AS REPRESENTATIVE OF ALL THE COMPOUNDS – THE FRACTIONS WHERE THE COMPOUND OF INTEREST IS PRESENT ARE HIGHLIGHTED IN BLUE

Fractions	Mobile phase	Test tubes	Fractions (mg)	Peak area
1	EtOAc:Hex (3:7 v/v)	00-09	17.69	0
2	EtOAc:Hex (3:7 v/v)	10-11	4.49	0
3	30 EtOAc:Hex (3:7 v/v) - EtOAc:Hex (1:1 v/v)	12-25	4.67	0
4	EtOAc:Hex (1:1 v/v)	26-32	2.95	0
5	EtOAc:Hex (1:1 v/v) - EtOAc:Hex (7:3 v/v)	33-50	5.24	0
6	EtOAc:Hex (7:3 v/v) – EtOAc	51-60	2.22	0
7	EtOAc – EtOAc:MeOH (8:2 v/v)	61-06	9.37	0
8	EtOAc:MeOH (8:2 v/v)	07-13	58.16	0
9	EtOAc:MeOH (8:2 v/v)	14-17	20.59	0
10	EtOAc:MeOH (8:2 v/v)	18-22	0	0
11	EtOAc:MeOH (8:2 v/v)	23-25	36.47	4.57E+07
12	EtOAc:MeOH (8:2 v/v)- EtOAc:MeOH (6:4 v/v)	26-33	112.52	1.21E+08
13	EtOAc:MeOH (6:4 v/v)	34-45	9.78	1.22E+08
14	EtOAc:MeOH (6:4 v/v) – MeOH	46-45	49.87	9.25E+09

From this fractionation 334.02 mg of extract were obtained for a yield of 77 %.

The fractions were analysed by LC-HRESIMS. However, as mentioned in the Materials and Methods section, the LC-HRESIMS program that was used to analyse these fractions was an optimized version of the original program, focused on the compound that was being pursued.

At this stage of the process, being able to obtain simpler fractions, it was observed that there were two analogues and one isomer of the original compound that were present in the same fraction as the compound that was being isolated.

After this fractionation step, the goal of the work changed from the isolation of one compound to the isolation of four compounds.

4.2.8. FC 03

After the last fractionation step, fraction 14 from FC 02 was used for one more FC, since the previous one did not provide the intended fractionation. For this a simpler program was used and a total of 11 factions were obtained (Table 10).

TABLE 9 - FC FRACTIONS WITH THE CORRESPONDING MOBILE PHASE, DIVISION BY TEST TUBES, MASS AND PEAK AREA OF THE PEAK WITH *M/Z* = 682.53660, AS REPRESENTATIVE OF ALL THE COMPOUNDS – THE FRACTIONS WHERE THE COMPOUND OF INTEREST IS PRESENT ARE HIGHLIGHTED IN BLUE

Fractions	Mobile phase	Test tubes	Fraction (mg)	Peak area
1	EtOAc	02-03	0.59	0
2	EtOAc	04-05	1.46	2.76E+06
3	EtOAc	06-07	1.92	8.23E+05
4	EtOAc	08-10	1.15	7.99E+06
5	EtOAc:MeOH (4:1 v/v)	11-13	8.75	1.59E+07
6	EtOAc:MeOH (3:2 v/v)	14-15	4.68	1.66E+08
7	EtOAc:MeOH (3:2 v/v) - EtOAc:MeOH (2:3 v/v)	16-19	18.35	6.95E+09
8	EtOAc:MeOH (1:4 v/v)	20-21	2.47	1.11E+10
9	EtOAc:MeOH (1:4 v/v) – MeOH	22-25	4.68	8.27E+09
10	DCM:MeOH (4:1 v/v)	26	0.98	3.35E+09
11	DCM:MeOH (4:1 v/v)	27-28	1.64	3.81E+09

From this FC 46.670 mg of extract were obtained for a yield of 94 %.

After the LC-HRESI analysis, as can be seen in table 10, the compounds were present in fractions 6 and 7.

4.2.9. HPLC 02

After the previous fractionation, fractions 6 and 7 were combined (23.03 mg) and used as the sample for the HPLC.

Three sub-fractions were obtained based on the program that was used and the chromatographic profile that was observed (Figure 11)

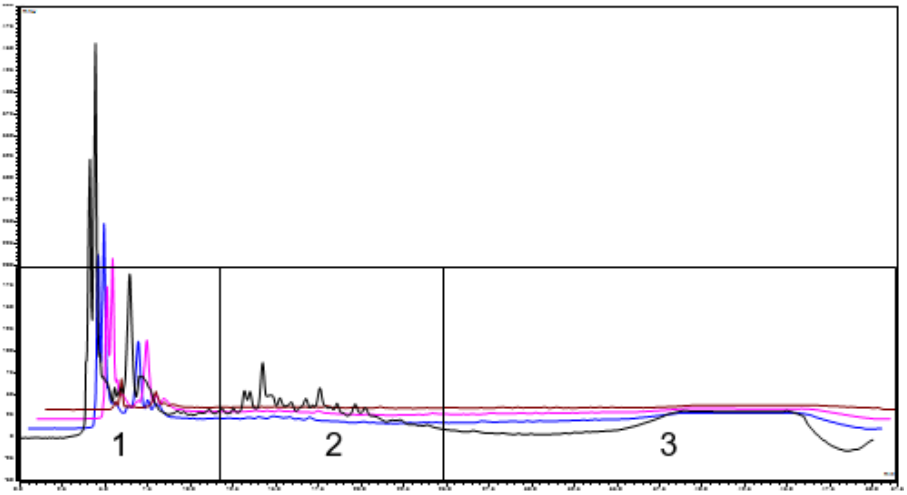


FIGURE 11 - – CHROMATOGRAPHIC PROFILE OBTAINED FOR THE HPLC (BLACK – 225 NM; BLUE – 254 NM; PINK – 280 NM; BROWN – 650 NM) – THE X AXIS CORRESPONDS TO THE RETENTION TIME WHILE THE Y AXIS IS REPRESENTATIVE OF THE INTENSITY

After analysis, it was determined that the compounds were present in sub-fraction 2 and that this fraction would be used to try to isolate the compounds. At this point, a decision was made to not join sub-fraction 2 of HPLC 02 with sub-fraction 4 obtained from the previous HPLC 01, since this new fraction had been simplified in comparison with the previous one, and so they were processed separately.

From this HPLC 18.01 mg of extract were obtained for a yield of 78 %.

4.2.10. HPLC 03

For this HPLC fraction 02 from HPLC 02 (12.44 mg) was used as the sample to be processed. It was not possible to isolate peaks so, based on the same criteria as for the previous HPLC fractionations, three different fractions were obtained (Figure 12).

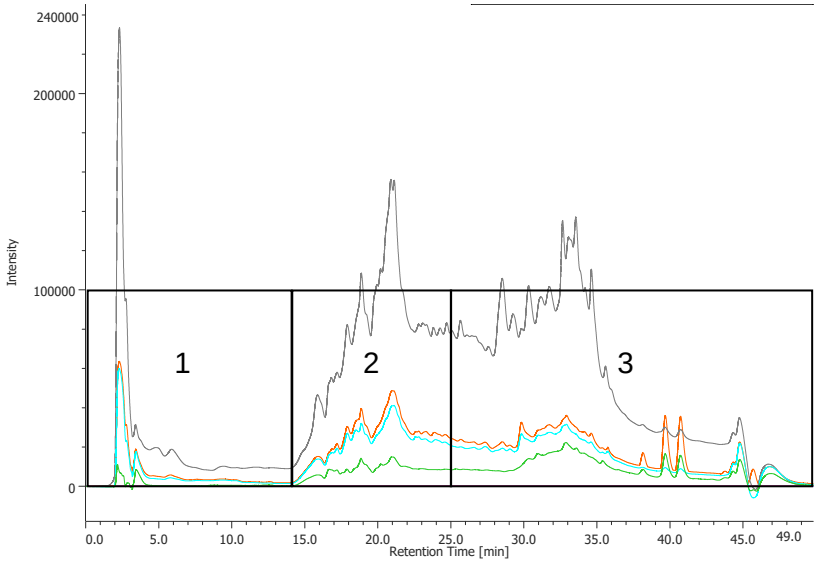


FIGURE 12 - CHROMATOGRAPHIC PROFILE OBTAINED FOR THE HPLC (GREY – 225 NM; ORANGE – 254 NM; BLUE – 280 NM; GREEN – 650 NM)

The extracts obtained after the processing of the sample corresponded to 7.4200 mg for a yield of 60 %.

After LC-HRESIMS analysis, the compounds were detected in sub-fraction 3 and, since peak isolation was not achieved in this fractionation and the sample was still complex, another HPLC was planned.

4.2.11. HPLC 04

Based on the program used for this fractionation the sample was divided in three sub-fractions (Figure 13).

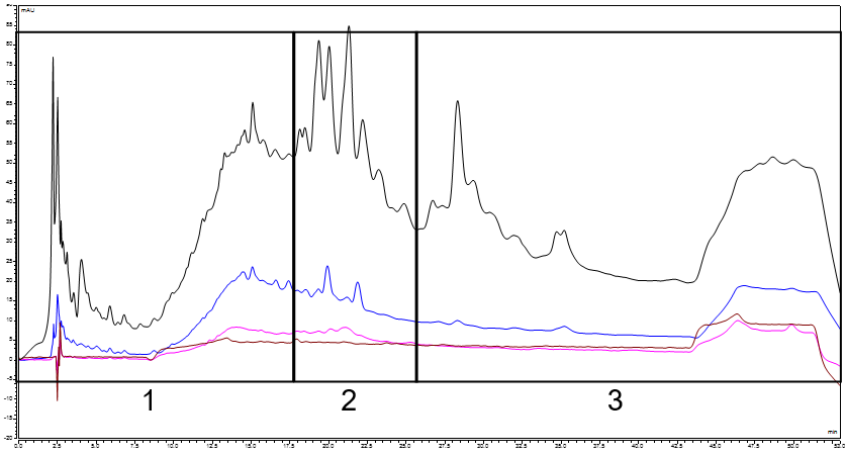


FIGURE 13 - CHROMATOGRAPHIC PROFILE OBTAINED FOR THE HPLC (BLACK – 225 NM; BLUE – 254 NM; PINK – 366 NM; BROWN – 450 NM) - THE X AXIS CORRESPONDS TO THE RETENTION TIME WHILE THE Y AXIS IS REPRESENTATIVE OF THE INTENSITY

The compounds that were being pursued were present in sub-fraction 2. Due to having a simpler fraction but not a lot of mass (1.81 mg), it was decided that it would be better to isolate the peaks corresponding to the compounds as the next step for this sample.

After this fractionation, 6.69 mg of extract were obtained for a yield of 90%.

4.2.12. HPLC 5.0

In this HPLC fraction 2 from HPLC 04 was used to isolate several peaks, including the ones that correspond to our compounds of interest.
Based on the program that was used and the chromatographic profile, the sample was divided in 15 sub-fractions as demonstrated in Figure 14.

