Novel Natural Products from *Euglena*

by

Eve Tallulah Roxborough

Department of Chemistry
University of Nottingham

Thesis submitted in fulfilment for the requirement of Doctor of Philosophy

January 2025

Abstract

Discovering compounds from nature has equipped humanity to address dietary needs, protect crops, and combat diseases. As the environment changes, the population grows, and antibiotic resistance rises, finding new compounds is critical. Historically, natural sources have been a reliable route for antibiotic discovery, with most current antibiotics derived from natural products. While bacteria have been the primary focus, other organisms like algae, with complex and underexplored genomes and metabolic profiles, hold untapped potential for novel compounds. Traditional bioassay-led approaches often rediscover known compounds. In contrast, modern structure- or gene-first strategies reduce this risk and streamline the selection of novel compounds. To develop new antibiotics and treatments for infectious diseases and cancer, new sources and methods must be explored.

This thesis highlights *Euglena* as a promising source of novel natural products. Using analytical chemistry and cheminformatics, over 32 Euglena strains were grown under varied conditions to generate diverse metabolic profiles. High-performance liquid chromatography-mass spectrometry (HPLC-MS^2) and Molecular networking via the Global Natural Products Social Molecular Networking Database identified potential new compounds. Three compounds were semi-isolated, with preliminary structural suggestions made through NMR analysis.

Acknowledgements

I would like to thank Dr Ellis O'Neill both for his contagious enthusiasm for discovery and for his mentorship over the past three years, my thanks also to Mr Ben Pointer Gledhill for imparting a drop of your infinite mass spectrometry wisdom to me. Thank you to Dr Huw Williams for all your help when it came to the overlap of the NMR world and the algal one.

Thank you also to Michele Pan for chatting biochemistry with me over cups of tea and to Callum Southwood, a better lab bench partner there is none.

Abbreviations

ACN: Acetonitrile

BGCs: Biosynthetic gene clusters

BLAST: Basic Local Alignment Search Tool

BPC: Base peak count

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

EG: Euglena gracilis growth media

EIC: Extracted ion chromatogram

GNPS: Global Natural Products Social Molecular Networking

HAB: Harmful Algal Bloom

HPLC: High-pressure liquid chromatography

HPLC-MS/MS: High-pressure liquid chromatography tandem mass spectrometry

JM: Jaworski's Medium

LB: Lysogeny broth

MAAs: Mycosporine-like amino acids

MN: Molecular Network

MS: Mass Spectrometry

NMR: Nuclear Resonance Spectrometry

NPs: Natural Products

PC: Phosphocholine

PCR: Polymerase chain reaction

PepM: Phosphoenol pyruvate mutase

SE: Size exclusion

TIC: Total ion count

SWBi: Soil, water biphasic media

Contents

Abstract....2
Acknowledgements....3
Abbreviations....4
Chapter One: Introduction and motivation to studying *Euglena* as a source of natural products....10
1.1 Identifying natural products for drug development....10
1.2 The impact of cancers and infectious disease on global health....11
1.3 Structural diversity of natural products....13
1.4 Natural product-based drugs for infectious disease and cancers....15
1.5 Penicillin: the first natural product discovered as treatment for infectious disease....19
1.6 Evaluating natural products relevancy as novel antibiotic candidates....20
1.7 Availability and sustainability of natural resources....24
1.8 Screening for novel natural products....25
1.9 Metabolomics in natural product research....26
1.10 Transcriptomics in natural product research....28
1.11 Metagenomics in natural product research....29
1.12 Cheminformatics in natural product discovery....31
1.13 Introduction to Algae....33
1.14 Cyanobacteria....35
1.15 Rhodophyta....36
1.16 Cryptophyta....37
1.17 Dinophyta....39
1.18 Heterokontophyta....41
1.19 Chlorarachniophyta....41
1.20 Chlorophyta....41
1.21 Comparing plants and algae as a source of natural products....42
1.22 Advances in genetic engineering of microalgal species....42
1.23 An introduction to Euglena....44
1.24 Genetics of Euglena....46
1.25 Significance of *Euglena* in Natural Product Research....48
1.26 Unique natural Products Produced by Euglena....49
Chapter Two: Can molecular networking be employed with Euglena to select targets for natural product isolation?....54
2.1 Natural Product Databases....54
2.2 NMR and MS based databases....54

2.3 Global Natural Product Social Molecular Networking 55
2.4 Analysing Euglena with GNPS 56
2.5 Cytoscape Analysis of Molecular Network One 59
2.6 Cytoscape Analysis of Molecular Network One: Media 62
2.7 Cytoscape Analysis of Molecular Network One: Library Identification 63
2.8 Novel Compounds from Cytoscape Network One 67
2.9 Target 1.1 68
2.10 Cytoscape Network Two – 1 L Cultures 76
2.11 Cytoscape Network Two - Target 2.1 77
2.12 Cytoscape Network Two: Target 2.2.1 and 2.2.2 79
2.14 Cytoscape Network Two: Target 2.4 81
2.15 Target Compounds – Isolation through HPLC 82
2.16 NMR 82
2.17 Disc Diffusion Assays 83
2.18 Chapter Two Conclusions 84
Chapter 3: Does Coculturing *Euglena gracilis* with other organisms increase the diversity of natural products produced compared to monocultures? 85
3.1 Introduction 85
3.2 Coculturing bacteria and algae to initiate release of novel natural products 85
3.2 Coculturing algae and Streptomyces: the initial experiment 88
3.3 Coculturing algae and Streptomyces: the Molecular network 89
3.4 Cytoscape Network 2: Phosphocholines 94
3.5 Cytoscape Network 2: A comparison between algal and bacterial unique nodes 97
3.6 Differences to Cytoscape Network 1 98
3.7 Cytoscape Network 2: Selecting Targets 99
3.7.1 Targets 1A, 1B and 1C 100
3.7.2 Target Two 103
3.7.3 Target Three 105
3.7.4 Target Four 106
3.7.5 Target Five 107
3.7.6 Target Six 108
3.8 Overview of Targets 109
3.9 Metabolic differences between 10 mL and 1L coCultures 110
3.10 Comparison of 1L cultures 113
3.11 New Targets from Cytoscape Network Three 115
3.12 Cytoscape Network 3: Isolating the 1 L Targets 117

3.13 *S. coelicolor* 1L targets 117
3.14 *S. glaucescens* 1 L targets 119
3.141 S. glaucescens 1 L targets: butanol extract 119
3.142 *S. glaucescens* 1 L targets: methanol extract 122
3.15 Target 421 structural prediction 124
3.16 Target 421: Increasing culture volume to 6 L 125
3.17 Target 421: Final purification attempt 127
3.18 Chapter Three Conclusions 128
Chapter 4: Investigation of novel phosphonate compound in *Euglena gracilis* 130
4.1 Relevance of phosphonates as clinically relevant natural products 130
4.2 Biosynthesis of phosphonates 131
4.3 Phosphonate based drug examples 132
4.4 Potential for phosphonate production in *Euglena* 133
4.5 Screening *Euglena gracilis* for the presence of phosphonate 134
4.6 Activated charcoal for the initial purification of *E. gracilis* cultures 135
4.7 Further purification through ion exchange chromatography 136
4.71 Further purification through ion exchange chromatography: DEAE chromatography 136
4.8 Purification through size exclusion chromatography 139
4.81 Further purification through use of a Dowex ion exchange column 141
4.9 Purification through SE and semi-prep HPLC 142
4.10 Purification through SE and DEAE 143
4.11 Mass spectrometry analyses 145
4.12 NMR analysis of impure phosphonate containing fractions 146
4.13 Conclusions of Chapter Four 149
Chapter Five: Are Mycosporine-like amino acids (MAAs) produced by *Euglena gracilis?* 151
5.1 Mycosporine-Like Amino Acids: Structure, Biosynthesis, and Applications 151
5.2 Distribution of Mycosporine-Like Amino Acids in Nature 152
5.3 Identification and Isolation of Mycosporine-Like Amino Acids from Algae 152
5.4 MAAs as natural products 153
5.5 Analysis of the biomass of *E. gracilis* through UV-spectroscopy 153
5.6 Secondary purification of UV-active fractions through additional prep-HPLC 155
5.7 MS analyses of UV active fractions from secondary HPLC purification 155
5.8 MS analyses of all fractions eluted from HPLC purification of *Euglena* biomass 157
5.9 NMR study of fraction six 159
5.10 Modifying growth and purification techniques 161

5.11 Pursuing additional purification 163
5.12 Conclusions from Chapter Five 164
Chapter Six: Conclusions and Future Directions 167
Chapter Seven: Materials and Methods 172
7.1 General Methods 172
7.11 Media Preparation 172
7.12 Small-Scale Cell Culturing 172
7.13 Large-Scale Cell Culturing 172
7.14 Small-Scale Extraction 172
7.15 Large-Scale Extraction 172
7.2 High-Performance Liquid Chromatography (HPLC) and Mass Spectrometry (MS) 173
7.21 HPLC-MS/MS 173
7.22 GNPS and Molecular Networking 173
7.3 Semi-Preparative High-Pressure Liquid Chromatography (HPLC) 173
7.4 Compound Detection and Isolation 173
7.5 Nuclear Magnetic Resonance (NMR) Spectroscopy 173
7.6 Chapter One: Screening of Euglenoid Strains 173
7.61 Small-Scale Cell Culturing 173
7.62 Large-Scale Cell Culturing 173
7.63 Extraction 173
7.8 NMR 174
7.10 Chapter Two: Coculture and Streptomyces Strain Analysis 175
7.11 Small-Scale Culturing 175
7.14 HPLC Fractionation and Compound Isolation 175
7.18 Chapter Three: Phosphonate Compound Isolation 176
7.19 Culturing and Biomass Processing 176
7.32 Chapter Four: Algal Culturing and MAA Isolation 179
Appendix 184
References 189

Chapter One: Introduction and motivation to studying *Euglena* as a source of natural products

1.1 Identifying natural products for drug development

There is a pressing need for the development of new medications to find remedies for diseases that are currently untreatable and to improve existing therapeutic approaches.[1] Natural products, defined as a compound produced by an organism not necessary for its survival but which improves its chance of survival, have played a significant role throughout history in the treatment of diseases, and there is immense potential for novel natural products to continue to bring forward revolutionary advances in future disease treatments.[2] The subsequent chapter offers examples of natural products used to treat disease, the inherit properties of natural products which make them suitable for this, as well as a thorough examination of the challenges and opportunities encountered in the pursuit of new natural products. It also highlights different strategies and emerging technologies that can enhance the effectiveness of drug discovery campaigns based on natural products.

Traditional bioassay-guided approaches to identify novel natural products frequently result in the rediscovery of known compounds due to their reliance on detecting biological activity within extracts against specific targets or assays.[3] This method inherently favours the identification of compounds with well-characterized bioactivity that are commonly produced by organisms, often in significant quantities. Given the extensive research on natural products over the past several decades, these biologically active and abundant compounds are often re-identified. Furthermore, bioassay-guided approaches tend to focus on the most dominant or readily extractable metabolites, which are more likely to have been previously characterized.[4]

Modern structure- or gene-first strategies allow for the identification of known compounds or genes which can then be immediately eliminated from the discovery process. Improvements in high-throughput genetic sequencing coupled with the open-access and information rich gene databases such as NCBI provide an alternative route to bioassay-led natural product identification.[5] Known genes or BGCs which have natural product producing abilities can be compared to unknown genes or BGCs either within the same organism or in different organisms. This approach allows for novel identification within parameters that guide to unknown genes which share similarity to known “useful” genes. In addition, the development of metagenomics allows for

multiple genomes, typically from one niche, to be screened, including those from unknown microbes or microbes previously uncultured in a laboratory.[6] Cloning and analysis on DNA from microbes that have not been previously cultured in a laboratory

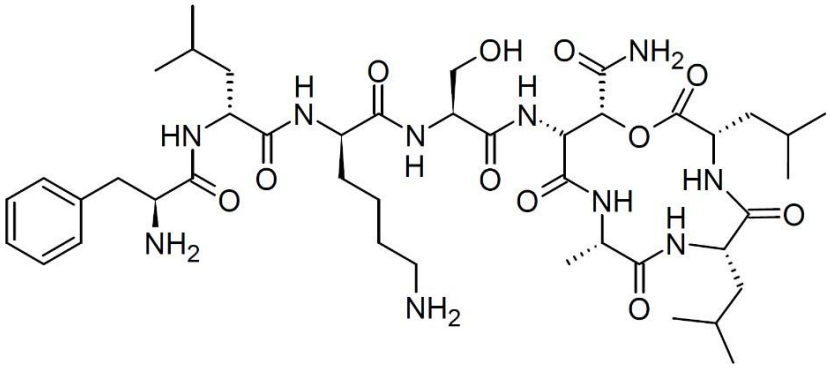


Figure 1.1 Structure of Clovibactin.

has been applied to locate natural products with antibacterial properties such as Clovibactin from a previously uncultured soil bacterium.[7,8] This peptide is effective against drug resistant *S. aureus*, it blocks peptidoglycan polymerisation which prevents formation of cell walls.[8]

In a similar approach to gene-guided approaches, the structure of compounds themselves can act as the initial guide from which a novel compound can be identified. This can be achieved by analytical techniques such as mass spectrometry. Mass-spectrometry based metabolomics is a field in which untargeted or targeted MS^2 data from large datasets can be screened through the use of algorithms which can identify known compounds from the dataset.[9] Both compound and gene guided techniques benefit from increased global which aids dereplication through the sharing of vast libraries and publications, and increased access to novel organisms and unique niches.

1.2 The impact of cancers and infectious disease on global health

Cancer was responsible for 29% of all deaths in the UK in 2022, and incidence of cancer diagnoses have been increasing in the UK since the early 1990s. Although treatment outcomes have improved over the last half century, cancer incidence is set to continue to increase. There were an average of 375,000 new cases of cancer in the UK from 2016 to 2018, which is predicted to raise to 506,800 diagnoses a year by 2038.[10,11] In a worldwide context, cancer emerged as the second most prevalent cause of mortality in 2020, ranking only below heart disease. Cancer claimed the lives of approximately 10 million individuals, accounting for nearly one-fifth of all deaths.[12]

Infectious diseases were responsible for 14% of global deaths, compared to 18% caused by cancer, or around one in seven deaths.[12] Of note, infectious disease was elevated to the highest level of international importance in 2019, when the infectious adenovirus Covid-19 triggered a global pandemic which caused over 7 million deaths worldwide, as well as exacerbating existing inequalities; worsening living standards, educational outcomes, limiting access to physical and mental healthcare and pushing tens of millions into poverty.[13,14] Although vaccines were developed for the effective treatment and prevention of Covid transmission, the societal and economic consequences of this global pandemic remain.[14]

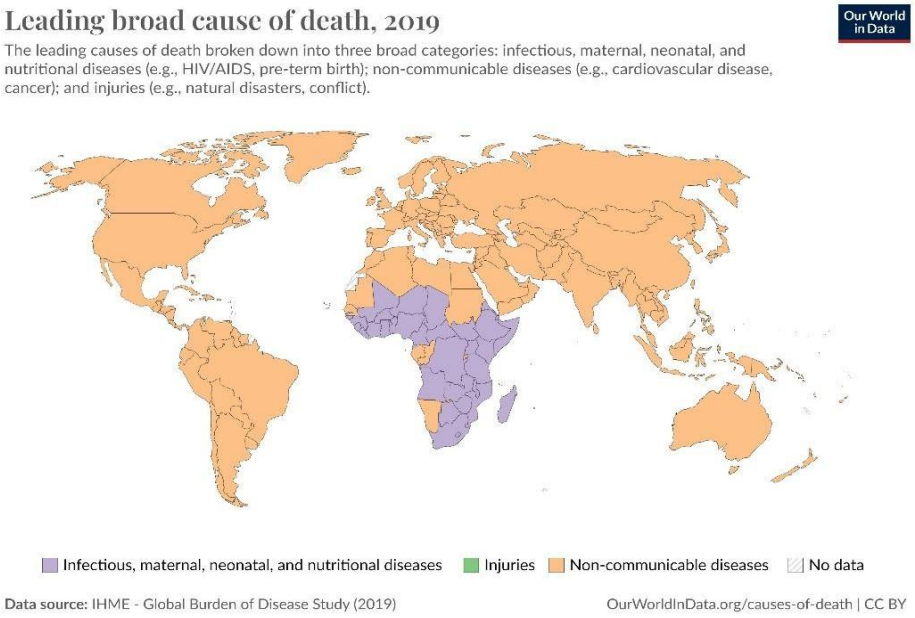


Figure 1.2 Leading cause of deaths worldwide. Non-communicable diseases such as heart diseases and cancer lead in all non-African countries. The leading cause of death in the majority of African countries is infectious disease.[12]

The high morbidity and mortality rates associated with infectious diseases and cancers underscore the need for newer and more effective treatment options. Developing such treatments, however, is challenging. The generation of a new drug typically takes around 12 - 15 years at a cost of over $1 billion. As well as the substantial costs and length of time required, many drugs fail in the research and development process, even after entering clinical trials.[15] Furthermore, many infectious diseases have developed resistance to existing drugs, rendering them ineffective. With the catastrophe of antimicrobial resistance a growing threat, the need for new and effective drugs has never been greater.[1]

1.3 Structural diversity of natural products

Natural products occupy a broad range of chemical space and exhibit greater structural diversity compared to existing synthetic compound libraries. While synthetic compounds tend to be more planar and hydrophobic, natural products are typically more three-dimensional and contain a higher proportion of heteroatoms, such as nitrogen, oxygen, and sulfur.[16] This three-dimensionality and chemical diversity make natural products excellent starting points for optimization, particularly in achieving the "key fit" Molecular shape and function needed to interact with biological targets, such as proteins. Furthermore, the highly specific nature of natural products often translates into increased effectiveness against drug targets with fewer off-target effects.[4] Both natural products approved for use as drugs and those derived from natural sources exhibit more structural variation than synthetic drugs. In biology, structure and function are intrinsically linked, which accounts for the wide range of biological functions displayed by natural products. Their inherent diversity is unparalleled by synthetic libraries, as natural products have evolved to bind to a variety of structurally diverse

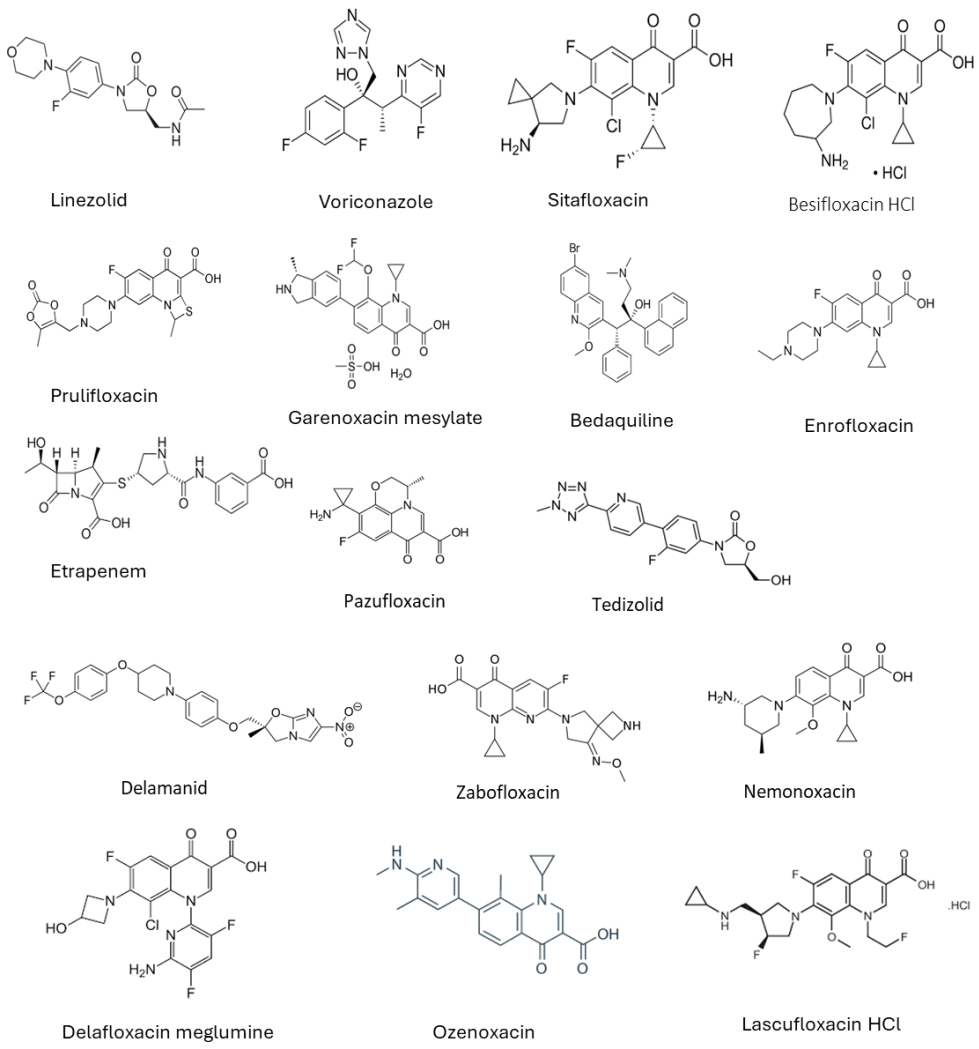


Figure 1.3 Structures of synthetic antibiotics

proteins and modulate their activities with high specificity. Although natural products evolved not as drugs for human use but as toxins or defence molecules, the evolutionary pressure to produce only biologically active compounds means they typically exhibit high bioactivity, offering significant potential for development as drugs for a variety of diseases.[17]

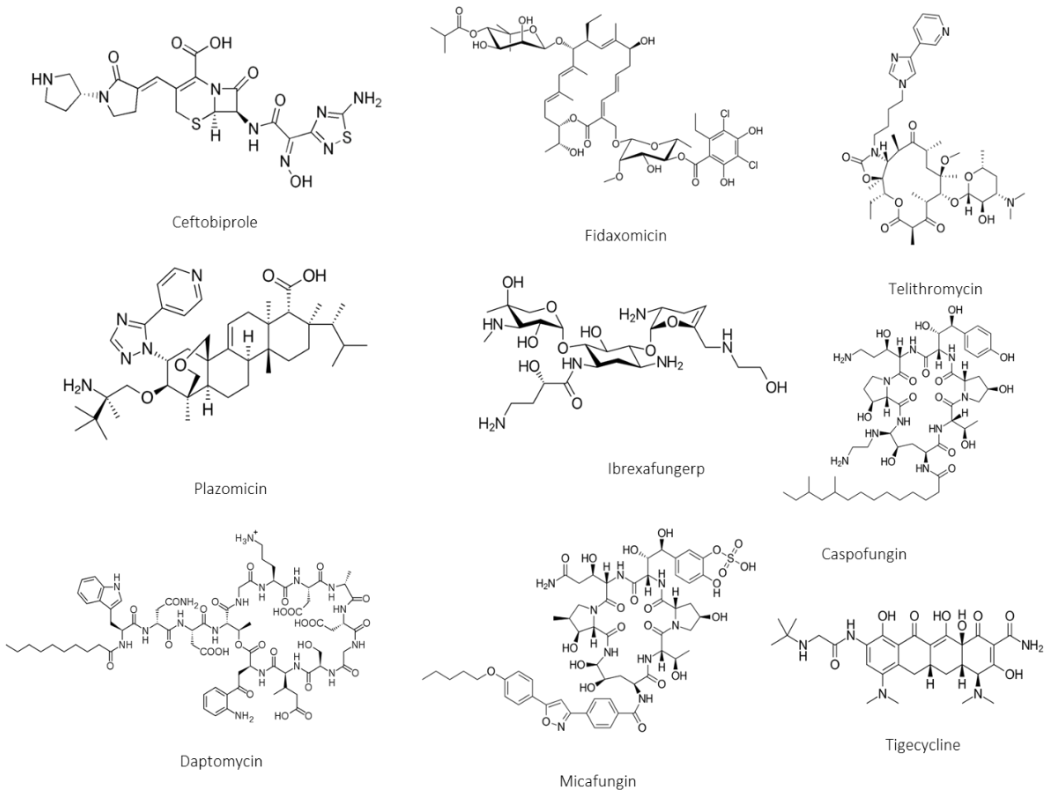


Figure 1.4 Structures of natural product antibiotics

Additionally, natural products sourced from traditional medicines and diets have been consumed by humans for thousands of years and, as such, have been "pre-screened" for human use.[18] Additionally many studies have reported that the synergistic interactions of multi-constituent natural products play a significant role in their therapeutic effects. With the study and development of modern extraction techniques, it is becoming increasingly easier to isolate and identify the active compounds present in natural medicines, even when these compounds exist in low abundance.[19]

1.4 Natural product-based drugs for infectious disease and cancers

The use of natural products, with the capacity to inhibit or destroy pathogenic organisms or malignant cells offers a potential treatment of infectious disease and cancers.[20] Phytochemicals utilised from higher plants reportedly regulate cancer progression through various pathways which include: increasing antioxidant status, carcinogen inactivation, inhibiting proliferation, induction of cell cycle arrest and apoptosis; and regulation of the immune system.[21] Paclitaxel, a tetracyclic diterpenoid from the Pacific yew tree, *Taxus brevifolia*, was reported to have anticancer activity in 1971 and is still used today for the treatment of ovarian cancer, oesophageal cancer, breast cancer, lung cancer, Kaposi's sarcoma, cervical cancer, and pancreatic cancer.[22]

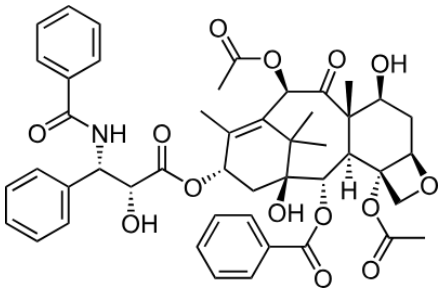


Figure 1.5 Structure of Paclitaxel

Drug resistance has become a major challenge in modern medicine and has arisen for a number of reasons: overuse, both in unnecessary prescriptions and in animal husbandry, and misuse: incorrect prescriptions, such as antibiotics for non-bacterial disease, or incompletion of a drug regimen or incorrect dose regiments, such as not finishing a full treatment course of antibiotics. Different mechanisms exist for drug resistance, depending on the drug in question. For example, resistance to antimicrobial drugs can occur due to a decrease in the concentration of a drug, an alteration in a drug's target, or an active drug being pumped out of the cell at a faster rate.[23]

Prior to treatment, drug-resistant strains of the causative agent are often present, so the application of the drug then acts as a selection pressure, favouring the proliferation of the drug resistant strain - this is true for bacteria, viruses and cancers.[24] For cancer, understanding the genetic makeup of specific cancer cells may one day lead to methods of identifying those who are at greater risk for drug resistance in an effort to custom tailor treatment strategies.[25] One way to break the cycle of drug resistance is the use of multiple drugs to treat a disease, which is referred to as drug combination therapy. This

type of therapy works due to the use of drug synergy, which is the cooperative effect resulting from the combined action of two or more drugs, and the probability that more than one drug may be effectively targeting the causative agent in a given patient. More specifically, a number of possible combined effects have been observed, including the ability of one drug to block the programmed resistance to another, the suppression of resistance by two drugs, or the presence of a synergistic effect between drugs which suppresses the causative agent more than either drug alone.[26]

It is worth mentioning that the application of synergistic components for the treatment of diseases extends beyond just the use of antibiotics. Anticancer medications, such as the combination of mitomycin C and tobramycin-ciprofloxacin, works to eradicate drug resistant Gram-negative bacteria. Furthermore, alternative approaches, like the incorporation of manuka honey in tandem with rifampicin, have demonstrated synergistic efficacy against *Staphylococcus aureus* biofilms.[27]

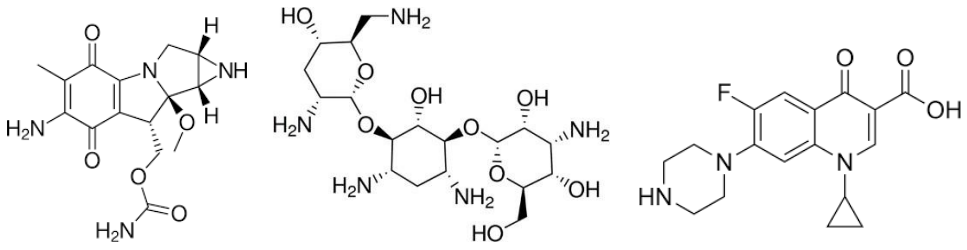


Figure 1.6 Structure of mitomycin C, tobramycin and ciprofloxacin (left to right)

Synergistic treatment with penicillin and streptomycin results in increased bactericidal effect compared to single drug treatment as penicillin breaks down the cell walls which typically pose as a barrier to streptomycin entry.[28] Additionally, streptomycin and cefotaxime produce a synergistic effect in cefotaxime resistant *Enterobacter cloacae* as streptomycin induces conformational changes in the enzyme responsible for cefotaxime hydrolysis which reduces its binding affinity to cefotaxime.[26]

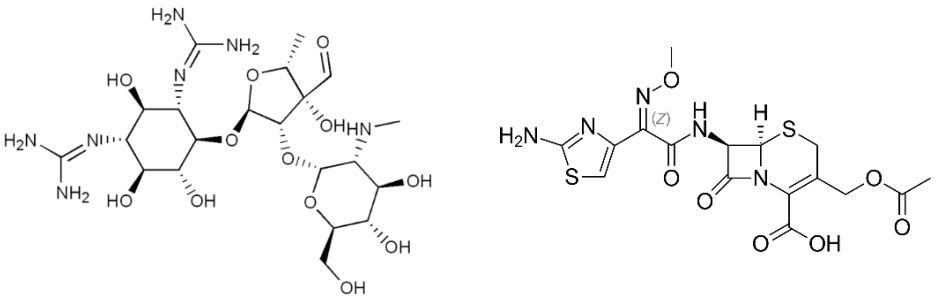


Figure 1.7 Structure of Streptomycin (left) and cefotaxime (right)

Natural products exert their therapeutic effects on infectious diseases through a multitude of potential mechanisms. Typically, antimicrobial natural products fall within a number of categories, including categories based on their chemical structure (such as alkaloids, glycosides, or terpenes) and categories based on the type of organism they target (such as antibacterial, antifungal, or antiviral).[29]

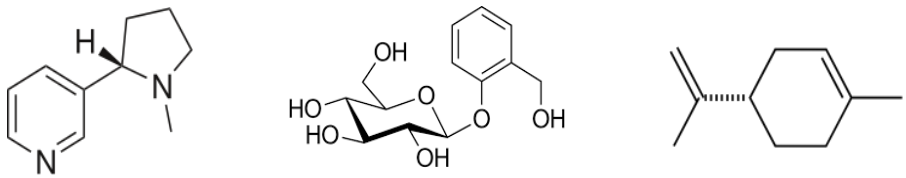


Figure 1.8 Terpene functional group, glycoside structure, basic terpene unit (left to right)

There are many natural products extracted from plants, fungi, and bacteria that can inhibit cancer growth by preventing the function of oncogenes. Many of the well-documented examples of natural product-based chemotherapies are alkaloid based. These are a group of nitrogenous chemicals which are mostly plant in origin. These nitrogenous chemicals are produced by the plants as a method of defending themselves against pests and diseases. They have also been found to have an effect on certain types of cancer, inhibiting the growth and spread of cancerous cells.[21]

One anticancer alkaloid is found in turmeric, a flowering plant in the ginger family Zingiberaceae. Curcumin can down-regulate the activity of certain oncogenes such as NF-kB, STAT3, etc., and up-regulate the activity of certain tumour suppressor genes including p53. These genes are responsible for cell division, apoptosis, and DNA repair, so the influence on these genes can result in the inhibition of cancer cell growth. Furthermore, curcumin has been found to inhibit the major angiogenesis stimulating factor VEGF in a variety of cancer cell lines. Therefore, curcumin is not only considered as a cytostatic agent that stops the growth of the tumour cells but also a cytotoxic agent that induces a direct killing effect on the cancer cells.[30]

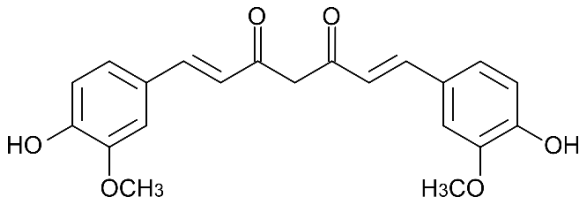


Figure 1.9 Structure of Curcumin

Another example are the Vinca alkaloids. These are a group of alkaloids found in the Madagascan periwinkle, a plant that had been used for centuries in herbal teas - though

its potential as a cancer treatment was not discovered until the 1950s. Over fifty alkaloids have been identified from the Madagascan periwinkle and there are four major vinca alkaloids in clinical use: vinblastine, vinorelbine, vincristine and vindesine. Two of which: vincristine and vinblastine, have been approved for the treatment of metastatic cancer, particularly of the lung or the breast. The Vinca alkaloids work by targeting the microtubules within the cancer cells. These are a component of the internal cellular structure and also critical in the movement of chromosomal DNA during mitosis and meiosis. By disrupting the microtubules, the alkaloids prevent the successful completion of mitosis and thus inhibit the growth of the cancer cells. [31]

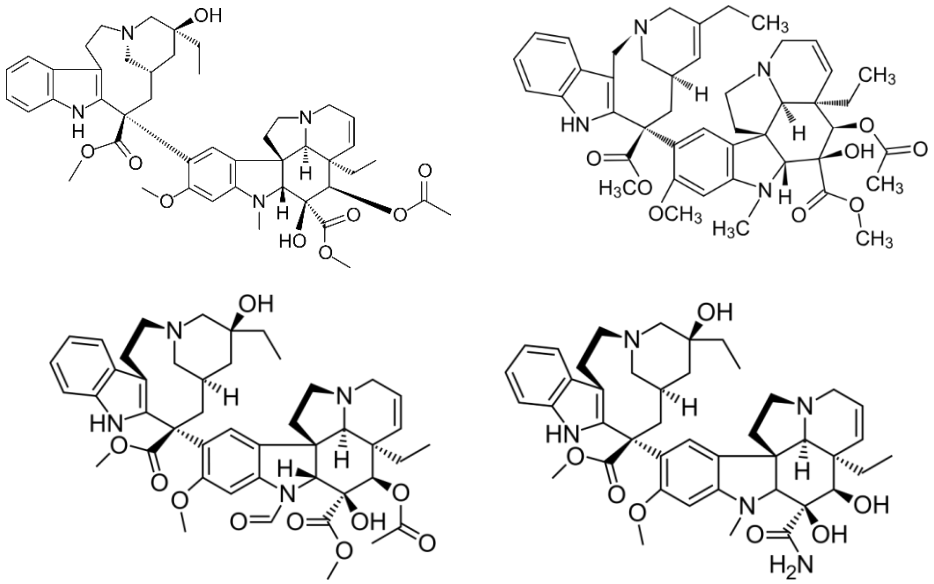


Figure 1.10 Structure of Vinca alkaloids. Vinblastine (top left), vinorelbine (top right), vincristine (bottom left) and vindesine (bottom right).

Terpenes also have activity against infection, such as the sesquiterpene lactone artemisinin, the active ingredient of *Artemisia annua*, a plant from traditional Chinese medicine, native to south-central Asia and also found in the Pacific. This natural product is very effective against malaria, including the malaria caused by drug resistant forms of *Plasmodium falciparum*. Artemisinin is first activated by intraparasitic heme-iron which catalyses the cleavage of the peroxide bridge. A resulting free radical intermediate may

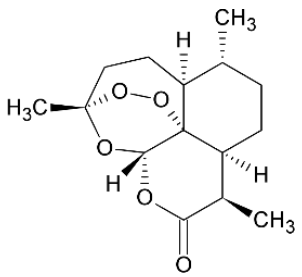


Figure 1.11 Structure of Artemisinin

then kill the parasite by alkylating and poisoning one or more essential malarial proteins.[32]

Cryptolepis, a herb commonly found in West Africa, it has been used for the treatment of malaria and diarrhoea. The antimicrobial action in the root extract of the *Cryptolepis sanguinolenta* plant was discovered when it was tested against a small sample of common microorganisms that can cause gastrointestinal problems. This extract has stronger antimicrobial properties than the clinically approved antibiotics Erythromycin and Streptomycin and is almost as strong as Ciprofloxacin.[33] Similarly, the natural product Berberine, sourced from many plants, a very effective drug used to treat intestinal parasites and also on every type of bacteria, virus, and yeast, has been found to be as effective as the prescription antibiotic drug chloramphenicol. It also displays antifungal activity against the pathogens causing athlete's foot, candida, and *Tinea capitis* (ringworm).[34]

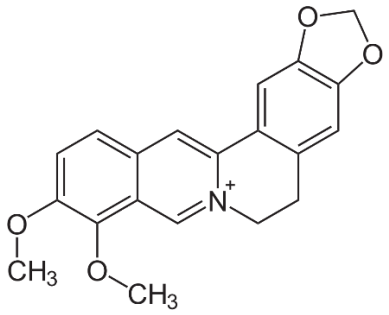


Figure 1.12 Structure of Berberine

1.5 Penicillin: the first natural product discovered as treatment for infectious disease
1.6

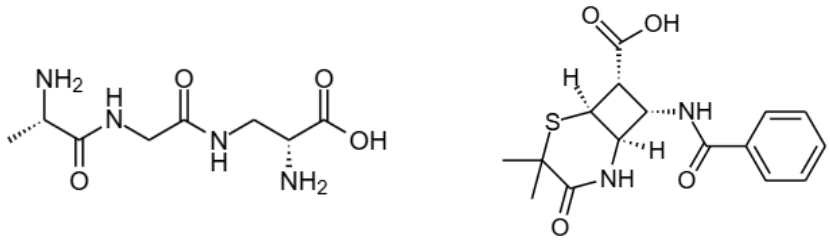


Figure 1.13 Structure of terminal D-alanyl-D-alanine (left) and structure of penicillin (right)

Perhaps the most famous of all natural products is penicillin. The basic structure of the penicillin consists of a thiazolidine ring connected to a β-lactam ring, to which is attached an acyl side chain. The beta-lactam ring in penicillin has the ability to inhibit an enzyme that helps in the synthesis of the cell wall in bacteria. Precisely, it inhibits

the enzyme DD-transpeptidase, which is crucial in the cross-linking process of the peptidoglycan cell wall. The beta-lactam ring closely resembles the chemical structure of the terminal D-alanyl-D-alanine amino acid sequence recognized by the enzyme in the bacteria. This allows penicillin to competitively bind to the enzyme's active site. As a result, the enzyme is no longer able to fully function as the active site is being blocked by the penicillin molecule. This interference disrupts the formation of the cross-linkages between the peptidoglycan chains, weakening the cell wall significantly. Consequently, when the bacteria try to divide and grow, the affected cell wall is unable to withstand the turgor pressure, which is the pressure of the cell contents against the cell wall. The cell wall ruptures, the bacteria dies ultimately due to the influx of water into the cell. The bacterial debris is then cleared from the body by the immune system. The patient receiving penicillin does not suffer as a direct action of the drug itself, penicillin cannot lyse the human cells because we do not have cell walls or DD-transpeptidase that are inhibited by penicillin. Therefore, unless allergic, the patient is able to recover efficiently without compromising their own cells while penicillin damages only the bacteria. However, the drug is not perfect, bacteria can quickly become resistant. Even working without resistance, penicillin's effectiveness requires the bacteria to be actively growing and synthesizing new cell walls. Additionally, gram-negative bacteria possess a protective outer membrane which prevents penicillin from entering the cell, thus penicillin is only effective against gram-positive bacteria. Some bacteria possess an enzyme called penicillinase is type of β-lactamase which breaks down penicillin, neutralising the antibiotic. Bacteria can become resistant to penicillin due to diminished permeability of the bacterial cell to the antibiotic; alteration of the penicillin-binding proteins, or production of β-lactamases.[35]

Penicillin was discovered by Alexander Flemming when he allowed a petri dish to go mouldy, and observed the resulting fungus to have a ring of lysed bacteria surrounding it. Alas, it seems the discovery of the next famous natural product-based antibiotic will not happen so serendipitously. Overuse of antibiotics over the decades has led to the development of antibiotic-resistant bacteria. This means that infections caused by the resistant bacteria are more serious and harder to treat. This is not just true for penicillin, but all current clinical antibiotics. It is predicted that by 2050 antibiotic resistance will be the 5th leading cause of death in the world.[1] For this reason, we need novel class of antibiotics unless we want to a return to the pre-antibiotic world in which a small cough or cut could lead to an infection that becomes life threatening.

1.6 Evaluating natural products relevancy as novel antibiotic candidates

The potential within the natural product research community is vast, there are an estimated 5,400,000 natural products produced by land plants alone, of which only 349 have been isolated.[36] Multiple compounds from terrestrial and marine organisms show inhibitory activity on a range of targets, but the majority remain in very early research stages.[2] A natural product needs to be much more than just inhibitory to a target under laboratory conditions before it can become a therapeutic drug applied in a clinical environment. In order to meet this threshold, it needs to be able to meet a range of standards: high efficacy, stable in storage, have an effective half-life in the human body, low toxicity with high specificity - no or limited off-target effects, soluble in water, as well as being economically viable.[37]

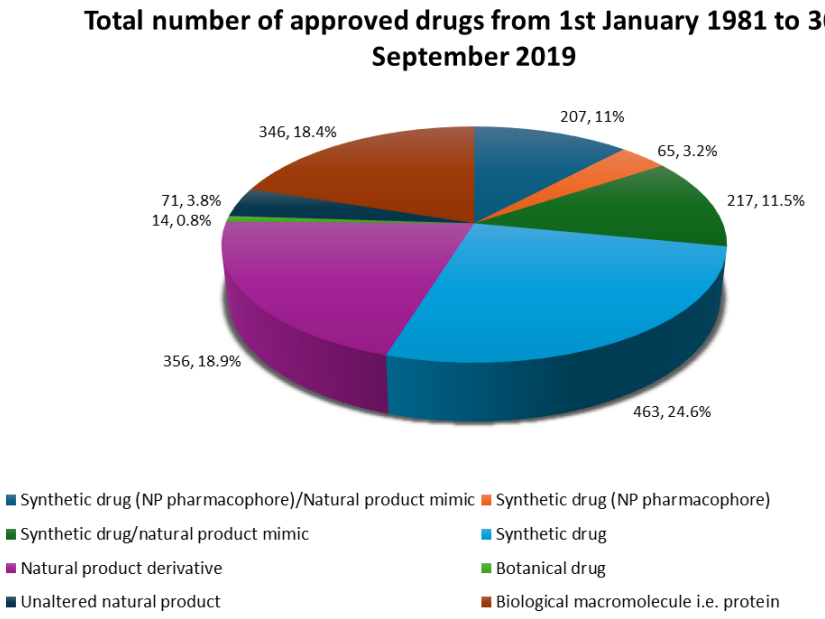


Figure 1.14 Total number of improved drugs from 1st January 1981 to 30th September 2019. Replicated from report by D.J Newman, 2020.

Over the past century, natural products have contributed significantly towards drug discovery and development. According to the recent analysis conducted by David J. Newman and Gordon M. Cragg, it was discovered that 32% of all small-molecule drugs that received approval from January 1981 to September 2019 were either natural products or derived from them. Furthermore, between 1981 and 2014, 51% of the 1211 new small molecule drugs approved globally were derived from natural sources. Additionally, a study conducted by Eric Patridge and colleagues which was published in 2016, reported that the US Food and Drug Administration (FDA) approved 547 natural products and their derivatives for medicinal use over a span of almost two centuries (1827 to 2013).[38] The same study revealed that 68% of the 136 small-molecule

anticancer drugs available from 1940 to 2014 were derived from natural sources. Natural products continue to contribute significantly towards new drugs approved by the FDA, in 2017, 36% of the 39 approved new drugs were natural product or natural product derived, in 2018 22% of 46 and in 2019, 32% of 48.[29, 39]

There are many advantages to natural products as drugs such as a wider range of pharmacophores, evolutionary optimization – natural products have effectively been pre-screened" for bioactivity, as well as increased Molecular complexity and density, chiral and three-dimensional structures - which are generally more effective to biological targets in a human body.[2] The use of natural products in society is not new. Quinine obtained from the bark of the cinchona tree has been used for centuries to treat malaria.[40] Similarly, morphine which is derived from the opium poppy has been used for pain relief for over 400 years.[41] The Nobel prize for medicine was awarded in 1908 for the discovery of bacterial natural product arsphenamine,[42] and has also been awarded for work on natural products over 100 years later in 2015 for the discovery of natural products ivermectin and artemisinin.[43]

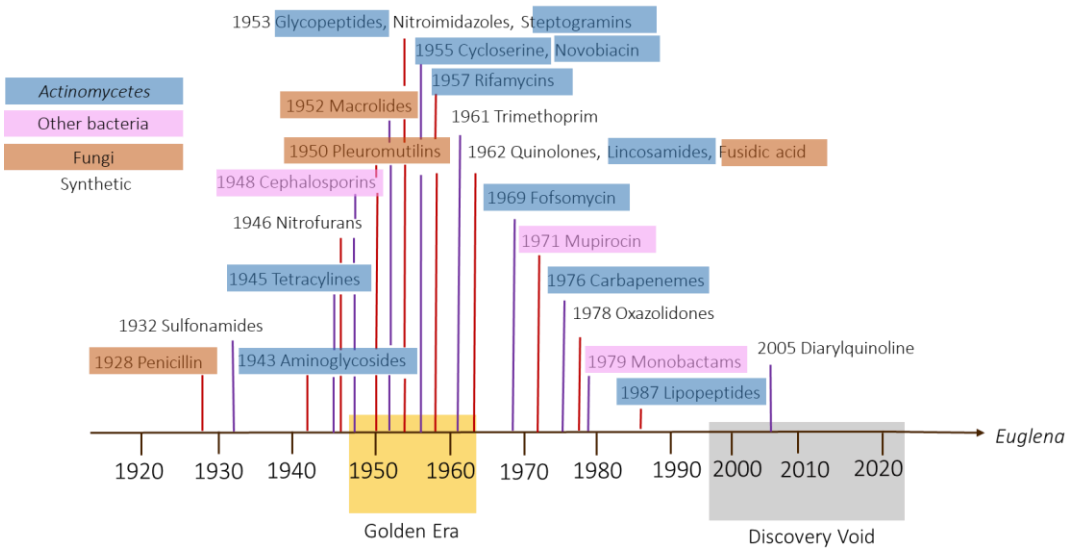


Figure 1.15 Classes of antibiotics from 1920 to 2005 and their source

Historically, Actinomyces, particularly the genus Streptomyces contributed the bulk of natural products to the field of antibiotics.44 Whilst mining natural products from these and other soil-based bacteria and fungi lead to the golden age of antibiotic discovery in the 20th century, this was followed by a discovery void. All new antibiotics approved for clinical use since 2014 are from classes discovered between 1950 and the late 1970s, predominantly cephalosporins, carbapenems, aminoglycosides or tetracyclines.[45] This is not to say that the natural world is no longer of relevance to drug discovery, there have been new classes of natural products identified with antibiotic activity in recent years. However, the approved clinical antibiotics from 2000-2001 were developed through modifications or optimisations of previously identified natural products, or synthesised with a natural product inspired pharmacophore.[46]

When we consider all the antibiotics discovered in said golden age were sourced from a few genera of microbes it seems obvious that to find more we need not look any further than the natural environment. It isestimated that less than 1% of the microbial world has been cultured in laboratories, and an estimated 18 million natural products are predicted to exist meaning that the opportunity to find novel natural products from uncultured or not-yet-cultured microorganisms is almost unimaginably large.[36]

Year	Drug Name	Type	Class
2000	Linezolid	Synthetic	Oxazolidinone
2001	Telithromycin/Caspofungin/Etrapenem	Natural product derived/ Modified Natural product /Synthetic	Macrolide/lipopeptide/carbapenem
2002	Biapenem/Voriconazole/Micafungin	Synthetic/synthetic/Modified Natural product	Carbapenem/triazole/lipopetide
2003	Daptomycin/Gemifloxacin mesylate/Balofloxacin	Natural product/Synthetic*/Synthetic*	Lipopetide/fluroquinolone/fluroquinolone
2004	Prulifloxacin	Synthetic	Flurouinolone
2005	Dorpenem/Tigecycline	Synthetic/Modified Natural product	Carbapenem/Glycycycline
2006	Anidulafungin	Modified natural product	Echinocandin (cyclic lipoopetide)
2007	Garenoxacin mesylate/Retapamulin	Synthetic/ Modified natural product	Fluroquinolone/pleuromutilins
2008	Sitafloxacin hydrate/Ceftobiprole/Medocaril	Synthetic/ Modified natural product/Modified natural product	Fluroquinone/cephalosporin/cephalosporin
2009	Tebipenem pivoxil/ Besifloxacin HCl/ Antofloxain	Natural product derived/synthetic/synthetic	Carbapenem/Fluroquinolone/ fluroquinolone
2010	Ceftaroline fosamil	Natural product derived	Cephalosporin
2011	Fidaxomicin	Natural product	Macrolide
2012	Bedaquiline	Synthetic	Diarylquinoline
2014	Pazufloxacin/Tedizolid/Posaconazole/Finafloxacin/Ceftolozane/Tazobactam/Delamanid	Synthetic/synthetic/synthetic/synthetic/Modified natural product /Modified natural product /synthetic	Fluroquinolone/oxazolidinone/azole/fluroquinolone/cephalosporin/cephalosporin/nitroimidazole
2015	Zabofloxacin HCl/Isavuconazole/Ceftazidime/Avibactam	Synthetic/synthetic/ Modified natural product /Synthetic	Quinolone/azole/cephalosporin/Cephalosporin
2016	Nemonoxacin	Synthetic	Fluroquinolone
2017	Delafloxacin meglumine/Ozenoxacin/Meropenem	Synthetic/synthetic/Natural product derived	Fluroquinolone/quinolone/carbapenem
2018	Omadacycline/Eravacycline/Plazomicin	Natural product derived	Tetracycline/tetracycline/aminoglycoside
2019	Lascufloxacin HCl/Cefiderocol/Alalevonadifloxacin mesylate/Nadifloaxin/Sarecycline HCl/Lefamulin/Imipenem	Synthetic/Modified natural product/synthetic/synthetic/synthetic/Modified natural product /Modified natural product	Flurouinolone/cephalosporins/quinolone/quinolone/tetracycline/pleuromutilin/thienamycin
2021	Ibrexafungerp/Contezolid	Modified natural product /synthetic	Hydroxysteroid/oxazolidinone

Table 1. Approved antibiotics from 2000 to 2021 and their origin and class.

In 2024, a novel natural product with antibacterial activity was identified from a completely new class of compounds—macrocyclic peptides. This discovery was made by screening over 45,000 small molecules using whole-cell phenotypic screening, which initially led to the identification of a promising lead compound. While whole-cell phenotypic screening can be challenging to interpret due to potential false positives or ambiguous results arising from nonspecific interactions or unrelated physiological changes within the cell, it offers a significant advantage. The complexity of this approach more closely mirrors the biological environment of its intended application, making it potentially more reliable than traditional in vitro studies for assessing a drug's efficacy.

This lead compound was then optimised to generate Zosurabalpin. What is particularly significant about Zosurabalpin is that is shows *in vitro* activity against a species of gram-negative bacteria, *Acinetobacter baumannii,* despite most antibiotics being effective only against gram-positive bacteria and that it does so through a novel mode of action – inhibiting the transport of bacterial lipopolysaccharide from the inner membrane to its destination on the outer membrane. Some strains of *A. baumannii* have developed a resistance to carbapenems and as such infection by carbapenem-resistant *A. baumannii* (CRAB) can be life threatening, with up to 70% of patients succumbing to infection. The novel mode of action by Zosurabalpin acts means that risk of CRAB developing cross-resistance is extremely low. [47] The success of this approach in discovering a novel antibiotic class highlights its potential when combined with innovative natural product identification processes. By testing a library of novel natural products through whole-cell phenotypic screening, the bioactivity of these compounds can be evaluated in a more biologically relevant context. This method offers a more realistic setting for assessing drug efficacy, potentially reducing the likelihood of candidate drugs failing in traditional in vitro studies.

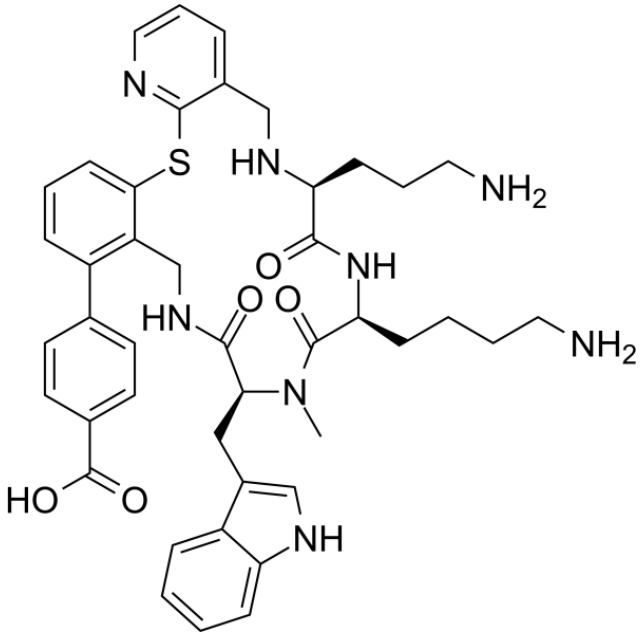


Figure 1.16 Structure of Zosurabalpin

1.7 Availability and sustainability of natural resources

Discovery of natural products has faced pitfalls due to the frequent discovery of known compounds from the same small pool of highly studied organisms. One way to overcome this problem would be to shift focus from the microbes already well-studied to the discovery of compounds from the remaining 99% of the microbial world yet to be investigated.[36] This can be achieved through the exploration of untapped environments that are rich in microbial biodiversity. For example, extreme habitats such

as deep sea, polar regions, caves, and deserts have been known to host microbes capable of producing novel natural products with unique bioactivities.[48] Moreover, the compounds isolated from these conditions might possess unusual chemical scaffolds that are not commonly found in the literature. However, it is important to consider the surrounding ecosystem when conducting such explorations and to prioritize sustainable practices.

Currently, a large number of natural product-based drug discovery projects are ongoing and are being explored clinically as potential therapeutics.[4] However, the potential new drugs arising from these discovery programs also bring an added pressure on natural resources. The increasing demand for natural product-based therapeutics can lead to overexploitation and extinction of the species producing these valuable natural resources.[49] Careful monitoring of any local or global impact on biodiversity triggered by natural product research, particularly medicinal plant research, will be critical to ensure successful and sustainable drug discovery from bioresources.

Given the fact that marine organisms are constantly exposed to varied biotic and abiotic stresses, there is an increasing belief that the marine environment could provide a rich source of pharmacologically active metabolites. Additionally, by comparing the biodiversity and the natural products between marine and terrestrial organisms, it is interesting to note that the marine natural diversity is much higher than the terrestrial natural diversity. Also, 75% of the known natural products are from terrestrial organisms, while 15% are from marine organisms.[36] The development of "Marine Microbial Drugs" project, which involves the extraction, isolation, and screening of natural products from pure Cultures of marine microorganisms, will certainly shed light on new discoveries of bioactive compounds that have the ability to treat human diseases.[50]

1.8 Screening for novel natural products

Another important factor to emphasize is the inefficiency of dereplication in natural product research, that is preventing the rediscovery of chemistries that are already described in the literature. When screening for novel natural products, the isolation process has typically been led by the biological activity of a singular target compound. This means a huge variety of interesting compounds from the same organism or niche are discarded due to the lack of high-throughput techniques in the early screening stages that could measure an array of biological activities simultaneously.[51] The use of these classic techniques for the isolation process, like bioactivity assays under the control of bio-guided isolation protocols, often leads to the re-isolation of known compounds. Lately, a new revolution in natural product discovery has started thanks to

the utilization of modern technologies for the study of nature's biosynthetic potential. These new techniques focus not on bioactivity of a single compound but genetic information of the organism (genomics, transcriptomics) or a community of organisms (metagenomics), or structural information of the natural product (cheminformatics), or a combination of these techniques.[52]

The application of metabolomics, transcriptomics, metagenomics and cheminformatics has enabled the development of new screening parameters and selection criteria to prioritize the isolation of novel natural products from a producing microorganism. This knowledge-driven approach eliminates the time-consuming and resource-heavy element of activity-guided isolation of known compounds and significantly decreases the possibility of re-isolation. For instance, metagenomics is a powerful tool which allows the analysis of genomic material recovered directly from environmental samples. A more comprehensive picture about the genomic potential of whole microbial communities can be achieved by this strategy, which reduces the possibility of a novel natural product being missed during the early stages of drug discovery.[53]

Both cheminformatics and bioinformatics can be applied to metabolomic screening. High throughput metabolomic screening involves the extraction of metabolites from cultures. The diversity of metabolites could be increased through modifications of the culture conditions, a technique known as the one-strain-many-compounds hypothesis (OSMAC). This technique induces metabolic changes to the organism by means of changing pH, salt concentration, nutrient availability, temperature and so on without the need for inducing specific changes to the organism's genome or genetic makeup. This means more diverse natural products can be expressed without laborious and costly genetic studies and manipulations.[54,55]

1.9 Metabolomics in natural product research

In 1998, Olivier et al. first introduced the term "metabolome" to encompass the entirety of metabolites produced by an organism, much like the definitions of genome and proteome. However, this definition has since undergone a process of refinement, now referring to "the quantitative complement of all low Molecular weight molecules (<1 kDa) found within cells during a specific physiological or developmental state.".[56] The terms metabolomics and metabonomics both refer to studying metabolites of a biological sample. Metabonomics is the study of the interactions of these metabolites over time in a complex system. Although metabonomics is in fact a subcategory of metabolomics, many use these terms interchangeably.[57]

Metabolomics has been proposed as the most informative among the various omics technologies due to the limitations associated with transcriptomics and proteomics, as

changes in the transcriptome and proteome may not always manifest as altered biochemical phenotypes, unlike the metabolome. In fact, the metabolome represents the final level of omics in a biological system, with metabolites serving as functional entities, in contrast to messenger RNA molecules that make up the transcriptome. Hence, metabolites have a clear and vital role in the functioning of a biological system, and they are also influenced by the surrounding environment. Therefore, the metabolome can be likened to a looking glass that provides insights into the physiological, developmental, and pathological state of a biological system.[58]

Traditional methodologies for analysing metabolites encompass metabolite fingerprinting, target analysis, and metabolite profiling. Metabolic fingerprinting is an untargeted analysis directed towards defining clinically relevant differences rather than identifying all molecules present in a sample.[56] The primary aim of metabolite fingerprinting is to rapidly categorize a multitude of samples using statistical methods that are typically devoid of distinguishing individual metabolites or quantifying them. Alternatively, metabolic profiling involves a preselection of a set of metabolites, or a specific class of compounds, that might participate in a targeted pathway, either with known or unknown identities, through a particular analytical procedure. Target analysis is limited solely to qualitatively and quantitatively examining a specific metabolite or group of metabolites. Consequently, this approach focuses only on a small portion of the metabolome, disregarding signals from all other components. These techniques are often used in sequence within natural product discovery.[58]

Metabolomic fingerprinting has been applied to a range of investigations. For example, to root extracts of *Arabidopsis thaliana* to detect 103 compounds, of which 90 were identified including: nucleosides, deoxynucleosides, aromatic amino acids, anabolites and catabolites of glucosinolates, dipeptides, indolics, salicylic and jasmonic acid catabolites, coumarins, mono-, di- and trilignols, hydroxycinnamic acid derivatives and oxylipins. This was achieved through solid-phase extraction of the root samples following by ultra-high pressure liquid chromatography (UHPLC)-quadrupole time-of-flight tandem mass spectrometry (QTOFMS/MS) in electrospray ionisation mode. The metabolites were then identified from their Molecular fragmentation patterns.[59]

Metabolomic profiling was employed effectively for the discovery of a new class of di- and trichlorinated acyl amides, henceforth called columbamides, from cyanobacteria. These compounds were discovered to have cannabimimetic activity. Firstly, the cyanobacteria were genetically screened to identify a regulatory domain upstream of a previously reported BSG. The genetic information juxtaposed with mass spectrometric profiles of metabolite extracts allowed scientists to identify a family of functionally

expressed novel metabolites. Two compounds from these extracts, columbamides A and B, were evaluated for cannabinoid receptor CB1 and CB2 binding efficacy and found to be the most potent analogues yet isolated from the marine world.[60]

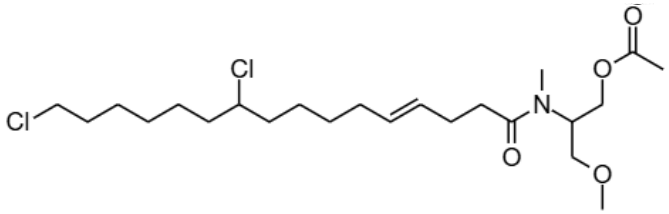


Figure 1.17 Structure of columbamide A

1.10 Transcriptomics in natural product research

Investigations into transcriptomics commenced in the early 1990s, and subsequent technological advancements since the late 1990s have transformed the field of transcriptomics into a highly relevant and sophisticated field. There exist two pivotal contemporary methodologies within this domain: microarrays, which quantify a predetermined set of sequences, and RNA sequencing (RNA-Seq), which employs high-throughput sequencing to capture the entirety of sequences. The evaluation of gene expression within an organism's various tissues, conditions, or temporal points yields insights into gene regulation and unveils intricate facets of an organism's biological makeup. Furthermore, it can facilitate the annotation of unknown genes. [61]

Transcriptomics can be employed within natural product research in order to predict their synthesis from biosynthetic gene clusters, and to connect gene expression of a natural product to the organisms environment or growth stage. Additionally, transcriptomics and metabolomics can be employed together to elucidate a fuller picture of the biochemistry of a natural products. For example, the biosynthesis of the bioactive diterpenoid tanshinones from the Chinese medicinal herb, *Salvia miltiorrhiza*, was investigated using metabolomics and transcriptomics. Axenic *Salvia miltiorrhiza* Cultures were induced with Ri T-DNA bearing *Agrobacterium rhizogenes* to induce hairy root disease. The hairy root Cultures of *Salvia miltiorrhiza* were subjected to untargeted metabolomics analysis through ultra-performance liquid chromatography coupled with diode array detection and quadrupole time-of-flight mass spectrometry (UPLC-DAD-QTOF-MS) analysis as set time points from induction.[59]

This revealed that tanshinone production was the predominant metabolic response, and that production increased at later time points. The transcriptional response of *S. miltiorrhiza* hairy root Cultures to induction was determined by an RNA-seq approach. This not only established a comprehensive transcriptome consisting of 20,972 non-redundant genes but also explored its response to induction. This analysis identified 6,358 genes exhibiting differential expression, with a significant enrichment for up-

regulation of genes involved in stress, stimulus, and immune response processes. Consistent with the findings from the metabolomics analysis, there appeared to be a gradual but sustained increase in transcript levels of known genes associated with diterpenoid and, more specifically, tanshinone biosynthesis.[62]

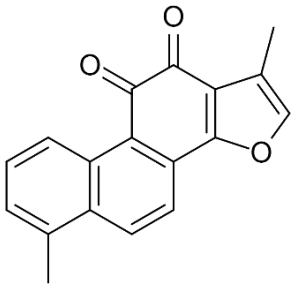


Figure 1.18 Structure of tanshinone

1.11 Metagenomics in natural product research

Metagenomics, involves the extraction of DNA directly from environmental samples. It works on the principle that if the DNA of these uncultured bacteria can be isolated then the genes responsible for the production of the compounds can be identified, bypassing the need to attempt to find a way to culture unculturable bacteria. Because the majority of environmental bacteria are not easily culturable, access to many bacterially encoded secondary metabolites will be dependent on the development of improved functional metagenomic screening methods. So far, metagenomics has shown promise in discovering genes for the production of bioactive compounds or enzymes, especially those involved in more complex biosynthetic pathways.[53]. For example, an initial screen of a 1.5 million-membered metagenomic library constructed in *Streptomyces albus*, led to the identification of the novel natural product metatricycloene.[63]

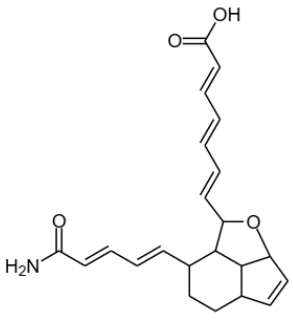


Figure 1.19 Structure of metatricycloene

Additionally, the first cyanobacterial trans-AT polyketide biosynthetic pathway in the cyanobacteria symbiont *Nostoc* of the lichen *Peltigera membranacea* was discovered through metagenomics, the resulting compound was called nosperin. The biosynthetic gene cluster and the structure of nosperin are related to those of the pederin group previously known only from non-photosynthetic bacteria associated with beetles and marine sponges.

This was achieved through extraction and whole genome sequencing of lichen DNA from field samples collected in Iceland. The sequencing revealed approximately equal contributions from the fungal component, the cyanobacterial component, and the community of non-photosynthetic bacteria that comprise lichen. Bioinformatic mining of the initial metagenome assembly yielded 18 candidate clusters containing genes that encode PKS enzymes. Two were members of the *trans*-acyltransferase (AT) PKS family in which AT domains are not encoded by the PKS genes but rather by a separate gene elsewhere. These gene clusters in the lichen are most likely derived from the cyanobacteria, as diagnostic markers of the *Nostoc* genome, such as *hgl* were identified with the domains. The structure was predicted based on the PKS gene cluster and then confirmed through NMR.[64]

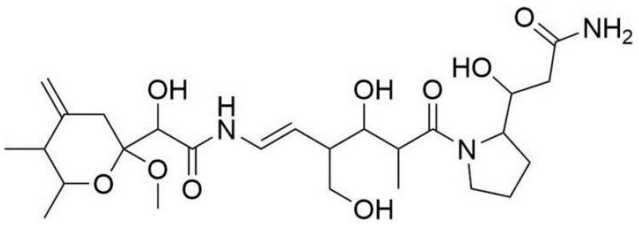


Figure 1.20 Structure of nosperin

Metagenomics can also be targeted to find particular functionalities or structures, such as to guide the discovery of tryptophan dimer (TD) gene clusters that are rare in the environment and are thus likely to encode for compounds with unprecedented structures. TDs have also been shown to bind diverse Molecular targets as TD biosynthesis diverges to give rise to a variety of distinct core structural classes that are then often highly modified by pathway-specific collections of tailoring enzymes. This was achieved in a metagenomics study in 2015 by Brady et al. soil samples were collected from across the United States and the genes responsible for upstream TD biosynthesis (CPAS sequences) were amplified by PCR. Previously unknown sequences

were identified by phylogenetically clustering the environmental sequences with known CPAS sequences to generate a library. ultimately, this library yielded a sequence that groups with the known pyrrolinone indolocarbazole clusters (hys tag) and a sequence (red tag) that groups away from all CPAS found in known clusters, which was predicted to be associated with the biosynthesis of a new TD core. These genes were induced into a new host and two new natural products with TD cores were identified: hydroxysporine and reductasporine.[65]

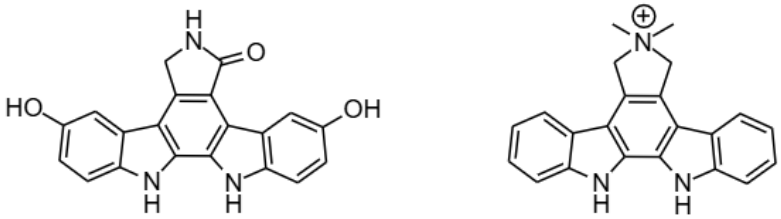


Figure 1.21 Structure of hydroxysporine and reductasporine

Although metagenomics tools are useful, they rely on sequenced DNA being available for comparison and identification of BGCs, and thus do not function for unsequenced microbes or to the same capacity for organisms which do not possess BGCs. In these instances, chemoinformatic tools can be utilised not to look for unknown DNA but instead for unknown compound structure. Looking for new structures without knowing their bioactivity might first appear counterproductive if the seeker is looking for new drugs, however if you consider one of the main principles of biology - that structure equals function - new structures will have new functions. This combined with the fact that if they are natural products have evolved with bioactivity, makes seeking novel structures whilst bypassing DNA sequencing and bioassay screening an attractive workflow for discovering novel natural products.

1.12 Cheminformatics in natural product discovery

Cheminformatics is a relatively new discipline that involves the use of computer technologies to process chemical data. It emerged from several older disciplines such as computational chemistry, computer chemistry, chemometrics, quantitative structure-activity relationship (QSAR), chemical information, etc.[66] Several research groups have recently used computational methodologies to organize data, interpret results, generate and test hypotheses, filter large chemical databases before the experimental screening, and design experiments.[67] An important component in this interplay of technologies are databases providing measured analytical data (e. g. bioactivities, chromatographic data, mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy data) for known NPs and their interrogation with computational methods. However, even the largest of these databases cover only a small fraction of the known NPs, for which reason computational methods are

increasingly being employed also for the prediction of MS fragmentation and NMR spectra, sometimes in combination with structure generators.[68]

One of the most popular recent examples of this is the global natural product social molecular networking database (GNPS). GNPS is an online database and molecular networking tool in which users can submit HPLC-MS2 data and compare their data sets with published data sets to identify novel compounds within their data. By using GNPS, MS2 data from natural products can be quickly analysed and structurally annotated, allowing for efficient dereplication ahead of isolation and/or bioassay-guided studies. GNPS may not only be used for the analysis of known groups of natural products but also those with unique structural fragments. For example, the 'feature-based Molecular networking' workflow on the GNPS website can group and visualize tandem mass spectrometry data based on spectral similarity. Such a resource can be very useful for researchers who explore the metabolic profiles of uncharacterized biological samples and hope to discover new natural products.[69]

By using the 'molecular networking' function on the GNPS website, researchers can map the fragmentation patterns of all the known members of a family of natural products and reveal non-ribosomal peptide synthetase (NRPS) extension units, which are building blocks that are incorporated into the growing natural product chains. Such an approach is featured in a recent publication about the discovery of potential new lipopeptides in the study of marine sediments. The researchers first performed liquid chromatography-tandem mass spectrometry (LC-MS/MS) on the environmental extracts to acquire the raw mass spectral data before the equipment of molecular networking was utilized. By interrogating the resultant molecular network and leveraging knowledge of the biosynthetic logic of NRPS assembly lines, the study successfully identified a new lipopeptides family defined by the presence of a 3-amino fatty acid chain.[70] As a result, through the aid of GNPS, numerous potential drug candidates have been discovered in different kinds of projects. This powerful and versatile technology will undoubtedly facilitate and advance the future of drug discovery, driving the development and innovation of medicine in years to come.

Also, GNPS has shown great potential for the discovery of natural product-based personalized therapeutic agents. In a recent research study, scientists use bioinformatics tools and GNPS to analyse the human microbiome data generated by the Human Microbiome Project, which is a project that aims to sequence the genetic material of microbial communities that inhabit various locations of the human body.[71] The researchers hope to discover how the microbes and their genetic materials interact and contribute to the health of the human host. By taking advantage of the famous

'molecular networking' workflow on the GNPS website, the scientific team was able to visually present the similarities and differences found within the small molec ules obtained from the human body. This not only provides a map of the human natural product exposome but also demonstrates the power of GNPS as an effective approach in translational bioinformatics research, further substantiating its role in the current and future landscape of drug discovery.[72]

As well as new technologies for discovering natural products, it is also necessary to look to new sources. For most of natural product based antibiotic history the focus has been on bacteria, and for natural product based chemotherapies the focus has been on plants.[38, 46] For this thesis, the source of natural products is a genus of algae known as *Euglena*. The following section will provide a background to algae to provide a context for *Euglena*, and highlight the reasons why this particular organism is of particular interest to the field of natural product research.