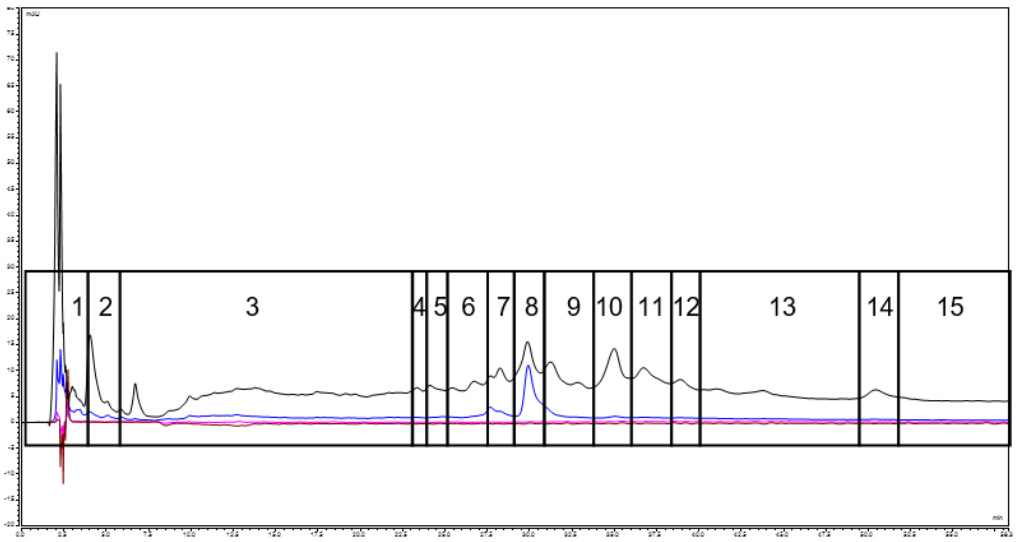


FIGURE 14 - CHROMATOGRAPHIC PROFILE OBTAINED FOR THE HPLC (BLACK – 225 NM; BLUE – 254 NM; PINK – 366 NM; BROWN – 450 NM) - THE X AXIS CORRESPONDS TO THE RETENTION TIME WHILE THE Y AXIS IS REPRESENTATIVE OF THE INTENSITY

After this fractionation a LC-HRESIMS run was done with fractions and peaks 1 to 5 and 11 to 15, the ones with more mass.

It was confirmed that none of those corresponded to any of the compounds of interest. Peak 8 was analysed by NMR since it was one of the main peaks of the sample and so, there was a possibility that it corresponded to one of the compounds that were being isolated. However, the NMR signal was very low due to mass limitation, and it was not possible to elucidate the structure of the compound.

4.2.13. HPLC 06

In this fractionation, sub-fraction 4 from HPLC 01 (5.73 mg) was used to try to isolate the compounds of interest.

This was a more complex sample, since it had only gone through one HPLC, while the sample used for HPLC 02 had been the result of two FCs.

HPLC 04 and 06, as well as HPLC 05 and 07 were done in parallel and that is why approximately the same program was used in each pair to simplify the samples and isolate the compounds.
The chromatographic profile was divided in three sub-fractions (Figure 15), knowing the sample still had to be further simplified to try to isolate peaks.

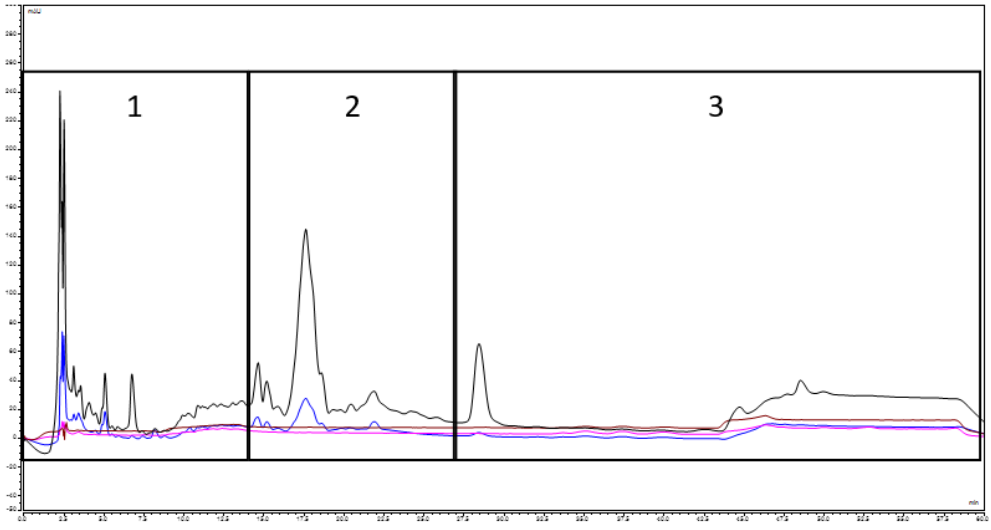


FIGURE 15 - CHROMATOGRAPHIC PROFILE OBTAINED FOR THE HPLC (BLACK – 225 NM; BLUE – 254 NM; PINK – 366 NM; BROWN – 450 NM) - THE X AXIS CORRESPONDS TO THE RETENTION TIME WHILE THE Y AXIS IS REPRESENTATIVE OF THE INTENSITY

After this fractionation 8.6 mg of extract were obtained for a yield of 164%.

The compounds of interest showed to be present in sub-fraction 2 which was then used for further fractionation.

4.2.14. HPLC 07

For this HPLC sub-fraction 2 from HPLC 06 (3.58 mg) was processed. It was not possible to optimize the program in a way that would allow for the isolation of compounds due to the very low absorbance of the compounds in several different wave lengths.

Because of that, one more division in fractions had to be done to process this sample (Figure 16).

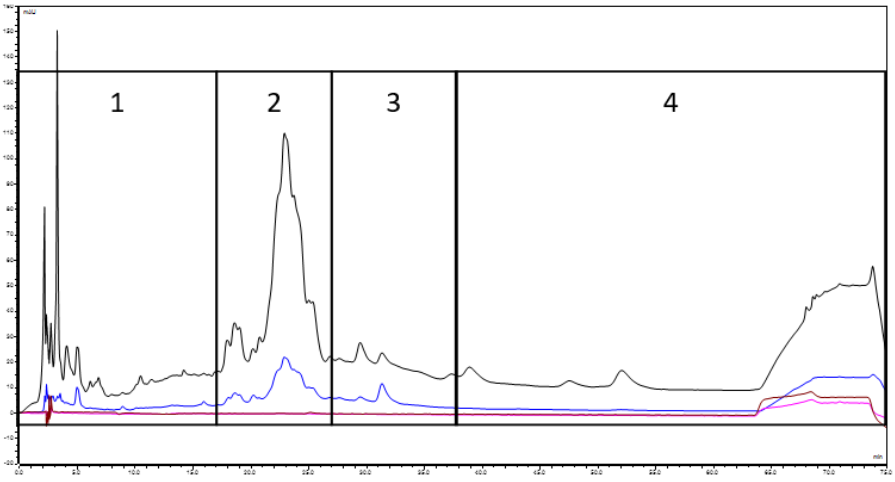


FIGURE 16 - CHROMATOGRAPHIC PROFILE OBTAINED FOR THE HPLC (BLACK – 225 NM; BLUE – 254 NM; PINK – 366 NM; BROWN – 450 NM) - THE X AXIS CORRESPONDS TO THE RETENTION TIME WHILE THE Y AXIS IS REPRESENTATIVE OF THE INTENSITY

From this fractionation, 2.86 mg of extract were obtained for a yield of 80%.

After the LC-HRESIMS analysis, it was observed that the compounds were all present in sub-fraction 3 (0.850 mg). This fraction will be used for a final fractionation that will hopefully lead to the pure compounds.

5. Discussion

During this project, ten different strains from different cyanobacterial species present in the LEGEcc were used for supplementation assays. This type of supplementation allowed for the labelling of the lipidic secondary metabolites present in each strain.

Not every compound that was detected in the cyanobacterial strains was an undiscovered secondary metabolite. Also, some of the compounds present in the LC-HRESIMS analysis were primary lipids, which is not the type of compound that was being pursued.

With that in mind, not every strain that was supplemented led to the discovery of a new natural product.

Nonetheless, during the course of this work, a compound was discovered in *Roholtiella* sp. LEGE 12411 with a $[M-H]^-$ peak at *m/z*= 1005.41147 that, after preliminary analysis, did not appear to be a primary lipid and it also was not present in any of the databases for cyanobacterial NPs that were queried.

With the aim of isolating this compound, the culture was scaled-up during the work of this dissertation, which will allow for future mass-guided isolation.

Meanwhile, another compound was pursued for isolation. It was present in *Kamptonema formosum* PCC 6407, and it had been previously identified using the same methodology. For this isolation process, enough dry biomass had already been collected.

For the isolation of this compound that, as explained before, then turned into the isolation of four different compounds, several fractionation steps were required with several different programs.

The low UV absorbance of these compounds made the process of choosing the right fractionation conditions complex.

For the final steps of the isolation process, not a lot of extract remained. Because of that, even though peaks were isolated during HPLC 05, with some of them possibly corresponding to pure compounds, further analysis of those peaks was not possible.

Nonetheless, because the fractionation process followed two different directions after SPE 02, there was still other HPLC fractions where the compounds were present and that still had enough mass for a possible isolation.

This separation was done because of the doubt that the compounds present in fraction 5 of that SPE and the compounds present in fraction 7 and onwards were isomers and not the same compound.

Nevertheless, after HPLC 07 a fraction that might allow for the isolation of compounds of interest was obtained and it will be used in future work.

If that does not occur, these compounds will be reisolated, from a larger amount of biomass, using what has been learned from these fractionations as a guide, and starting the isolation with the program that was used for FC 03.

6. Conclusion and future perspectives

This work reinforced that the FA labelling strategy to discover new cyanobacterial NPs is a very straight-forward approach that allows for a fast processing of data and the discovery of novel compounds. The dereplication part of the process eliminates what is a very frequent problem of rediscovery of NPs.

It is not capable of covering every possible compound since, in the way it was designed, it does not allow for the identification of compounds other than lipid derivatives. Still, it presents advantages in comparison to other methods and is an important addition to the cyanobacterial NPs field.

In this work a new cyanobacterial NP was discovered, and it will be isolated and characterized. Currently several litres of biomass are growing so that the isolation process of this new NP can begin.

Regarding the compounds from *Kamptonema formosum* PCC 6407, even though they reached the final steps for their isolation, the obtained amount was not enough to characterize them both in terms of chemical structure and biosynthesis.

Nevertheless, this work is in progress, and the remaining mass will be used to try to isolate these compounds and, if that is not possible, re-isolation will be attempted.

In any case, the fractionation process was optimized throughout this work, which will make the re-isolation of these compounds much more straightforward.

In general, during this work, many different techniques were used, some of the goals for this project were completed and a new NP was discovered.

The discovery of a new compound through the successful use of a recent and less traditional approach are examples of how NPs discovery, in particular in cyanobacteria, is evolving.

CHAPTER II – Study of fatty acid degradation in cyanobacteria

1. Introduction

One of the premisses that led to the development of the strategy used in Chapter I was that cyanobacteria do not have a functional β-oxidation pathway. However, during that work, after supplementation with deuterium-labelled FAs, a wave-like pattern was observed in the mass spectra showing an individual abstraction of deuterium. This observation put into question the lack of β-oxidation or if there are other types of catabolic processes in cyanobacteria.

β-oxidation is considered the most significant degradative metabolic pathway for FAs in living organisms. FA β-oxidation is one of the major pathways for energy metabolism, producing in each cycle acetyl-CoA, NADH and $FADH_2$, which then can lead to ATP production[40]. In this way, FAs allow for the storage of energy by the organisms.

Cyanobacterial FAs and their derivatives also have a high potential for being used as an alternative source of fuels, due to the high energy of their carbon-carbon and carbon-hydrogen bounds[41].

FAs together with other building blocks, such as glycerol, can form more complex lipids such as glycerophospholipids that are some of the main classes of complex lipids. These lipids are the main component of cellular membranes. Other classes of lipids exist depending on the nature of the building blocks that bind to the FAs, which gives the diversity of lipidic molecular species[42]. For example, in cyanobacteria the main lipids are monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), sulfoquinovosyldiacylglycerol (SQDG) and phospholipid phosphatidylglycerol (PG)[43].

In the membrane, the FA composition of the lipid bilayer influences the fluidity of the membrane, which can modify protein interactions and activities, influencing cell function. Membrane lipids can also be substrates for the production of lipidic secondary messengers of the signal transduction pathways[42].

The bacterial FA degradation pathway is catalysed by enzymes encoded by the fatty acid degradation (fad) regulon. These genes are responsible for the transport and activation of long-chain FAs and their β-oxidative cleavage into acetyl-CoAs. For the β-oxidation pathway, FAs are transported with the help of an outer membrane protein, FadL, and an inner membrane protein, FadD, an acyl-CoA of broad substrate specificity that activates the free fatty acids (FFA) for β-oxidation. The β-oxidation cycle, then, starts with a dehydrogenation by FadE, hydration and dehydrogenation by

FadB and a thiolation by FadA, until one turn of the cycle is completed and a shortening of the acyl-CoA by two carbons to acetyl-CoA occurs (Figure 17) [44,45].

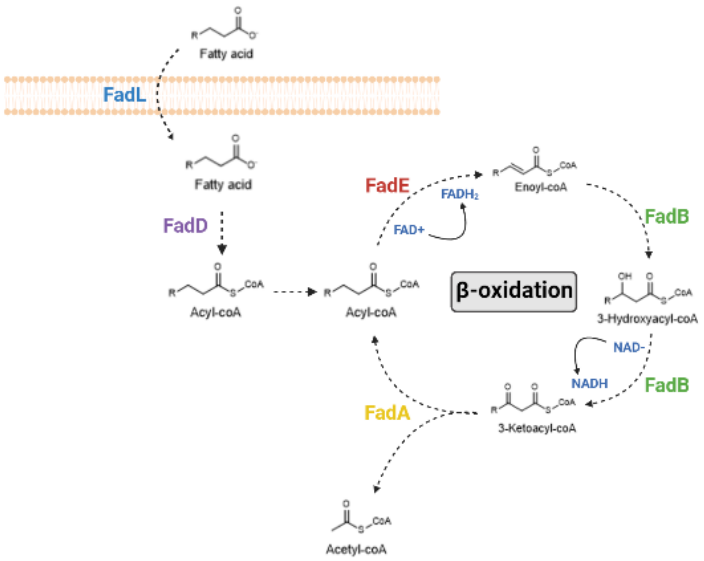


FIGURE 17 - SCHEMATIC REPRESENTATION OF THE B-OXIDATION PATHWAY (CREATED WITH BIORENDER.COM)

This pathway has been described in all organisms, except for cyanobacteria, where evidence of a β-oxidation pathway is lacking. In spite of there being some cyanobacterial strains with all of the β-oxidation genes annotated in their genome, there is no experimental evidence of an active β-oxidation pathway for FFA degradation in cyanobacteria[46]. One explanation for the lack of this important pathway in these organisms could be the fact that they can easily obtain energy from photosynthesis, and so, the presence of this pathway as a source of energy could be unnecessary[47].

The *fadD* gene in particular deservers further observation. All cyanobacterial genomes known to date encode the *fadD* gene, which encodes the first step of the β-oxidation pathway and works as a FA activating enzyme Due to its importance, without the presence of this gene the occurrence of β-oxidation is not possible. The sequences of fadD are annotated as acyl-CoA synthetases and seem to be members of the superfamily of AMP-binding proteins[48]. However, it has a much higher degree of similarity to other AMP-binding proteins of plant origin with a less established function than to acyl-CoA synthetase[48].

Although there is not a described alternative for a catabolic pathway of FAs in cyanobacteria, some hypotheses have been considered during this work, for alternative catabolic pathways that could explain the previous observations.

One of those is *α*-oxidation, which results in the oxidative removal of the first carbon, the carboxyl group, of the FA, which yields carbon dioxide (CO_2) and a FA shortened by one carbon atom. This is a process that occurs in other organisms when the FA cannot be degraded by β-oxidation[49]. On the other hand, with this pathway a loss of one carbon would be observed, but two deuterium, so it does not completely match our observations.

Another possible explanation for the observed labelling pattern is the alka(e)ne formation pathway from FAs. The alka(e)ne can then, under high-light conditions, be converted into alcohols and aldehydes by an aldehyde-deformylating oxygenase, ADO[50].

In more detail, this mechanism begins with FFAs being activated to form acyl-ACP by the action of the acyl-ACP synthetase (Aas) that catalyses the ATP-dependent esterification of FAs to the thiol of ACP. Then, the acyl-ACP is reduced by an acyl-ACP reductase (Aar), forming an aldehyde, a reactive chemical species. The aldehyde can form alka(e)nes by the action of ADO that is responsible for a deformylation of the aldehyde to an alkane. This is a particularly important step for this specific project since it is when a loss of one carbon occurs and, consequentially, the shortening of the fatty acyl chain.

The alkanes are extremely non-reactive. However, ADO is able to oxidize it into an alcohol, and then, in a second oxidation, turns the alcohol into an aldehyde. After, another enzyme, an aldehyde dehydrogenase (ALDH) can convert the aldehyde back to a FFA, now one carbon shorter.

This mechanism could explain the degradation pattern observed in the results of the supplementation assays. Still, this hypothesis, just like all the others, needs to be further explored in order to confirm if this is, indeed, the mechanism at play.

Another possibility that was considered was lipid peroxidation. The hydrogen atoms that are nearby olefinic bonds are very susceptible to oxidative damage, especially the ones that are present among unconjugated olefinic bonds. Lipids are rich in these types of bonds, which makes them targets for oxidative reactions[51]. The reaction is triggered by the action of reactive oxygen species that abstract an allylic hydrogen from polyunsaturated FAs[52].

2. Objectives

This work is part of an exploratory project – CyanoBox- that aims to gain further understanding of FA metabolism in cyanobacteria and answer some questions that have not been looked at in detail until now. In the end, it is intended, through a multidisciplinary approach, to reveal if there is, and if so, which is the FA catabolic pathway used by cyanobacteria.

To achieve that goal, a strategy divided in three tasks was designed, with each one of them focusing on one of the major parts of the work. This dissertation focused, mainly, on two of those tasks. The first one aimed at questioning if and how the FAs are degraded in cyanobacteria and to, definitively, explain the degradation pattern that is generated and the mechanisms that are responsible for it. The second task revolved around the study of the *fadD* gene, which encodes the first protein of the β-oxidation pathway and is the only one that seems to be present and annotated in all cyanobacterial genomes. For this the goal was to learn more about this protein, mainly its enzymatic activity and confirm if it is correctly annotated in the genome.

3. Materials and Methods

3.1. *In vivo* metabolic evidence for beta-oxidation

3.1.1. Supplementation assays

For these assays several cyanobacterial strains were chosen, namely *Synechocystis* sp. PCC 6803 (from here on PCC 6803); *Cyanothece* sp. PCC 7425 (from henceforth PCC 7425) and *Cyanothece* sp. PCC 7424 (from now on PCC 7424).

The first step was to grow the strain in 500 mL (day 01) in conditions mentioned in section **3.2.2.** of Chapter I.

After this, on day 04 of the experiment, the culture was split into 9 Erlenmeyers with 50 mL each – 3 for the control conditions and 3 for each of the experimental conditions. Each of the control cultures was supplemented with a 0.5 mM final concentration of palmitic acid (C_{16}:0) and each of the experimental cultures was supplemented with a 0.5 mM final concentration of perdeuterated palmitic acid (d_{31}-C_{16}) and palmitic acid uniformly labelled with carbon 13 ($U^{13}C_{16}$), from a 500x concentrated solution of each acid in DMSO.

At day 11, after seven days of supplementation, the biomass from each Erlenmeyers was collected through a centrifugation at 7500 g for 10 minutes. Then the supernatant was discarded, and the pellet with the biomass was washed and went through another centrifugation at 5000 g for 10 minutes. Afterwards the pellet was freeze dried for 24 to 48 hours.

This same assay was also done with octanoic acid (C_8:0), perdeuterated octanoic acid (d_{15}-C_8) and octanoic acid uniformly labelled with carbon 13 ($U^{13}C_8$).

3.1.2. High light assay

According to Qiao et al. an assay was constructed, in which two different white light conditions were used: a standard condition of 30 µmol m^{-2} s^{-1} and high light of 300 µmol m^{-2} s^{-1}.

For this assay PCC 6803, a cyanobacterial model strain, was used. The strain was diluted to an O.D. of 1 at 730 nm and divided in 4 tubes with 60 mL in each. The cultures in tubes 1 and 3 were supplemented with C_{16}:0 and the cultures in tubes 2 and 4 were supplemented with $U^{13}C_{16}$. The assay went on for 24 h of constant light with agitation.

This assay was performed in the Multi-Cultivator MC 1000-OD MIX (Figure 18).

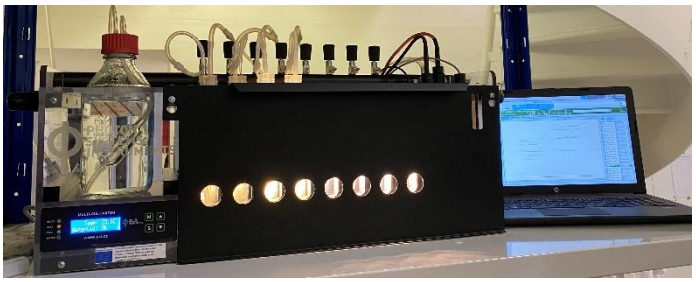


FIGURE 18 - MULTI-CULTIVATOR MC 1000-OD MIX

3.1.3. Lipidomic analysis

For the lipidomic analysis two types of extractions were performed.

The first type of extraction was the same as described in section **3.2.2.** of Chapter I.

A lipid extraction was also used, after the high light assay with PCC 6803, in order to remove non-lipid contaminants.

For this the biomass was transferred to test tubes, to which 2 mL of DCM:MeOH (2:1 v/v) were added to the dry biomass and vortexed for one minute. Afterwards the mixture was centrifuged at 2000 rpm for 10 minutes at 4° C and the extracts that were obtained were transferred to a new test tube. These steps were repeated two more times and the extracts were always transferred to the new test tube.

Next, one last centrifugation was done to eliminate any biomass that might be present, and the extracts were transferred to RB flasks. The extracts were completely dried in the rotavapor.

To remove any non-lipid contaminants, the extracts were once again dissolved in 2 mL of DCM:MeOH (2:1) to a new test tube and vortexed for one minute. Then, 0.75 mL of distilled H_2O was added followed by two minutes of vortex. The mixture was centrifuged in the same conditions as before and, after, the organic phase was collected to a new RB flask. The aqueous phase was re-extracted with 2 mL of DCM, followed by vortex and centrifugation two times after the initial extraction. The organic phase was always collected to the RB flask.