1.13 Introduction to Algae

Algae describes a large and diverse polyphyletic group of typically autotrophic organisms, ranging from unicellular to multicellular (seaweed) forms.[73] Algae are found in various ecosystems, mostly marine and freshwater, and to a lesser extent in some terrestrial environments, in snow, ice, deserts and in lichen.[74] In recent years, there has been growing interest in the potential use of algal natural products, with researchers working on different species of this highly diverse group of photosynthetic organisms. This is in part due to the relatively recent discovery of promising bioactive compounds from algae and organisms associated with them, renewing interest in a research area that had been in development since the late 1950s.[75] Algae are eukaryotic organisms with more complicated biosynthetic pathways compared to bacteria, which translate into more complicated natural products.[76] Most possess more genomic information than bacteria which means cell for cell they contain more biosynthetic potential than bacteria. With an estimated 72,500 algal species yet to be discovered and 50,000 held in culture collections, the untapped metabolic potential of these organisms is significant.[77]

Despite being distinct organisms, algae are often thought of as “little plants”, indeed algae are plant’s evolutionary ancestor, through a series of endosymbiotic events occurring 1.5 billion years ago algae obtained the ability to photosynthesise and some later evolved into land plants.[78] Microalgae possesses a higher photosynthetic capacity than land plants however, at 10-50 times a higher efficiency.[79] As well as the ability to photosynthesise, most algae also produce some form of starch and some signalling

molecules derived from macroalgae can also activate defence pathways in plants, such as carrageenans.[80],[81] Algae, however, lack specialized roots, stems, vascular bundles, and a diploid embryo stage. Although incredibly varied, the characteristic features of algae are the presence of an autotrophic chlorophyll-bearing thalloid plant body, the absence of sterile tissue around its reproductive structures, and zygote development by mitosis or meiosis but not via embryo formation. Algae are polyphyletic, no single common evolutionary ancestor has been identified, and although chlorophyll a is present in all algae, as well as in plants and cyanobacteria, secondary photosynthetic pigments vary.[73]

Additionally, although there are more than 7 times as many known plant species than algal, 372,000 compared to 50,589,[77,82] the inherit diversity in algae could be argued to be greater of that than within plants: evolutionary distance between different algal groups is incredibly large and the phenotypic variety is remarkable.[73] For example, the chlorophyte lineage contains both the smallest and the largest known free-living single-celled eukaryotes, *Ostreococcus tauri* (less than 1 um) and *Caulerpa taxifolia* (5 – 65cm),[83, 84] and multicellular forms that range in size from the alga *Tetrabaena socialis* consisting of only four cells,[85] to the largest alga, *Macrocystis pyrifera*, a heterokont which grows in underwater beds and is commonly compared to redwood forests due to its capacity to span over 60m.[86]

The classification of algae has changed over time and remains in flux, initially based on shape, colour and the absence or presence of flagella, then taking into account the evolutionary clade then as genetic tools improved, the sequence similarities. At present, most researchers accept nine divisions of algae: Cyanophyta, blue-green algae or cyanobacteria; Rhodophyta, red algae; Cryptophyta, cryptomonads; Dinophyta, dinoflagellates; Heterokontophyta, heterokonts; Haptophyta, haptophytes; Chlorarachniophyta, chlorarachniophytes; Euglenophyta, euglenoids and Chlorophyta, green algae.[87]

1.14 Cyanobacteria

Cyanobacteria are the only prokaryotic algal division, there are up to 8000 predicted cyanobacteria species of which 2968 are confirmed species.[88] Cyanobacteria are also responsible for the primary rise of atmospheric oxygen around 2.3 billion years ago and are known as the key organisms for fixing nitrogen.[89] They are also widely known for their role in the formation of harmful algal blooms (HABs), predominantly in freshwater environments: HABs occur due to the over proliferation some of blue-green algae which are capable of producing toxins.[90]

Microcystis is a common toxin-producing blue-green alga that is responsible for many HABs. This can cause toxic effects on other organisms; for instance, in some fish species, exposure to the toxins produced by *Microcystis* can cause liver damage. On rare occasion, HABs affect humans (e.g. skin irritation, nausea), animals (e.g. pets that swim in water containing a HAB) and other organisms in the environment (e.g. by reducing the oxygen available to other species). It is common for HABs to occur in nutrient-rich, stagnant water, and this has led to increasing concern over the impact of nutrient pollution on water bodies, particularly in countries with intensive agriculture.[91]

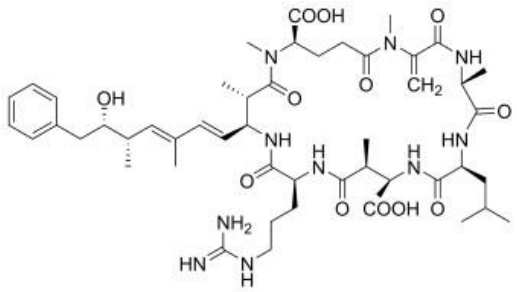


Figure 1.22 Structure of cyanobacterial monocyclic heptapeptide.

Like gram-negative bacteria, cyanobacteria produce lipopolysaccharides which are known to be weakly toxic. The most common and more efficacious toxins produced by cyanobacteria are isolated from *Microcystis* - Microcystins; a family of monocyclic heptapeptides, of which more than 70 variants have been identified in the Molecular weight range of 909 to 1115 Da.[92] Also produced by cyanobacteria is the neurotoxin anatoxin-a (ATX), a water-soluble secondary bicyclic amine, structurally related to alkaloids. Various analogues of ATX have been identified, such as dihydroanatoxin-a (HTX-a) which has similar toxicity level to ATX, and dihydrohomoanatoxin-a (dhATX-a) which is 10 times less potent than ATX.[93]

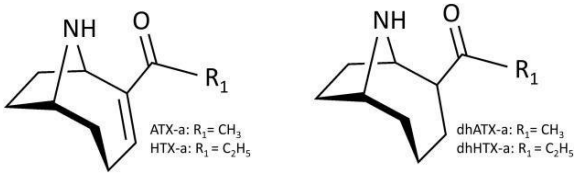


Figure 1.23 Structure of ATX analogues.

ATX is found worldwide, but the predominant analogue varies depending on the geographical location. The most common in New Zealand is dhATX-a, followed by dhHTX-a, where ATX-a is most prevalent in surface waters from Asia–Pacific, European and North American regions. The native role of ATX is unclear, another toxin produced by filamentous cyanobacteria Cylindrospermopsin (CYN), a tricyclic alkaloid, is thought

to have an allelopathic activity as it is has been observed to restrict the growth of green alga *Chlorococcum in vitro.*[93]

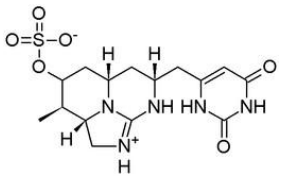


Figure 1.24 Structure of dolastatin 10 synthetic analogue

Various natural products from cyanobacteria are also known to have beneficial properties, a synthetic analogue of dolastatin 10 is currently used against Hodgkin's lymphoma and more than 90 genera of cyanobacteria have been observed to produce compounds with potentially beneficial activities. The diverse cyanobacterial metabolites possessing beneficial bioactivities belong to 10 different chemical classes (alkaloids, depsipeptides, lipopeptides, macrolides/lactones, peptides, terpenes, polysaccharides, lipids, polyketides, and others) that exhibit 14 major kinds of bioactivity. However, no direct relationship between the chemical class and the respective bioactivity of these molecules has been demonstrated. Most of the active cyanobacterial molecules are considered as being produced either through the non-ribosomal peptide (NRP) or the hybrid polyketide-NRP biosynthetic pathways, or by the ribosomal synthesis of pro-peptides that are post-translationally modified (RiPP).[94]

Peptides make up the largest group of compounds with beneficial activity recorded in cyanobacteria at 25.4%, this is followed by peptide subclasses depsipetides (15%), lipopetides (10%) then alkaloids (9.2%), macrolides and lactones (8.5%), terpenes (6.9%), polyketides (3.8%), polysaccharides (1.2%) and 10% having no clear classification. In terms of bioactivities, cytotoxicity is the most common observed making up 30.4%, antibacterial (11.9%), antigfungal (7.7%) and enzyme inhibitory (14.3%) are also more common relative to the other bioactivities observed.[95]

1.15 Rhodophyta

Rhodophyta are macroalgae in which the photopigments phycoerythrin and phycocyanin are dominant, these pigments give the seaweed its characteristic red colour and are used in the textile industries. Additionally, of all the marine seaweeds Rhodophta contains the highest amount of proteins. As a general rule, the protein content of algae usually ranges between 5 and 20%, but species of red algae can achieve greater proportions, with maximum values reaching 47% of total dry weight. There are an estimated 7000 different species of Rhodophyta, and they are a rich source of

hydrogel compounds that are used widely in the food, pharmaceutical and cosmeceutical industries as gelling agents. Additionally, they contain high amounts of Mycosporine-like-amino acids (MAAs) which are used in moisturizers or for their antiphotoaging effects. Rhodophyta species vary in their composition but in general a singular cell is 40-50% polysaccharides in dry weight, and the polysaccharides agar and carrageenan, are some of most studied and commercially applied compounds extracted from seaweeds, being used as stabilizers, emulsifiers, and homogenizers.[96]

Recent studies have also explored the therapeutic potential of compounds from Rhodophyta as it contains a very high amount halogenated compounds, which exhibit diverse biological activities including anti-bacterial, antifungal, anti-inflammation, cytotoxic, and insecticidal activity.[97] In one study, the activity of two different (methanolic and dimethyl sulfoxide) extracts of *Gracilaria corticata* and *Gracilaria edulis* (methanolic and dimethyl sulfoxide (DMSO) extracts), against pathogenic bacteria such as *E. coli*, *Bacillus subtilis*, *B. cereus*, *S. aureus*, *Photobacterium* sp., and *Pseudomonas fluorescens* were tested. GC-MS analysis revealed the presence of numerous bioactive metabolites such as sulphurous acid, 2-ethylhexyl isohexyl ester, eugenol, benzene, and phthalic acid in both red strains. Eugenol is also found in clove, where it was also shown to exhibit antibacterial activity via the disruption of the cell structure of the lipopolysaccharides layer of the cell membrane.[98]

Crude extracts of *Asparagopsis armata*, *Brongniartella byssoides*, Heterosiphonia plumosa, Plo-camium cartilagineum, and Ceramiumciliatum were tested against three cancer cell lines.[99] Significant activity against the Jurkat cell line (T-lymphocytes) was demonstrated by *Asparagopsis armata, Brongniartella byssoides* and Heterosiphonia plumosa, only *Brongniartella byssoides* had cytoxic activity against K562 cell lines (human leukaemia cell line). No correlation between phenol content and anticancer activity was observed in this study despite the well documented antioxidant activity of phenols. Rhodophyta is rich in phenols, extracts of which show the ability to scavenger ROS and inhibit lipid peroxidation. [99]

1.16 Cryptophyta

Whilst Rhodophyta and cyanobacteria are both well-researched organisms which produce commercially valuable organisms, cryptophyte are comparatively less understood and utilised. Cryptophyte algae form one of the major groups of phytoplankton, with more than 20 genera composed of 200 species. They are unicellular microalgae, found in marine, freshwater and brackish environments. They evolved from secondary endosymbiosis event in which a single eukaryote engulfed a red alga. This has resulted in a cell with two nuclei, two cytoplasts, one of each is in the

chloroplast, which is covered with four membranes, and with unique content and distribution among algae of harvesting-light pigments. Depending on their accessory pigments, cryptophytes are bluish, reddish, brownish or green in colour, they have chlorophylls *a* and c_2, phycocyanin (PCY) or phycoerythrin (PER). Like most algae, they are a rich source of omega-3 fatty acids eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), they also produce high quantities of sterols and amino acids which contribute to their function as primary producers.

They lack a cell wall, instead possessing an outer and inner periplast layer. This makes them easy to break down making them highly bioavailable, more so than dinoflagellates, but they grow very slowly, less than 0.8 divisions per day, which is commercially undesirable. Cryptophyta do contain the beta-cryptoxanthin pigment which has been shown to have effectiveness against lung cancer, but this pigment is also found in cyanobacteria, rhodophyta and certain vegetables such as carrots.[100]

1.17 Dinophyta

Dinophyta or dinoflagellates are a large group of motile phytoplankton, many species of which possessing bioluminescence. Like cyanobacteria, they can also form HABs and produce toxins harmful to both humans and aquatic life – such as saxitoxin which is 1000 times more deadly than cyanide by mass. Not all dinoflagellates are photosynthetic, and a variety of plastids are distributed throughout the division from distinct evolutionary endosymbiotic events. Most photosynthetic species contain chlorophylls *a* and c_2, the carotenoid beta-carotene, and a group of xanthophylls that appears to be unique to dinoflagellates, typically peridinin, dinoxanthin, and diadinoxanthin. These pigments give many dinoflagellates their typical golden-brown colour. However, some dinoflagellates have acquired other pigments through endosymbiosis, including fucoxanthin. The nuclear structure of Dinophyta is unique, very few or in some cases no nucleosomes are associated with the DNA and the chromosomes are always visible, they do not go through any condensation stage in mitosis. [101]

Dinoflagellates are found in marine, freshwater, brackish waters, snows and ice, they are also found in symbiotic relationships with corals and jellyfish. The symbiotic marine dinoflagellate *Symbiodinium* sp. is the dominant algal symbiont in reef-building corals, and *Cladocopium sp* is harboured by its jellyfish host *Mastigias papua*. These endosymbiotic dinoflagellates are of interest to researchers because they produce long carbon-chain polyol compounds which show high bioactivity. However, it can be difficult to culture the endosymbiont in the absence of its host, attempts to achieve this with *Cladocopium* have been unsuccessful. *Symbiodinium* has been cultured axenically

and from it novel natural products were isolated with anticancer activity: amphoteric iminium alkaloids, symbioimines and symbiospirol suppress differentiation into osteoclasts, and thus would be a candidate for treating osteoporosis. From *Amphidinium klebsii,* which is an endosymbiont of the flatworm-like invertebrate *Amphiscolops langerhansi,* three groups of antibacterial polyketides have been isolated: the Amphidinins, the Amphinidols and the Amphidinolides.[102]

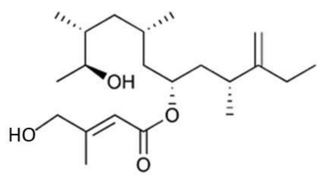


Figure 1.25 Amphidinin E

Amphidinin E showed antibacterial activity against *Staphylococcus aureus*, and *Bacillus subtilis* with an MIC of 32 µg/mL. The Amphinidols also displayed antifungal activities against *Aspergillus sp.* and *Candida sp.* and 34 of the amphidinolides displayed anticancer activity. Amphidinol-22 was isolated from crude extracts of *A. carterae* and was found to exhibit moderate cytotoxic activity against lung, liver, and pancreas cancer cell lines. The amphidinolides most effective against cancer cells are Amphidinolide N, which was found to exhibit potent cytotoxic activity against human cervix adenocarcinoma cell, and Amphidinolides B and H, which were revealed to be effective against murine leukemia (IC_{50} = 0.14 and 0.48 ng/mL, respectively, In addition to amphidinolides, Amphidinium spp. possess other long-chain compounds such as luteophanols, colopsinols, and caribenolide, the latter of which possesses cytotoxic properties. The most potent anticancer compounds identified to date are two macrolides (isocaribenolide-I and chlorohydrin-2) recently isolated from a free-living *Amphidinium* sp. (strain KCA09053) and found to possess high cytotoxic activity against human cervix adenocarcinoma cells. [102]

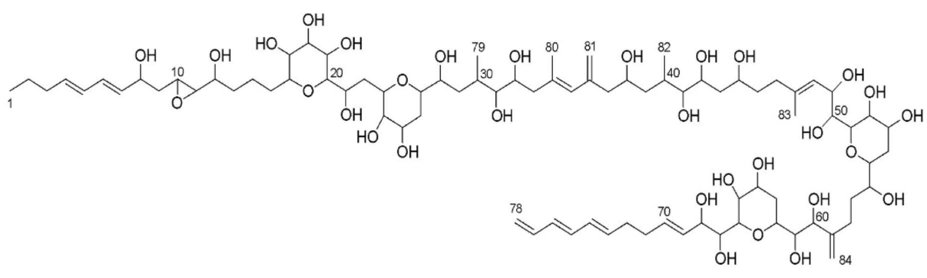


Figure 1.26 Amphidinol 22

1.18 Heterokontophyta

A diverse group of photosynthetic, from the secondary endosymbiosis of a red alga, and heterotrophs, they are a monophyletic group containing 17 classes and represents a diverse group of marine, freshwater, and terrestrial algae. Contained within heterokontophyte are the brown seaweeds, and the Chrysophyceae - "golden algae" – which are typically freshwater, planktonic unicells but also contain the multicellular Xanthophyceae and the Eustigmatophyceae, many strains of which are soil-dwelling, although some are marine.[103] They are all eukaryotic and are characterized by the presence of stiff tripartite external hairs, which can be present on the flagella or cell surface.[104] In some cases the hairs have been lost but they are still considered Heterokontophyta due to genetic evidence that their ancestors once possessed stiff tripartite external hairs.[105]

Despite being a highly varied division there is not as much interest in the study of Heterokontophyta for the purpose of discovering bioactive compounds. One study showed that crudes extract of *Halopteris filicina* displayed antimicrobial activity against *B. subtilis* in disc diffusion assays at 11mg/disc in methanol and 6.5mg/disc in dichloromethane. Hexane, methanol and dichloromethanol extracts from *H. filicina* all showed antifungal activity against *C.albicans*. However, the green algae measured in the same study showed more potent antibacterial and antifungal activity at lower concentrations on average and one red alga strain, *Sphaerococcus coronopifolius*, displayed antifungal activity against *C. albicans* comparable to the antifungal medication amphotericin b.[106]

1.19 Chlorarachniophyta

Only found in marine environments, and most commonly located in tropical and temperate waters, Chloroarachniophyta contain mainly mixotrophic algae – the photosynthetic capability arising in this case from the endosymbiosis of a green alga. Chloroarachniophyta are morphologically and genetically similar to filose amoebae and sarcomonads, and are able to “catch” pray through use of a net-like structure formed from actin microtubules. As many as 150 individual Chloroarachniophytes can come together to combine these net structures into a superstructure called a reticuloplasmodial continuum. Although the evolution history of Chloroarachniophyta is of great interest to researchers, they are not investigated as a source of natural products.[107]

1.20 Chlorophyta

There are an estimated 4500 species of Chlorophyta which are commonly referred to as “green algae” and which are closely related to land plants.[108] The core Chlorophyta include three major classes, Chlorophyceae, ulvophyceae, and Trebouxiophyceae, plus

the two smaller lineages, the Chlorodendrophyceae, comprising the scaly quadriflagellates *Tetraselmis* and *Scherffelia* from freshwater, brackish water, marine, and hypersaline habitats and Pedinophyceae, which include asymmetric, uniflagellate, mostly naked green algae from marine, freshwater, or soil habitats.[109]

250 new natural products have been identified from 1798 marine Chlorophyta species between 1965 to 2012, but thus far none have proved to possess useful bioactivities. Ethanolic and methanolic extracts from *Chlamydomonas reinhardtii* have shown antibacterial activity against *Staphylococcus aureus*, *Streptococcus iniae* and *Pseudomonas aeruginosa*, and chloroform extracts have shown activity against *Bacillus subtilis*.[110]

1.21 Comparing plants and algae as a source of natural products

Although plants contain the highest number of known natural products,[36] there are advantages possessed by algae in the field of natural product discovery that plants lack. Firstly, microalgal cells multiply faster than larger plant cells, with numerous species achieving high cell densities in less than two weeks, compared to months or even years taken by plants.[79] This means algal natural products can be produced at a faster rate than plants which expedites discovery workflows and mass production. Secondly, algae can be cultured using waste or saline water in areas unsuitable for the cultivation of larger plants.[111] This means that algae cultivation does not compete with land for agricultural purposes, nor does it require the additional costs attributed to providing a clean water supply, or the environmental consequences in areas with limited water accessibility. Thirdly, algae are already cultured for a wide range of natural products commercially, including antioxidants, fatty acids, enzymes, and food pigments, as well as a source of nutraceuticals, health products and cosmetics which means there is already existing infrastructure for the exploitation of algal species. [112] According to a global market research, the market for algal products across various sectors is expected to grow at a compound annual growth rate of 4.2% from 2018 to 2025 by which point it will possess a total market value of more than 3.4 billion USD.[113]

1.22 Advances in genetic engineering of microalgal species

Despite the apparent advantages of pursuing algae as a source of natural products there exist barriers to its exploiting its full potential at present. Fully sequenced and annotated microalgal genomes were almost completely absent at the turn of the century.[114] Until 2008, only three microalgal species, *Chlamydomonas reinhardtii, Thalassiosira pseudonana*, and *Phaeodactylum tricornutum*, had been sequenced.[115] However, as microalgal research has garnered mounting interest the number of sequenced microalgal genomes has rapidly increased, as well has the number of ongoing microalgal sequencing projects. The first algal genome to be sequenced was

the diatom *Thalassiosira pseudonana* in 2004,[116] followed by green algae model organism *Chamydomoas reinhardtii* in 2007,[117] then coccolithophore *Emiliania huxleyi* in 2013.[118]

As of January 2020, 222 de novo genome assemblies have been published for 149 species from 97 genera. These assemblies represent 32 of the approximately 50 algal classes mostly Chlorophyte (131/222, 59.0%) or Stramenopiles (Heterokontophya) (48/222, 21.6%).[119] Although multiple genome projects are generating gene models, they are often retained in private collections. As of 2022 the number of published, fully annotated and openly accessible microalgal genomes has risen to 64.[120] There is ongoing work to increase the available microalgal genome sequences. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) aims to sequence nearly 700 marine microbial species of 17 phyla, 140 of which are marine microalgal species.[121] Whole-genome sequencing was carried out by the Nelson et al in 2021 on 107 different species of microalgae from the UTEX and NCMA culture collection centres and the New York University Abu Dhabi (NYUAD) isolate collection, with representatives from all divisions of algae.[122]

Eukaryotic algae contain at least three separate genomes, from the nucleus, chloroplast, mitochondria, chromoplasts, and/or leucoplasts which further complicates genetic sequencing.[123] The genomes can also be very large, contain highly repetitive regions and in some cases such as with *Euglena* contain uncommon nucleotides, modified base J in this case,[124] which hinder sequencing.[120] In the case of *Euglena* additionally, annotation is difficult due to the lack of fully sequenced genomes for comparison and due to the absence of biosynthetic gene clusters. [124]

Various transformation techniques have been developed for microalgae, including glass bead agitation, electroporation, Agrobacterium tumefaciens-mediated transformation, biolistic bombardment, CRISPR-Cas, and *E. coli conjugation*. These methods rely on integrating expression cassettes into the nuclear or plastid genome. Transformation efficiencies are often low, especially in the nuclear genome, due to cell physical barriers and random integration, which causes "position effects" leading to variable expression and the need for extensive screening.[125]

Genetic engineering through CRISPR/Cas9, TALENs (transcription activator-like effector nucleases), and ZFNs (Zinc finger nucleases) has been used in certain strains of microalgae to target specific genes involved in key metabolic pathways.[126] TALENs has been used to target uridyl diphosphate (UDP)-glucose pyrophosphorylase in *Phaeodactylum tricornutum*, leading to enhanced lipid accumulation.[127] Furthermore,

the use of TALENs to disrupt the urease gene has been carried out successfully in *P. tricornutum*, albeit with a low gene disruption efficiency of only 16%.[128]

Although ZFNs was the first genetic tool to used to knockout COP3 and COP4 genes in *Chlamydomonas reinhardtii*, effective knockouts were only observed in certain strains whereas CRISPR/Cas9 has been shown to be more specific and adaptable. The application of the CRISPR/Cas9 system in microalgal species was initially demonstrated in *Chlamydomonas reinhardtii* to disrupt a starch synthesis pathway, leading to enhanced lipid accumulation for biofuel production.[129, 130] However, the continuous expression of Cas9 exhibited cytotoxic effects in C*. reinhardtii*, this was ameliorated by the transient delivery of the in-vitro assembled Cas9/sgRNA ribonucleoprotein (RNP) complex via electroporation. Moreover, the transient expression of the Cas9/sgRNA RNP complex substantially reduces off-target effects and cytotoxicity, thereby enhancing the overall efficiency of gene editing.[131]

Furthermore, the CRISPR/Cas9 system was employed to successfully edit the urease gene in diatom*, Thalassiosira pseudonana*, achieving disruption efficiency exceeding 60%.[132] The initial application of the CRISPR system to the industrially significant oleaginous marine microalga *Nannochloropsis oceanica* involved the disruption of the nitrate reductase gene; however, this was accomplished with relatively low efficiency, around 1%. Subsequently, a Cas9 editor line of *Nannochloropsis gaditana* was developed, which exhibited constitutive expression of Cas9 and was utilized for the editing of targeted transcription factor genes, achieving efficiencies as high as 78%.[133]

Advances in molecular tools, such as CRISPR-Cas9, have enabled transformation of several algal species, though challenges remain in achieving consistent results across diverse taxa. Despite progress, genomic data for many algal lineages are limited. For example, *ulva mutabilis* is one of the few multicellular green algae with robust genetic tools, while transformation in green algae species species like *Chromochloris zofingiensis* and *Chlorella* spp. remains inconsistent.[115] Continued development of genomic resources and transformation protocols is essential for unlocking algae's full potential in biotechnology and environmental applications.

1.23 An introduction to Euglena

Euglenids (*Excavata, Discoba, Euglenozoa, Euglenida*) describes a group of free-living, single-celled flagellates living in aquatic environments. The uniting and unique morphological feature of euglenids is the presence of a protective proteinaceous structure around the cell membrane called a pellicle, they lack a cell wall like that observed in plants and bacteria. Euglenids exhibit diverse modes of nutrition, including phagotrophy and photosynthesis. Photosynthetic species (*Euglenophyceae*) constitute

a single subclade within euglenids. Their plastids embedded by three membranes arose as the result of a secondary endosymbiosis between a phagotrophic eukaryovorous euglenid and the *Pyramimonas*-related green alga. Within photosynthetic euglenids, three evolutionary lineages can be distinguished. The first, most basal lineage is formed by one mixotrophic species, *Rapaza viridis*. Other photosynthetic euglenids are split into two groups: predominantly marine *Eutreptiales* and freshwater *Euglenales*. *Euglenales* are divided into two families: *Phacaceae*, comprising three monophyletic genera (*Discoplastis, Lepocinclis, Phacus*) and Euglenaceae with seven monophyletic genera (*Euglenaformis, Euglenaria, Colacium, Cryptoglena, Strombomonas, Trachelomonas, Monomorphina*) and the focus of this thesis: the poly-phyletic genus *Euglena*.[134]

Euglena get their name from the Greek words "eu," meaning "good," and "glena," meaning "eyeball" or "eye." This name is attributed to the organism's distinctive red or orange eyespot, in which carotenoid globules containing photosynthetic pigments astaxanthin and zeaxanthin aggregate, and the function of which is to aid phototaxis during locomotion. *Euglena* are protists, their cells range in size from 15 to 500 micrometres in length, the average cell length is between 40 to 80 micrometres and they are generally elongated and spindle-like in shape.[135] *Euglena* predominantly move via locomotion of their long anterior flagella, one of two flagella the Euglena possess although the short flagellum does not extend beyond the pellicle.[136] When under capillary stress *Euglena* can switch from locomotion to a series of pellicle contractions and expansions to push the cell through characteristic movement known as metaboly, this coupled with its phototaxis ability makes Euglena gracilis an excellent model for nanobots which need to navigate tight environments as displayed in the development of the bio-microbot “Ebot”.[137]

In 2011, *Euglena* acquired a new subclass known as Euglenophycidae, in total the diversity contained within *Euglena* is extensive, there are over three thousand different species of *Euglena* distributed globally in both freshwater and saltwater environments.[138] The model organism for the species is *Euglena gracilis*, a freshwater species common in pond ecosystems in the UK.[139] *Euglena* were previously classified amongst the green algae (*Chlorophyta*) due to their morphological similarities and the endosymbiosis of green plastids. However, more recent genetic evidence has shown that they are actually much more closely related to the Kinetoplastids (Euglenozoa), a class which also includes protazoan parasites *Trypanosoma,* responsible for African sleeping sickness, and *Leishmania* which causes leishmaniasis.[140]

1.24 Genetics of Euglena

In order to fully determine the full evolutionary origin of the *Euglena* genome, it is necessary for sequencing to be completed. Slow progress in genome elucidation for *E. gracilis* has been due to the organism possessing a large genome (estimated at >1.5 Gbp), with a high prevalence of repetitive sections and the hypermodified 'base J' (β -D-glucopyranosyloxymethyluracil).[140] Base J is found across the phylum Euglenophyta and is thought to prevent local transcription. *Trypanosome* and *Euglena* genomes both contain similar rates of base J modification (replacing approximately 1% of all thymidines) located in inactive telomeric regions in trypanosomes but spread more uniformly across the *Euglena* genome.[141] As of 2022, the Euglena International Network (EIN) intends to sequence all known species of Euglenids over the next decade which should complement development of biosynthetic toolkits in *Euglena*.[142]

The chloroplast genome and transcriptome of *Euglena gracilis* has been fully sequenced, and the genome has been partially sequenced. From this it has been discovered that although less than 1% of the *Euglena gracilis* genome is comprised of coding sequences, this still totals 36, 562 proteins -16,000 more than the human genome.[140] Open reading frames (ORFs) are interspersed along the genome between long repetitive sequences with high AT content, causing a disparity between transcriptome GC content (~60%) and the genome GC content (~50%).[143] Indication of polyketide synthase genes and non-ribosomal genes have been identified in *Euglena gracilis* but this far no corresponding natural products has been identified as expressed from these genes.[144] This indicates the presence of novel natural product potential, additionally, preliminary analysis has identified several enzymes in *Euglena* associated with secondary metabolite synthesis, whilst BLAST hits show the presence of non-redundant proteins with no homology. Attempts to predict the structure of these proteins using tools such as PredictProtein, Proteus or Alphafold yield unrealistic results or fail completely, further cementing the unique chemistry present within *Euglena's* metabolome.[139]

Unlike most organisms, euglenozoan protein abundance is not regulated by gene expression, but predominantly through post-transcriptional factors. This may be linked to other genetic traits observed in *Euglena* such as polycistronic genes, which are usually uncommon in eukaryotes, and the presence of non-conventional introns, which have only been found in euglenozoans and confer unique stability onto secondary RNA structures by the formation of circular structures. The function of polycistronic genes in *E. gracilis* is unknown, comparison to polycistronic genes in trypanosomes encode proteins with unrelated biosynthesis pathways or Molecular function. Trypanosomes also process their polycistronic genes post-transcriptionally using trans splicing to

produce monocistronic mRNA. This process could also occur in *E. gracilis* as conventional and non-conventional, cis and trans splicing have been observed,[125] giving *E. gracilis* a large degree of optionality regarding its mRNA processing which could explain the large adaptability displayed by *Euglena* to changing environmental conditions.

Due to the endosymbiotic events in *Euglena* evolutionary history, horizontal gene transfer (HGT) from foreign species, and endosymbiotic gene transfer (EGT) from endosymbiont to host have introduced foreign genes into the *Euglena* genome which further add to its complexity. As stated, *Euglena'*s photosynthetic capabilities began with a *secondary* endosymbiosis event in an ancestor over one billion years ago, one branch of this common ancestor then evolved into the kinetoplasts such as *Leishmania* and *Trypanosoma*, another underwent a secondary endosymbiosis event and evolved into *Euglena*.[144] Both events were accompanied by EGT, with the endosymbiont being reduced from an independent organism to a DNA-containing organelle as genes were transferred from endosymbiont to host. The chloroplast genome is only made up of 70,000 base pairs and 51 genes - making it one of the smallest chloroplast genomes.[145] The third genome, the *Euglena's* kinetoplast, a mass of tightly packed circular DNA within the mitochondria of kinetoplasts which coordinates the cell's flagella and it is a site for the extranuclear RNA editing. Much research on *Euglena's* kinetoplast centres around trying to understand these RNA editing processes. with four endosymbiotic genomes: *Euglena*-specific genes, Kinetoplastida-specific genes, eukaryotic genes that are spread in other eukaryotes, and genes acquired during the secondary endosymbiosis, which, in addition to evidence of extensive HGT, make *Euglena* particularly genetically unique.[146]

One notable aspect of the kinetoplast in *Euglena* is its association with the organism's bifunctional lifestyle. *Euglena* has the capacity to switch between autotrophy and heterotrophy, making it a mixotrophic organism. During periods of light availability, *Euglena* engages in photosynthesis. The kinetoplast contributes to this versatility by housing genes associated with both photosynthetic and non-photosynthetic metabolic pathways.[134] This dual functionality further showcases the adaptability of *Euglena* to varying environmental conditions.

There are also considerations to made of the “metagenome” of *Euglena*, as evidence suggests that certain species may live in symbiosis with non-Euglenid microbes. The term phycosphere has been used to describe phytoplankton–bacteria relationships in which metabolites and infochemicals are used to facilitate interactions between an alga and other organisms inhabiting the same environment which may enable organisms to

gain competitive advantages.[147] This benefit of Coculturing *Euglena* with bacteria has been utilised within biofuels research to increase *Euglena* biomass yield.[148] In natural niches, *Euglena* have been found to co-occur with specific microorganisms. Bacterial communities have been associated with *Euglena sanguinea* and its toxic blooms, with 16S rDNA amplicon sequencing identifying a significant number of *Deinococcus–Thermus* bacteria.[147] Furthermore, a bacterium was isolated from the cell surface and cytostome of *E. gracilis*, and whole-genome sequencing revealed it to be a *Paenibacillus* sp., however the exact nature of these two relationships has not been determined.[149]

Euglena species show incredibly high tolerance to acid stress and heavy metal contamination, exceeding that of other microalgal species, which has pushed *Euglena* to the forefront of bioremediation research.[150] Thus far, this tolerance has been attributed at least in part to the increased expression of antioxidant species in *Euglena* and its use of metal chelating enzymes such as phytochelatin.[151] Symbiotic relationships may also improve the stress tolerance of *Euglena* as isolation of *E. mutabilis* from mine waste water runoff revealed it existed in mutualistic relationships with *Cryptococcus* spp. that appeared to increase the acid and heavy metal tolerance of the alga. Full-length amplicon sequencing and targeted Sanger sequencing identified the fungus *Talaromyces* sp. and the bacterium *Acidiphilium acidophilum*, which combine to form a fungal, algal, and bacterial (FAB) consortium with *E. mutabilis* that has been shown to improve its cadmium tolerance.[152]

Euglena expresses a wide range of carbohydrate-active enzymes, suggesting an unexpectedly high capacity for the synthesis of complex carbohydrates for a single-celled organism. Lectin- and antibody-based profiling of whole cells and extracted carbohydrates revealed a complex galactan, xylan and aminosugar based surface on the *E. gracilis* pellicle.[153] Inspection of the transcriptome of *Euglena* showed that there are more CAZymes than in most other sequenced algae, but fewer than in the land plants, which require complex cell walls to support their growth. There are also many more CAZymes encoded than in the other sequenced Euglenozoa, such as the human pathogens *Trypanosoma brucei* and *Leishmania braziliensis*.[154]

1.25 Significance of *Euglena* in Natural Product Research

Two advantages of isolating natural products from *Euglena* include its fast growth rate and it is non-pathogenicity. The extraction of natural products from microbial culture normally requires large quantities of biomass, and a short generation time is favourable. *Euglena* has a doubling time of 8-10 hours, which is approximately 24 times faster than other microalgae used commercially, such as *Nannochloropsis sp*. or *Phaeodactylum tricornutum*. Moreover, being classified as a class I microorganism, there are fewer

safety controls on the handling of *Euglena*. This is important as cultivation of microorganisms in the open may lead to contamination of surrounding areas or waterways. The regulations and oversight required when working with pathogenic microorganisms will increase the cost of cultivation and harvesting of natural products.[155]

Euglena, was initially of great commercial interest due to its ability to produce large volumes of wax esters that can be used as biofuels.[148] Wax esters are described as "low-value, high-volume products", more recently, *Euglena* has been shown to produce "high-value, low-volume" products such as vitamins A, C and E.[156] The most efficient system of extraction from *Euglena* would therefore be comprised of processes which extract both these categories from one *Euglena* growth. There are some complications to this, as where the wax esters produced for biofuel are a fermented product of the high-value, low-volume compound paramylon. Whilst biofuels from *Euglena* are attractive to due *Euglena's* position as a photosynthetic microbe that can be cultured aquatically, therefore leaving more space for land to be used for crops rather than for biofuel production at current the cost of *Euglena* biofuel production far exceeds that of sugar cane biofuels.[157] In addition, paramylon has unique physicochemical properties and can be used in a wide range of applications such as in the food industry as a thickening agent. Additionally, under laboratory conditions it has antitumor, radical scavenging, and immunomodulatory effects, and thus potential pharmaceutical applications. Moreover, *Euglena* can be induced to produce paramylon in large enough quantities to be commercially viable under dark and anaerobic conditions.[158] As well as paramylon, *Euglena* produces a variety of bioactive compounds with potential pharmaceutical and industrial applications. Several *Euglena* species such as *Euglena gracilis* and *Euglena mutabilis* have been reported to synthesize polyphenolic compounds, which have strong antioxidant properties and could be used as preventive supplements against aging and cancer.[151] These compounds also have applications in cosmetics, it has been found that the extracts of *Euglena gracilis*, which is known to contain the xanthophyll astaxanthin, can protect the skin against damage from UV radiation.[159] However, other than paramylon, these compounds are not unique to *Euglena*, despite its extensive metabolic potential only three unique natural products have been identified from *Euglena*.