Lastly, the extract present in the RB flask was completely dried in the rotavapor and transferred with DCM to pre-weighted glass vials, which were then completely dried and kept at -20° C until further analysis.

After all the previous extractions a LC-HRESIMS analysis of each sample was done using the same equipment and conditions as described in section **3.2.4.** of Chapter I.

3.2. Characterization of the fadD

3.2.1. Isolation of *fadD* and *ACP*

3.2.1.1. *In silico* analysis

For the isolation of the necessary genes the first step was an *in silico* analysis, using Geneious Prime and the NCBI blast website[53] to blast different *fadD* and *acp* sequences from different organisms. The goal was to obtain both the sequences that were needed to design the primers for the further steps, and to align these sequences against each other using the Clustal Omega tool[54], to try to confirm if the *fadD* gene was annotated correctly.

3.2.1.2. Gene of interest and vector design

PCR was used for the cloning of the *fadD* and *acp* genes, amplified from cyanobacterial genomic DNA of *Calothrix brevissima* NIES-22, from the National Institute of Environmental Studies (henceforth NIES-22). All constructs were coned into a pET28a plasmid (5369 bp), containing a C-terminal His-Tag for subsequent purification on a Ni-NTA and kanamycin (Kan) resistance cassette for selection. For both clonings, the multiple cloning site (MCS) of the pET28a plasmid was analysed as well as the restriction sites in the inserts, and NcoI and XhoI were selected as the restriction enzymes to be used.

For the *fadD* cloning, 2 bp were added to be aligned with the reading frame of the gene (T and G base pair that transcribed into a Valine) and the ATG was removed from the gene of interest as NcoI restriction site of pET28a already contained an ATG. The following primers were obtained:

TABLE 10 - PRIMERS FOR THE FADD CLONING

Primer Fw	ggctgCCATGGTG CCAGTAAATCTCCAGTTAATTAC
Primer Rev	agtgcCTCGAG TCGAGATGCGCGAAATAG

These primers gave a PCR product of 1508bp and a final product of 6734bp.

After looking at these primers using NEB TM calculator[55] and selecting the Q5 as the Master Mix, a calculated annealing temperature of 59° C was obtained. Then the primers were checked with OligoCalc[56] and OligoEvaluator[57] and it gave a warning that the Rev primers might create a self-dimer.

For the cloning of the *acp*, 2 bp were added to be aligned with the reading frame of the gene (T and G base pair transcribed into a Valine) and the ATG was removed from the gene of interest as NcoI restriction site of pET28a already contain an ATG. This resulted in the following primers:

TABLE 11 - PRIMERS FOR THE ACP CLONING

Primer Fw	ggagcCCATGGTG AGCAAAGAAGAA TGTTTGAAAAAG
Primer Rev	attagCTCGAG GGCAGATGTAGTAGGTTCTCC

These primers gave rise to a PCR product of 282bp and a final product of 5519bp.

These primers were evaluated using the same protocol as for the *fadD* primers and the calculated annealing temperature was 63° C. This time no warning was given after verifying with OligoCalc[56] and OligoEvaluator[57].

3.2.1.3. Extraction of gDNA

The gDNA of NIES-22, was extracted from a 200 mL culture with the NZY Plant/Fungi gDNA Isolation Kit, according to the manufacturer's instructions.

The extraction was done in duplicate, with a final concentration of 20.329 ng μL^{-1} and 14.728 ng μL^{-1}. The sample with the biggest concentration was used for the following steps.

3.2.1.4. Amplification of genes with PCR

With the gDNA that was obtained, a PCR reaction was done with the NEB Q5 2x High-Fidelity Master Mix (Table 12), for both *fadD* (Table 13) and *acp* (Table 14), using the primers designed for both as described in Tables 10 and 11. For this first PCR, which was used for TM optimization, several PCR reactions were done to test different TM for the annealing. Those TM, for both genes, are described in Table 15. The controls were done with the calculated TM.

TABLE 12 - PCR REACTION

Component	Volume (50 µL)
Q5 Master Mix	25 µL
10 µM Fw Primer	2.5 µL
10 µM Rev Primer	2.5 µL
DNA	1.5 µL
H_2O	18.5 µL

This reaction was divided in 5 PCR tubes of 10 µL each to test all the different TM (Table 15) and to have one negative control without DNA and 20 µL of H_2O.

TABLE 13 - PCR PROGRAM FOR FADD AMPLIFICATION

Step	Temperature/Time	Cycles
Initial Denaturation	98° C / 30 s	1x
Denaturation	98° C /10 s	
Annealing	Table 15 / 30 s	35x
Extension	72° C / 45 s	
Final Extension	72° C / 2 min	1x

TABLE 14 - PCR PROGRAM FOR ACP AMPLIFICATION

Step	Temperature/Time	Cycles
Initial Denaturation	98° C / 30 s	1x
Denaturation	98° C /10 s	
Annealing	Table 15 / 30 s	35x
Extension	72° C / 15 s	
Final Extension	72° C / 2 min	1x

TABLE 15 - DIFFERENT TM USED FOR THE PCR PROGRAMS

fadD	56° C	59° C	62° C	65° C
acp	58° C	63° C	65° C	67° C

After these PCR reactions, an electrophoresis was done, with a 1.5 % agarose gel for the *acp* and a 1% agarose gel for the *fadD*. The PCR products were run into a BioRad electrophoresis machine, at 90 V for 35 minutes. The gel was then put into a Gel Doc EZ Imager (BioRad) an analysed using Image Lab 6.0.

After this electrophoresis it was concluded that the optimal TM for the *fadD* was 62° C and for the *acp* 63° C.

Next, the PCR was repeated with the same final volume and with the same program, as described before, but using only one PCR tube with the 50 µL and the best TM for each gene, determined after analysing the results from the previous PCR.

PCR samples were run into an agarose gel as previously described to check if the band was clear. Next, the PCR product was purified with a gel cleaning kit (BioLabs).

The DNA concentration was measured after the purification and the DNA concentration for *fadD* was of 118.058 ng μL^{-1} and 130.837 ng μL^{-1} for the *acp.*

3.2.2. Cloning in *E. coli*

After amplification, using these concentrations, a reaction with the restriction enzymes (NcoI and XhoI) for both the inserts and the plasmid was prepared, using the rCutSmart buffer (BioLabs). The reactions are described in Tables 16 and 17 for the inserts and the plasmid, respectively.

TABLE 16 - MIXTURE FOR THE REACTION WITH THE R.E AND THE INSERTS

Component	Volume (50 µL)
Buffer	5 µL
R.E.	1 µL each
DNA	8.47 µL (f*adD*) / 7.64 µL (*ACP)*
H_2O	35.53 µL (f*adD*) / 36.36 µL (*ACP)*

TABLE 17 - MIXTURE FOR THE REACTION WITH THE R.E AND THE PLASMID

Component	Volume (50 µL)
Buffer	5 µL
R.E.	1 µL each
DNA	15 µL
H_2O	28 µL

For the plasmid triplicates reactions were performed. Reactions were put to incubation at 37° C for 30 minutes. After that time, an electrophoresis was done using big wells to be able to remove the bands and purify them with the purification kit mentioned before. Then the DNA concentration was measured (Table 18).

TABLE 18 - DNA CONCENTRATION AFTER ELECTROPHORESIS GEL BAND PURIFICATION

DNA Concentration – ng/µL			
43.335	42.995	22.069	*pet28*
	33.283		*fadD*
	37.065		*ACP*

Afterwards, ligation between the plasmid with the highest concentration and the inserts was done, using the T4 DNA ligase (BioLabs). The volume of insert and plasmid to be used was calculated using the NEBioCalculator for a 3:1 ratio respectively. The reaction mixtures are described in Tables 19 and 20.

TABLE 19 - MIXTURE FOR THE LIGATION REACTION BETWEEN PET28A AND FADD

Component	Volume (10 µL)
T4 ligase buffer	1 µL
T4 ligase	0.5 µL
fadD	1.39 µL
pet28a	1.15 µL
H_2O	5.96 µL

TABLE 20 - MIXTURE FOR THE LIGATION REACTION BETWEEN PET28A AND ACP

Component	Volume (10 µL)
T4 ligase buffer	1 µL
T4 ligase	0.5 µL
ACP	0.22 µL
pet28a	1.16 µL
H_2O	7.12 µL

The mixture was left overnight at room temperature, and further incubated at 65° C for 10 minutes to inactivate the restriction enzymes.

The last step was to do a transformation with *E. coli* competent cells. For this *E. coli* XL1 competent cells were used. Two reactions were done, one where 5 µL of the ligation with the *fadD* was added and another with 5 µL of the ligation with the *acp.* The mixtures were left on ice for 30 minutes, incubated at 42° C for 45 seconds and put back on ice for two minutes. After that, 250 µL of Luria Broth (LB) medium was added to the mixtures that were then incubated at 37° C for one hour. At the end of that time 150 µL of the transformed competent cells were platted onto LB plates with 50 µg mL^{-1} of kan. Plates were further incubated at 37° C overnight.

On the next day, some colonies were picked up to perform a colony-PCR with the NZTaq II 2x green master mix (NZYTech). The PCR mixture and program are described in Tables 21 and 22, respectively. For this reaction T7 forward and reverse primers were used.

TABLE 21 - MIXTURE FOR THE COLONY-PCR

Component	Volume (50 µL)
Primer	2 µL
DNA	-
Master Mix	25 µL
H_2O	23 µL

TABLE 22 - PCR PROGRAM FOR COLONY-PCR

Step	Temperature/Time	Cycles
Initial Denaturation	95° C / 3 min	1x
Denaturation	94° C /30 s	35x
Annealing	Table 15 / 30 s	
Extension	72° C / 15 s (*ACP)* / 45 s (*fadD*)	
Final Extension	72° C / 10 min	1x

The colonies chosen for the PCR were put in liquid LB medium with kan 50 µg mL^{-1} and left to grow overnight at 37° C.

On the next day, positive colonies, meaning the colonies containing the inserts, were used to perform a miniprep with the NZYTtech kit. Afterwards, the DNA concentration was measured, a glycerol stock was done, and the colonies were sent for sequencing.

Both the *fadD* and the *acp* constructs were confirmed by sequencing.

4. Results

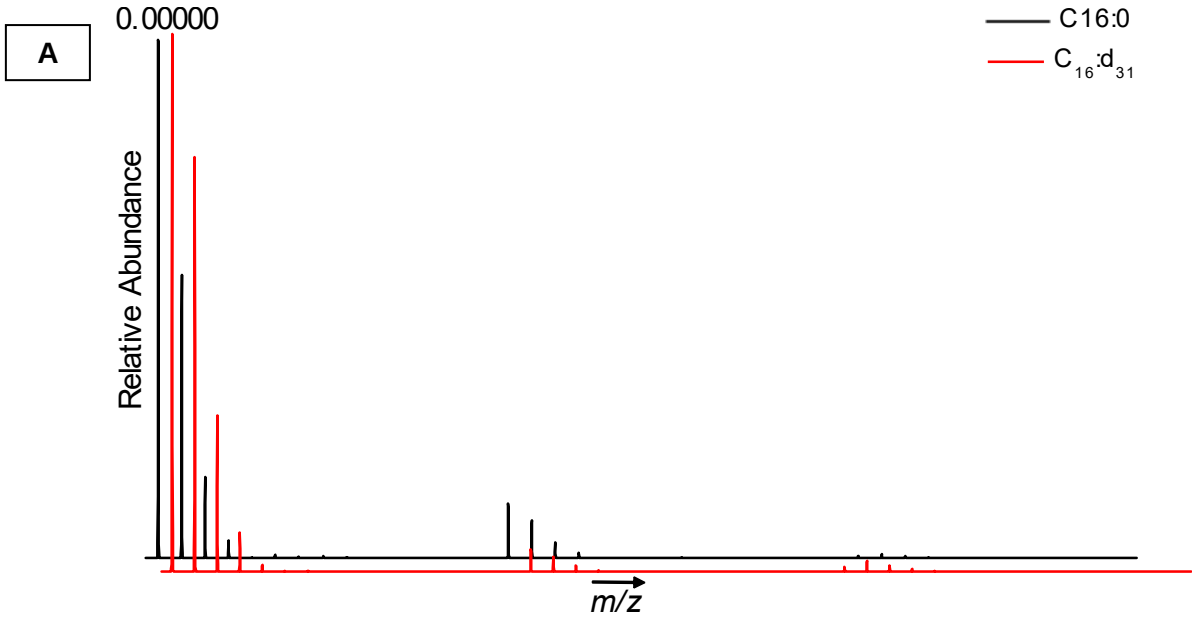
4.1. *In vivo* metabolic evidence for β-oxidation

In order to study the catabolism of FAs in cyanobacteria and to determine if they indeed lack the mechanism of β-oxidation, PCC 6803, PCC 7425, and PCC 7424 were used for supplementation assays with isotopic-labelled long-chain (C_{16}) and short-chain (C_8) FAs.

As a model cyanobacteria, the data from a previous seven days supplementation assay on PCC 6803 served as a guide for the studies that were done.

In this supplementation C_{16}, d_{31}-C_{16} and $U^{13}C_{16}$ were used. An intense wave-like pattern of degradation of deuterons was not observed, unlike with other supplemented strains (Figure 19 – A).

From the supplementation with $U^{13}C_{16}$ it was possible to observe an individual abstraction of carbons in a wave-like pattern (Figure 19 – B), with different intensities in different lipids.



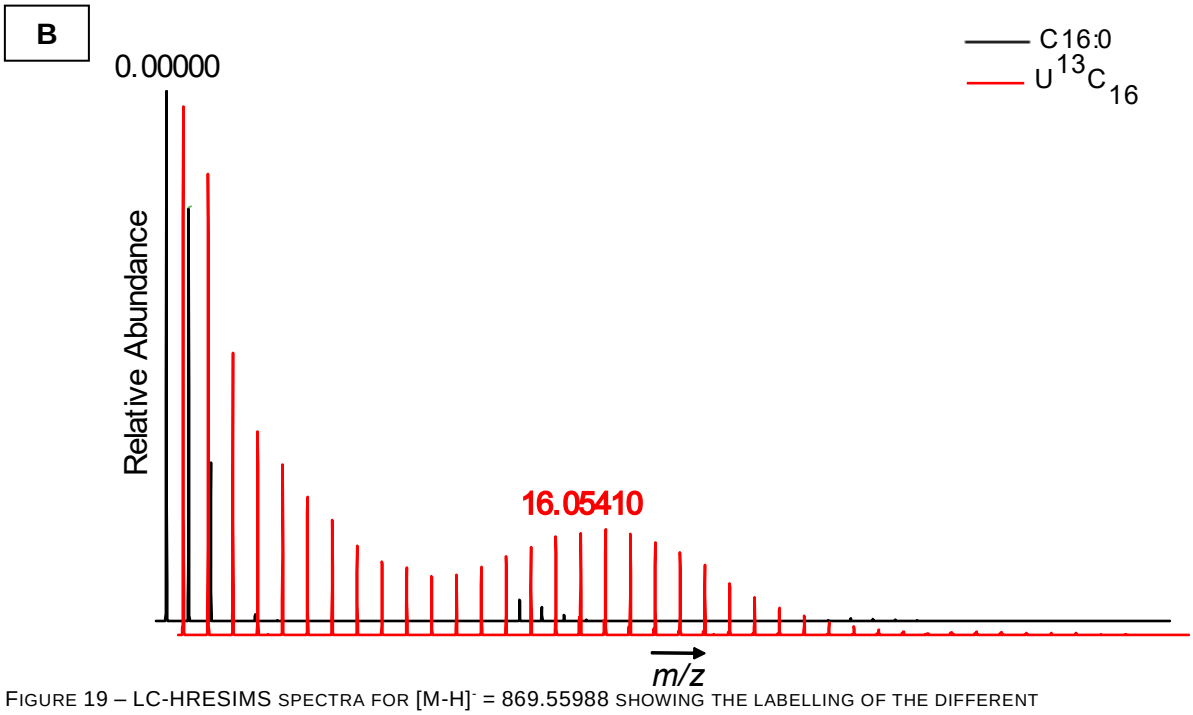


FIGURE 19 – LC-HRESIMS SPECTRA FOR $[M-H]^-$ = 869.55988 SHOWING THE LABELLING OF THE DIFFERENT EXPERIMENTAL CONDITIONS. A – CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH DEUTERIUM (D_{31}-C_{16}) (RED); B - CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH CARBON 13 ($U^{13}C_{16}$) (RED)

PCC 7425 was also tested, which is a strain that possess homologs for all of the previously described β-oxidation genes annotated in its genome. This strain was supplemented with the long-chain FAs.

After analysis it was concluded that, for PCC 7425, the labelling pattern was different from what was described before for PCC 6803.

In case of Figure 20, it shows the example of SQDG 32:0, a lipid with a peak at $[M-H]^-$ = 793.51593 *m/z*. In this case, there is an intense peak that corresponds to $+d_{31}$ after supplementation with d_{31}-C_{16}, that corresponds to the same lipid with an intact labelling of 31 deuterons (Figure 20 – A). The same happened after supplementation with $U^{13}C_{16}$. In this case, there is a clear labelling of $+C_{16}$ and $+C_{32}$, corresponding to one or two fatty acyl chains, respectively (Figure 20 - B). Contrary to what happened with PCC 6803, in this strain there was no detection of an individual abstraction of neither deuterium nor carbon-13.

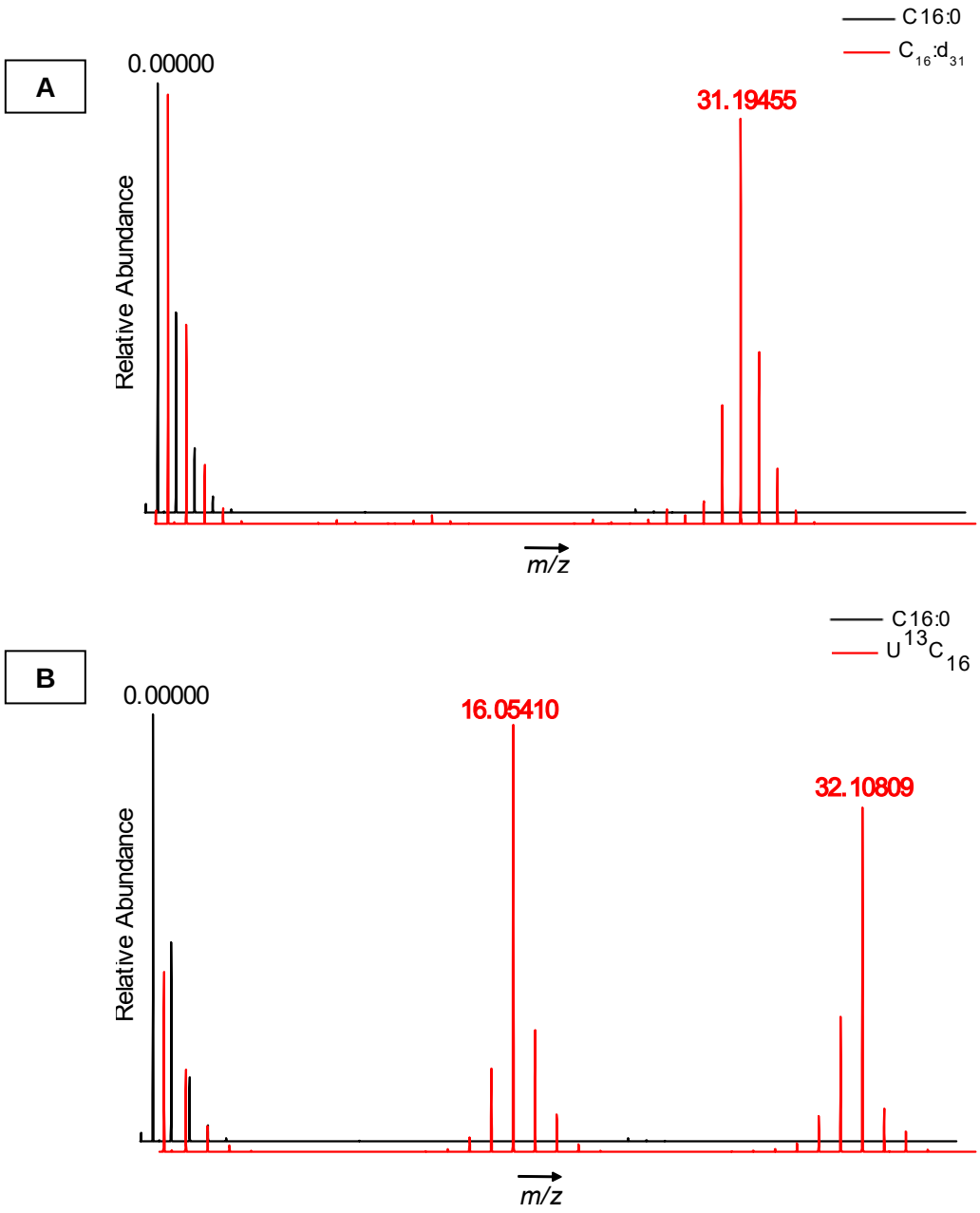


FIGURE 20 – LC-HRESIMS SPECTRA FOR SQDG 32:0 SHOWING THE LABELLING OF THE DIFFERENT EXPERIMENTAL CONDITIONS. A – CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH DEUTERIUM (D_{31}-C_{16}) (RED); B - CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH CARBON 13 ($U^{13}C_{16}$) (RED)

Interestingly, in other lipids, such as in the case of PG34:0 with a $[M+HCO_2]^-$ peak at *m/z* 795.52801, a $+C_6$ and a $+C_{22}$ labelling was observed after supplementation with $U^{13}C_{16}$ (Figure 21 – B), which had not been detected in the previous example. Also, this lipid had a $+C_{29}$ labelling instead of a $+C_{31}$, which is indicative of an unsaturation (Figure 21 – A).