1.26 Unique natural Products Produced by Euglena

Euglenapterin

The first novel natural product to be identified from *Euglena* was the heterocycle Euglenapterin reportedly isolated from *E. gracilis* in 1980. It was the first discovered

naturally occurring pterin with a 2-dimethylamino group at the 2-position of the pyrimidine, which is a unique structural feature among the known pterins. Euglenapterin is a byproduct stored within the cells after the biosynthesis of biopterin. It is thought pterin compounds play a role in the photoreception occurring in the eye spot of Euglena although further research in this area is required. Unfortunately no bioactivity has been reported for Euglenapterin.[139]

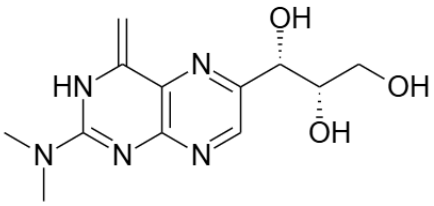


Figure 1.27 Structure of Euglenapterin

Euglenophycin

The first natural product to be isolated from *Euglena* with bioactivity was euglenophycin, a toxic alkal oid that has been isolated from certain species of euglenoids, such as *Euglena sanguinea* and *Lepocinclis fusca*. Euglenophycin has been found to have acute toxic effects on fish, and is responsible for the 'watermelon blood condition' in lakes, which is a foaming disease of aquatic organisms. This toxin also has potential for medicinal use and recent studies have discovered that euglenophycin has anti-trypanosomal activities, indicating that it could be used as a new lead compound against trypanosome parasites.[160]

Euglenophycin is a disubstituted piperidine alkaloid structurally similar to the alkaloid solenopsin, the venom found in fire ants. This structural similarity allows researchers to predict the mechanism of action behind Euglenophycin's cytotoxic abilities. Solepsin is known to have an inhibiting effect on PI3K/AKT in the mTOR pathway, which may explain why Euglenophycin has shown anticancer activity in mammalian cell lines. The Euglenoid toxin has been shown to have inhibitory characteristics toward cancers including against production and proliferation of cancer cells and of the vasc ular endothelial growth factor. In vivo studies with Euglenophycin demonstrate its ability to arrest colorectal cancer cells in the G1 phase of growth and diminish their migratory abilities. Euglenophycin further showed anti-tumour activity in mouse models, both decreasing tumour size and production of pro-inflammatory markers.[139]

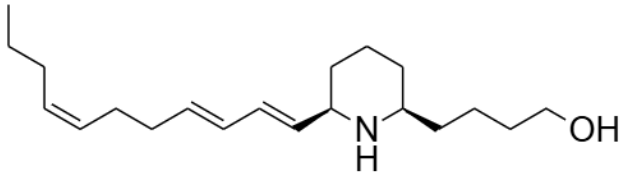


Figure 1.28 Structure of Euglenophycin

Euglenatides

The Euglenatides were discovered recently in 2020, these compounds are cyclic lipopeptides produced by several strains of *Euglena*, containing two rare non-proteinogenic amino acids, β-aminoisobutyric acid and dihydroxynorvaline, and a novel hydroxylated lipid tail. They are only produced at appreciable levels when the microalgae are cultured in a minimal media with single amino acids as the nitrogen source, and so far have been detected in three *Euglena* species including *E. gracilis*. While the role of these compounds within the organism is unknown, since over forty Euglenatide-like metabolites have been detected within the three species it is expected to be of important biological relevance. The Euglenatides showed excellent cytotoxic activity in MCF-7 breast cancer cells, with only one tenth of the value of the clinically approved drug vorinostat. They also show antifungal activity against Aspergillus and antihelminthic activity against *C. elegans*.[139]

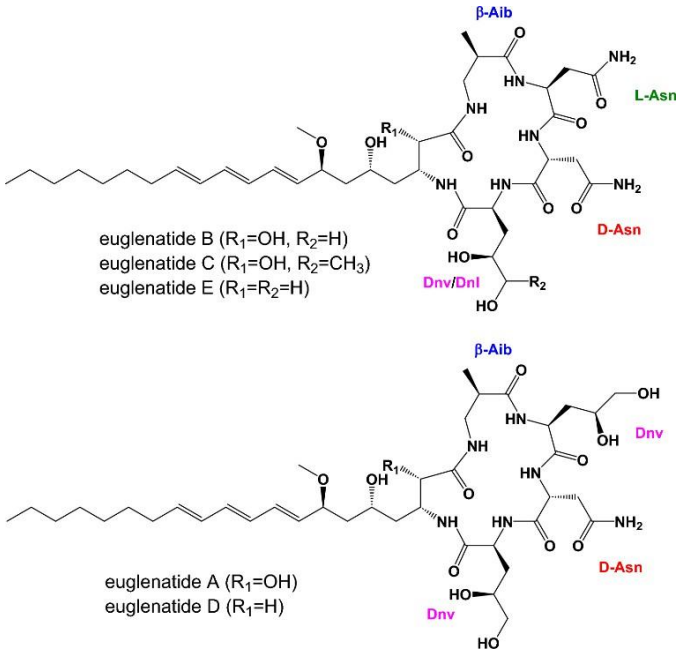


Figure 1.29 Structure of Euglenatides A-E

This chapter has aimed to emphasize the pressing need for novel natural products for a wide range of applications, but in particularly for the treatment of infectious disease and cancers which are of particular relevance in the global scope of human health. Two major barriers to novel natural product discovery have been identified: the re-isolation of known compounds and the re-investigation of well-studied microbes. Two solutions have been explored to ameliorate these issues; the use of "bottom-up" approaches to natural product discovery – gene-first or compound-first as opposed to "top-down" bioassay approaches; and the investigation of new sources, with focus on microalgae. Further detail into *Euglena* was explored as a particularly attractive source for novel natural products due to the apparent mismatch between the minimal unique natural products isolated from this microalga when compared to its large, complex genome, and diverse, adaptable metabolome. This coupled with its ease of growth and industrial scalability make it highly promising as a source of novel natural products. This thesis seeks to evaluate *Euglena* as a source of natural products by employing Molecular networking to assay the number of "purifiable" compounds from *Euglena*. In addition, as Coculturing has been shown to increase biomass in *Euglena* this thesis aims to discover if Coculturing of *Euglena gracilis* further increases natural product diversity. Finally, this thesis will aim to discover a purification process suitable for the extraction of a novel natural product from *Euglena.* These aims are summarised below as four research questions.

Research Questions

Based on the need for novel natural products outlined and *Euglena's* potential as a source for novel natural products this thesis aims to answer the following research questions.

1) Can Molecular networking be used to identify candidates for novel natural product identification from *Euglena*?
2) Does Coculturing *Euglena gracilis* with other organisms increase the diversity of natural products produced compared to monocultures?
3) Is there potential within the *Euglena* genome from which a novel compound can be identified?
4) Can a novel natural product be isolated from *Euglena*?

Contributions

All work is my own unless otherwise stated, in which case contributions are noted when necessary.

Chapter Two: Can molecular networking be employed with *Euglena* to select targets for natural product isolation?

2.1 Natural Product Databases

In order to carry out compound-first identification of novel natural products an extensive and reliable database of natural products has to exist so that known natural products from new analyses can be accurately identified and removed from consideration. In addition, in order to ensure a maximum of natural products are assayed for novelty, large complex datasets are preferred that encompass the scope of a organisms metabolic potential.

There is no global consensus database in which all published Natural Products (NPs) are entered and stored, instead there exists a variety of NP databases with varying degrees of accessibility and functionality. 120 different NP databases have been published between 2000 and 2020: 98 of them remain publicly accessible, 50 are free to use, and 43 of them have metadata communicating information such as application and source of the NP. NP databases without a theme exist alongside those with a focal theme such as "plants", "food", "drug-like", or for pharmacologically active components of alternative medicines in China, India and Africa. They also vary extensively in the size of their libraries: from the smallest, the Indofine Chemical Company Inc. natural products, containing just 56 NPs, to the largest, the chemical structure lookup service (CSLS), which contained over 27 million unique structures in 2007. However, the CLSL is no longer maintained, and the lookup service is no longer running, thus requiring manual search through the data.[65] The ability to rapidly compare multiple data points against a large database library is one of the main advantages of computer analysis in online NP databases. Despite this, NP databases vary in their ability to both accept and compare large data files; some allow the upload and mass searching of multiple data points while others require individual input and output.[161]

The loss of accessibility, poor nature of user interfaces, and paywall barriers to some databases have all resulted in the loss of information from the NP research community – and as such, a lack of practical application. For example, the Novel Antibiotics database contains 5430 NPs with antibiotic activity, though no structures are available to download.[162] ChemIDplus returns over 9000 queries in a search for "natural product" but has no option to bulk download the results of said query, thus requiring the user to undergo individual analysis of all entries.[163]

There currently exists no themed microalgal database.[9] Four of these NP databases curate marine NPs. Two are no longer accessible: the Marine Natural Product

Database (MNPD) and the Marine Compound Database (MCDB), both of which contained a few hundred entries. The Dragon Exploration System on Marine Sponge Compounds Interactions (DESMCI) is technically accessible, but none of the data visible, and the Seaweed Metabolite Database (SWMD), is accessible and maintained but contains only 423 unique structures.[164]

2.2 NMR and MS based databases

Databases do not just exist for final structures of NPs; some also accept raw, unprocessed data, such as NMR or MS, and some accept a combination. NMRshiftDB is an open database for the entry of organic molecules and their NMR spectra. It contains 1875 peer-reviewed molecules, and a large number of spectra that makes it excellent for NMR-based dereplication.[9] MassBank of North America (MoNa) is the most popular database for dereplication based MS search queries, as it accepts data from more sources than either the European MassBank or the Japanese MSSJ MassBank, and contains rich community-curated metadata.[165] The European MassBank currently contains 86,575 unique spectra, and 14,788 unique compounds.[166] MoNa contains 692,367 spectra, and 227,044 unique compounds.[167]

Mass Spectrometry databases are of significant use to NP researchers as all NP scientists use MS in their research –as part of the discovery workflow and/or as part of a larger catalogue of evidence for publication. MS is a highly sensitive analytical tool, with top instruments displaying 24 million resolution and higher with sub parts per billion (ppb) mass accuracy, and more general MS instruments displaying high resolutions of greater than 40,000 and mass accuracies below 1 ppm.[69] This makes them ideal for the identification of NPs from microbes, which are typically produced in tiny amounts unless genetically overexpressed.

NMR is the most effective tool for final structural elucidation, but MS is more suitable to the start of a workflow being more sensitive and able to be used in conjunction with chromatographic techniques such as HPLC. Furthermore, MS/MS fragmentation data can be used to allocate compounds to most-likely structures in the preliminary stages of investigation, as well as to compare in databanks – such as METLIN and GNPS.[168] Additionally, MS^2-acquisition can be coupled with chromatography, which helps to more accurately assign putative identifications to compounds with similar Molecular weights.

METLIN is a free to use MS database of over 1 million molecules equivalent to >1% of Pubchem's 93 million compounds, and over 4,000,000 curated MS^2 spectra.[169] It contains over 850,000 Molecular standards with MS^2 data generated in both positive

and negative mode. Furthermore, the use of MS^2 data in addition to precursor m/z values reduces the number of matches in METLIN searches from tens to hundreds of compounds down to only a few.[167] However, METLIN limits data analysis to just a few LC-MS files at a time, the data is not easy to download once analysed, and its libraries are not freely available.

2.3 Global Natural Product Social Molecular Networking

The global natural product social Molecular networking (GNPS) webpage began as a collaborative project in 2011 between the Bandeira and Dorrestein research groups in attempt to provide an open-source database for raw, unprocessed and unknown mass spectrometry datasets. It became available to the wider scientific community in 2014 has now established itself as global network for researchers to contribute and share MS/MS data.[169]

GNPS is a free-to-use database which provides a continual peer review of data, rich metadata, and easily downloadable NP structures. GNPS has now expanded to 49 US states and over 150 countries worldwide. The database is used by scientists in industry, academia and government within fields of biomedical research, environmental science, ecology, forensics, microbiology and chemistry.[168]

GNPS provides the ability to analyse large datasets and compare them to all publicly available data on the Mass Spectrometry Interactive Virtual Environment (MassIVE) data repository. The GNPS database contains 221,083 reference MS^2 spectra, including all provided from MassBank. The user can choose whether to make their data publicly accessible, all public data is continually reviewed each month against the NIST 2017 spectral library, and high-confidence spectral matches are annotated as such, creating a "living data set" of continual NP identification. Whilst new, and not as big as other well-established NP databases, GNPS is rapidly growing and regularly accessed, with over 200,000 page views a month. In the same time frame, it performs over 6,000 analysis jobs with Molecular networking being the most popular.[9]

By compiling this vast array of MS/MS data, GNPS serves as an extensive database that aids in dereplication – the process of identifying known compounds from complex mixtures. This feature is particularly important in natural product research, where rapid identification of known compounds allows scientists to focus their efforts on discovering novel molecules.[170] Dereplication is primarily achieved through its advanced algorithm for spectral similarity searching called "Molecular networking".

Molecular networking is a computational analysis that aids both visualisation and interpretation of the complex data that arises from MS^2 analysis. By exploiting the assumption that structurally related molecules will produce similar fragmentation patterns, MS^2 data can be displayed in graphical form where each ion is represented by a node and connected by edges to spectrally similar nodes, forming a cluster. This aids dereplication by preventing re-isolation of both known compounds and their unknown-but-similar analogues, which will be present in the same cluster. By the same logic, Molecular networking could also aid the discovery of unknown analogues of antibiotic agents if they were found to be connected in the same cluster. In recent metabolic profiling experiments, it has been established to function successfully as a tool in the identification of novel compounds.

Looking ahead, there is immense potential for further developments within GNPS. For instance, incorporating additional analytical techniques such as nuclear magnetic resonance (NMR) spectroscopy could offer complementary information about compound structures. Furthermore, expanding the scope beyond natural product research to include other fields like metabolomics or clinical diagnostics would maximize the utility of GNPS.

The utilisation of GNPS in the work discussed in this thesis has allowed the metabolites produced by *Euglena* to be viewed in terms of their structural novelty when referenced against the large and living GNPS database. The following chapter will layout natural produced by 32 *Euglena* strains, with description of their novelty, the conditions of their expression and predicted structures.

2.4 Analysing Euglena with GNPS

Euglena gracilis has great potential for natural product biosynthesis, as described in Chapter 1. Due to the difficulties in genome sequencing that persist in *Euglena*, an alternative route to novel product discovery is to interrogate the secondary metabolite structures directly. Currently there does not exist an untargeted analysis of the metabolic profile of *Euglena*, this chapter highlights the use of Molecular networking to produce such a profile. The use of HPLCMS-MS to generate spectra for the Molecular network restricts the metabolic profile in two ways which are beneficial to the selection of novel natural products for downstream drug discovery. The first; a m/z range of 300 – 1500 m/z is applied which restricts the metabolites assayed to those in the "small molecule" range which are most applicable in drug development. The second is the use of a chromatography column, which identifies a metabolites retention time on a C18 column. This information can support the purification process through HPLC in the subsequent isolation process and excludes

compounds which may be novel but are not readily isolated through chromatographic techniques.

For this aim, 32 different Euglenoid algal strains were obtained from the c ulture collection of algae and protoxoa (CCAP) and National Institute for Environmental Studies, Japan and cultured (10 mL) in three different liquid media of increasing complexity from the most minimal, Af6, to an equal ratio of JM and soil extract, JM:SE, and to the most complex: a biphasic mix of soil in water: SWBi. After 6 weeks of growth under continuous light at 25°C, half of each culture was harvested. Whole culture extractions were performed through sequential liquid-liquid extraction with solvents of increasing polarity: ethyl acetate, butanol and methanol to extract increasingly polar metabolites. The extractions were then dried, reconstituted in methanol, centrifuged and analysed using HPLCMS-MS. The raw spectra were then processed and uploaded onto GNPS to produce a Molecular network which consists of 1504 nodes with 2094 edges when visualised in Cytoscape.

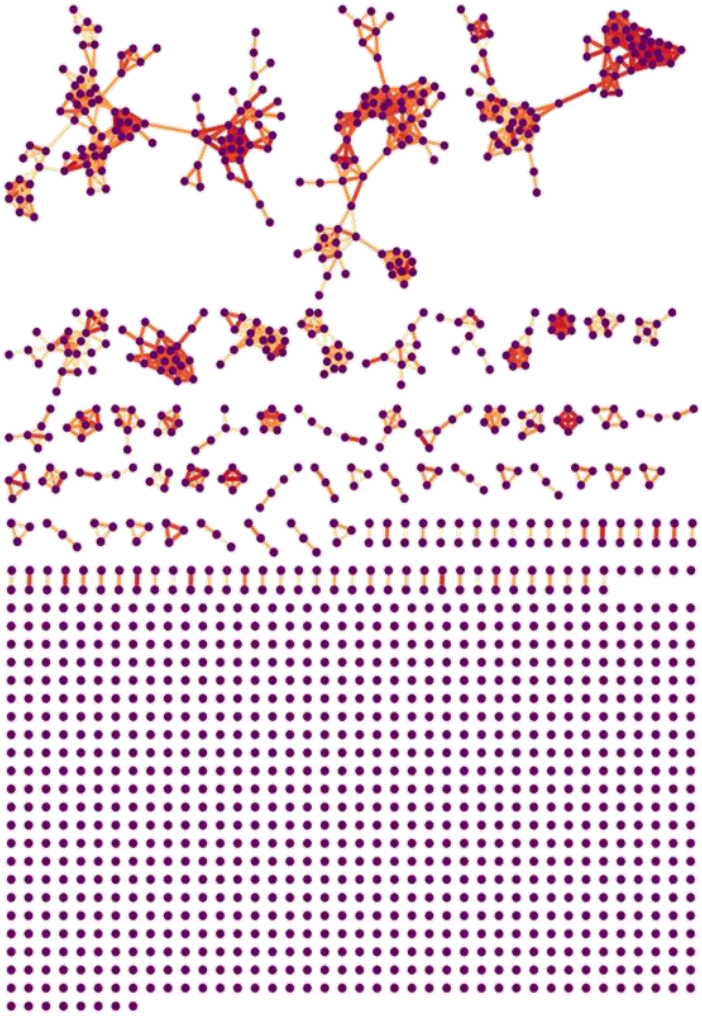


Figure 2.1 Molecular Network One: 32 Euglenoid strains and three media controls, JM:SE, SWBi, Af6 extracted into EtOH, BuOH, MeOH

This molecular network was generated under default GNPS settings for a large dataset which processes the open MS^2 spectra files as follows: first, GNPS aligns each MS^2 spectrum to the other spectra in the dataset using a modified cosine spectral similarity algorithm. The identical spectra, with a cosine score of 1, are then combined into a single node. GNPS then searches for similar compounds which have a modified cosine score of $\geq 0.7 < 1$, the stringency of which can be varied for further analysis. Nodes are known to be structurally related by their comparable but not identical MS^2 spectra, and these are connected by edges to form a cluster. A cluster therefore is a set of two or more nodes connected by edges, and all nodes in a cluster can be considered structurally similar molecules

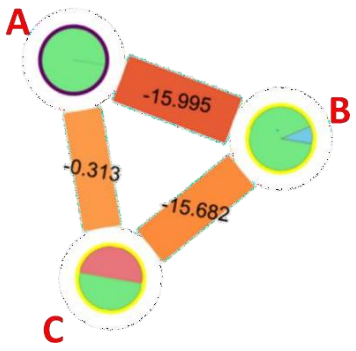


Figure 2.2 Cluster of 3 nodes: "A, B, C", taken from Cytoscape Network 1. Edge labels represent mass dofference between nodes. Edge colour represents cosine score. Node colours represent contribution by different cultures

One of the features available in Cytoscape allows user to modify edge annotations to visualise the differences between nodes in the same cluster. When the mass difference between nodes is shown along the connecting edge, it can be possible to make an estimation of an unknown compounds structure by comparing it to related compounds in the same cluster.

Node	m/z	RT / s	Species, Media, and Extraction Method
A	667.576	858.708	*Euglena viridis,* JM:SE, BuOH
B	651.583	819.708	*Trachelomonas pertyi*, JM:SE, BuOH *Astasia klebsii*, SWBi, MeOH *Euglena viridis,* JM:SE, BuOH
C	667.263	654.046	*Astasia ocellata var provasolii, Af6, EtOAc* *Euglena viridis,* JM:SE, BuOH

Table 2.1 Cluster of 3 nodes: "A, B, C", taken from Cytoscape Network 1. Euglena virids contributes to all three nodes.

For example, in the cluster of Figure 2.2 , both A differs from B and B differs from C by a mass difference of ~16 – this could indicate the structures differ by an oxygen atom. As A and C differ by a m/z difference of ~0, they could be isomers of each other. If the structure of A or C is elucidated, it is then possible to predict that B is a hydroxylated analogue of this structure.

Mass difference and clustering are not the only data points to take into consideration. Additionally, the similar retention time for A and B further indicates a high degree of structural similarity. Furthermore, all 3 nodes are produced by *Euglena viridis* when grown in JMSE and are all extracted by butanol.

However, B is additionally produced by CCAP1283.13 when grown in JMSE and CCAP1204.15 when grown in SWBi, but neither strains produce A or C. Moreover, C has a retention time 200 seconds slower than A and is also additionally produced by a different strain, CCAP1204.9 in AF6, that does not produce B or C.

Additionally, a thicker darker red coloured edge indicates a higher cosine score, which means A is considered more closely related to B than it is to C, even if there is a greater difference in mass and retention time. What this exemplifies is that whilst the Molecular network is data-rich and provides clues on structural information, to be certain of the chemical structure more information by means of isolation and structural analysis is required. However, Molecular networking can provide a guide from which putative structures can be estimated.

2.5 Cytoscape Analysis of Molecular Network One

It is easier to make predictions on unknown nodes if it is connected to a library match in the same cluster. For example, if a peptide is identified by a GNPS library match, and if the identified peptide is connected to an unidentified node by say 147.0723 Da, it could be assumed it was a phenylalanine conjugate of the peptide. This means that if a few compounds from a metabolite family have been identified within a cluster, the user can assume unidentified nodes connected within the same cluster are likely to be undiscovered members of the same family.[69] However, if they are connected to a library match then they are not a novel class of natural products and therefore unlikely to be suitable for purification – unless connected to a particularly relevant known compound.

In this network, GNPS has identified some nodes from library matches, including in the large clusters. It is possible that the unknown nodes in these clusters are part of the same chemical family as the known nodes. If this is the case it is possible

to predict an estimated structural backbone for the unknown nodes.

For example, in the top left cluster shown in Figure 2.3, which has manifested as two tightly connected clusters connected by one longer edge, a 2,6-dibromo-4-(2,4-dibromophenoxy)phenol has been identified. This could indicate this cluster shares a structural relatedness specific to brominated diphenyls. In the same cluster but in the rightmost section, we can see another diphenyl compound, Isosulochrin. From this we could predict that the cluster is constituted by diphenyl compounds, with the leftmost being bromo-substituted or halogen- substituted in one way, and the right one being predominantly organic.

In the cluster below it, Sarmentoside B has been identified, a common plant biological natural product and could be part of a larger cluster of glycosides present in this cluster.[171] N,N-Bis(3-aminopropyl)butane-1,4-diamine, a bacterial metabolite, was also identified in the bottom row of connected clusters. It is connected to only one other node with a m/z difference of 0.02, and a retention time difference of 2 seconds which indicates that this unknown node is either very similar, or the same compound.

In the third cluster in of the second row, Homonataloin B was identified, possibly indicating the presence of more Anthrones within this cluster. In the 4th row,4-o- was identified caffeoylquinic acid, a plant metabolite acid has shown to be an active inhibitor of PTPN1,[172] indicating its potential in anticancer therapies. 9-Octadecenamide is another plant metabolite found in a nearby cluster, which could be connected to similar fatty amides – it itself has been shown to have function as a plastic additive and a sleep-inducing drug.[173]

Although this appears a promising start to the network analysis, none of the aforementioned compounds will be considered for the rest of this report as the aim is to discovery novelty. The negative control in this experiment was sterile media extracted with the same solvent-solvent method into ethyl acetate, butanol and methanol. Of all 1504 nodes, 473 are present in the negative control. However, all clusters containing at least one negative control node can also be dismissed as part of the negative control, as they are likely media products that have been modified by the algae, which means the number of negative control nodes is higher.

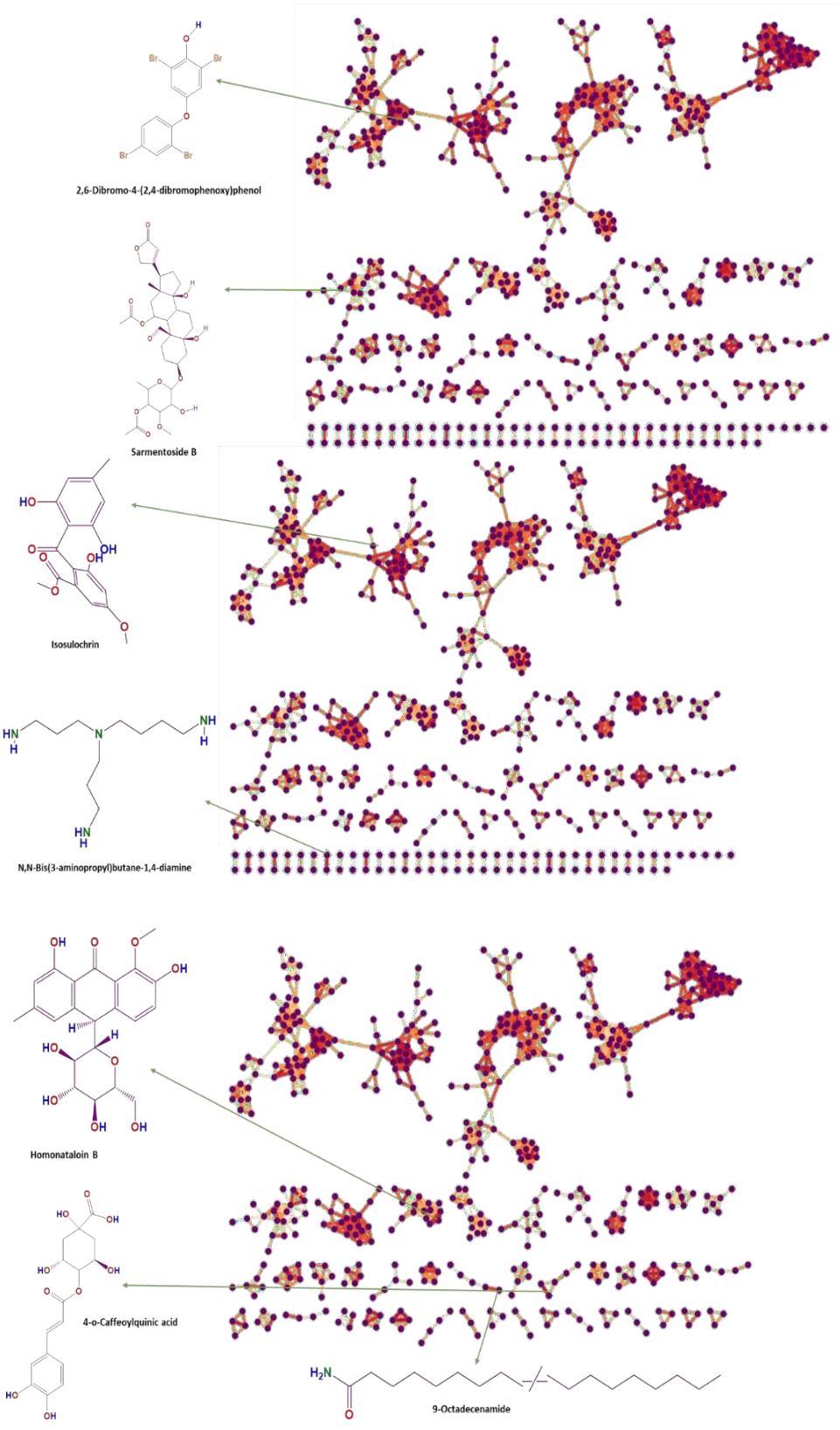


Figure 2.3 Structure of known compounds, taken from Cytoscape Network 1.

Apon adjusting to consider the negative control, most of the large clusters become "negative clusters", and all the previously mentioned compounds to negative control nodes. This explains the presence of substances that you would not expect algae to be producing. For example, both 4-o-Caffeoylquinic acid and 9-Octadecenamide are plant metabolites. Both are only observed in JM:SE and SWBi media, as are all the connecting nodes in both clusters. Both media were made using soil from the same source, which indicates that both are components from said soil, and which have likely been metabolised by the algae into varying analogues.

2.6 Cytoscape Analysis of Molecular Network One: Media

Having excluded all control and control-linked nodes we are left with a network containing 1031 nodes. The fewest compounds are produced in Af6, with only 324 nodes (out of total non-negative nodes), then JMSE with 391, and the highest in SWBi with 589. The biggest overlap between nodes is between JMSE and SWBi, 90% of JMSE nodes are also present in SWBi, but only 11% of JMSE nodes are present in AF6. AF6 shares 26% of its' nodes with SWBi, and 13% with JMSE. SWBi shares nearly half, 49% of its nodes with JMSE and just 14% with AF6. Only 3% of all nodes are found in all three media.

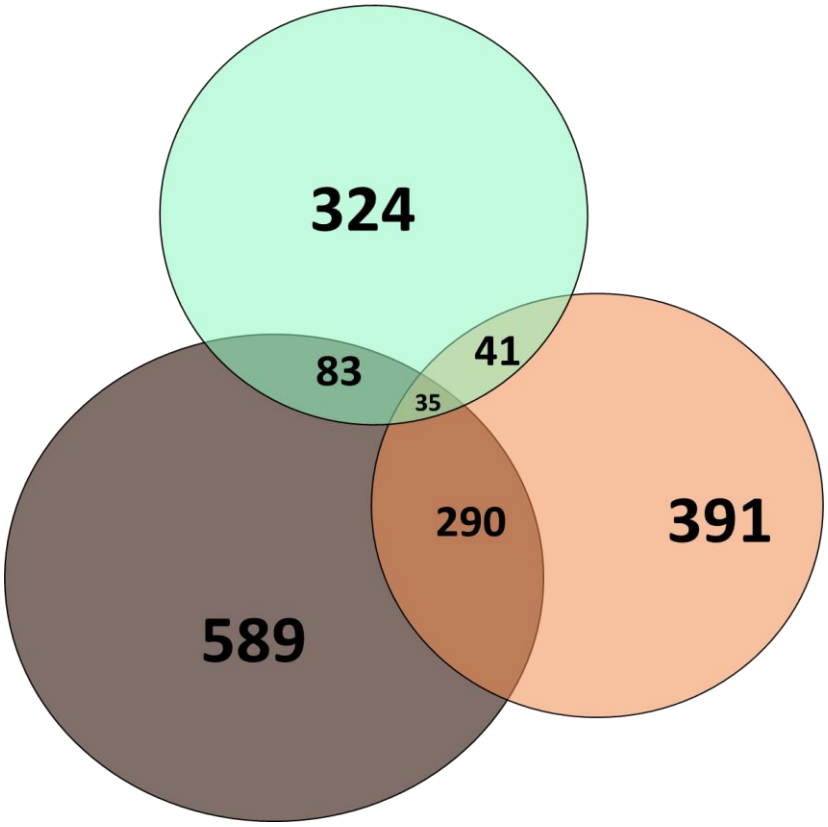


Figure 2.4 Number of nodes in media components: Brown: SWBi, Orange: JMSE, Green: Af6.

A microorganism, particularly one with the metabolic adaptability of *Euglena*, would expectedly produce more compounds in a more diverse environment – more metabolites present in richer media offer more compounds for diverse biological pathways. This hypothesis is supported by the observation that across the three media, compound number rises as the media composition increases. It could be possible that using soil as a media component would introduce a lack of reproducibility in the results due to innate variation in a complex mixture. This does not to be considered in the scope of this analysis as all the 10 mL Cultures were cultured from the same batch of media. However, it could be a factor to consider in future experiments and highlights the importance of noting the number of negative nodes identified in each media so that changes can be identified between future batches. From this analysis; in the total number of negative control nodes is 473: 236 are in Af6, 349 are in JMSE and 441 are in SWBi.

2.7 Cytoscape Analysis of Molecular Network One: Library Identification

When accounting for the negative control, most of the large clusters (>6 nodes) are now effectively in the negative control. Of the 66 clusters, 36 are in the negative control which means only 45% have the potential to be novel compounds. Additionally, of these clusters, 77% are clusters of only two nodes. Two-node clusters tend to have a very small mass difference and when inspected manually are sometimes revealed to be the same compound that has been mistakenly reported as two separate compounds.

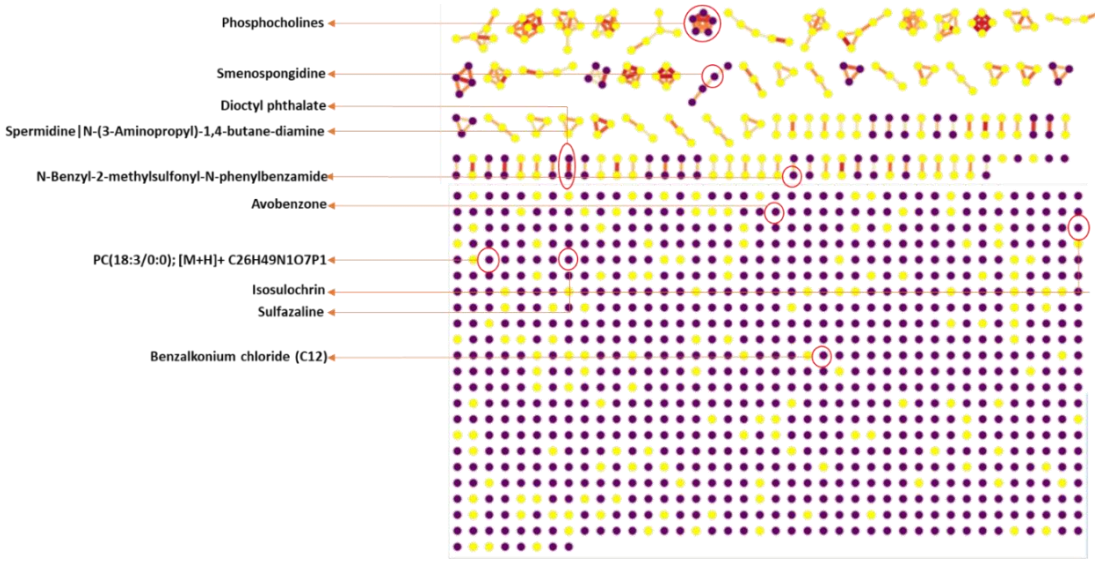


Figure 2.5 Compounds identified in the GNPS library highlighted by red arrow, red circle in Cytoscape Network One. Yellow nodes are in the negative controls or connected to negative control nodes.

Of the remaining network, 60% (893) of the nodes are isolated single nodes, 153 of which are negative control-containing nodes. This means 83% of the unconnected nodes have the potential to be novel compounds.

The presence of so many unconnected nodes both speaks to the high structural diversity of the metabolite profile of these Cultures and also to the lack of reported data on the chemical space of *Euglena* natural products. 25 nodes have library identification matches in the GNPS spectral library. 14 of these nodes are not connected to any negative control node, and none of which are algal-specific metabolites. These are shown in Figure 2.5.

Smenospongidine, a marine sponge antagonist of the Wnt/β-catenin signalling pathway with anti-cancer activity against melanoma, is identified but has not been reported in fresh-water algae.[174,175] It is found in 38% of the spectra, indicating that "Smenospongidine" might be a misidentification of a common metabolite.

Spermidine is a polyamine and microalgal metabolite derived from putrescine that is involved in many biological processes, including the regulation of membrane potential, the inhibition of nitric oxide synthase (NOS) the induction of autophagy, and a key part of trypanothione – an antioxidant unique to *Euglena* and some Trypanosomes .[176, 177] As a common microalgal and bacterial metabolite it is likely this is a correct identification by the GNPS network. However, Spermidine is connected, by a cosine score of 0.9517, to a compound which is a spectral match to Dioctyl phthalate, a common man-made plasticiser. The presence of this Dioctyl phthalate is likely an experimental contaminant, as is the presence of Avobenzone, a common sunscreen ingredient, and Benzalkonium chloride - which is used in cleaning agents.

Isosulochrin is a metabolite produced by some fungal species such as *Aspergillus wentii.*[178] Possible fungal contamination could be the cause, although all Cultures were checked for contamination through light microscopy before extraction and none was observed. Sulfasalazine was also identified, a synthetic arthritis medication that is therefore unlikely to be present in any of the cultures. Both these compounds are antioxidant species, as *Euglena* has reported high oxidative stress capabilities these compounds could be incorrectly identified polyphenols (Isosulochrin) or azobenzenes (Sulfasalazine).

GNPS identified the largest non-negative cluster to be comprised of 6 phosphocholines, these are ubiquitous algal metabolites and this library match can be considered a sensible cluster assignment.[179] Some algal

lysophosphatidylcholines have been reported to possess anti-inflammatory activity, whilst some have reported pro-inflammatory activity.[180] PC(18:3/0:0); [M+H]+ $C_{26}H_{49}NO_7P$ is the only phosphocholine out of the ones identified not to be connected in a cluster to the others. This is unexpected, as it has shares a high degree of strucutral similarity to PC(18:2/0:0), in that they are identical except for the additional of an alkene bond within PC(18:3/0:0). Indeed this cluster contradicts expectations in other aspects as well; it assigns PC(0:0/16:0) to be more strucutrally similar to PC(18:2/0:0) than PC(18:2/0:0) is similar to PC(0:0/18:1) even though the latter both contain the same number of carbons and possess unsaturated fatty tails not saturated ones.