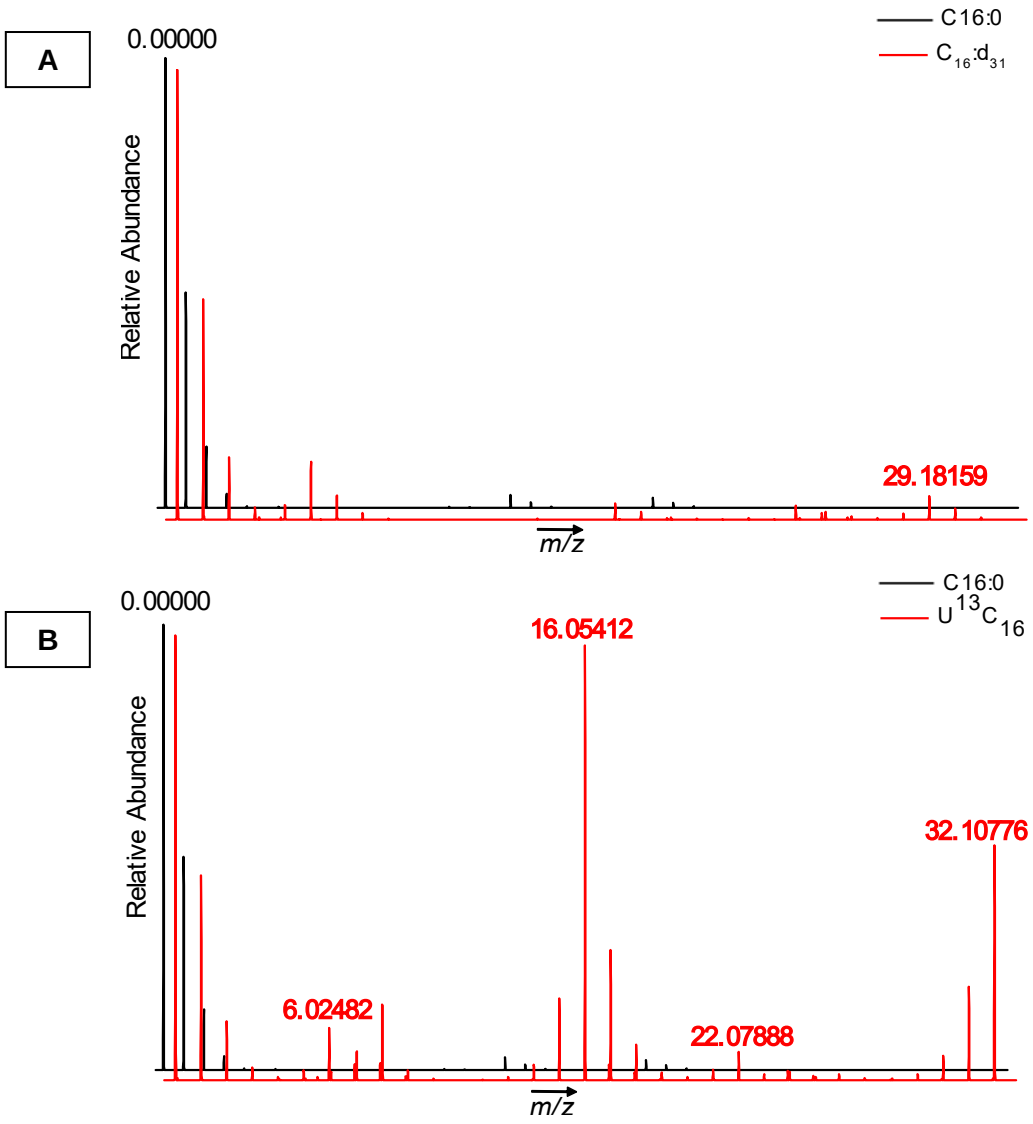


FIGURE 21 - LC-HRESIMS SPECTRA FOR PG 34:0 SHOWING THE LABELLING OF THE DIFFERENT EXPERIMENTAL CONDITIONS. A – CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH DEUTERIUM (D_{31}-C_{16}) (RED); B - CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH CARBON 13 ($U^{13}C_{16}$) (RED) FROM PCC 7425

In another supplementation, different light conditions were used to test for FA degradation in stress conditions in PCC 6803, following the work from Qiao et al.. In this case, after 24 hours, and contrary to what was observed for in the seven days assay (Figure 19), it was possible to see an intact labelling of $+C_{16}$ for the supplementation with $U^{13}C_{16}$ (Figure 22). However, during this time, there was no degradation pattern, and the labelling is not very abundant. Besides, there was no significant difference, after 24 hours, between the low light condition (Figure 22 – A) and the high light condition (Figure 22 – B).

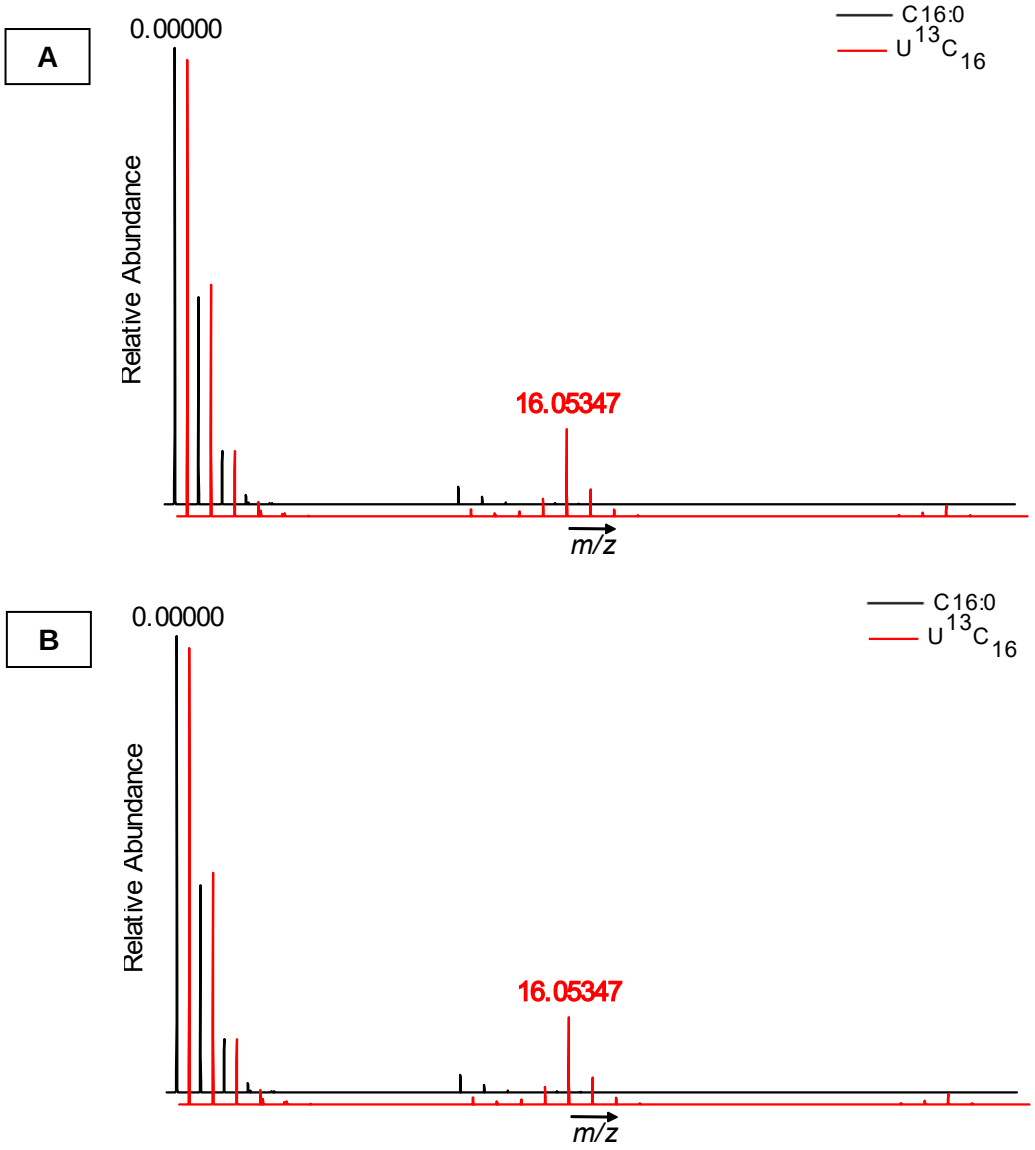


FIGURE 22 -LC-HRESIMS SPECTRA FOR $[M-H]^-$ = 745.50262 SHOWING THE LABELLING OF THE DIFFERENT EXPERIMENTAL CONDITIONS. A – CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH CARBON 13 ($U^{13}C_{16}$) FOR STANDARD LIGHT (RED) (RED); B - CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH CARBON 13 ($U^{13}C_{16}$) FOR HIGH LIGHT (RED)

Another strain used for a supplementation assay was PCC 7424, which does not have the *ado/aar* genes that are related to the alkane degradation pathway, previously mentioned as a hypothesis for the observed FA degradation in cyanobacteria.

For this strain both long-chain and short-chain FAs were used for supplementation.

In Figure 23 there is an example of PG 32:1, a lipid with a peak at $[M-H]^-$ *m/z* = 719.48792 that was supplemented with long-chain FAs. For this lipid, there was a degradation wave in the supplementation with $U^{13}C_{16}$ that had been typical for other strains supplemented in the same manner, such as PCC 6803 (Figure 19). There was

not, however, a presence of the $+C_{31}$ labelling in some of the lipids supplemented with d_{31}-C_{16}, which had also been noticed for other strains before.