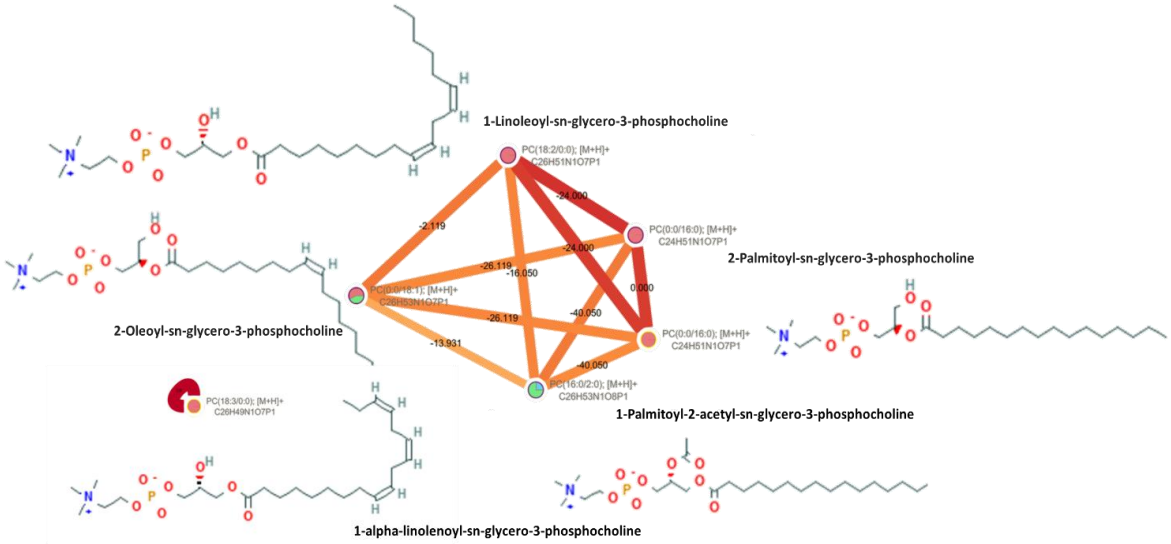


Figure 2.6 Chemical Structures of the phosphocholines identified in Cytoscape Network 1

Additionally, this cluster seems to contain two of the same compound. Although the node pertaining to PC(0:0/18:1); [M+H]+ C26H53N1O7P1 "2" has a 2 second shorter retention time than PC(0:0/18:1); [M+H]+ C26H53N1O7P1 (556 s compared to 558 s) and is present in 14 more samples, they have identical m/z to 3.d.p: 496.334 and are identified with the same chemical formulae. It appears that although they have the same structure, there is sufficient difference in their MS^2 spectra for GNPS to consider them discrete compounds.

The "Raw Spectrum" of the node PC(0:0/18:1); [M+H]+ C26H53N1O7P1 and PC(0:0/18:1); [M+H]+ C26H53N1O7P1 "2" can be seen alongside the "Consensus Spectrum" from the GNPS website [see appendix; the reference spectra in this case were both submitted by Thomas Mertz]. They are both considered to be spectra correlating to the same compound: PC(0:0/18:1); [M+H]+ C26H53N1O7P1, although they are themselves different from each other.

The major peaks for both nodes are the same, but it observable that there are a high degree of "noise" peaks in the spectra of PC(0:0/18:1); [M+H]+ C26H53N1O7P1 "2". This could be that the same compound has been assigned as two different nodes because in some analyses it fragmented more heavily than others causing a divergence in fragmentation patterns that led to GNPS assigning the same compound as two nodes. This would explain also why there are two different consensus spectra that can be used for the same compound. Additionally, this infidelity of splitting patterns could be the reason why so many compounds are in two nodal clusters with themselves.

Compound	Cosine Score	Shared Peaks	GNPS Library Score
PC(0:0/16:0); [M+H]+ C24H51N1O7P1	0.95	15	Gold
PC(18:2/0:0); [M+H]+ C26H51N1O7P1	0.92	11	Gold
PC(0:0/16:0); [M+H]+ C24H51N1O7P1	0.92	9	Gold
PC(0:0/18:1); [M+H]+ C26H53N1O7P1 "2"	0.82	8	Gold
PC(16:0/2:0); [M+H]+ C26H53N1O8P1	0.79	7	Gold
PC(18:3/0:0); [M+H]+ $C_{26}H_{49}NO_7P$	0.76	11	Gold
Smenospongidine	0.71	6	Bronze
Spermidine	0.77	6	Bronze
Dioctyl phthalate	0.86	6	Bronze
Avobenzone	0.84	6	Bronze
Benzalkonium chloride	0.80	9	Bronze
Isosulochrin	0.70	6	Bronze
Sulfasalazine	0.75	15	Gold

Table 2.2 Cosine score, shared peaks and GNPS library score of compounds identified in the GNPS library Cytoscape Network 1.

The GNPS network is not infallible when it comes to assigning library matches. As there are criteria that must be fulfilled for nodes to be connected in a cluster, likewise there are a set of default parameters engaged for GNPS to make a match.

These are: a parent mass peak tolerance of 2.0, an MS2 peak tolerance of 0.5, a minimum cosine score of 0.5 (whereas structures are only connected in a cluster if they have a minimum cosine score of 0.7) and a minimum number of matched peaks of 6. The reference spectra used to make these matches are given a score correlating to the estimated reliability of the data, starting at Bronze and moving up to Silver and then Gold.

Therefore, if there is heavy fragmentation in the MS2 spectra, or only low quality reference spectra on the database then the GNPS match is more likely to be inaccurate. However, as GNPS continually updates and reviews its databases, the accuracy of the matches is likely to increase with time and more user data input.

2.8 Novel Compounds from Cytoscape Network One

Analysis then moved to the next stage: to identify possible novel compounds for isolation and structural elucidation. These targets had to meet certain conditions:

- they could not be part of any negative control containing cluster
- they could not have been given a library identification
- they must have a retention time between two and ten minutes
- they must be present in few Cultures (<10) in order to increase the likelihood of their interesting or uncommon chemistry
- they must have a reasonable intensity in the LCMS/MS chromatogram and have a MS2 fragmentation pattern free of a high degree of noise

The initial 4 conditions can be filtered for on Cytoscape, this leaves us with 387 nodes as possible targets. 49 of these were considered to be a replication of the same compound if they had the same mass to 3 decimal places and similar (within <10 s) retention times and thus dismissed. The number of replications decreased as the mass of the compounds increased, possibly due to higher levels of fragmentation making the matching of MS patterns more reliable. This left 338 possible targets.

These targets were then prioritised from those with the least number of spectra to those with the highest and their masses manually searched for using an EIC on Compass software for each LCMS/MS spectra. Many did not appear at a high enough level above the noise to be considered and were discarded. This also provided opportunity for manual review of the target MS spectra to confirm it matched the GNPS report. This lead to the final selection of six targets.

2.9 Target 1.1

There are seven non-control-node containing clusters with 3 or more nodes in the network. Two of these contain library ID compounds: the phosphocholines and the Smenospongidine-containing cluster. Of the remaining five, four have nodes with retention times outside the HPLC gradient at over 631 s. The remaining cluster was selected as Target 1.1.

Target 1.1 is a cluster consisting of four nodes 134624, 131736, 131733 and 135525 corresponding to m/z's: 488.26, 474.39, 474.391 and 490.386, and to retention times of 601.74 s, 581.26 s, 584.49 s and 582.20 s.

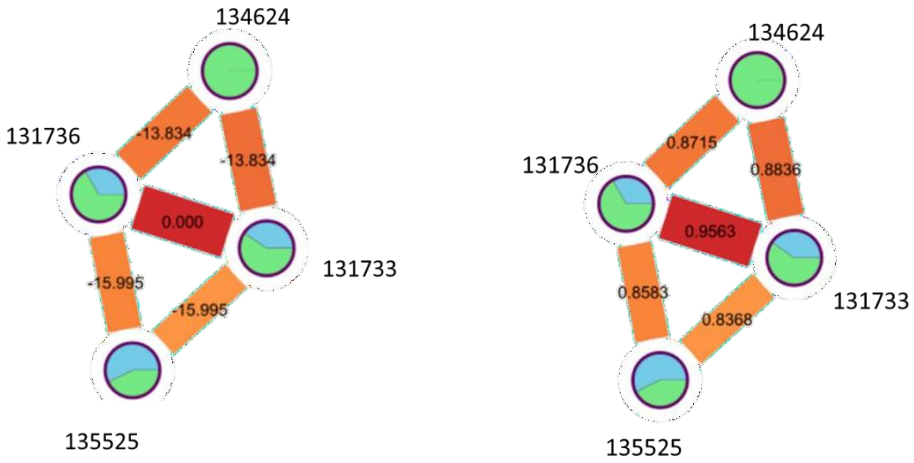


Figure 2.7 Cluster One, Edges on: Left: m/z difference, Right: Cosine score

Although these retention times are close to the 10-minute cut off, which could hamper purification, this is the largest cluster to almost meet all the criteria in this Molecular network which makes it unique in the metabolic profile of the *Euglena* assessed and as such of interest. The most commonly occurring node in this cluster is 131733 which is produced in 17 different Cultures conditions. This may at first appear high but is only 18% of the total 96 cultures, and only 9% of the total extractions. Node 131736 is produced at an intensity in the 10^4 region, whereas the others aren't expressed above an intensity of 5000 which puts them just above the baseline and therefore might not be produced in enough concentration to be isolated from the mixture. However, even if 131736 is the only node from this cluster to be structurally identified the structures of the others could be predicted from the cluster. . All nodes in Target 2.1 were produced by CCAP1261.6 when

grown in JMSE and extracted into methanol in this cluster, this was therefore selected to be grown in large scale in the next stage.

There is a high degree of structural similarity between the nodes in this cluster; there is a cosine score of 0.8715 and a ~14 m/z difference between 131736 and 134624, and a 0.8836, ~14 m/z difference between 134624 and 131733; indicating a possible [CH_2] difference between them. Similarly, there is a cosine score of 0.8583 and 16 m/z between 131736 and 1355525 and of 0.8368 and 16 m/z between 135525 between 131733, indicating a possible [O] difference. The highest cosine score is between 131736 and 131733 which is 0.9563, and these also share a 0 m/z difference, indicating that they could be the same compound or structural isomers.

Although you would expect the similar compounds to be extracted into the same solvent regardless of the species in which it was produced, the compounds in Target 1.1 are seen in all three solvent extractions. However, out of the total 42 extracts below, ethyl acetate is only observed 9 times, but into butanol and methanol both 17 and 16 time respectively, indicating a higher affinity of Target 1.1 compounds for more polar solvents.

Node	m/z	RT /s	Species	Media	Extracts
134624	488.260	601	CCAP1261.6	JMSE	EtOAc, MeOH
			CCAP1283.8	JMSE	BuOH
131736	474.390	581	CCAP1204.3	JMSE	EtOAc
			CCAP1261.6	JMSE	MeOH
			CCAP1271.1	SWBi	MeOH
			CCAP1283.13	JMSE	BuOH
131733	474.391	584	CCAP1204.15	SWBi	MeOH
			CCAP1204.17B	SWBi	BuOH
			CCAP1216.3C	SWBi	BuOH
				JMSE	BuOH, MeOH
			CCAP1224.27	SWBi	
				JMSE	EtOAc, BuOH,
			CCAP1224.33	JMSE	MeOH
				SWBi	BuOH, MeOH
			CCAP1224.50	JMSE	

				SWBi	EtOAc, BuOH,
			CCAP1261.6	SWBi JMSE	MeOH BuOH
					MeOH
			CCAP1271.1	SWBi JMSE	MeOH
					BuOH
					EtOAc, MeOH
			CCAP1283.8	JMSE	
					BuOH, MeOH
			CCAP1283.13	JMSE	EtOAc, BuOH
			CCAP2149	SWBi	
					BuOH
					EtOAc, BuOH
					BuOH
135525	490.386	582	CCAP1224.27	SWBi	MeOH
			CCAP1204.15	SWBi	BuOH, MeOH
			CCAP1261.6	JMSE	EtOAc, MeOH

Table 2.3 Molecular weight, retention time, producing species, media conditions and extracting solvent for Target 2.1.

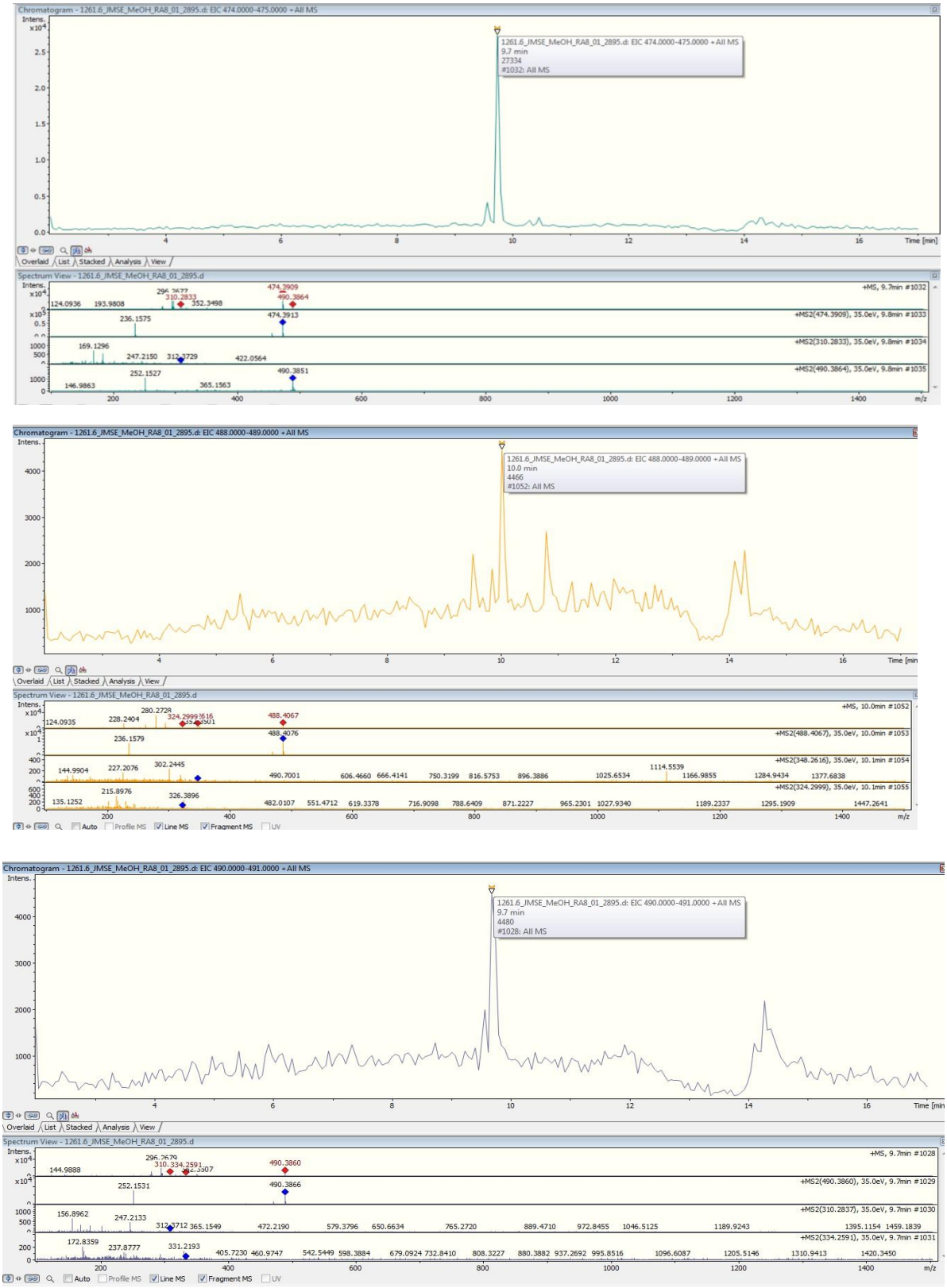


Figure 2.8 EICs for Target 1.1

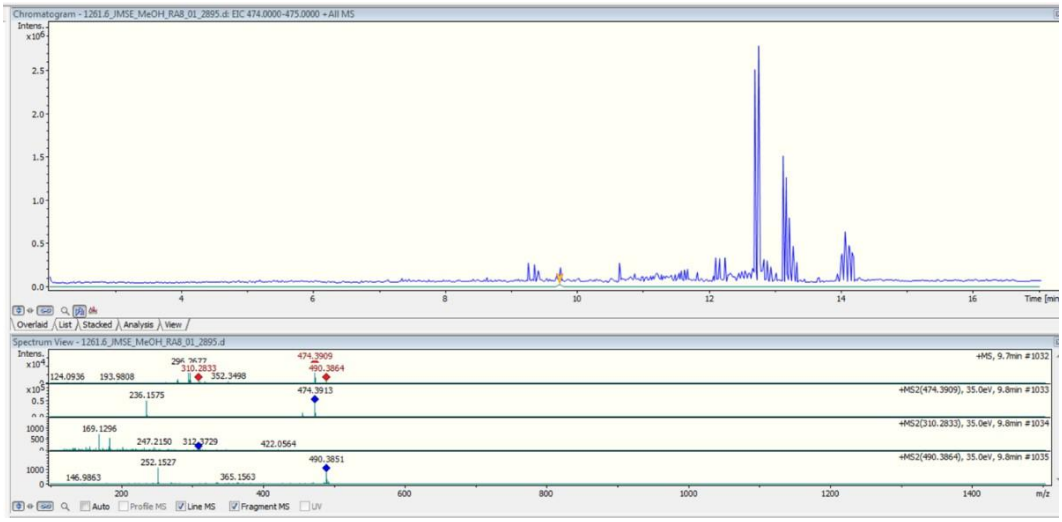


Figure 2.8 HPLCMS-MS spectra CCAP1261.6, grown in JMSE and extracted into MeOH. Target 1.1 highlighted by yellow arrow on green trace.

By subtracting the difference between the major peak fragments in the MS2, it is possible to estimate possible form ulas for the fragment. In some cases, this can help build up a loose structural prediction. This is most easily done with peptides, as they fragment well in MS2. Therefore, all fragment differences for the major fragments were calculated by subtracting the mass of the major fragments from the parent mass and the resulting difference was compared to the m/z of amino acids (-H2O) to identify and matches. Most fragment differences were too high to match a fragmenting amino acid, some were equal to the loss of water.

Node	Parent m/z	Major Fragments m/z	Δm/z	Predicted Fragment Formula
134624	488.26	470.00	18.20	H_20
		236.16	252.26	$C_{17}H_{33}N$ $C_{16}H_{31}N_2$

Table 2.4 Fragments for Node 134624.

Predicting possible structures can also be attempted by using the parent mass to predict possible Molecular formulas for the compound. For node 134624 and all other Molecular formula predictions were generated using the ChemCalc MF-Finder tool. Selections were devised with the elemental parameters: C1-100 H1-200 N0-2 O0-20 S0-2 P0-2. The Molecular formulas output were then narrowed down manually by comparing m/z to 4 d.p., choosing Molecular formulas with 0 to low unsaturation (<10) and a low error range: -20 < 0 < 20 ppm.

Possible formulas for 134624	unsaturation
$C_{20}H_{41}NO_{12}$	1
$C_{15}H_{31}N_{14}O_5$	7.5
$C_{16}H_{37}N_7O_{10}$	2
$C_{14}H_{35}N_{10}O_9$	2.5
$C_{13}H_{29}N_{17}O_4$	8

Table 2.5 Possible Formulas for Node 134624.

At current it is not possible to use in silico prediction tools to generate realistic Molecular structures from Molecular formulas. In order to confirm the structure of these targets NMR of the pure compounds must be obtained.

The same process was then completed for all nodes in Target 1.1. In Table 2.3 the major fragments are displayed for Node 131733 and Node 131736 which are impossible to distinguish in the raw spectra.

Node	Parent m/z	Major Fragments m/z	Δm/z	Predicted Fragment Formula
131733/6	474.3909	456.38	18.20	H_20
		236.16	238.23	$C_{14}H_{28}N_3$ $C_{16}H_{30}O$

Table 2.6 Fragments for Node 131733/6.

Possible Molecular formulas for 134624	Unsaturation
$C_{20}H_{47}N_{11}O_2$	3
$C_{18}H_{45}N_{14}O$	3.5
$C_{24}H_{51}N_5O_4$	2

Table 2.7 Possible formulas for Node 131733/6.

Node	Parent Mass	Major Fragments m/z	Δm/z	Predicted Fragment Formula
135525	490.386	334.2156	156.17	$C_{10}H_{22}N$ $C_9H_{20}N_2$
		252.1529	238.23	$C_{14}H_{28}N_3$ $C_{16}H_{30}O$

Table 2.8 Fragments for Node 135525.

135525	Unsaturation
$C_{20}H_{47}N_{11}O_3$	3

Table 2.9 Possible formula for Node 135525.

The potential Molecular formula shown here for Target 1.1 are not convincing; none differ by an oxygen atom as would be expected from the networking data. It could also be necessary to include more elements into the predicted formulae, such as sulphur and phosphorus. Once purified, predicted structures can be rejected based on NMR and new structures generated as required. Three more targets were selected from the Molecular network for isolation, the overview of which are shown below, additional information can be found within the supplementary information.

Node	m/z	RT /s	Species	Media	Extracts
1.2	597.473	504	*Euglena gracilis, Colacium vesiculosum*	Af6	BuOH
1.3	578.322	405	*Astasia ocellata var provasolii*	JMSE	BuOH
1.3	596.334	361	Astasia ocellata var provasolii, Astasia longa	JMSE	BuOH
1.4	381.241	613	*Astasia longa, Khawkinea quartana, Colacium vesiculosum, Distigma proteus, Euglena geniculata, Euglena gracilis, Euglena clara*	Af6	BuOH
1.5	635.422	274	*Distigma proteus Euglena clara*	Af6	MeOH

Table 2.10 Overview of targets 1.2-1.5

Once these targets had been selected, it was necessary to culture the conditions required for their expression in a higher volume in order to obtain enough material for isolation. The spectra in which all the target compounds were found were compared, and the conditions selected based on which produced the compound in the highest intensity.

2.10 Cytoscape Network Two – 1 L Cultures

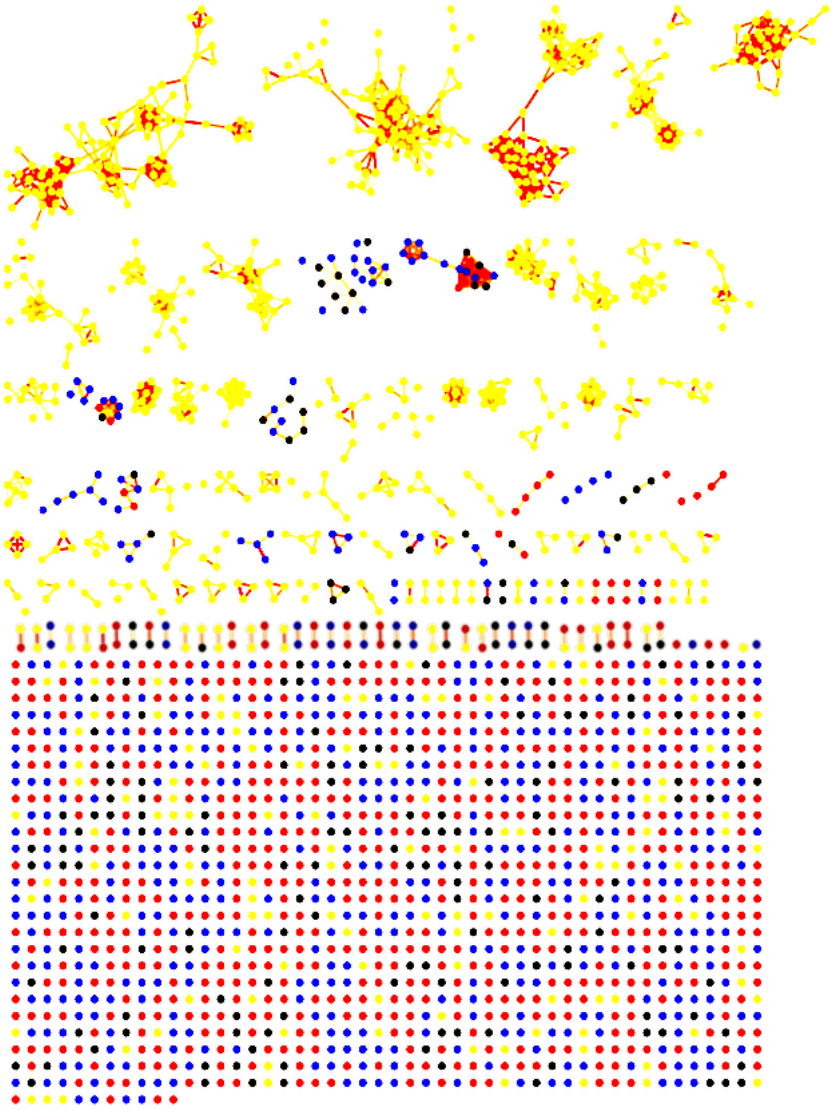


Figure 2.9 Cytoscape Network Two, generated through GNPS, and displayed in Cytoscape. The addition of the replicate extractions from the 1 L batches is added to all the data from Cytoscape Network One. Red nodes: only present in original 32- Algal 10 mL cultures; Blue nodes: only present in new 1 L extractions, Black nodes: present in both 10 mL and 1 L cultures; Yellow: either a negative-control node or a cluster containing a negative control node.

When the 1L Cultures had finished growing after 6 weeks, 3 x 5 mL samples were taken from each 1 L culture and sequentially extracted into ethyl acetate, butanol

and methanol as previously. They were dried, reconstituted in methanol and analysed through LCMS-MS.

This generated a new Molecular network of 2124 nodes and 2914 edges. 482 of these nodes were present in the negative control. There are 137 clusters, 62% of which are negative control. There are 1259 unconnected nodes, 10% of which are in the negative control. Of the non-negative control nodes, 553 are exclusive to new cultures, 280 are found in both the new and old network, 696 are only exclusive to the original 10 mL Cultures and 113 are from the methanol control. It appears that Culturing a higher volume is sufficient an environmental change to induce metabolic changes in the algae. As a result of this, only Target 1.1 from *Phacus pusillus* was found in the 1 L cultures.

Therefore, new targets were to be selected. These were chosen based on the same criteria as previous, with the exception that they must be present in the 1 L Cultures which were grown to produce the previous targets, these being: *Astasia ocellata* var *provasolii* (CCAP1204/9)*, Distigma proteus* (CCAP1216/3A)*, Colacium vesiculosum* (CCAP1211/3)*, Euglena geniculata* (CCAP1224/4E) and *Khawkinea quartana* (CCAP1204/20A). Additionally, they were to be present in at least two of the triplicate repeat extractions from these 1 L cultures.

2.11 Cytoscape Network Two - Target 2.1

Target 2.1 is a single node with a m/z of 352.238 and a RT of 470 s which was found both in the original 10 mL culture of *Astasia ocellata var provasolii* grown in JMSE , and the first, second and third repeat samples of the 1 L culture of *Astasia ocellata var provasolii* grown in JMSE, when extracted into ethyl acetate. In order to predict the fragments it was necessary to find the spectra in which the MS2 was cleanest, this was in repeat three of the 1 L culture, shown in Figure 2.1.

Target	m/z	RT /s	Species	Media	Extraction
2.1	352.238	470	CCAP1204.9	JMSE	EtOAc

Table 2.11 Molecular weight, retention time, producing species, media conditions and extracting solvent for Target 2.1.

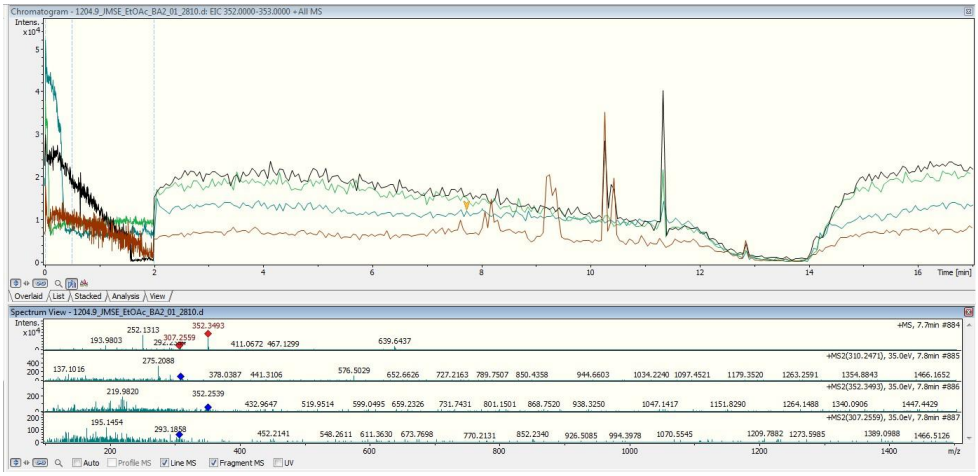


Figure 2.10 EICs of HPLCMS-MS spectra for Target 2.1. Blue: original 10 mL culture; Extractions of the 1 L culture shown in: Green: Repeat 1; Black: Repeat 2; Brown: Repeat 3.

Node	Parent m/z	Major Fragments m/z	Δm/z	Predicted fragment formula
77267	352.238	236.1013	114.1176	Asparagine
		151.1511	201.0869	Unknown

Table 2.12 Fragments for Node 77267

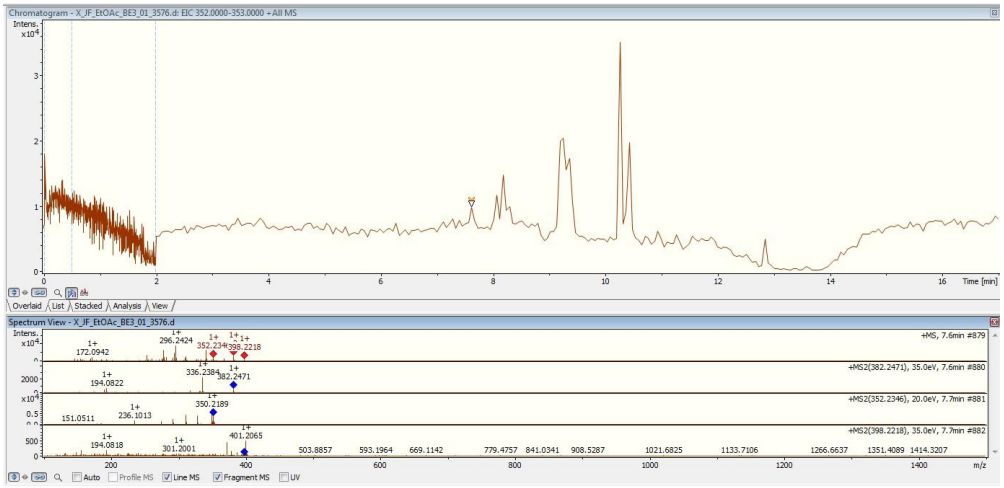


Figure 2.11 EIC of HPLCMS-MS spectra for Target 2.1, Repeat 3 extraction from the 1 L culture.

Node	Parent m/z	Major Fragments m/z	Δm/z	Predicted fragment formula
77267	352.238	236.1013	114.1176	Asparagine $C_6H_{14}N_2$ $C_7H_{16}N$
		151.1511	201.0869	$C_8H_{13}N_2O_4$ $C_6H_{11}N_5O_3$ $C_9H_9N_6$ $C_4H_9N_8O_2$

Table 2.13 Fragments for Target 2.1

135525	Unsaturation
$C_{16}H_{29}N_7O_2$	6
$C_{18}H_{32}N_4O_3$	6
$C_{15}H_{33}N_3O_6$	1
$C_{14}H_{27}N_{10}O$	6.5
$C_{18}H_{31}N_4O_3$	5.5
$C_{13}H_{31}N_6O_5$	1.5
$C_{12}H_{25}N_{13}$	7

Table 2.14 Predicted Molecular formula for Target 2.1

2.12 Cytoscape Network Two: Target 2.2.1 and 2.2.2

Two targets were selected from the 1 L CCAP1216.3C culture, the ~3-minute RT difference between them thought sufficient that they could be separated from each other. The first, 154677, has a m/z of 475.322, a RT of 297s,was observed in equal amounts in both original Cultures and 1 L cultures.

The second, 187613, has a m/z of 645.557 and a RT of 462 s. This target showed up in two out of the three repeats. It is found in all three media and in the original 10 mL culture of CCAP1204.3.

Target	m/z	RT /s	Species	Media
2.2.1	475.322	297	CCAP1216.3C	Af6
2.2.2	645.557	462	CCAP1216.3C	Af6

Table 2.15 Molecular weight, retention time, producing species, media conditions and extracting solvent for Targets 2.21 and 2.22.

Neither 2.2.1 or 2.2.2 have fragmentation patterns that convey a lot of information: they are clear of noise but haven't split into informative peaks in the MS2, which does not allow for prediction of the fragments.

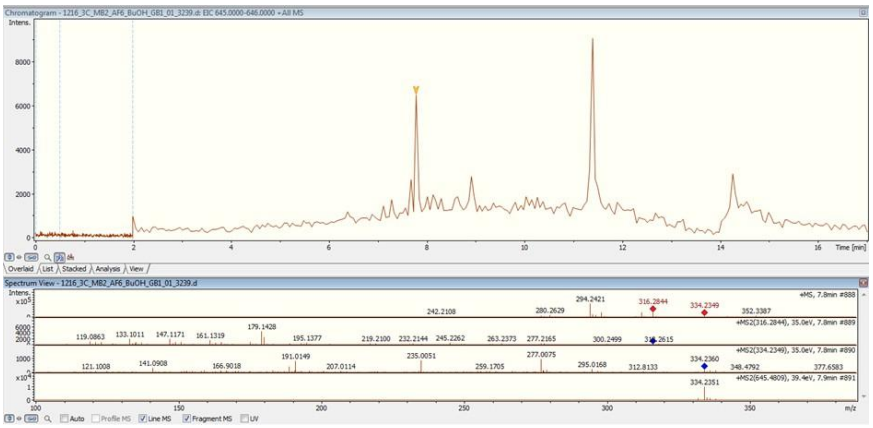


Figure 2.12 EIC of HPLCMS-MS spectra for Target 2.2.1, Repeat 2 extraction from the 1 L culture

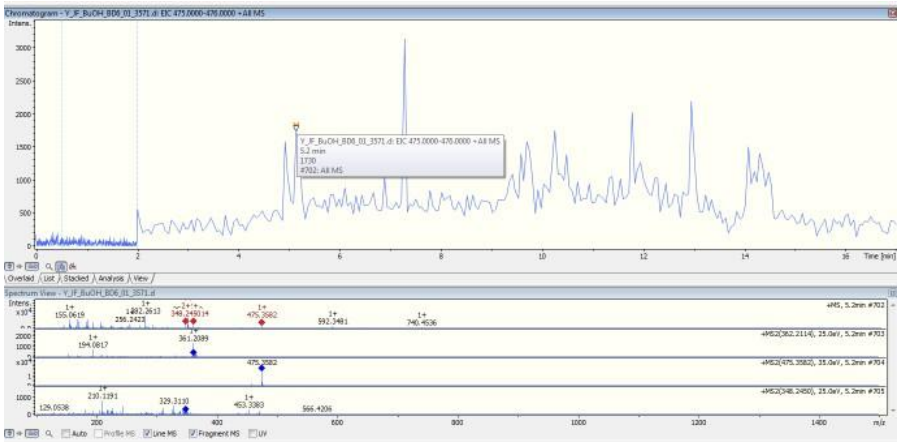


Figure 2.13 EIC of HPLCMS-MS spectra for Target 2.2.2, Repeat 3 extraction from the 1 L culture.

Target 2.2.1	Unsaturation
$C_{12}H_{34}N_{20}O$	6
$C_{13}H_{40}N_{13}O_6$	0.5
$C_{11}H_{38}N_{16}O_5$	1
$C_{26}H_{42}N_4O_4$	8
$C_{14}H_{36}N_{17}O_2$	5.5

.

Table 2.16 Predicted Molecular formula for Target 2.2.1

Target 2.2.2	Unsaturation
$C_{34}H_{72}N_6O_5$	2

Table 2.17 **Predicted Molecular formula for Target 2.2.2**

2.14 Cytoscape Network Two: Target 2.4

Target 2.4 has a m/z of 458.357 and a RT of 470s, it is exclusively found in the 1 L culture of 1224.4E grown in Af6. It shows up in 2 out of the three repeats, extracted into butanol for both, as well as into methanol in the third repeat.

Node	m/z	RT /s	Species	Media
149607	458.357	470	CCAP1224.4E	Af6

Table 2.18 Molecular weight, retention time, producing species, media conditions and extracting solvent for Targets 2.4.

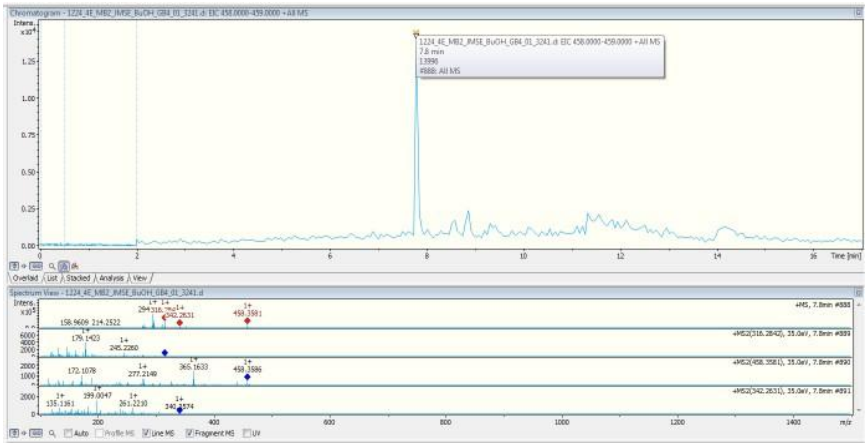


Figure 2.15 EIC of HPLCMS-MS spectra for Target 2.4, Repeat extraction 2 from the 1 L culture.