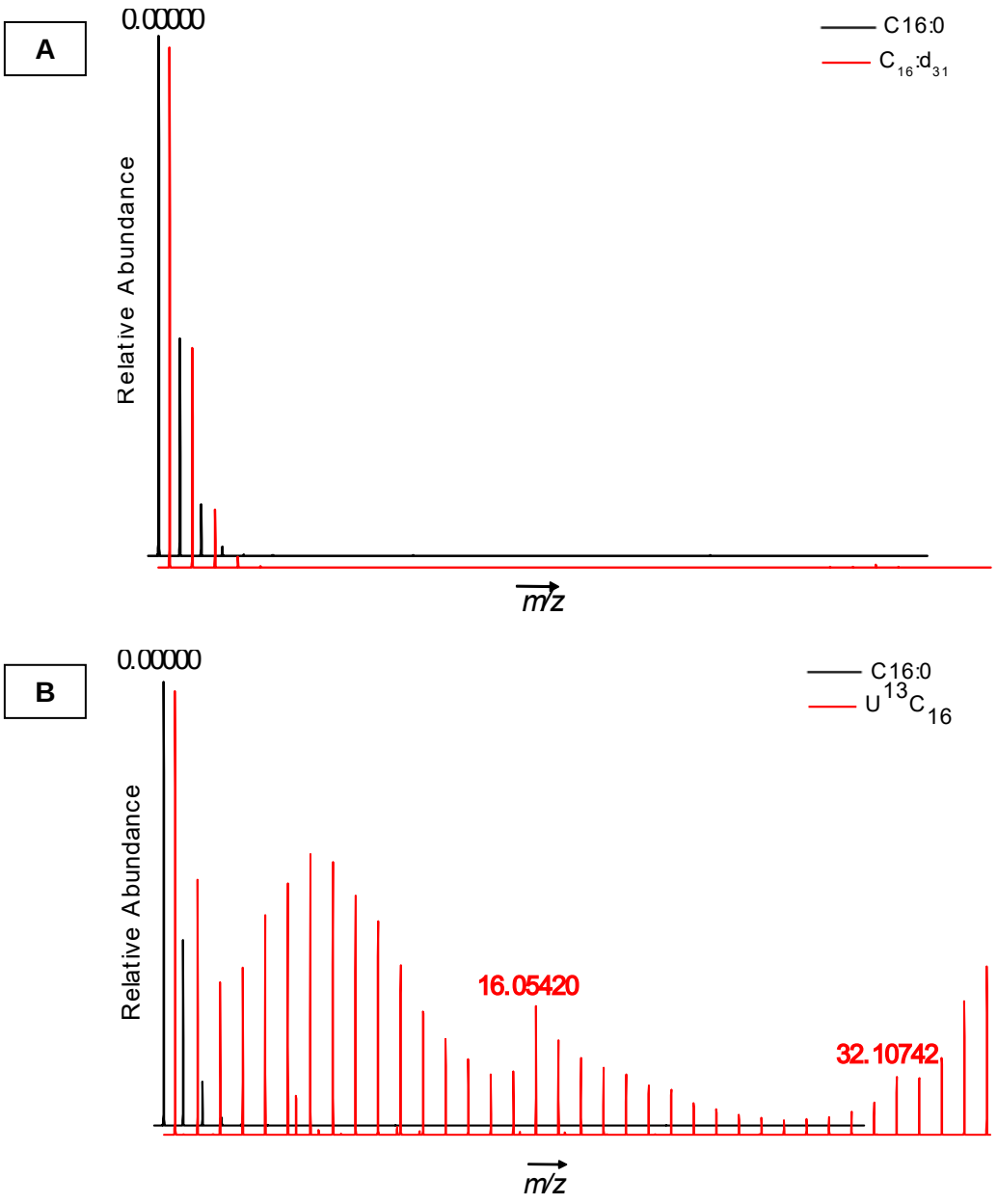


FIGURE 23 - LC-HRESIMS SPECTRA FOR PG 32:1 SHOWING THE LABELLING OF THE DIFFERENT EXPERIMENTAL CONDITIONS. A – CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH DEUTERIUM (D_{31}-C_{16}) (RED); B - CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH CARBON 13 ($U^{13}C_{16}$) (RED)

The same pattern with a wave-like degradation pattern was present in the samples supplemented with C_8. This pattern was visible for, for example, PG 32:1 (Figure 24 – B), which, unlike what was observed for some of the lipids in the supplementation with C_{16}, always showed a deuterium labelling, in this case of $+d_{15}$ (Figure 24 – A)

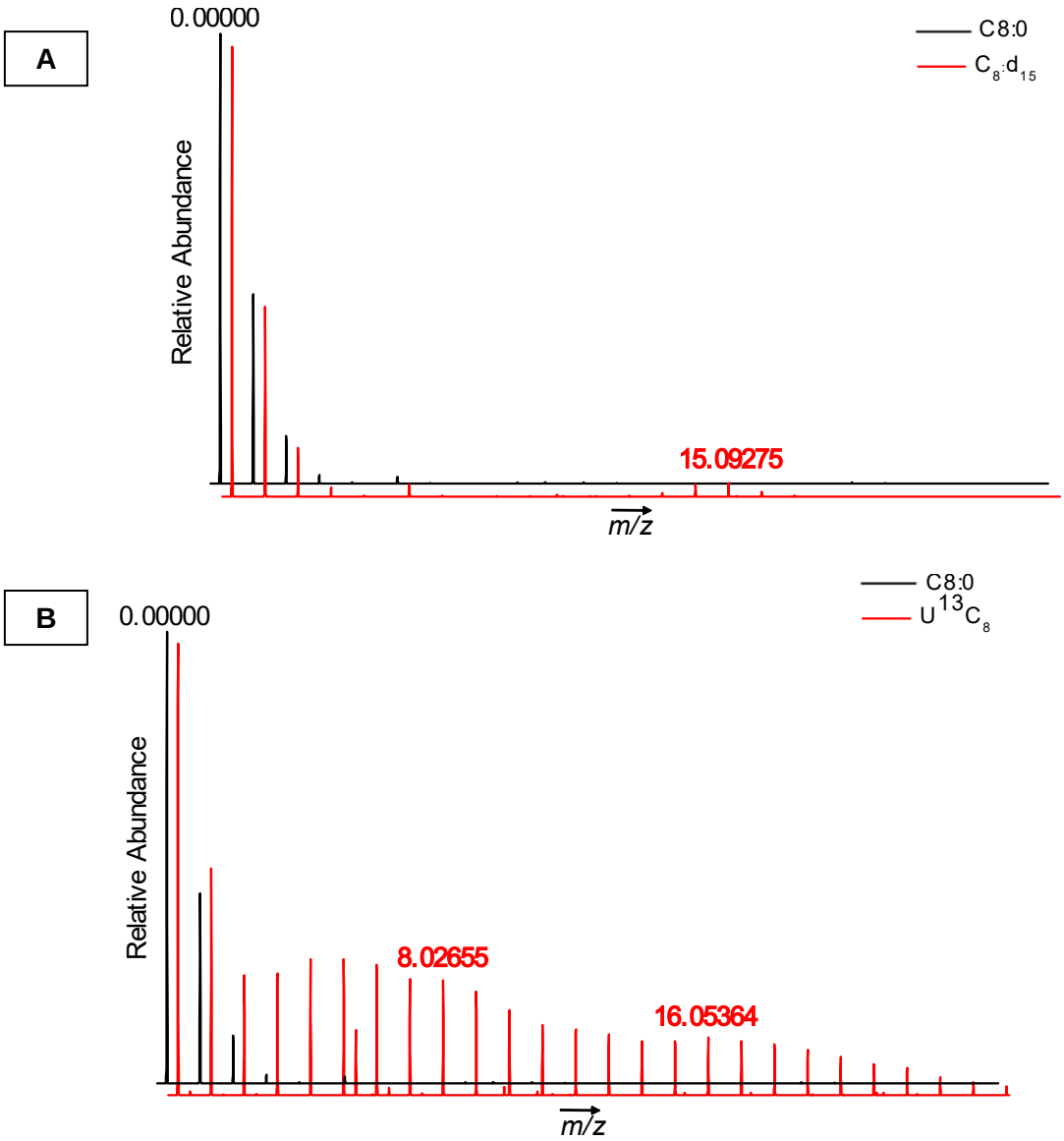


FIGURE 24 - LC-HRESIMS SPECTRA FOR PG 32:1 SHOWING THE LABELLING OF THE DIFFERENT EXPERIMENTAL CONDITIONS. A – CONTROL CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH DEUTERIUM (D_{31}-C_{16}) (RED); B - CONDITION (C_{16}:0) (BLACK) AND SUPPLEMENTATION WITH CARBON 13 ($U^{13}C_{16}$) (RED)

4.2. Characterization of the *fadD* gene

Both *fadD* and *acp* genes from Nies-22, which has all the *fad* genes, were successfully cloned into *E. coli* pET28a containing a C-terminal histidine tag for further purification. The agarose gel results of the plasmid cut with NcoI and XhoI showed a band of 1508 bp corresponding to the *fadD* insert and a band of 282 bp corresponding to the *acp* from Nies-22.

The sequencing results for both genes showed that they had been inserted into pET28a and confirmed the integrity and fidelity of each cloned sequence (Figure 25).

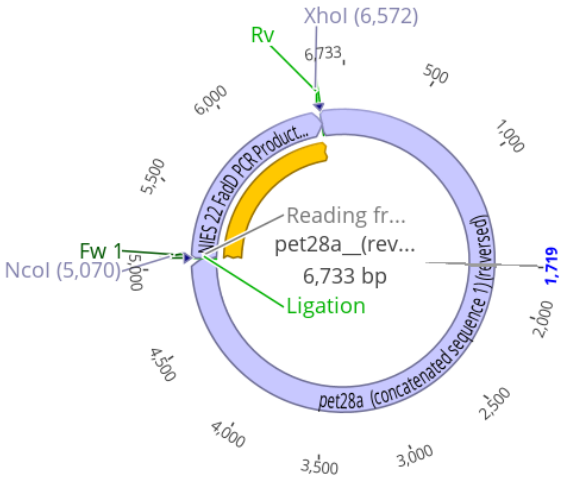


FIGURE 25 - PET28A WITH FADD INSERT

5. Discussion

For this work different cyanobacterial strains were chosen to be supplemented in order to evaluate the different hypothesis related to FA catabolism.

Previous to this dissertation, the first seven days assay using $U^{13}C_{16}$ was done on PCC 6803. This supplementation showed that, besides the individual abstraction of deuterium that had been observed in previous studies, there was also an individual abstraction of carbon. With this, the hypothesis of lipid peroxidation being the cause of the pattern that had been observed was disproven, since for that to be the case, the aliphatic chain would have to remain unchanged.

From the supplementation of PCC 7425, a strain with homologs for all β-oxidation genes annotated, there was no evidence of individual abstraction of carbon or deuterium. This may be due to the interference of the *fad* genes when they are all present in the genome.

At the same time, there were lipids with a presence of a $+C_6$ and a $+C_{22}$ labelling, which is consistent with β-oxidation. The $+C_{22}$ might represent a double chain in which one chain remained with 16 carbons and the other was shortened two carbons at a time, to a molecule with six labelled carbons.

Related or not with β-oxidation, this cyanobacterial strain that has all the genes related to that pathway seems to present a different FA degradation pattern.

In the 24 hours high-light assay it was only possible to observe the intact labelling both with deuterium and with carbon-13 and there was no labelling that could indicate degradation. Thus, it cannot be inferred if the high light that would induce the alka(e)ne biodegradation is related or not with the FA degradation pattern observed in the lipid analysis.

Nonetheless, this experiment proved to be relevant since it showed that the labelled FAs are in first instance incorporated in the lipids without suffering any degradation. Consequently, this data allowed for the conclusion that the degradation pattern observed after seven days of supplementation derived from an intact labelling.

A supplementation result that was particularly interesting was the one from PCC 7424. At the moment of this supplementation assay the hypothesis that the alka(e)ne degradation pathway could be responsible for the observed wave-like degradation pattern was being considered. And so, it was expected that this strain would not show

a degradation pattern, since PCC 7424 lacks two of the enzymes involved in the alka(e)ne degradation pathway.

Unlike what was expected, the supplementation results showed the wave-like pattern of degradation that had been observed in other strains, both with C_{16} and C_8 supplementation. This may be indicative that the alka(e)ne pathway is not responsible for the abstraction of carbon molecules in FA.

As for the cloning results, they showed that both genes of interest were successfully isolated, to now continue to the next steps of the work.

6. Conclusion and future perspectives

This work has helped gain important insight into FA catabolism in cyanobacteria. Although there are still several questions that remain unanswered, some hypotheses have been rejected and there is a clearer idea of the fate of the FFAs after supplementation. Moreover, after analysing these results, future experiments were planned, which will lead to more definitive answers on how FAs are catabolized in cyanobacteria.

With PCC 7425 longer assays will be done to confirm if the lack of degradation of the $+C_{16}$ and the $+d_{31}$ labelling continues, and also to analyse possible changes in the $+C_6$ labelling. The results of this assay may lead to a supplementation with different timepoints to pinpoint the time frame in which the degradation of the FAs or the appearance of the $+C_6$ labelling occurs.

For PCC 6803, a regular supplementation assay must be done with more time and different timepoints, in order to observe the evolution of the wave-like pattern. After those results are obtained, another assay with different light intensities should be done in order to determine the effect of stress conditions, which induce the alka(e)ne pathway, in the labelling and, consequentially, in the FA degradation.

After the results showed for PCC 7424, a knockout mutant for the alka(e)ne pathway genes will be constructed to definitively conclude if this pathway is related to FA catabolism.