Target 2.4	Parent Mass	Major Fragments m/z	Δm/z	Predicted fragment formula
Target 2.4	458.357	365.163	93.194	C_6H_7N $CH_7N_3O_2$
		277.215	181.142	$C_2H_{15}N_9O$ $C_9H_{17}N_4$
		172.108	286.249	$C_{15}H_{32}N_3O_2$ $C_{13}H_{30}N_6O$

Table 2.19 Fragments for Target 2.4

Target 2.4	Unsaturation
$C_{20}H_{44}N_9O_3$	3.5

Table 2.20 Molecular formula for Target 2.4

2.15 Target Compounds – Isolation through HPLC

The litre Cultures were evaporated down *in vacuo* to <25% of their original volume. They were then washed twice with an equal volume of the extracting solvent, the organic layer was then washed with DI H_2O and the organic layers recombined and removed *in vacuo* to leave a white residue.

An initial HPLC purification was run as described in the Materials and Methods Section. After this initial purification, the initial HPLC-MS/MS analysis was run once more to identify target ion peaks, which were observed for 2.1, 2.3, 2.5 1.1, The chromatograms remained heavily contaminated with non-target ions, particularly in the 10 to 15 minute region. The large amount of organic material present even in polar fractions with low RTs is possibly a consequence of there being too much organic material for the column to fully adsorb. It was clear further purification would be required, and so the target containing fractions were combined, dried, and redissolved in the minimal amount of MeOH for a second purification through preparative HPLC.

After the second purification, target ion peaks were observed for 2.1, 2.3, and 2.5. They were observed at increasingly lower intensity, and the spectra remained contaminated with peaks in the 10-15 minute region that did not appear to be removed through HPLC. The spectra shows many non-target peaks present in the chromatogram, and the relative abundance of the target peaks appeared so small it was likely that as for the other targets in the second HPLC purification, a third attempt at purification would result in loss of targets. However, NMRs of these fractions were run in an attempt to observe any peaks that could reject hypothetical structures, and disc diffusion assays of the fractions were performed to assess if these fractions possessed antimicrobial activity.

2.16 NMR

The fractions containing 2.1, 2.3, 2.5, were dried, the 2.1, 2.3, 2.5 containing fractions were dissolved in 100% DI H_2O. ^1H NMR spectra were obtained using the 800 MHz NMR. The resulting spectra showed very weak peaks with intensity barely above the noise, predominantly in the sp3 region – likely corresponding to the

compounds which have the most intense peaks in the HPLC chromatograms: those in the 10 – 15 minute range.

As a result, they provided no structural information and it is clear that further purification is needed before NMR analysis can take place.

2.17 Disc Diffusion Assays

Disc diffusion assays to assess the extracts activity against *E. coli, K. rhizophila* and *S. cerevisiae* were carried out. No bioactivity observed against *E. coli*, *K. rhizophila* or *S. cerevisiae* for any of the below extracts, but the positive control showed zones of inhibition across all plates.

Extracts Tested:

Target 1.1 containing fraction 30 of the first purification of CCAP1261.6.

Target 2.1 containing fractions after two HPLC purification steps of CCAP1204.9, pooled, dried and redissolved in DI H_2O (5 ul).

Target 2.3 all fractions from first purification step of CCAP1216.3C, pooled, dried and redissolved in DI H_2O (5 ul).

Target 2.4 Fractions 1-29 of first purification step of CCAP1211.3 pooled, dried and redissolved in DI H_2O (5 ul). Target 2.4 containing fraction 30 of previous purification step after purification through second HPLC pooled, dried and redissolved in DI H_2O (5 ul).

Target 2.5 containing fraction 30 of the first purification of CCAP1204.20A.

The disc diffusion assays do not suggest that it is likely that the above extracts display antimicrobial activity against the tested strains at this concentration. However, they do not completely disprove the possibility, if repeated using a higher concentration of material it is possible that bioactivity could be displayed. Additionally, disc diffusion assays are not the only method of testing for antimicrobial activity and these compounds could have bioactivity against other microorganisms or against therapeutically relevant human targets.

2.18 Chapter Two Conclusions

The comprehensive High-Performance Liquid Chromatography-Tandem Mass Spectrometry (HPLC-MS/MS) analyses conducted on the myriad natural products derived from 32 Euglenoid species cultivated in three distinct media have unequivocally illuminated the vast presence of hitherto undiscovered compounds within *Euglena*. The dynamic variations in both the quantity and types of compounds observed across diverse growth media underscore the profound impact of environmental changes on *Euglena's* biochemical repertoire. Such alterations in expression patterns unveil a rich source of varied chemistry, thereby offering a nuanced perspective for natural product chemists exploring the untapped potential of this remarkable group of organisms.

Moreover, the integration of Molecular networking emerges as a powerful tool, enhancing the efficiency and facilitating the streamlined selection of compounds from the intricate metabolic profile of *Euglena* for subsequent purification. The natural product library on the Global Natural Products Social Molecular Networking (GNPS) platform not only sheds light on the lack in collective knowledge regarding *Euglena's* natural product chemistry but also serves as a valuable resource for dereplication. By pinpointing known compounds that prevent redundancy in discovery efforts, GNPS aids natural product chemists in navigating the expansive landscape of *Euglena*'s chemical diversity.

While this chapter did not culminate in the isolation of specific compounds, it laid the foundation for envisioning potential Molecular structures. As artificial intelligence (AI) and computational algorithms continue to advance in both reliability and speed, the future holds promise for generating predicted structures from proposed formula. Such studies could assess the novelty of hypothetical structures to indicate the value of targets for the extensive purification required for extraction from complex mixtures.

Furthermore, this chapter not only supports the trajectory of future work within the thesis but also underscores the viability of this approach as a route for the discovery of novel compounds. However, it is imperative to acknowledge that the realization of this potential hinges upon the scalability of culture volumes. Addressing this limitation becomes a focal point in the subsequent chapter, paving the way for a more extensive exploration of *Euglena's* chemical landscape.

Chapter 3: Does Coculturing *Euglena gracilis* with other organisms increase the diversity of natural products produced compared to monocultures?

3.1 Introduction

The difficulty of stimulating microorganisms to release their full potential of metabolites under laboratory conditions has been well-documented. Despite genetic sequencing indicating a multitude of unknown BGCs in microorganisms, these are not always expressed under laboratory conditions.[181] If sufficient genetic tools exist these genes can be activated through increasing gene expression, however this requires exact knowledge of the gene of interest to be targeted. In the case of *Euglena* in which case a lack NP producing target genes have been identified alternative methods can be employed to activate gene expression.

To this end the one-strain-many-compounds or "OSMAC" approach can be employed; adjustments to the culture conditions; pH, medium salinity, media composition, addition of different salts, different temperatures or agitation conditions can be employed to activate the production of different metabolites.[54] Furthermore, microorganisms can be treated with epigenetic modifiers such as histone deacetylase inhibitors or DNA methyl transferase inhibitors to modulate histones and promote epigenetic changes for the activation of silent BGCs.[182] Culture changes which have a negative effect on the microorganism, such as nutrient depletion or high saline levels, are referred to as "stressing" the microorganism.[183] The previous chapter touched on the different compounds produced through simple media changes. This chapter will expand on this through investigation of coculturing , otherwise known as mixed fermentation, in which two or more microorganisms are cultured together. This has been used successfully for the production of high value compound in *Euglena* such as fatty acids and antioxidants, but has not been explored for the purpose of undiscovered natural products.[184]

3.2 Coculturing bacteria and algae to initiate release of novel natural products

Coculturing *in vitro* attempts to mimic the ecological scenarios in which complex micro-communities continually interact. Competition for limited resources and space create both symbiotic and antagonist relationships between microbes which are mediated through the production of secondary metabolites. Purposeful inoculation of multiple species into one culture can simulate these natural relationships so that the compounds produced both from interspecies symbiosis and competition may be isolated. Deliberately stressing microbes under these

conditions has been shown to activate silent biosynthetic genes of fungi, bacteria and algae.[185]

Several co-cultivation studies provide evidence that this is a viable experimental approach for enhancing the chemical diversity of compounds produced by microorganisms grown *in vitro*. As well as this, it has been shown to reduce the chance of environmental contamination of the culture, and in some cases enhance microalgal growth. For example, coculturing microalgal species *Tetraselmis chuii*, *Cylindrotheca fusiformis* and *Nannochloropsis gaditana* with marine bacteria strains of the *Muricauda* genus for 33 days showed that the final cell density of *T. chuii* and *C. fusiformis* was higher than that of controls (21.37–31.18 and 65.42–83.47 %, respectively).[186]

In aquatic environments, microalgae and bacteria are not only integral to nutrient recycling and energy flow through the ecosystem but also have very direct influence on each other, positively as well as negatively. Bacteria can promote microalgal growth through reducing dissolved oxygen concentration and consuming organic materials excreted by algae. They additionally secrete crucial compounds for the growth of algae: biotin, cobalamine and thiamine.[187] In order to aid this symbiotic relation in culture, the addition of necessary compounds continually is required to reduce competition between the species.[188] However, for the purposes of deriving antibacterial agents from cocultures, it is the antagonistic relationship that is mostly likely to divulge compounds with antimicrobial activity. As a result, the decision was made to not introduce further nutrients to the growing cultures in this chapter to promote increasing competition for diminishing nutrients between the two species.

Examples in which coculturing has lead to the production of valuable natural products include the coculturing of two mangrove fungi together which lead to the production of not just one known antibiotic, neoaspergillic acid, but also a new xanthone derivative, 8-hydroxy-3-methyl-9-oxo-9H-xanthene-1-carboxylic acid methyl ether or "Aspergicin" which displayed inhibitory activity against five microorganisms, including the fungal pathogens of plants *Gloeasporium musae* and *Peronophthora cichoralearum*.[189]

Secondly, coculturing proteobacteria *Thalassospira* CNJ-328, with a marine deuteromycete *Pestalotia* CNL-365 isolated from the surface of a brown alga *Rosenvingea* sp resulted in the synthesis of a new benzophenone antibiotic named

"Pestalone". Pestalone was not observed in the monocultures of either species, and shows potent antibiotic activity against methicillin-resistant *Staphylococcus aureus* (MIC = 37 ng/mL) and vancomycin-resistant *Enterococcus faecium* (MIC = 78 ng/mL). [190]

Two novel compounds emericellamides A and B were also produced in the co-culture of marine-derived fungus *Emericella* and marine actinomycete *Salinispora arenicola.* Emericellamides A and B show antibacterial activities against methicillin-resistant *Staphylococcus aureus* with MIC values of 3.8 and 6.0 µM, respectively.[191]

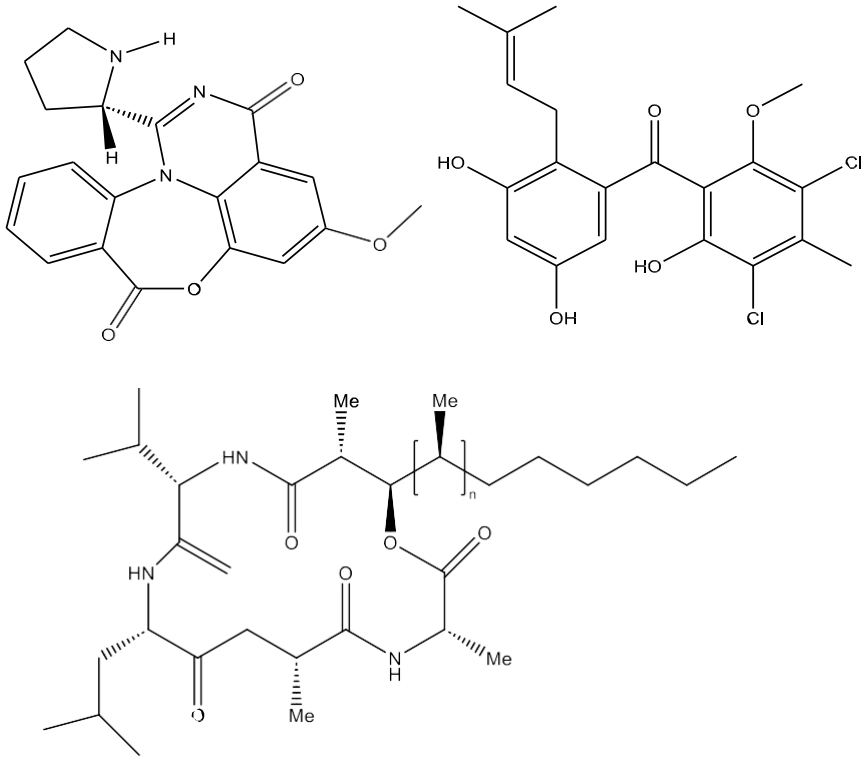


Figure 3.1 Structures of Aspergicin (top left), Pestalone (top right) and emericellamides A (n=1) and B (n=2)

In order to investigate whether coculturing *Euglena* with bacterial species increases novel natural product expression compared to monocultures, several cultures were grown with *Euglena*, and *Streptomyces* as well as *Chlorella vulgaris* and cyanobacteria *Synechococcus* sp.7002 for comparison. *Streptomyces* was selected as the coculturing bacteria due to its history in the production of antibiotics. This chapter details the results from the initial experiments and the subsequent attempts at isolation of novel compound grown from coculture of *Euglena gracilis* and a *Streptomyces*.

3.2 Coculturing algae and *Streptomyces*: the initial experiment

In order to isolate compounds produced from co-cultures of a competing microorganism with microalgae, co-culture experiments were carried out between algal and streptomyces strains. Five strains of *Streptomyces:, S. coelicolor A3, S. glaucescens, S. limosus, S. griseus, and S. netropsis* were cocultured axenically and with three algal species: two freshwater algae *Euglena gracilis* and *Chlorella vulgaris,* and marine cyanobacteria *Synechococcus* sp.7002. To further increase the diversity of the metabolite profiles they were also cultured in three different media: EG-JM, LB and AF6.

After *Streptomyces* strains were cultured on solid oatmeal agar for 1 week from frozen colony stores, they were then inoculated into 10 mL of liquid media along with either coculture of algae or as a monoculture. These were grown for 4 weeks and then extracted through a liquid-liquid extraction with increasingly polar solvents: ethyl acetate, butanol and methanol. The separate extracts were then analysed through HPLCMS-MS. Initial growth, extraction and preparation for HPLC-MS/MS was done by PhD student Beth Brown, the following analyses are my own

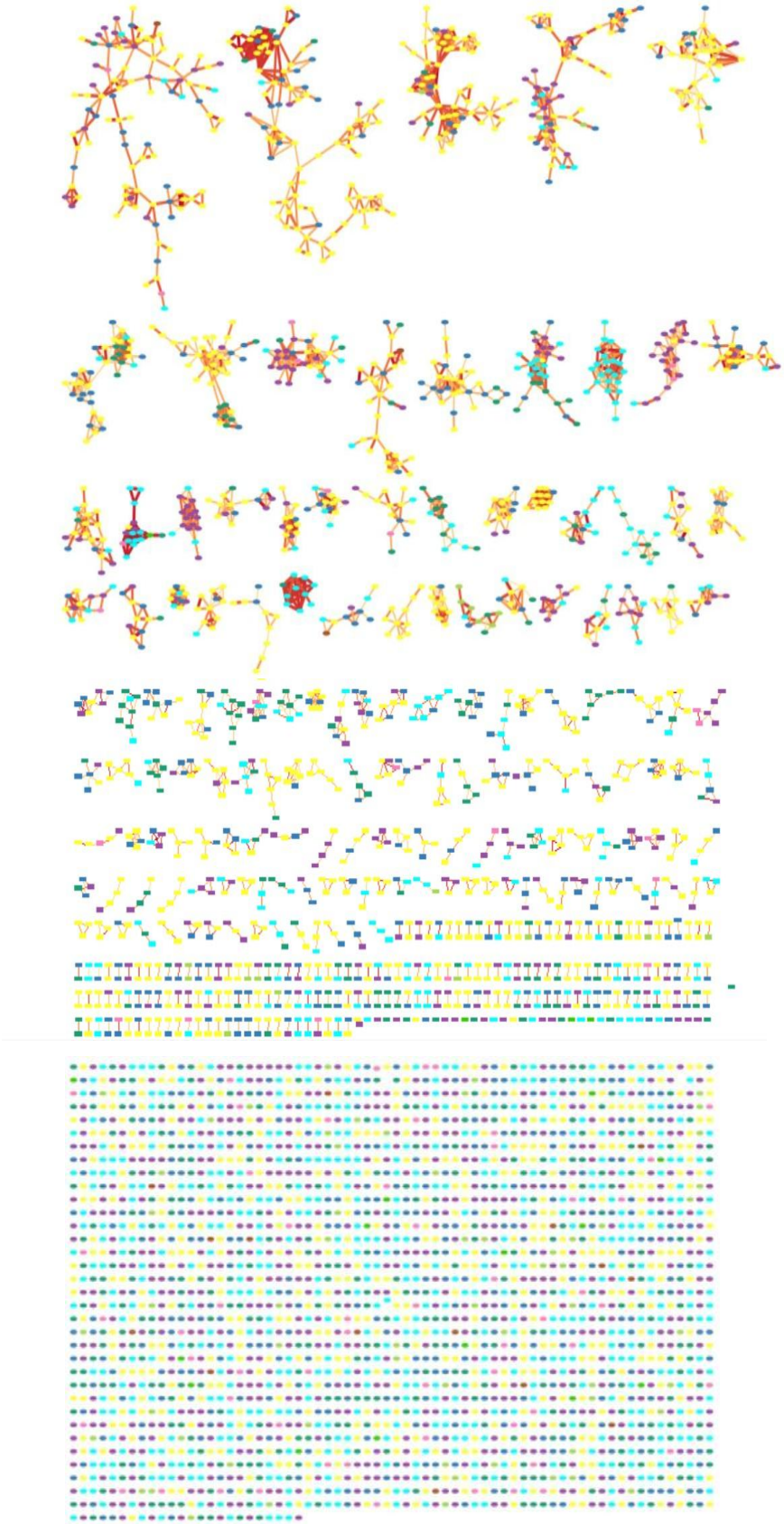


Figure 3.2 Molecular network of coculturing experiment. Yellow: control, cyan: coculture, purple: bacteria and coculture, pink:

The same process of HPLCMS/MS analysis was applied here as described in Chapter Two, the raw files were converted and then uploaded to GNPS and submitted to Molecular networking. The network was then analysed in Cytoscape, it has 5124 nodes and 4143 edges, and will henceforth be referred to as Cytoscape Network 2. Within these culture constraints and experimental parameters, and removing the negative control nodes, 13 more nodes were present in algal monocultures than bacterial monocultures. 41 nodes are present in both bacterial and algal monocultures but not cocultures, perhaps indicating that the introduction of the coculturing species negated the need for the expression of some metabolites. 769 nodes were unique to cocultures alone, broadly highlighting that coculturing increases natural product production. 604 nodes were present across all conditions, which can therefore be deemed unsuitable for study in this chapter as they are not produced as a result of coculturing interactions.

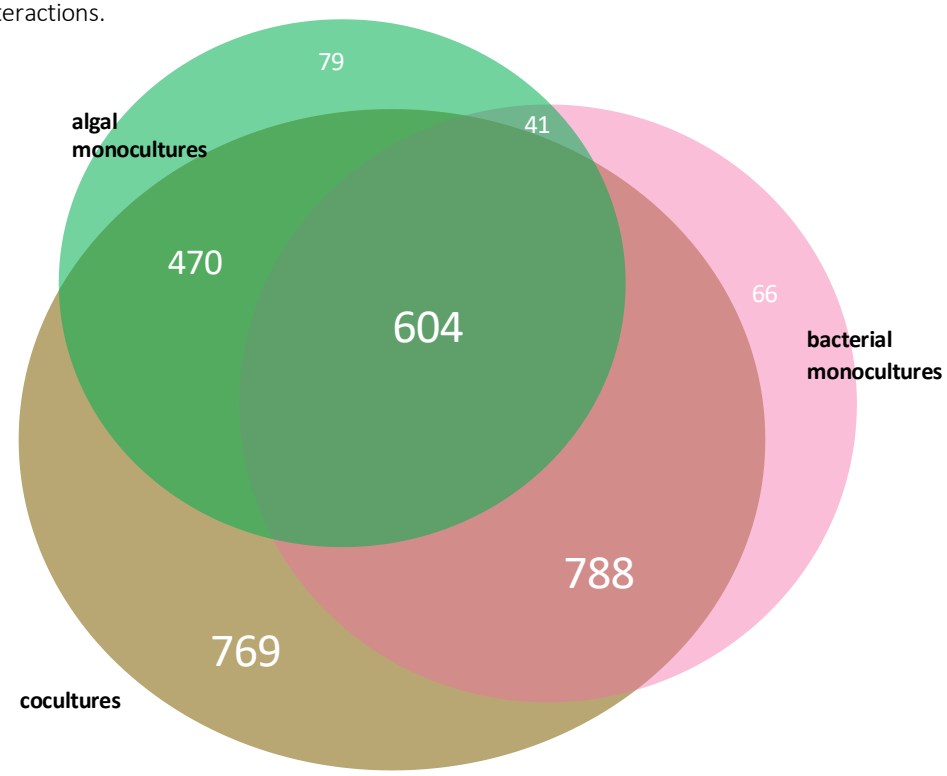


Figure 3.3 Total nodes: algal monocultures (79) bacterial monocultures (66) bacterial and algal monocultures but no cocultures (41), cocultures and no monocultures (769), in cocultures and both bacterial and algal monocultures (604), in cocultures and bacterial monocultures but not algal monocultures (788), in cocultures and algal monocultures but not bacterial monocultures (470).

Taking together the exclusively algal monocultures and algal cocultures there are a total of 549 nodes, repeating the same for bacteria gives 854 nodes. At first one might expect to see an elevated number of nodes observed in bacteria overall as a result of there being two more bacterial monocultures than algal. Despite this, there are 13 more unique compounds to the algal monocultures than the bacteria. When taken as an average per culture, this makes just over 13 nodes per monoculture for bacteria and just under 24 nodes per monoculture for algae. With the caveat that only metabolites between the 300 – 1500 m/z range are detected, and only metabolites that are extracted from the three solvents ethyl acetate, butanol and methanol, an increased amount of compounds from the algal cultures are observed compared to the bacterial cultures. This can be attributed to increased biochemically complexity within algal cells compared to bacterial cells due to the presence of diverse genes from larger genomes, in the case of *Euglena* and *Chlorella*, and plastids which *Streptomyces* lack.

When it comes to nodes per individual organism, 5 were completely unique to *S. glaucescens* monocultures, 12 to *S. coelicolor*, 14 to *S. griseus*, 29 to *S. limosus* and 6 to *S. netropsis*. The most overlap could be expected to exist between *S. limosus*, *S. coelicolor* and *S. griseus* as they are in the same clade in the *Streptomyces* 16S rRNA gene tree, clade 112.[192] However, when excluding algal cultures, cocultures and other *Streptomyces* strains there were no nodes only expressed by these three strains. Only 1 node was produced by both *S. limosus* and *S. coelicolor* – an unidentified single node only expressed in Af6 media. 2 nodes were shared between *S. limosus* and *S. griseus* – one expressed in only Af6 media, and one expressed between both EG:JM and LB. There were more shared nodes between *S. coelicolor* and *S. griseus* who both shared 9 nodes, 7 of which were only produced in LB, 1 of which was produced in LB and EG:JM and one of which was produced only in Af6.

In general, *S. limosus* shared low node counts with all other *Streptomyces* strains analysed: 1 between *S. limosus* and *S. coelicolor,* 1 between *S. limosus* and *S. netropsis* and 3 nodes between *S. limosus* and *S. glaucescens. S. glaucescens* is part of a very small clade, clade 99, with only 1 other species, *S. pharetrae*.[192] It displayed slightly elevated shared node counts to other species compared to *S. limosus,* but still low overall: 1 node with *S. netropsis*, 4 with *S. coelicolor,* 3 with *S. griseus* and 3 with *S. limosus. S. netropsis* is in clade 52, with *S. distallicus*, *S. flavopersicus, S. kentuckensis* and *S. syringium*. [192] It shares two nodes with *S. coelicolor;* 1 produced in both Af6 and EG:JM and the other produced in both Af6 and LB. *S. netropsis* shares 1 node with *S. glaucescens* and *S. griseus.* No nodes

were produced in any monocultures of *Streptomyces* exclusively, 32 nodes were produced both in cocultures and in monocultures of *Streptomyces*. This means that there were no compounds ubiquitously produced by all *Streptomyces* monocultures which were no longer produced when in coculture with algal strains.

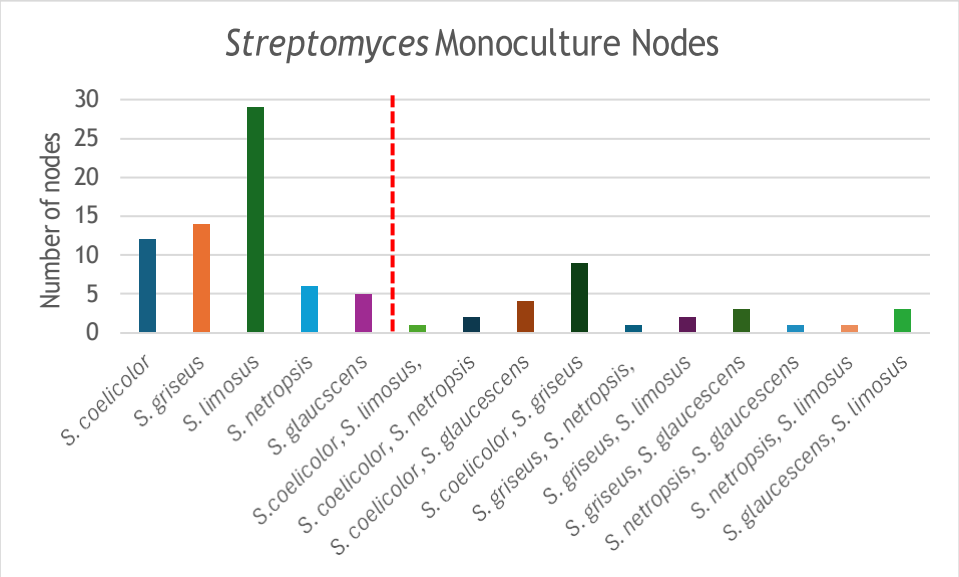


Figure 3.4 Number of nodes for *Streptomyces* by species

The data suggests that each *Streptomyces* strain has a largely distinct metabolic profile, with minimal overlap in metabolite production. No compounds (nodes) were ubiquitously produced across all monocultures, indicating that each strain synthesizes unique secondary metabolites. Interestingly, phylogenetic relatedness does not appear to strongly correlate with shared metabolite production. Despite *S. limosus*, *S. coelicolor*, and *S. griseus* belonging to the same 16S rRNA clade, they did not share any exclusive compounds. However, *S. coelicolor* and *S. griseus* displayed the highest number of shared nodes (9), suggesting that while genetic similarity may influence metabolic output, it is not the sole determinant.

Among the strains analysed, *S. limosus* demonstrated the most distinct metabolic profile, producing the highest number of unique compounds (29) and exhibiting the lowest level of shared nodes with other strains. This could indicate a specialized biosynthetic capability or regulatory differences that limit metabolite overlap. Additionally, shared compounds were highly dependent on the culture medium. For example, the majority of shared nodes between *S. coelicolor* and *S. griseus* were found in LB medium, emphasizing the role of environmental factors in secondary metabolite expression.

Despite the observed metabolic differences between monocultures, coculturing did not suppress the production of all compounds found in monocultures. The presence of nodes in both cocultures and monocultures suggests that Coculturing can expand metabolic diversity without necessarily inhibiting strain-specific metabolites. These findings highlight the importance of optimizing culture conditions to enhance metabolite production and suggest that further exploration of different media and coculture strategies could lead to the discovery of novel bioactive compounds. Additionally, the unique metabolic output of *S. limosus* warrants further investigation to determine whether its distinct profile results from regulatory mechanisms or strain-specific biosynthetic gene clusters.

Euglena gracilis, *Chlorella vulgaris*, and *Synechococcus sp. 7002* each produce a distinct set of metabolites, with *S. 7002* generating the highest number of unique compounds (37), followed by *E. gracilis* (21) and *C. vulgaris* (13). This contradicts the expectation that eukaryotic algae would produce more unique metabolites than prokaryotic cyanobacteria, however this could be due to *S. 7002*'s significant genetic and biochemical divergence from both the other algae and *Streptomyces*. Unlike *E. gracilis* and *C. vulgaris*, which share more similarities as eukaryotic algae both possessing a green plastid, *S. 7002* is a photosynthetic prokaryote, making it evolutionarily distinct from both the algae and the heterotrophic *Streptomyces* strains. Additionally, genome size does not correlate with metabolite uniqueness in this dataset, as *E. gracilis* possesses the largest genome (2.25 Gb) but does not produce the most unique compounds, while *S. 7002*, with the smallest genome (3.4 Mb), exhibits the highest number of unique nodes.

When considering shared metabolites, no compounds were found exclusively within algal monocultures that were absent from *Streptomyces* monocultures or cocultures, reinforcing the idea that algal metabolic outputs are not diminished by *Streptomyces* coculturing , only expanded. *E. gracilis* and *C. vulgaris* share the most compounds (7), as expected given their evolutionary proximity, while *C. vulgaris* and *S. 7002* share only 2, as do *E. gracilis* and *S. 7002*. However, a deeper analysis reveals that when cocultures are included, the number of shared metabolites between *E. gracilis* and *C. vulgaris* increases dramatically to 195, and further to 540 when *Streptomyces* monocultures are also considered. Similarly, the number of shared nodes between *C. vulgaris* and *S. 7002* rises from 2 to 74 in cocultures and further to 196 when bacterial cultures are included. The same pattern is observed for *E. gracilis* and *S. 7002*, where shared nodes increase from 2 to 115 in cocultures and 180 with bacterial cultures. This could indicate that the relationship

between the algal species and the *Streptomyces* is similar independent of eukaryotic or prokaryotic considerations.

In summary, whilst 79 nodes are produced in algal monocultures, 751 nodes appear exclusively in cocultures, demonstrating that microbial interactions from these *Streptomyces* strains significantly enhance natural product biosynthesis compared to monocultures.

3.4 Cytoscape Network 2: Phosphocholines

Of the 37 compounds unique to cyanobacteria s.7002, two had library identifications these being Vitamin B12 and a ceramide. All 37 compounds were displayed as single nodes, with one exception: the ceramide. The ceramide was contained within a cluster of related compounds all of which had library identifications, and which mostly consisted of phosphocholines. The other compounds within this cluster are found within both algal and bacterial Cultures of*: E. gracilis, C. vulgaris*, *S. glaucescens, S. coelicolor* and S*. limosus.* However, only certain combinations of these organisms lead to the production of particular phosphocholines.

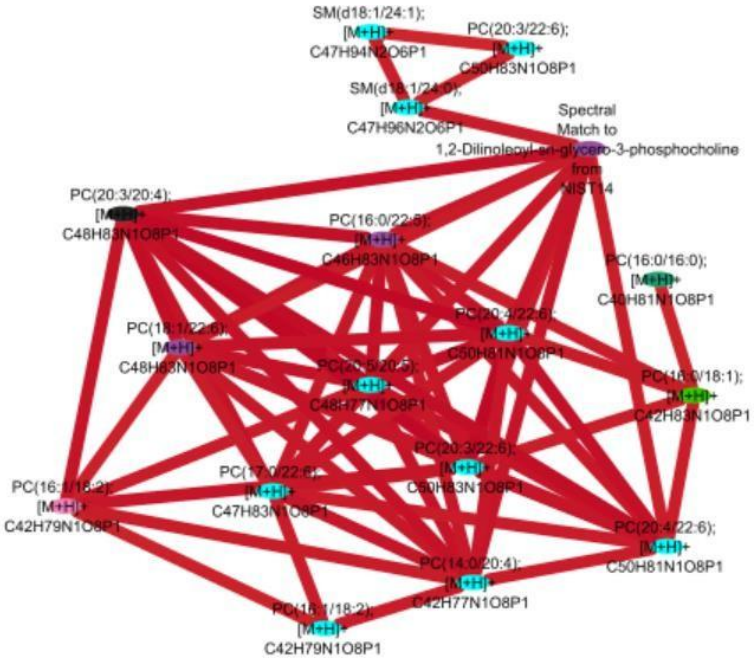


Figure 3.5 Phosphocholine cluster in Cytoscape Network 2

Library Identification	Culture	Species
SM(d18:1/24:0); [M+H]+ C47H96N2O6P1	Coculture	*C. vulgaris, S. coelicolor, cyanobacteria*
PC(14:0/20:4); [M+H]+ C42H77N1O8P1	Coculture	*C. vulgaris, S. coelicolor*
PC(20:5/20:5); [M+H]+ C48H77N1O8P1	Coculture	*C. vulgaris, S. coelicolor, E. gracilis, S. glaucescens*
PC(17:0/22:6); [M+H]+ C47H83N1O8P1	Coculture	*C. vulgaris, S. coelicolor*
PC(16:0/18:1); [M+H]+ C42H83N1O8P1	Coculture	*C. vulgaris, S. coelicolor*
PC(16:0/16:0); [M+H]+ C40H81N1O8P1	Coculture, Algae Monoculture	*Cyanobacteria, C. vulgaris, S. nepotis*
PC(16:0/22:5); [M+H]+ C46H83N1O8P1	Bacterial monolculture	*S. glaucescens, S. coelicolor*
PC(16:1/18:2); [M+H]+ C42H79N1O8P1	Bacterial monolculture	*S. glaucescens, S. coelicolor*
PC(16:1/18:2); [M+H]+ C42H79N1O8P1	Coculture	*C. vulgaris, S. coelicolor*
1,2-Dilinoleoyl-sn-glycero-3-phosphocholine	Coculture, Bacterial monolculture	*S. glaucescens, cyanobacteria, S. coelicolor, C. vulgaris, S. glaucescens*
PC(20:3/20:4); [M+H]+ C48H83N1O8P1	Coculture, Control	*C. vulgaris, S. coelicolor, C.vulgaris, S. glaucescens*
SM(d18:1/24:1); [M+H]+ C47H94N2O6P1	Coculture	*cyanobacteria*
PC(18:1/22:6); [M+H]+ C48H83N1O8P1	Coculture, Bacterial monolculture	*C. vulgaris, S. coelicolor, S. glaucescens, S. lumosum, E. gracilis, S. glaucescens, S. coelicolor*
PC(20:4/22:6); [M+H]+ C50H81N1O8P1	Coculture	*C. vulgaris, S, coelicolor*
PC(20:4/22:6); [M+H]+ C50H81N1O8P1	Coculture	*E. gracilis, S. coelicolor*
PC(20:3/22:6); [M+H]+ C50H83N1O8P1	Coculture	*C. vulgaris, S. coelicolor*
PC(20:3/22:6); [M+H]+ C50H83N1O8P1	Coculture	*C. vulgaris, S. coelicolor*

Table 3.1 Overview of the different phosphocholines, and their producing strains and environment.

Phosphocholine (PCs) are a crucial component of cell membranes involved in cell signalling,[193] an increased variety and number of PCs are produced in cocultures, which could be a direct response to the changes in cell signalling due to the presence of the coculturing organism. The most phosphocholines (11) were produced in cocultures of *Chlorella vulgaris* and *Streptomyces coelicolor*, with 8 of these appearing exclusively under coculture conditions. Cyanobacterial cultures produced four phosphocholines, but three of these only emerged in cocultures

with *S. coelicolor*, *S. netropsis*, or *S. glaucescens*, suggesting that bacterial interactions directly influence phosphocholine biosynthesis.

A similar pattern was observed in *Euglena gracilis*, which produced three phosphocholines (*PC(20:5/20:5), PC(18:1/22:6), PC(20:4/22:6)*), none of which overlapped with the phosphocholines identified in Cytoscape Network 1. This discrepancy may initially seem attributable to differences in media conditions; however, further analysis indicates that coculturing conditions, rather than media composition alone, dictate phosphocholine production. For example, *PC(20:4/22:6)* was produced in Af6 media but only in coculture between *E. gracilis* and *S. coelicolor*, while *PC(20:5/20:5)* was exclusive to cocultures of *C. vulgaris* with *S. coelicolor* or *E. gracilis* with *S. glaucescens* or *S. coelicolor*. Similarly, *PC(18:1/22:6)* was observed in monocultures of *S. glaucescens*, *S. coelicolor*, and *S. limosus*, as well as in cocultures of *S. glaucescens* with *E. gracilis* and *S. coelicolor* with *C. vulgaris*, but not in any cocultures involving *S. limosus*. This suggests that while *S. coelicolor* and *S. glaucescens* retain their ability to produce *PC(18:1/22:6)* in coculture, *S. limosus* may experience an inhibitory effect when grown with algal strains.

Interestingly, phosphocholines identified in Cytoscape Network 1 also appear in Network 2, but under different conditions. For instance, *PC(18:3/0:0)*, previously produced by *Khawkinea quartana* and *Colacium vesiculosum* in Af6 media with butanol extraction, is now detected in cocultures of *Synechococcus 7002* with *S. coelicolor, S. netropsis,* or *S. glaucescens*. Three phosphocholines were produced by *E.* gracilis in Cytoscape Network Two: PC(20:5/20:5), PC(18:1/22:6), PC(20:4/22:6). However none of the same phosphocholines reported to be produced in Cytoscape Network One are present in this cluster: PC(18:3/0:0); PC(18:2/0:0),PC(0:0/16:0), or PC(0:0/18:1).This suggests that not only do coCultures induce the production of phosphocholines absent in monocultures, they also inhibit the production of phosphocholines.

The variation in phosphocholine composition across different cocultures suggests that cell-to-cell signalling mechanisms change when *Euglena* is placed in a competitive microbial environment through a phosphocholine-mediated cell signalling response, but do not provide evidence to support that this is the only change in cell signalling pathways as a response to coculturing.

3.5 Cytoscape Network 2: A comparison between algal and bacterial unique nodes

The unique compounds identified in *Euglena gracilis*, *Chlorella vulgaris*, and cyanobacteria s.7002 offer important insights into their natural product diversity and potential applications. None of the 21 compounds unique to *E. gracilis* had library matches, suggesting they are novel metabolites. Additionally, all were single nodes, meaning they do not cluster with related compounds, which may indicate they are chemically distinct or produced in low abundance. *C. vulgaris* initially had 13 unique compounds, but two were discounted due to their connection to negative control nodes, reducing the confirmed number to 11.

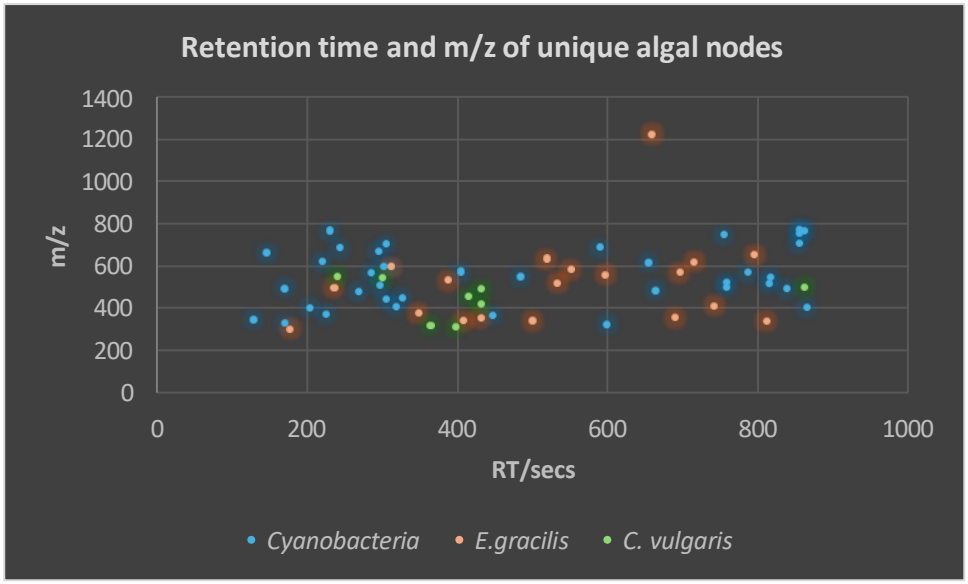


Figure 3.6 Mass to charge ratio of unique algal nodes and their corresponding retention times

A clear trend emerges when analysing the retention times of these unique compounds. While *E. gracilis* and cyanobacteria exhibited a broad range of retention times, most of the *C. vulgaris*-specific compounds eluted around 400 seconds. This suggests that the unique *C. vulgaris* compounds possess similar polarity and molecular properties whereas the unique *E. gracilis* and cyanobacteria compounds possess more inherit diversity.

The majority of unique compounds from both algae and *Streptomyces* species were below 800 m/z, indicating a general trend toward smaller molecules. *E. gracilis* produced one notably high-mass compound (1213 m/z), which could represent a complex macromolecule, such as a glycolipid or a high-molecular-weight polyketide. Additionally, *S. griseus* and *S. coelicolor* produced a few larger metabolites, suggesting that certain *Streptomyces* strains generate structurally complex secondary metabolites unique to their strain, potentially with antibiotic or signalling functions. All unique algal compounds eluted before 900 seconds, whereas *Streptomyces* compounds had retention times extending up to nearly 1000 seconds. This implies a slight trend of unique algal metabolites towards more polar and hydrophilic structures than unique *Streptomyces* compounds.

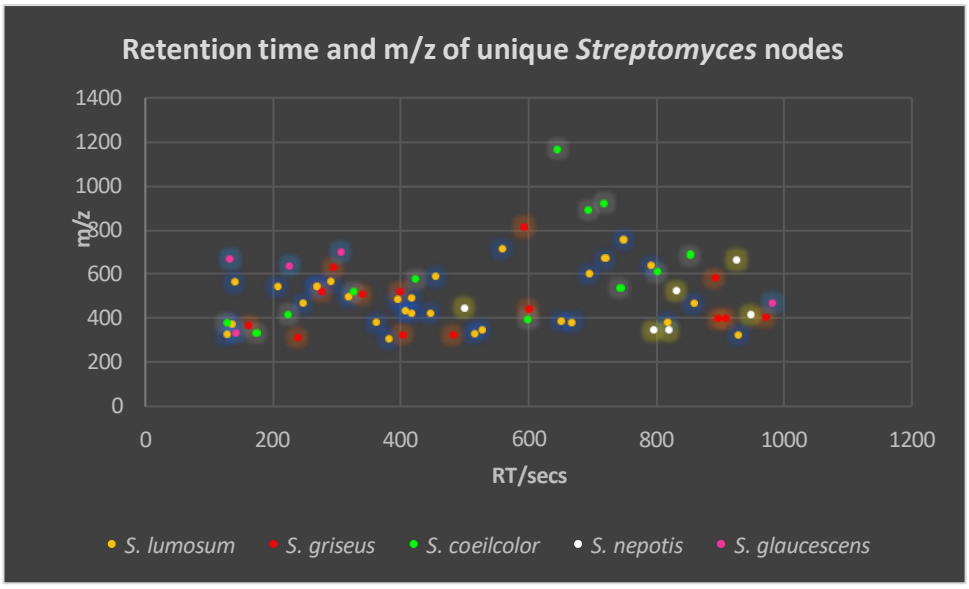


Figure 3.7 Mass to charge ratio of unique *Streptomyces* nodes and their corresponding retention times

These could be explained by considering the contextual natural environments of these organisms. The algae studies here are all aquatic microorganisms, therefore producing metabolites that are more polar and water-soluble aids in their dispersion and bioavailability in aqueous surroundings. In contrast, *Streptomyces* are soil-dwelling bacteria, with less need to introduce secondary metabolites in an aqueous environment.

3.6 Differences to Cytoscape Network 1

The comparison between Cytoscape Network 1 and Cytoscape Network 2 displays a significant increase in the number of detected compounds when algae are cocultured with *Streptomyces*. In Cytoscape Network 2, which consists of both axenic cultures and cocultures, a total of 5,124 nodes were detected, with 1,096

of these identified as media components or present in the methanol used for extraction and HPLC-MS/MS preparation. In contrast, Cytoscape Network 1, which analysed only Euglenoid monocultures, contained 2,124 nodes and 2,914 edges, with 482 nodes present in the negative control.

A key observation is that, despite performing 81 fewer extractions in Cytoscape Network 2 (216 vs. 297 extractions in Network 1), the total number of detected nodes increased by 981. When removing negative control nodes, Cytoscape Network 2 contained 4,028 non-control nodes, while Network 1 contained only 1,624—an increase of 2,404 additional non-control nodes in the coculture dataset. This provides strong evidence that the stress of coculturing algae and *Streptomyces* induces the expression of additional metabolic pathways, leading to the production of a greater number of secondary metabolites.

3.7 Cytoscape Network 2: Selecting Targets

Targets for isolation were then selected from this network for isolation. The same parameters were used as described in Chapter 2 with the following additional requirement that they only be found when *Euglena gracilis* was cocultured with *Streptomyces* species, not when either culture was grown axenically. This was to ensure the compounds selected were first; a result of interaction between the two species which increases the likelihood of the compound possessing an antimicrobial application, and second; either the product of *Euglena* or induced by indirect effects of *Euglena*. This left a total of 568 nodes to choose from.

3.7.1 Targets 1A, 1B and 1C

Target 1A, 1B and 1C were chosen from an LB co-culture of *S. limosus* and *Euglena gracilis*, A and B were extracted into methanol and C into ethyl acetate, all three were present in intensity from the $x10^4$ to the $x10^5$ range as can be seen in the extracted ion chromatograms in Figure 3.10. However, there was some concern that they would be difficult to isolate due to the density of high intensity peaks around the same retention time, as can be seen in Figure 3.8 and 3.9.

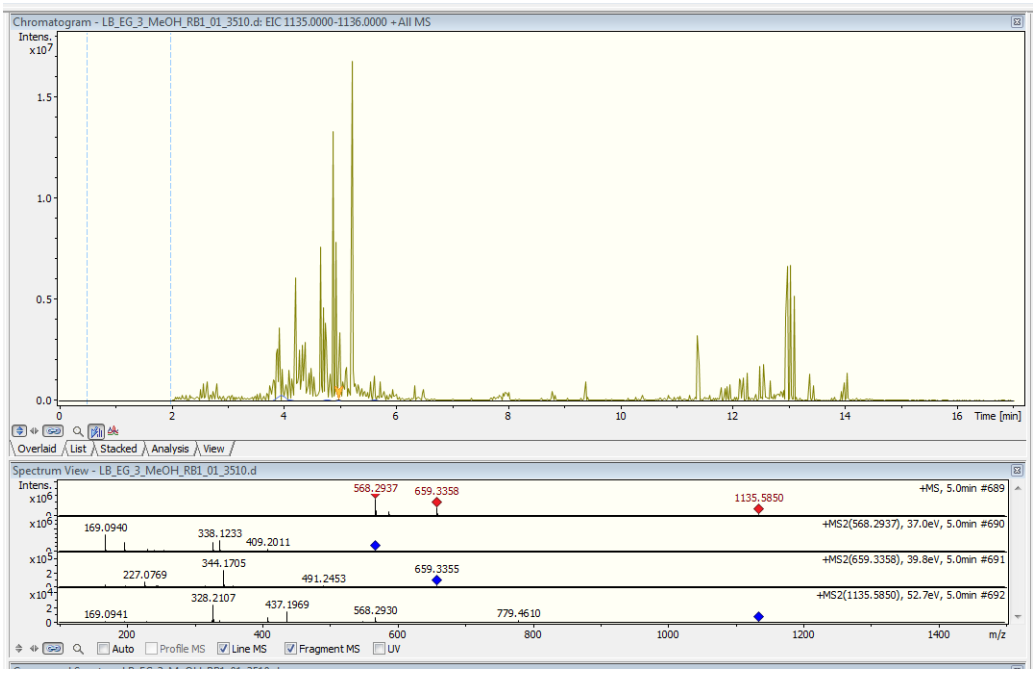


Figure 3.8 TIC of *S. limosus* and *E. gracilis* grown on LB and extracted into MeOH, 1A shown (yellow arrow)

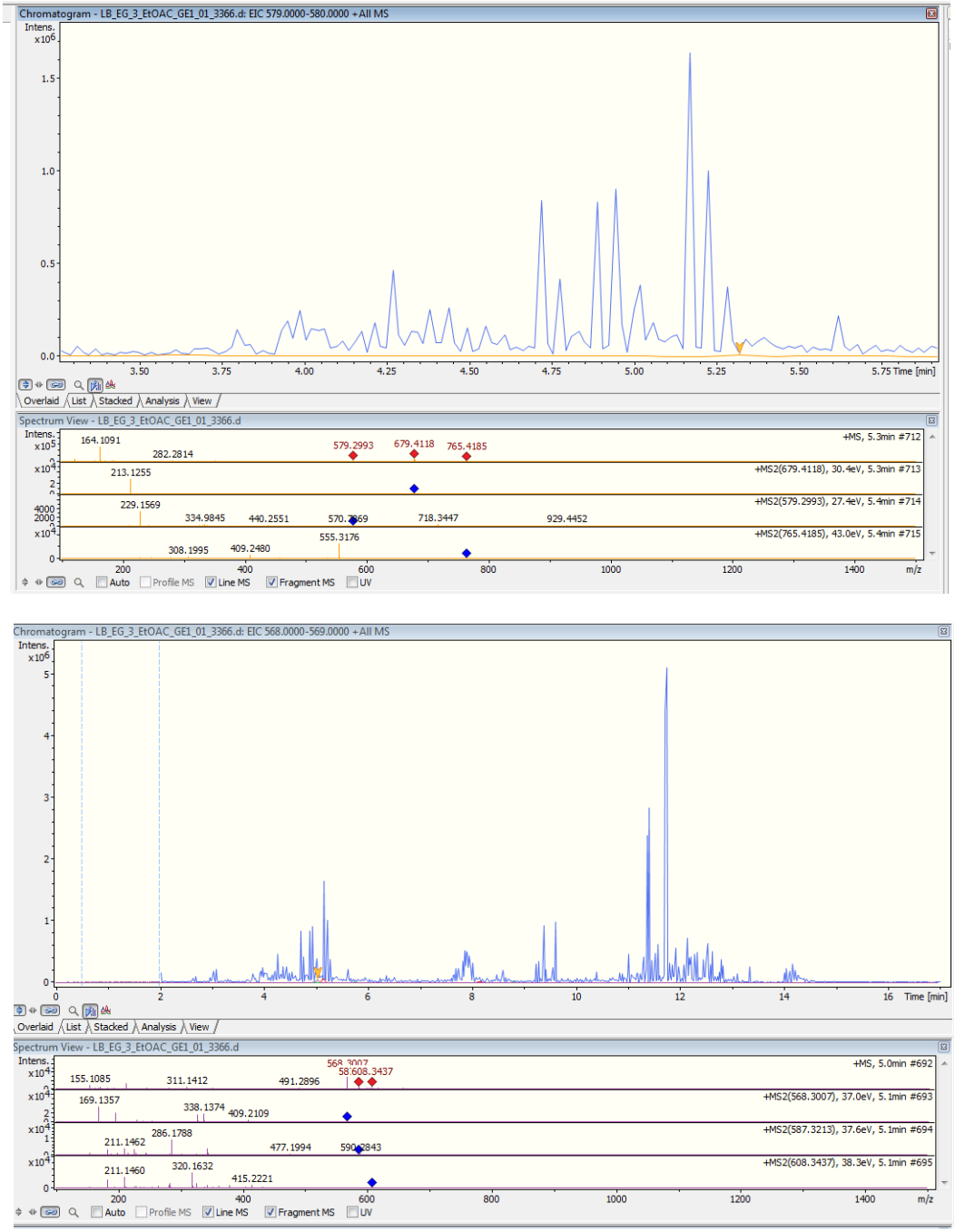


Figure 3.9 TICs of *S. limosus* and *E. gracilis* grown on LB and extracted into EtOAc, 1C shown (yellow arrow). Top image magnified to retention time of 1C in TIC.

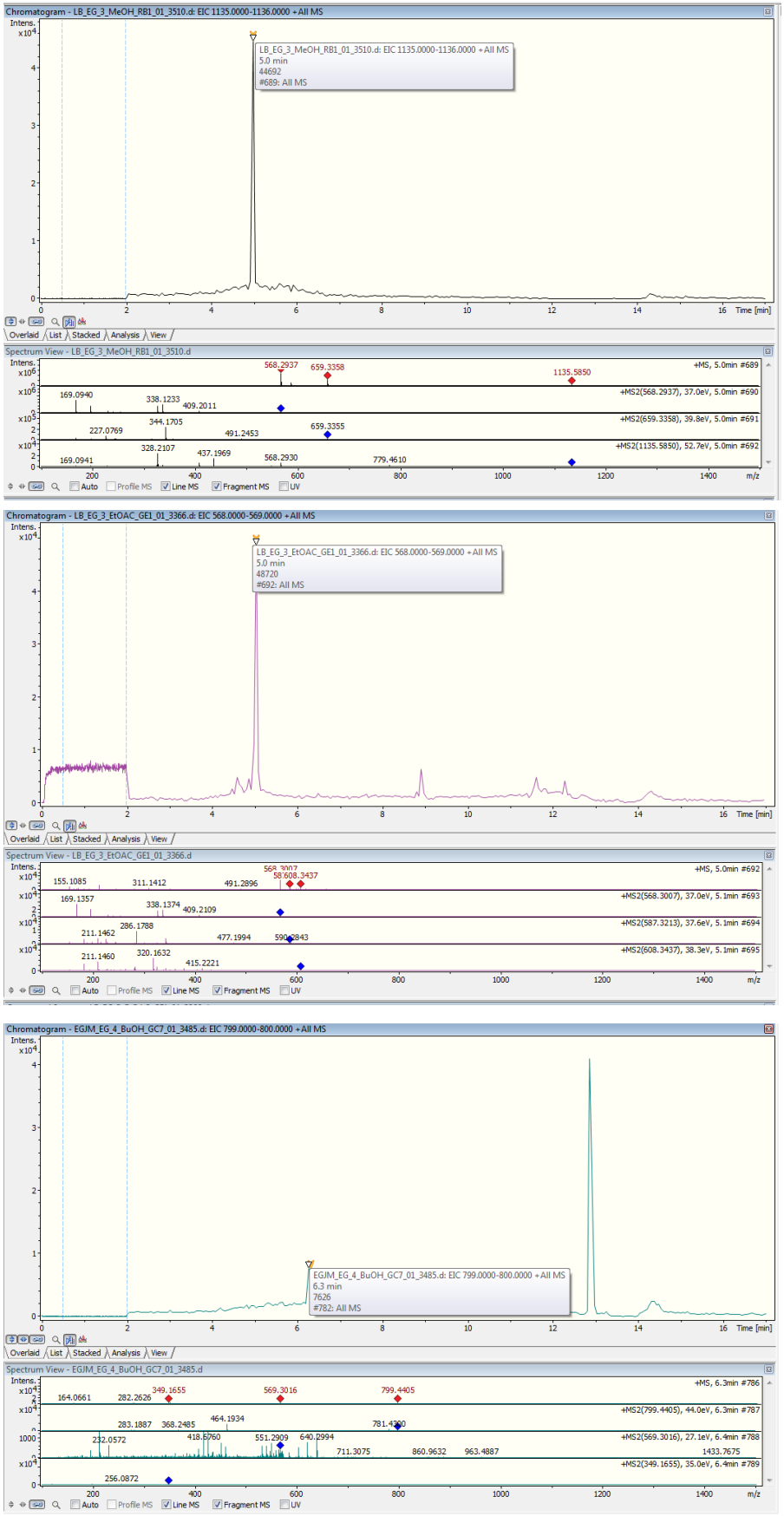


Figure 3.10 EICs of Targets 1A (top) 1B (middle) and 1C (bottom)

3.7.2 Target Two

Target 2A and 2B were a 2-nodal cluster coming off at effectively the same retention time at 6.55 and 6.56 minutes and a m/z difference of 84.254 between them. They were produced in coculture with *S. coelicolor* in EG:JM and extracted into butanol.

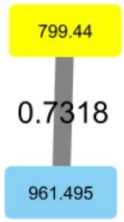


Figure 3.10 showing 2 nodal cluster of target 2A and 2B

	Molecular Formula	m/z	Unsaturation
1	C_6H_{12}	84.0939	1
2	$C_5H_{10}N$	84.0813	1.5
3	$C_4H_8N_2$	84.0687	2
4	C_5H_8O	84.0575	2
5	$C_3H_6N_3$	84.0562	2.5
6	C_4H_6NO	84.0449	2.5
7	$C_2H_4N_4$	84.0436	3
8	$C_3H_4N_2O$	84.0324	3

Table 3.2 Possible molecular formulas for difference between Target 2A and 2B

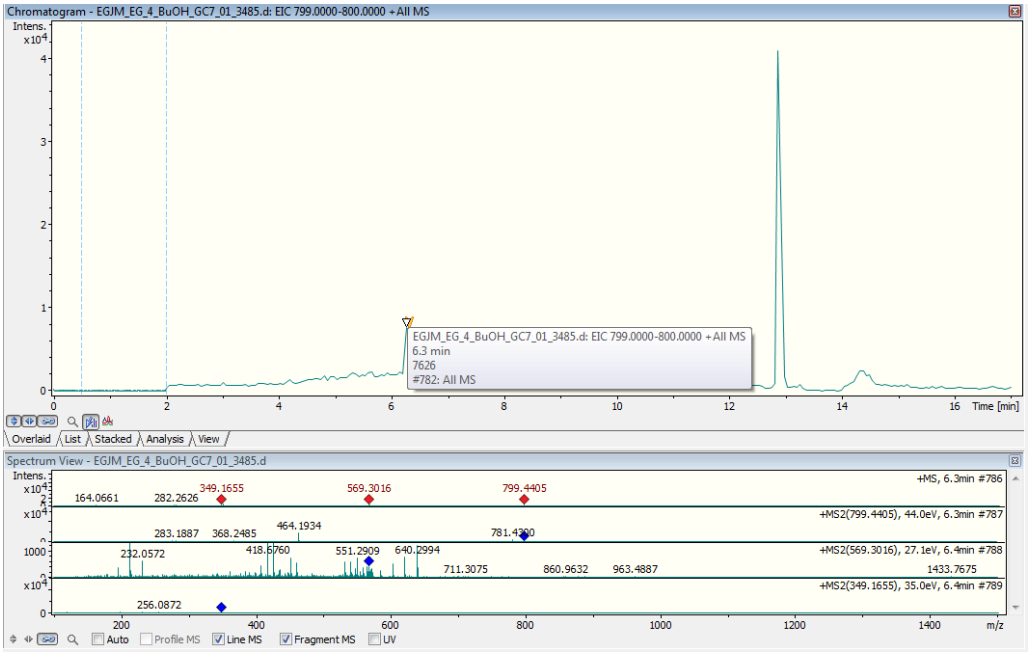


Figure 3.11 EIC of Target 2A

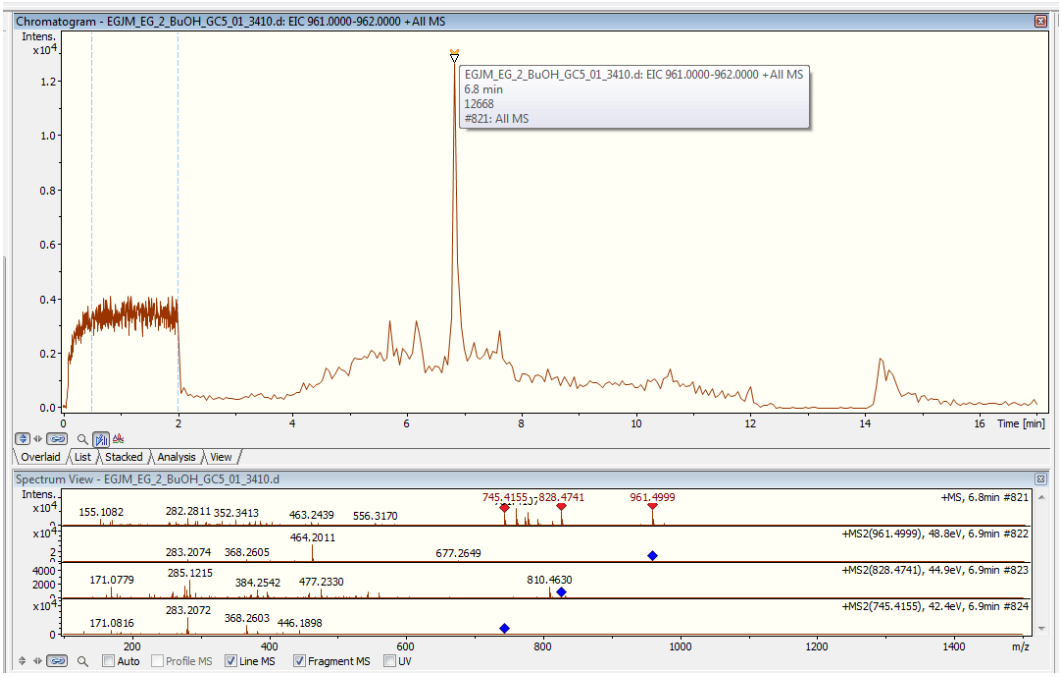


Figure 3.12 EICs of Target 2B

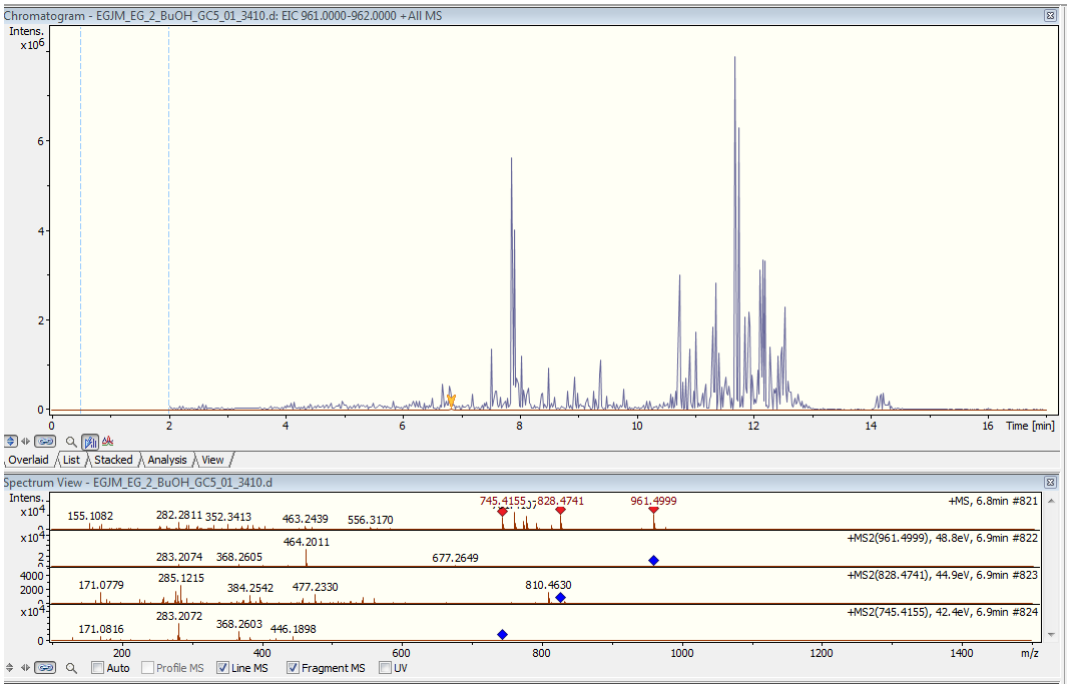


Figure 3.13 TIC of *Streptomyces coelicolor* and *Euglena gracilis* cultured on EG:JM and extracted into butanol.

Target 2A and 2B were both in the x10^4 intensity range and less masked by other peaks than Target 1A, B and C but still masked by large peaks as can be seen in Figure 3.13

3.7.3 Target Three

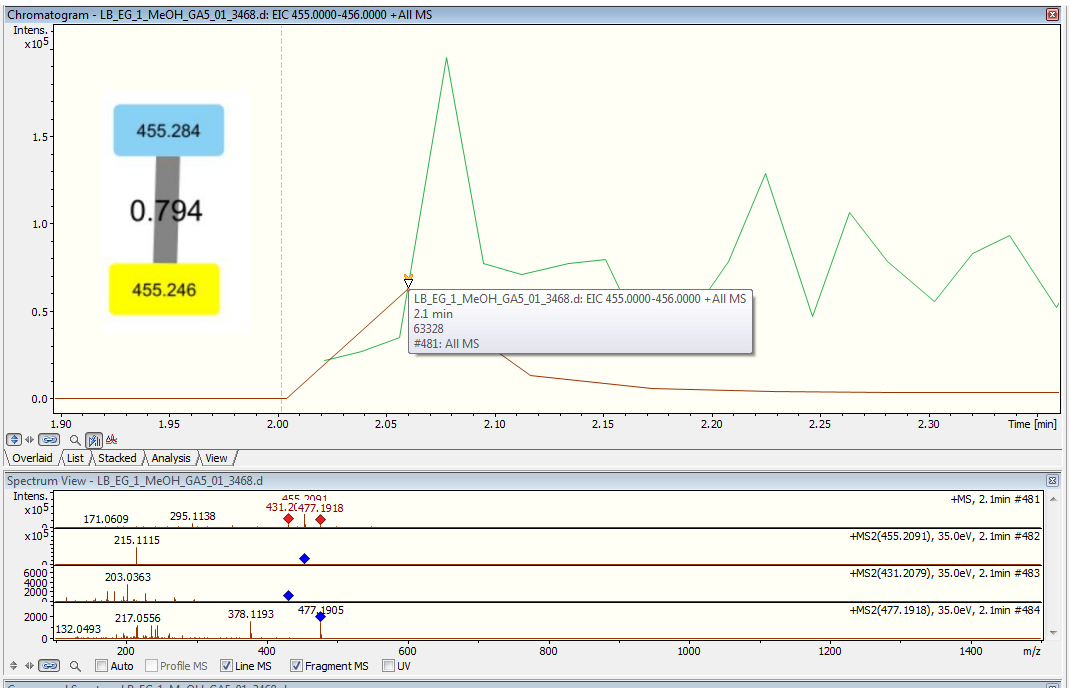


Figure 3.14 TIC (green) of *Streptomyces coelicolor* and *Euglena gracilis* cultured on LB and extracted into methanol. Magnified to target 3A (red)

Target Three was also derived from a coculture with *S. coelicolor* and *E. gracilis* but unlike Target Two it was from a culture grown in LB and extracted into MeOH. It is also a doublet, but there is no significant mass difference between the two, only a difference of retention time of 1.25 minutes. The hope for Target 3A, which eluted at 2.05 minutes was that by using a very polar solvent system it would be possible to isolate it fairly easily as the first compound eluting off the column.

3.7.4 Target Four

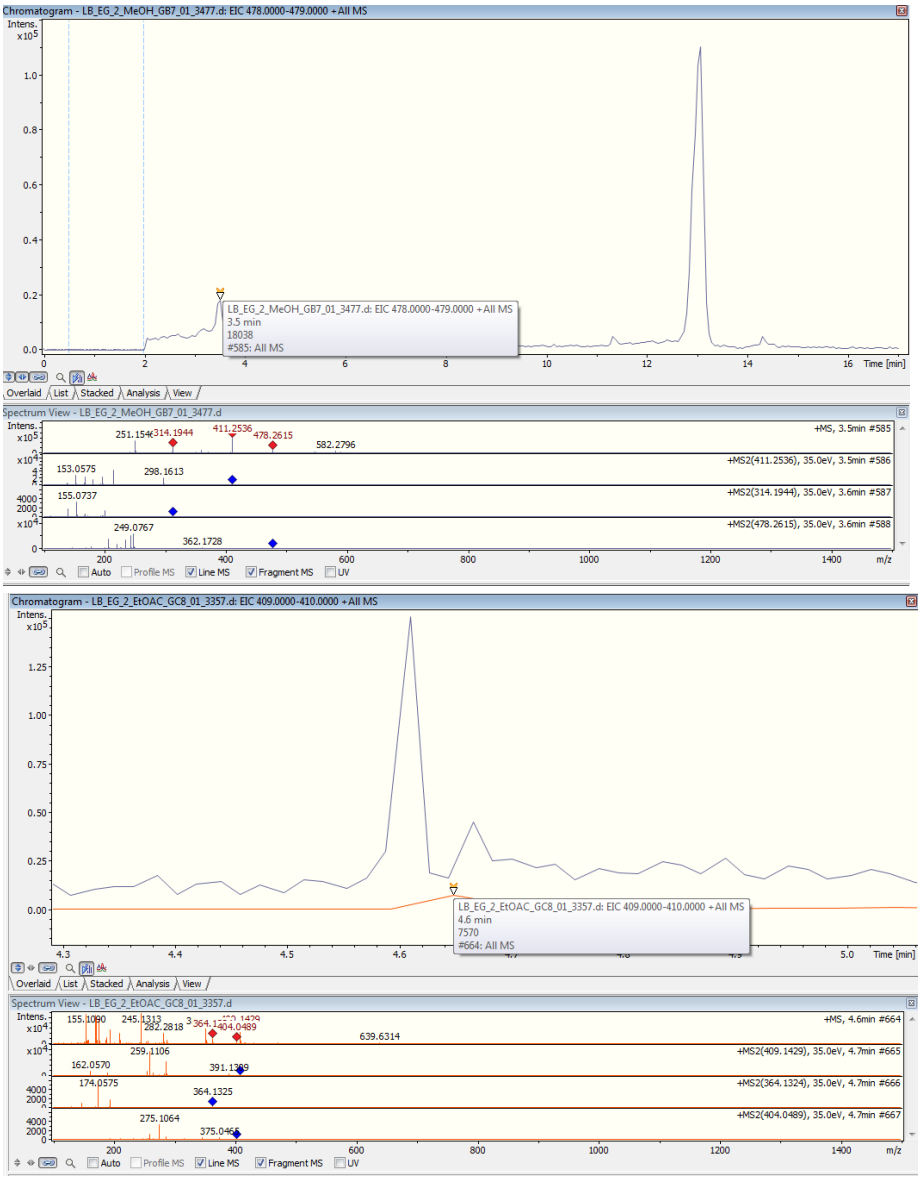


Figure 3.15 EIC of *S. glauscens* and *E. gracilis* coculture grown in LB and extracted into MeOH (top) TIC of *S. glauscens* and *E. gracilis* coculture grown in LB and extracted into MeOH (bottom) with target 3B shown in orange.

Similar to Target 3A, Target 4A came off the column at the beginning of the gradient but this time is derived from a coculture with *Streptomyces glauscens* grown on LB and extracted into MeOH. 4A eluted at 3.83 minutes and 4B at 4.67 minutes, less than a minute difference but sufficient enough to separate using HPLC.

3.7.5 Target Five

Like Target Four compounds, the two target five compounds are found in cocultures of *S. glauscens*. They were also a 2-nodal cluster with the same m/z but different retention times, 5B eluting just over a minute and a half after Target 5A. Target 5A and 5B were observed in x10^4 intensity when cultured in LB and extracted into butanol.

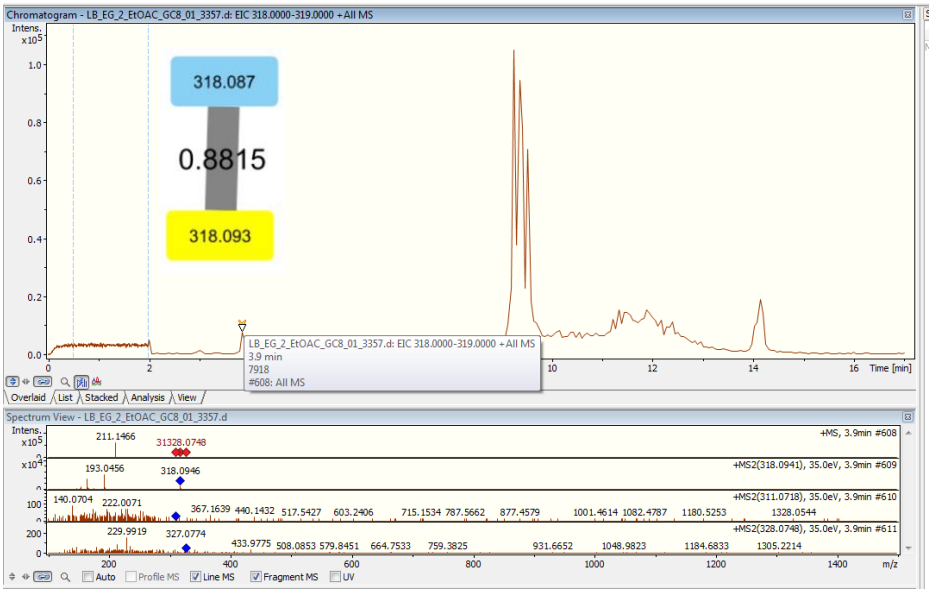


Figure 3.16 EIC of *S. glauscens* and *E. gracilis* coculture grown in LB and extracted into EtOAc with Target 5A highlighted with yellow arrow.

3.7.6 Target Six

Target Six is a singular target despite the fact that it is present in a two nodal cluster. Although Target 6A is only observed when a coculture of *S. limosus* with *E. gracilis* is grown in LB and extracted into butanol, the node it is connected to is much more prevalent; being found in seven different environments including both monocultures and cocultures *E. gracilis,* and cocultures of both *C. vulgaris* and cyanobacteria but not in the same conditions as 6A. The choice was therefore made to isolate the rarer of the connected compounds.

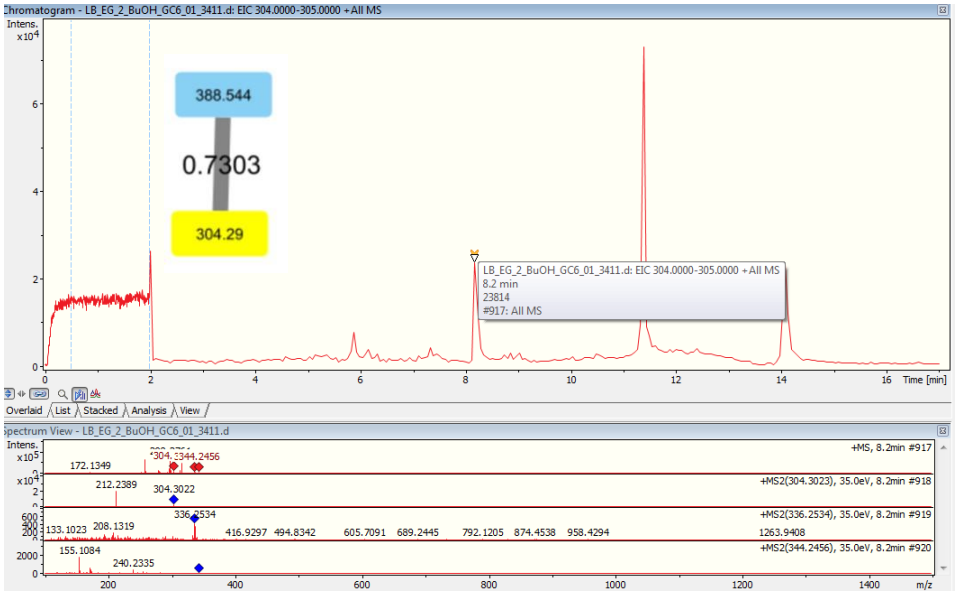


Figure 3.17 EIC of *S. limosus* and *E. gracilis* coculture grown in LB and extracted into BuOH with Target 6A highlighted with yellow arrow.