Besides this work, other cyanobacterial strains may be screened based on the future construction of phylogenetic trees with the *fad* genes where these genes will be searched for to determine, through time of evolution between strains and between a specific strain and its ancestral, which are the best cyanobacteria to use. This tool will assist with determining the relation between the FA metabolism and cyanobacterial evolution.

For the work related to the better understanding of the *fadD* gene, the protein will be expressed and *in vitro* assays will be performed to study its mechanism of action, in order to make clear the annotation of the gene and how it is related to the previous observations in cyanobacteria.

Conclusion

With this dissertation work two different projects have been tackled, both of which intend to increase knowledge about cyanobacteria and the compounds it produces. Great progress has been made during this time and there are several plans for future work.

The isolation process of the compounds from PCC 6407 will continue, to later proceed to the structural characterization.

In terms of the β-oxidation project, different theories have been studied and all the data that has been collected will be used to clarify all the questions that began this work. A better understanding of cyanobacterial FA metabolism will also be important to improve how FAs are used to discover new NPs in the future.

In general, this was a multidisciplinary work, with a focus on analytical chemistry, organic chemistry, molecular biology and even bioinformatics. This approach has allowed for the application of knowledge from different fields with the goal of not just contributing for the cyanobacterial NPs field, but also gain more insight into the metabolism of these organisms.

References

1. Schirrmeister, B. E., Antonelli, A. & Bagheri, H. C. The origin of multicellularity in cyanobacteria. *BMC Evol. Biol.* **11**, (2011).

2. Paerl, H. W., Pinckney, J. L. & Steppe, T. F. Cyanobacterial-bacterial mat consortia: Examining the functional unit of microbial survival and growth in extreme environments. *Environ. Microbiol.* **2**, 11–26 (2000).

3. Zahra, Z., Choo, D. H., Lee, H. & Parveen, A. Cyanobacteria: Review of current potentials and applications. *Environ. - MDPI* **7**, (2020).

4. Tan, L. T. & Phyo, M. Y. Marine cyanobacteria: A source of lead compounds and their clinically-relevant molecular targets. *Molecules* **25**, (2020).

5. Van Der Hooft, J. J. J. *et al.* Linking genomics and metabolomics to chart specialized metabolic diversity. *Chem. Soc. Rev.* **49**, 3297–3314 (2020).

6. Figueiredo, S. A. C. *et al.* Discovery of Cyanobacterial Natural Products Containing Fatty Acid Residues**. *Angew. Chemie - Int. Ed.* **60**, 10064–10072 (2021).

7. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).

8. Abdel-Aziz, S. M., Abo Elsoud, M. M. & Anise, A. A. H. *Chapter 2 - Microbial Biosynthesis: A Repertory of Vital Natural Products A2 - Grumezescu, Alexandru Mihai*. *Handbook of Food Bioengineering* (Elsevier Inc., 2017). doi:10.1016/B978-0-12-811372-1/00003-8.

9. Harvey, A. L. Natural products in drug discovery. *Drug Discov. Today* **13**, 894–901 (2008).

10. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **14**, 111–129 (2015).

11. Medema, M. H. *et al.* HHS Public Access. **11**, 639–648 (2016).

12. Hong, J. Role of natural product diversity in chemical biology. *Curr. Opin. Chem. Biol.* **15**, 350–354 (2011).

13. Jiménez, C. Marine Natural Products in Medicinal Chemistry. *ACS Med. Chem. Lett.* **9**, 959–961 (2018).

14. Machado, H., Tuttle, R. N. & Jensen, P. R. Omics-based natural product discovery and the lexicon of genome mining. *Curr. Opin. Microbiol.* **39**, 136–142 (2017).

15. Chiang, Y. M., Chang, S. L., Oakley, B. R. & Wang, C. C. C. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr. Opin. Chem. Biol.* **15**, 137–143 (2011).

16. Fox Ramos, A. E., Evanno, L., Poupon, E., Champy, P. & Beniddir, M. A. Natural products targeting strategies involving molecular networking: Different manners, one goal. *Nat. Prod. Rep.* **36**, 960–980 (2019).

17. Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst* **137**, 293–300 (2012).

18. Sindelar, M. & Patti, G. J. Chemical Discovery in the Era of Metabolomics. *J. Am. Chem. Soc.* **142**, 9097–9105 (2020).

19. Du, X., Smirnov, A., Pluskal, T., Jia, W. & Sumner, S. *Metabolomics Data Preprocessing using ADAP and MZmine 2*. (2021). doi:10.1007/978-1-0716-0239-3.

20. Quinn, R. A. *et al.* Molecular Networking [1 _ TD $ IF] As a Drug Discovery , Drug Metabolism , and Precision Medicine Strategy. *Trends Pharmacol. Sci.* **xx**, 1–12 (2016).

21. Rippka, R., Deruelles, J. & Waterbury, J. B. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **111**, 1–61 (1979).

22. Whitton, B. A. Ecology of cyanobacteria II: Their diversity in space and time. *Ecol. Cyanobacteria II Their Divers. Sp. Time* **9789400738**, 1–760 (2012).

23. Brito, Â. *et al.* Culture-dependent characterization of cyanobacterial diversity in the intertidal zones of the Portuguese coast: A polyphasic study. *Syst. Appl. Microbiol.* **35**, 110–119 (2012).

24. Abed, R. M. M., Dobretsov, S. & Sudesh, K. Applications of cyanobacteria in biotechnology. *J. Appl. Microbiol.* **106**, 1–12 (2009).

25. Medema, M. H. & et al. Minimum information about a biosynthetic gene cluster HHS public access author manuscript. *Nat Chem Biol* **11**, 625–631 (2015).

26. Chen, R. *et al.* Discovery of an Abundance of Biosynthetic Gene Clusters in

Shark Bay Microbial Mats. *Front. Microbiol.* **11**, 1–13 (2020).

27. Dittmann, E., Gugger, M., Sivonen, K. & Fewer, D. P. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. *Trends Microbiol.* **23**, 642–652 (2015).

28. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta - Proteins Proteomics* **1865**, 1587–1604 (2017).

29. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9259–9264 (2014).

30. Wang, C. *et al.* Chemometrics and genome mining reveal an unprecedented family of sugar acid-containing fungal nonribosomal cyclodepsipeptides. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–12 (2022).

31. Verma, S., Thapa, S., Siddiqui, N. & Chakdar, H. Cyanobacterial secondary metabolites towards improved commercial significance through multiomics approaches. *World J. Microbiol. Biotechnol.* **38**, 1–22 (2022).

32. Zhang, J. J. & Moore, B. S. Natural Product Reports Genetic platforms for heterologous expression of microbial natural products. (2019) doi:10.1039/c9np00025a.

33. von Berlepsch, S. *et al.* The acyl-acyl carrier protein synthetase from Synechocystis sp. PCC 6803 mediates fatty acid import. *Plant Physiol.* **159**, 606–617 (2012).

34. Beld, J., Abbriano, R., Finzel, K., Hildebrand, M. & Burkart, M. D. Probing fatty acid metabolism in bacteria, cyanobacteria, green microalgae and diatoms with natural and unnatural fatty acids. *Mol. Biosyst.* **12**, 1299–1312 (2016).

35. Kotai, J. Instructions for preparation of modified nutrient solution Z8 for algae. *Nor. Inst. Water Res.* 1–5 (1972).

36. Kai Blin, Simon Shaw, Alexander M Kloosterman, Zach Charlop-Powers, Gilles P van Weezel, Marnix H Medema, & T. W. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* (2021) doi:10.1093/nar/gkab335.

37. Dictionary of Natural Products 31.1. doi:https://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml;jsessionid=E28C89125473FB4AA7B679D7BEED0612.

38. van Santen, J. A.; Jacob, G.; Leen Singh, A.; Aniebok, V.; Balunas, M. J.; Bunsko, D.; Carnevale Neto, F.; Castaño-Espriu, L.; Chang, C.; Clark, T. N.; Cleary Little, J. L.; Delgadillo, D. A.; Dorrestein, P. C.; Duncan, K. R.; Egan, J. M.; Galey, M. M.; H, R. G. The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Central Science, 5, 11* 1824–1833 (2019) doi:10.1021/acscentsci.9b00806.

39. Sud M., Fahy E., Cotter D., Brown A., Dennis E., Glass C., Murphy R., Raetz C., Russell D., and S. S. LMSD: LIPID MAPS® structure database. *Nucleic Acids Research 35, D527-32* (2006).

40. Rogers, G. W., Nadanaciva, S., Swiss, R., Divakaruni, A. S. & Will, Y. Assessment of fatty acid beta oxidation in cells and isolated mitochondria. *Curr. Protoc. Toxicol.* **2014**, 25.3.1-25.3.19 (2014).

41. Choi, Y. N., Lee, J. W., Kim, J. W. & Park, J. M. Acetyl-CoA-derived biofuel and biochemical production in cyanobacteria: a mini review. *J. Appl. Phycol.* **32**, 1643–1653 (2020).

42. Burdge, G. C. & Calder, P. C. Introduction to fatty acids and lipids. *World Rev. Nutr. Diet.* **112**, 1–16 (2015).

43. Moore, A. L. <*10.1007-978-90-481-2863-1.Pdf*>.

44. Fujita, Y., Matsuoka, H. & Hirooka, K. Regulation of fatty acid metabolism in bacteria. *Mol. Microbiol.* **66**, 829–839 (2007).

45. Lennen, R. M. & Pfleger, B. F. Engineering Escherichia coli to synthesize free fatty acids Motivations for engineering fatty acid metabolism. *Trends biotechnol.* **30**, (2013).

46. Wang, L., Chen, L., Yang, S. & Tan, X. Photosynthetic Conversion of Carbon Dioxide to Oleochemicals by Cyanobacteria: Recent Advances and Future Perspectives. *Front. Microbiol.* **11**, 1–14 (2020).

47. Mulkidjanian, A. Y. *et al.* The cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13126–13131 (2006).

48. Kaczmarzyk, D. & Fulda, M. Fatty acid activation in cyanobacteria mediated by

acyl-acyl carrier protein synthetase enables fatty acid recycling. *Plant Physiol.* **152**, 1598–1610 (2010).

49. Schulz, H. *Fatty Acid Oxidation*. *Encyclopedia of Biological Chemistry: Second Edition* vol. 2 (Elsevier Inc., 2013).

50. Qiao, Y., Wang, W. & Lu, X. High Light Induced Alka(e)ne Biodegradation for Lipid and Redox Homeostasis in Cyanobacteria. *Front. Microbiol.* **11**, 1–12 (2020).

51. Singh, S. C., Sinha, R. P. & Häder, D. P. Role of lipids and fatty acids in stress tolerance in cyanobacteria. *Acta Protozool.* **41**, 297–308 (2002).

52. Maeda, H., Sakuragi, Y., Bryant, D. A. & DellaPenna, D. Tocopherols protect Synechocystis sp. strain PCC 6803 from lipid peroxidation. *Plant Physiol.* **138**, 1422–1435 (2005).

53. Genome Blast. doi:https://blast.ncbi.nlm.nih.gov/Blast.cgi.

54. Clustal Omega. doi:https://www.ebi.ac.uk/Tools/msa/clustalo/.

55. Tmcalculator. doi:https://tmcalculator.neb.com/#!/main.

56. Biotools. doi:http://biotools.nubic.northwestern.edu/OligoCalc.html.

57. Oligoevaluator. doi:http://www.oligoevaluator.com/LoginServlet.