3.8 Overview of Targets

	Streptomyces strain	Media	m/z	RT (mins)	Extracting Solvent
1A	*S. limosus*	EG:JM	560.324	5.60	MeOH
1B	*S. limosus*	EG:JM	1135.59	5.04	MeOH
1C	*S. limosus*	EG:JM	568.336	4.64	EtOAc
2A	*S. coelicolor*	EG:JM	799.440	6.55	BuOH
2B	*S. coelicolor*	EG:JM	961.495	6.56	BuOH
3A	*S. coelicolor*	LB	455.246	2.71	MeOH
3B	*S. coelicolor*	LB	455.284	3.95	MeOH
4A	*S. glaucescens*	LB	478.261	3.84	EtOAc
4B	*S. glaucescens*	LB	409.143	4.67	EtOAc
5A	*S. glaucescens*	LB	318.093	2.40	BuOH
5B	*S. glaucescens*	LB	318.087	3.85	BuOH
6A	*S. limosus*	EG:JM	304.290	8.22	MeOH

Table 3.3 Selected targets from Cytoscape Network Two

Targets were thus selected from this network to be grown up in 1 L cultures: 2 of *S. limosus* grown in EG:JM, 1 of *S. coelicolor* in EG:JM, 1 of *S. coelicolor* in LB, 2 of *S. glaucescens* in LB and 1 of *S. limosus* in EG:JM were cultured. Unfortunately, all the *S. limosus* cultures were contaminated by what appeared to be fungus and not taken forward. After culturing for 4 weeks, three 5 mL samples were taken from each the *S. coelicolor* and *S. glaucescens* 1 L cultures and extracted into ethyl acetate, butanol and methanol for HPLCMS-MS analysis. These were networked with the previous data from Cytoscape Network Two to produce Cytoscape Network Three, analysis of which revealed several differences between the 10 mL and 1 L cultures.

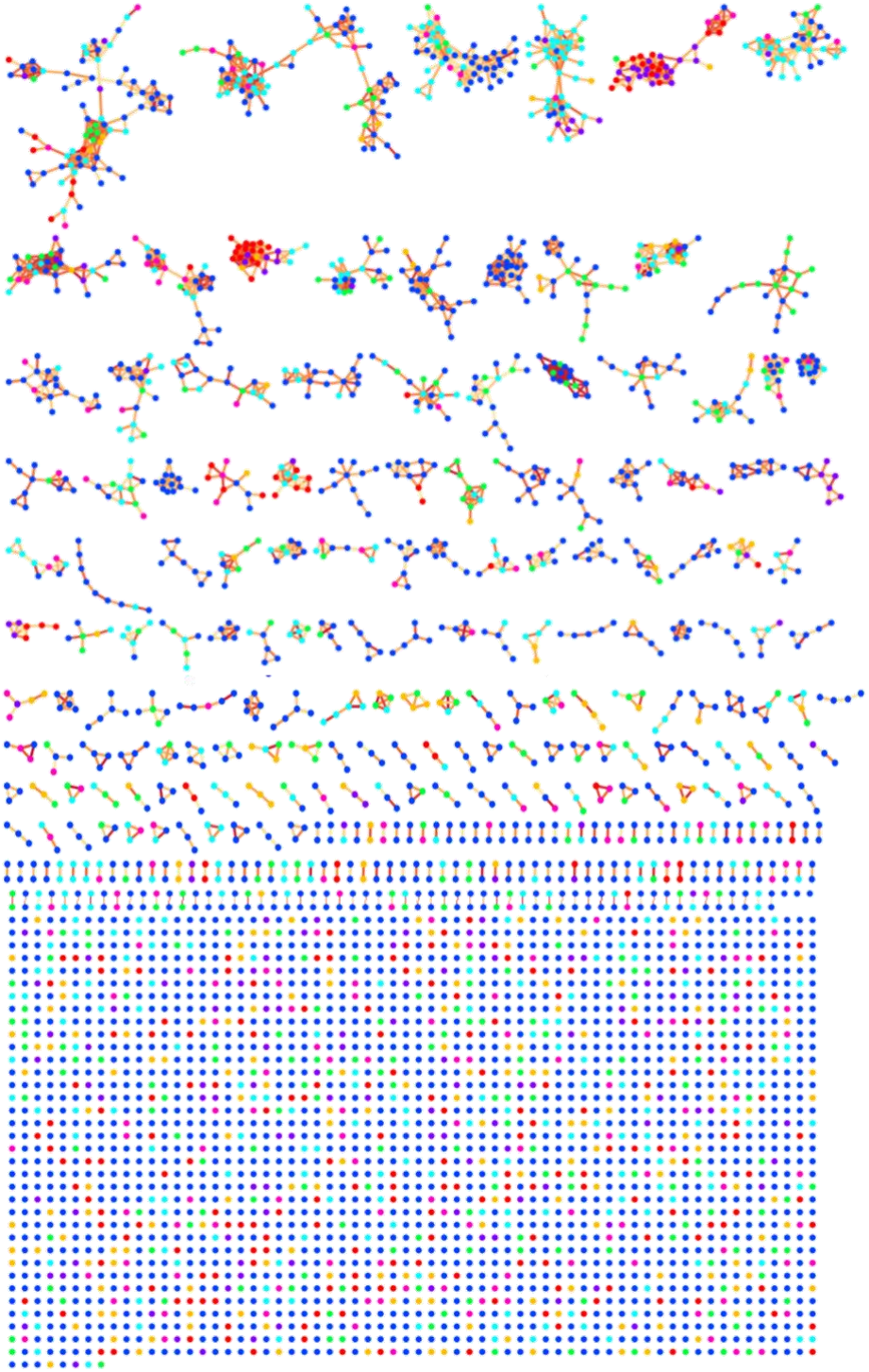


Figure 3.20 Cytoscape network of original cocultures (blue), large scale *S. glauscens* in LB (yellow), *S. coelicolor* in EG:JM and LB (red), original cocultures and *S. glauscens* (green), original cocultures and S. *coeilcolour* (pink), all three (cyan)

Observed differences in metabolite production between 10 mL and 1 L cultures suggest that culture volume significantly influences secondary metabolite profiles. In the case of *Streptomyces glaucescens*, two separate 1 L cultures, despite being prepared identically and inoculated with the same species, exhibited distinct coloration—one purple and one red—indicating variations in metabolic activity between the cultures. This visual difference was mirrored at the molecular level, with 240 nodes shared between the two cultures but 166 unique to one and 184 unique to the other. The reason for such a difference is unclear, it could be attributed to contamination of some kind, or differing bacteria/algae inoculation ratios. Such stark differences highlight the sensitivity of microbial metabolism to environmental factors, including changes in oxygen availability, nutrient gradients, and microbial interactions, all of which vary significantly with culture volume. These results emphasize that even when external conditions are kept constant, the metabolic landscape within a culture can diverge drastically due to small shifts in internal microenvironments.

Comparative analysis between 10 mL and 1 L cultures of *Streptomyces coelicolor* in different media (EG:JM and LB) further reinforced this trend, with larger cultures displaying a greater number of unique nodes. For example, in EG:JM medium, the 1 L culture exhibited 263 unique nodes compared to just 24 in the 10 mL culture, with only 40 nodes shared between them. A similar pattern emerged in LB medium, where the 1 L culture produced 193 unique nodes while the 10 mL culture had only 24, with 75 nodes shared. The substantial increase in unique nodes in larger volumes suggests that increasing culture size may provide conditions that either allow for the expression of previously silent biosynthetic pathways or induce stress-related metabolic adaptations that drive the production of novel compounds. This suggests that larger-scale cultures enable more extensive metabolic diversification, likely due to spatial and chemical heterogeneity that emerges at greater volumes.

Several factors could explain why increasing culture volume leads to such stark differences in metabolite production. First, oxygen diffusion varies significantly between small and large cultures. In smaller cultures, oxygen can more readily reach all parts of the medium, ensuring uniform aeration and supporting aerobic metabolic pathways. However, in larger cultures, oxygen transfer becomes limited, creating distinct aerobic and microaerobic zones. This can shift metabolic priorities within microbial communities, leading to the activation of different biosynthetic pathways. For instance, certain secondary metabolites may be produced only under low-oxygen conditions, while others require high oxygen availability.

Second, nutrient gradients develop more prominently in larger cultures. In small volumes, nutrients are relatively homogeneous throughout the medium, ensuring a consistent supply to all cells. In contrast, larger cultures experience localized depletion zones where cells consume nutrients at different rates, leading to metabolic stress in some areas. This stress can trigger the production of secondary metabolites that would not otherwise be produced under uniform, nutrient-rich conditions. It has been well-documented in *Actinomycetes* that secondary metabolism is often upregulated in response to nutrient limitation, as many bioactive compounds function as competitive molecules against other microbes when resources become scarce.

Additionally, waste accumulation differs between culture volumes. In small cultures, metabolic byproducts are more easily diluted, minimizing their impact on microbial growth and metabolism. In contrast, in larger cultures, localized accumulation of inhibitory compounds can occur, which may induce stress responses that alter metabolic activity. This could explain why *S. glaucescens* exhibited two distinct metabolic profiles in what were nominally identical cultures—slight differences in localized waste concentrations or diffusion patterns may have triggered different metabolic responses in each culture.

Another factor to consider is microbial interactions. In co-culture conditions, bacteria and algae may compete differently depending on culture volume. In larger cultures, spatial separation can occur, creating micro-niches where different microbial populations thrive independently. This can lead to variations in metabolic output, as some microbial species may dominate certain regions of the culture while others struggle to compete. The observation that bacteria outcompeted algae in the 1 L cultures, particularly in nutrient-rich media, further supports this idea. As the algal population declined due to competition, its contribution to secondary metabolite production likely diminished, shifting the overall metabolic output of the culture.

These findings align with previous research on the impact of culture conditions on metabolite production. Studies on actinomycetes and cyanobacteria have shown that altering culture volume can lead to significant changes in secondary metabolite profiles, sometimes unlocking novel compounds that remain undetected in small-scale cultures. For example, research on *Streptomyces* species has demonstrated that changes in culture volume can influence antibiotic production, with larger cultures often inducing stress-related pathways that enhance bioactive compound synthesis.[194] Similarly, in marine cyanobacteria,

increasing culture volume has been shown to enhance the production of bioactive compounds due to changes in light penetration, nutrient availability, and microbial competition.[195]

In conclusion, culture volume is a critical factor influencing the metabolic output of microbial cultures. Larger culture volumes introduce oxygen gradients, nutrient limitations, waste accumulation effects, and altered microbial interactions, all of which contribute to shifts in secondary metabolite production. Understanding these dynamics is essential for optimizing the cultivation of bioactive compounds, particularly in the context of drug discovery and biotechnology, where the ability to modulate metabolite production through culture conditions could enhance the yield of valuable natural products. Further research into the mechanistic basis of these volume-dependent metabolic changes will be key to developing more efficient strategies for microbial natural product discovery and bioproduction.

However, the driving force of this work was to isolate one of these novel compounds therefore to move forward, new targets were identified from Cytoscape Network Three.

3.10 Comparison of 1L cultures

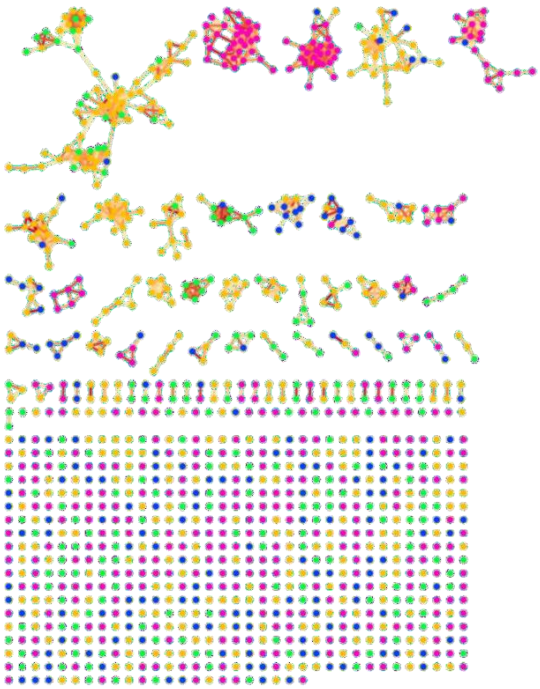


Figure 3.21 Cytoscape network of 1 L *S. coeilcolor* cultures with nodes present both culturing media (blue), nodes only in LB (green), only in EG:JM (pink), and control (yellow).

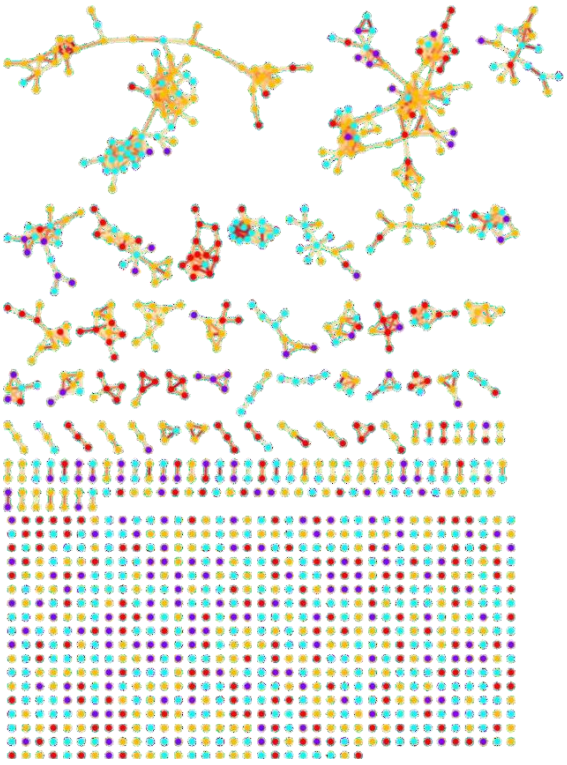


Figure 3.22 Cytoscape network of 1 L *S. glauscens* cultures with nodes in both cultures (blue), nodes in LB Flask 1 (purple), LB Flask 2 (red), control (yellow).

The molecular network analysis of *Streptomyces coelicolor* coculture grown in 1 L cultures revealed a total of 1129 nodes, 1499 edges, and 72 clusters, with 738 of those nodes being non-control. This network includes 365 nodes unique to EG:JM cultures, 203 nodes unique to LB cultures, and 170 nodes common to both culture conditions. The increased number of nodes produced in more complex media has been consistently observed throughout Cytoscape networks thus far. The observation of non-overlapping unique nodes highlights that culture conditions can complement coculturing in the production of increased novel product synthesis.

When compared to *S. glaucescens*, which produced 1299 nodes, 1475 edges, and 800 non-control nodes in its 1 L molecular network, it is apparent that there is a higher total number of nodes produced by *S. glaucescens* (170 more nodes than *S. coelicolor*). Despite this, the *S. glaucescens* network contained fewer edges (24 less) and more clusters (20 more). The difference in cluster count suggests that the compounds produced by *S. glaucescens* might exhibit more variation in their structure or properties, leading to the formation of more distinct groups (clusters). This could indicate that *S. glaucescens* produces a broader range of secondary

metabolites, or that its compounds are less chemically similar than those of *S. coelicolor*, indicating that whilst the network might contain fewer compounds, they are more chemically diverse.

When examining the cultures in Flask One and Flask Two for *S. glaucescens*, 171 nodes were unique to Flask One, 235 were unique to Flask Two, and 394 nodes were common between the two. The 64 additional unique nodes in Flask Two may reflect an increased presence of *Euglena* within the culture. The ethyl acetate extractions for Flask Two were notably greener in colour, indicating a higher concentration of chlorophyll thus suggesting a higher population of *Euglena gracilis*. The increase in algal biomass could increase the amount of cell signalling between *E. gracilis* and *S. glaucescens*, leading to the production of novel compounds not observed in Flask One.

3.11 New Targets from Cytoscape Network Three

New targets were selected from Cytoscape Network Three using the same conditions as previous, with the added parameter that they be present in at least 2 of the 5 mL repeats from the 1 L cultures. There were more compounds which fulfilled the criteria in the *S. glaucescens* and *Euglena gracilis* co-cultures than the *S. coelicolor* and *Euglena gracilis* co-culture. However, there was a lack of targets present in both 1 L cultures of *S. glaucescens* in LB - only the 503 m/z and 421 m/z target were present in both. An overview of the target compounds properties are shown in Table 3.4.

Euglena gracilis co-culture organism	Media	Parent mass	RT/min	Solvent
S.coelicolor	LB	635.240	2.8	MeOH
S.coelicolor	LB	778.225	7.9	BuOH
S.coelicolor	EG:JM	507.152	5.4	EtOAc
S. glaucescens	LB	421.285	3.6	BuOH
S. glaucescens	LB	503.291	5.1	BuOH
S. glaucescens	LB (Flask 1)	468.216	5.7	BuOH
S. glaucescens	LB (Flask 2)	991.552	4.8	MeOH
S. glaucescens	LB (Flask 2)	506.337	7.9	EtOAc
S. glaucescens	LB (Flask 2)	502.306	8.6	EtOAc
S. glaucescens	LB (Flask 2)	504.322	8.0	EtOAc
S. glaucescens	LB (Flask 2)	410.206	4.4	BuOH
S. glaucescens	LB (Flask 2)	486.238	6.0	BuOH
S. glaucescens	LB (Flask 2)	444.191	4.6	BuOH

Table 3.4 Selected targets from Cytoscape Network Three

3.12 Cytoscape Network 3: Isolating the 1 L Targets

3.13 *S. coelicolor* 1L targets

The LB culture was centrifuged and extracted into butanol as the highest relative intensity of 778.225 m/z was observed in this HPLCMS/MS spectrum compared to the other solvents. The organic layers were combined and washed with water before being concentrated down on the rotary evaporator. The aqueous layers were then frozen, and the water removed on the freeze drier to leave a mixture of brown and white precipitate. The precipitate was then extracted into MeOH, as in the methanol extraction the highest relative intensity of peak 635.240 m/z was observed. The large-scale extraction of EG:JM was carried out with EtOAc in order to begin the isolation of target 507.152 m/z. After extraction, the presence of the target peaks was confirmed using HPLCMS/MS, but unfortunately 507.152 m/z and 635.240 m/z were not identified. This could be due to degradation of the product due to a much lengthier extraction and concentration time due to the higher volume of media, or perhaps prolonged exposure to heat on the rotary evaporator.

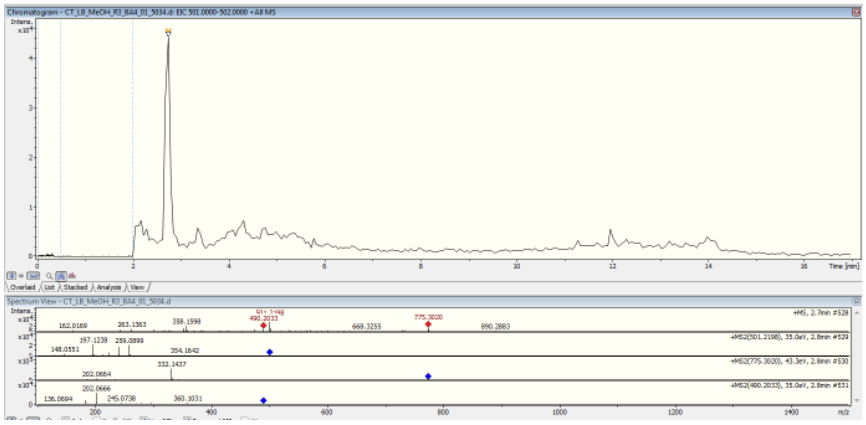


Figure 3.23 EIC of Target 778.225 m/z

After confirmation of the presence Target 778.225 m/z in the butanol extraction of the *S. coelicolor* in LB culture, attempts to purify the target were carried out by reverse-phase HPLC. The extract was dried *in vacuo* before reconstitution in MeOH for purification by HPLC. However, unlike in the small-scale extractions previously there were difficulties with solubility in methanol for the 1 L culture experiments. This could have been due to different compounds being produced in the 1 L cultures compared to the smaller scale cultures or due to a build-up of degradation products. Unlike the 10 mL samples or the 5 mL samples taken from the 1 L cultures, the 1 L cultures were left for an extended period of time before extraction begun due to the need to remove water from the cultures to make the volume more manageable. Despite repeated attempts at centrifuging and filtration, precipitate continually formed from the 100% MeOH. Therefore, a series

of preparative dilutions were trialled to separate the target compounds from the precipitate.

Figure 3.24 Cytoscape network of serial ACN dilutions Cultures with nodes in all dilutions (yellow), node in 100% acetonitrile (green), node in 70% acetonitrile (purple), node in 40% acetonitrile (cyan), node in 10% acetonitrile (pink), media

In order to identify a solvent system that would be suitable for the isolation of the targets in LB a series of dilutions in increasing acetonitrile concentrations in 10% steps up to 100% were carried out. Analyses in Cytoscape showed that 64 nodes were present in all dilutions. Only one node was unique to 100% acetonitrile, one to 70%, one to 40% and one to 10%. In addition, 778.225 m/z was not identified within dilutions. The lack of segregation of compounds based on polarity indicated poor solubility in the volume of acetonitrile solution used, was attributed to the high concentration of metabolites present in a small volume. It was also thought that not all material had been successfully diluted into the volume used which explained the absence of 778.225 m/z, therefore the volume of 10% acetonitrile in water used to solvate the residue was increased from 1 mL to 10 mL and syringe filtered rather than centrifuged in order to ensure full solubility before semi-prep HPLC.

From the HPLCMS/MS analysis, Target 778.225 m/z eluted at 7.5 minutes which corresponds to an estimated 60% acetonitrile. A similar gradient was applied when the mixture was purified on the semi-prep HPLC, and Target 778.225 m/z was found in 3 HPLC fractions at around 60%, confirmed by HPLC-MS/MS analysis. These fractions were combined, dried, and went through another HPLC run at 45-70% acetonitrile concentration gradient. Target 778.225 m/z was not detected in this second HPLC run, likely due to the fact that it was present in too low a concentration.

In order to attempt some structural prediction on the target, the fractions were then pooled, dried and reconstituted in 90:10 H_2O:D_2O solvent mix and submitted to 400 MHz NMR. No conclusions could be made from the NMR due low concentration of analytes.

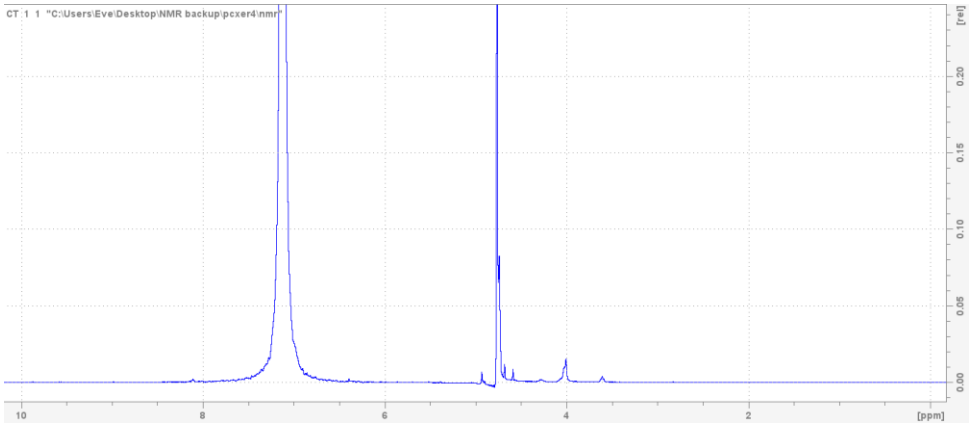


Figure 3.25 ^{1}H NMR of combined target 778.225 m/z containing fractions

3.14 *S. glaucescens* 1 L targets

In order to ensure sufficient solubility of target compounds from *S. glaucescens*, residues from each solvent extract were dissolved in 1 mL solutions of acetonitrile and water and submitted for LC-MS/MS analysis in order to check solubility in the volume used.

3.141 *S. glaucescens* 1 L targets: butanol extract

Target 421.285 m/z was found in a high intensity in the 10% ACN solution but also in lower intensities in the 20-60% ACN solutions. As Target 421.285 m/z elutes in less than 20% Acn, its presence in higher volumes suggests it was not completely dissolved in the 10% solution so the volume was increased. 503.291 m/z was found in low intensities in solutions of 20-50% ACN, and targets 468.216 m/z, 410.206m/z, 486.238m/z, and 444.191 m/z were not found in any of the solutions. Target 503.291m/z was found in a high relative intensity in the 10% acetonitrile

solution, and also in lower relative intensities in the 20-60% acetonitrile solutions. The same conclusions were drawn for 503.291 m/z as were for 421.285 m/z. Although 421.285 m/z was observed at a higher intensity in the ethyl acetate extract, attempts were made to isolate 421.285 m/z from the butanol extract as well in order to increase the total amount of product isolated. After purification, Target 421.285 m/z was found in HPLC fractions collected from 4 to 7 minutes – corresponding to 12-19% ACN. However, compound 503.291m/z was not found in any of the HPLC fractions. Given that 503.291m/z was soluble in the solvent system at the gradient used, it was likely not found following purification because it was too dilute to identify its peak.

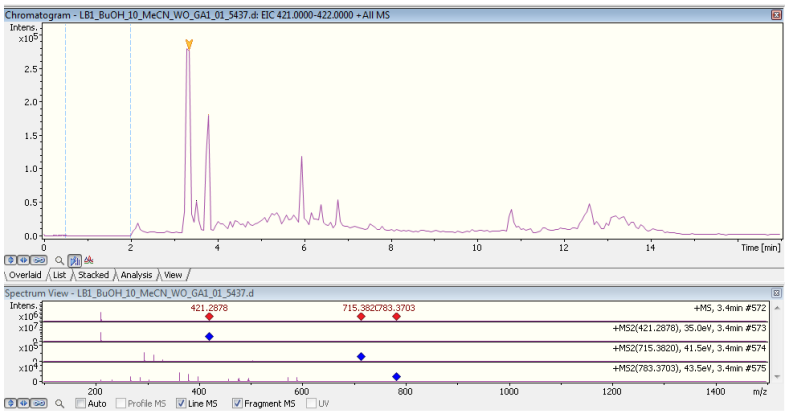


Figure 3.26 EIC of 421.2885 m/z in 10%

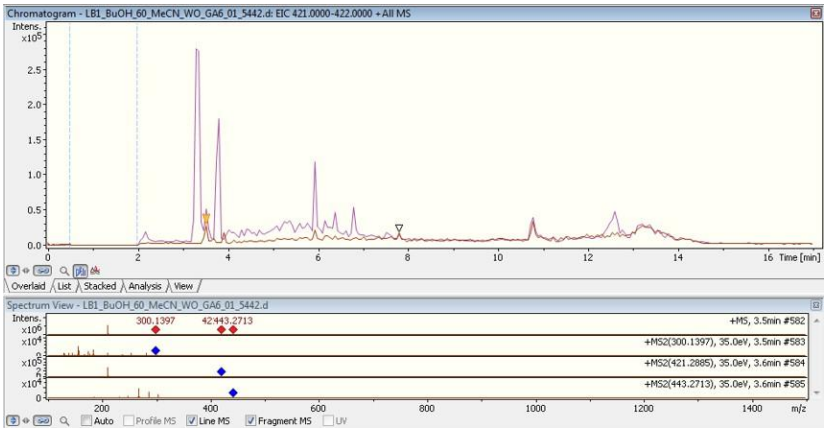


Figure 3.27 EIC of 421.2885 m/z in 10% acetonitrile (pink) and 60% acetonitrile (orange)

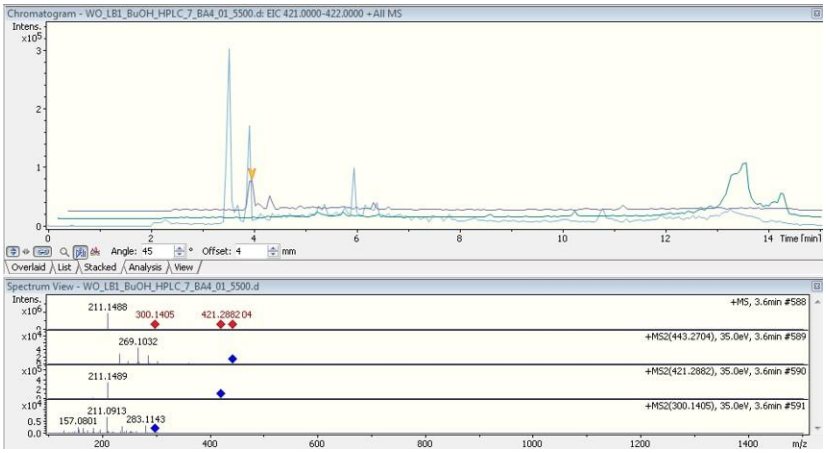


Figure 3.28 EIC of 421m/z (orange arrow) and 503.291m/z (turquoise) in fraction 6 after HPLC.

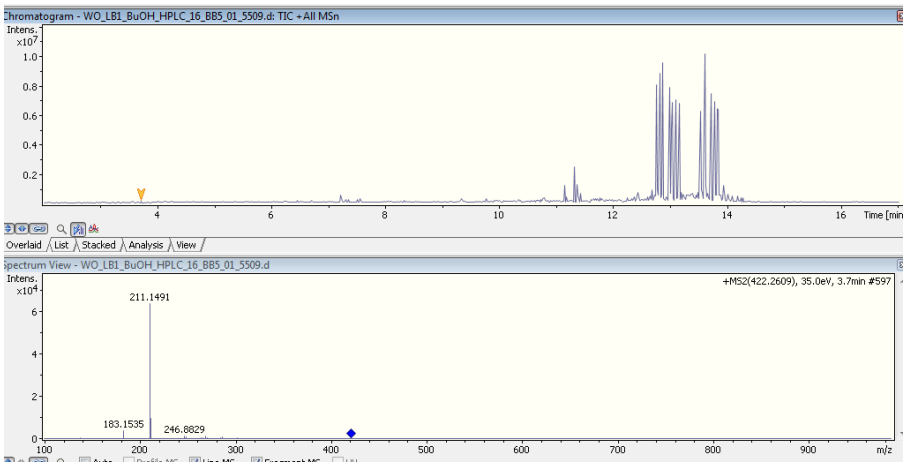


Figure 3.29 TIC of 421 m/z (yellow arrow) fraction 16 after second HPLC run.

The fractions containing 421 were pooled at run on a shallower column, from 0 – 30% acetonitrile over 30 minutes. Subsequent analysis identified 421 in one fraction – fraction 16 corresponding to 15% acetonitrile. However lipophilic contaminants remained in the mixture, and the relative intensity of 421 was very low.

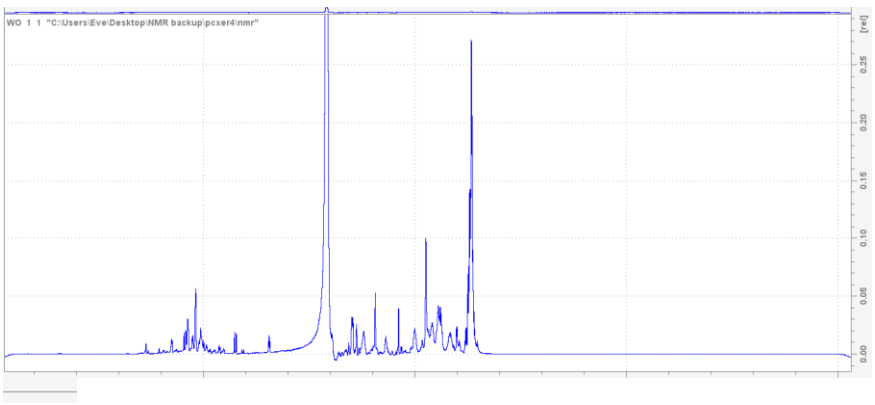


Figure 3.31 H^1 NMR of fraction 16 containing target 421

In order to attempt some structural prediction on the target, the fractions were then pooled, dried and reconstituted in 90:10 H_2O:D_2O solvent mix and submitted to 400 MHz NMR. In this case, the mixture contained too many compounds to make any structural annotations.

3.142 *S. glaucescens* 1 L targets: methanol extract

The same method was used to test the solubility in the HPLC solvent system as described for the BuOH extracts. However, the MeOH extracts dissolved in mixtures of acetonitrile and water rapidly separated into two layers. Separation did not occur when the BuOH extracts were dissolved in the same solvent mixtures, which indicates that this separation was caused by a compound or salt only present in a high concentration in the dried LB2 MeOH extractions. NaCl is the most likely cause of the solvent separation observed; given that the LB media used to cultivate the coCultures contain a high amount of salt (10 g/L), and NaCl is insoluble in EtOAc and BuOH and thus will have remained in the aqueous phase for the MeOH extractions, in which it is weakly soluble (14 g/L). Furthermore, NaCl is highly soluble in water (360 g/L) but insoluble in acetonitrile, which would explain why the two solvents separate and also suggests that the white solid which precipitated over time on the boundary in between the two solvents was mainly if not entirely salt.

All of the samples were cloudy, although there was more precipitate observed in the aqueous layers than in the organic layers. The layers were separated, filtered and analysed separately using LC-MS/MS, to check for the presence of the target compounds either in the aqueous or in the organic layers but none were observed. The LB2 MeOH targets would be expected to be found in solutions of 40-70% ACN to water, but due to the solvent separation in these samples the extracts were divided into layers of 100% water or layers of 100% ACN. Therefore, it is possible that the target compounds are not soluble in the separated solvents and would not be found in the mass spectrometry data.

For lack of ability to find a suitable solvent system, the MeOH extracts were not analysed further; however, if this target was to be pursued in future work, an attempt could be made to remove the salt through a size exclusion column. Alternatively, the aqueous layers could be demineralised by ion exchange chromatography utilising a cation exchange resin in the hydrogen form followed by an anion exchange resin in the hydroxide form. This demineralised solution could then be dried down and reconstituted in the required acetonitrile and water solvent systems and analysed using LC-MS/MS, before purifying by HPLC if the target compounds are found in solution.

Figure 3.30 Molecular Network of 100% water (yellow) and 100% acetonitrile (blue) extracts of LB Flask 2.

Target 421.285 m/z was the only compound explored that displayed sufficient stability during handling and chromatography therefore it was selected as the sole target for further analysis. The lack of success in purification of Target 421.285 m/z was attributed to a lack of material and the need for an additional chromatographic technique to remove the lipophilic components in the mixture. Therefore, it became clear that a much larger amount of material would be required so that it could withstand several purification steps. Compound 421.285 m/z, henceforth known as compound 421 was the focus of the subsequent increase in culturing volume. As well as remaining observable by HPLCMS/MS even after heating and two rounds of HPLC purification, indicating that it was stable enough to withstand a lengthy purification purpose. Additional features of 421 which made it an attractive candidate for isolation was its production in both cultures of *S. glaucescens* despite their largely different metabolic profiles. Additionally, the fragmentation pattern was consistent across all analyses, always showing a dimer at 211 m/z, and it was produced in relative high intensity compared to the other compounds in the same extract as observed in the HPLCMS/MS spectra.

3.15 Target 421 structural prediction

LC-MS/MS data obtained for Target 421.2885 m/z shows only one fragment peak at m/z 211.1414, which is half of the parent mass at m/z 421.2885, indicating that the target is a dimer. Both of these m/z values include one H^+ ion, meaning that the exact masses are 210.1336 for the fragment monomer and 420.2807 for the dimer. A possible chemical formula for the monomer is $C_{11}H_{18}N_2O_2$ with an unsaturation value of 4; this equates to valine (V) and leucine (L) or isoleucine (I) bonded in a ring with an extra unsaturation in the side chain of either amino acid.

There are 6 possible positions for the unsaturation in the VL combination of amino acids, and a further 8 possible arrangements for the VI combination – including 2 pairs of E/Z stereoisomers. These are shown in Figure 3.29. Furthermore, if epimerase enzymes are present in the synthesis of compound 421.2885 m/z either or both of V and L (or V and I) could exist as D-enantiomers. Two monomers could be held together to form the parent compound observed in mass spectroscopy either by hydrogen bonding or by a H^+ ion generated during positive ionisation (ESI). Typically, COSY, TOCSY and HSQC NMR spectra would be utilised to elucidate the structure and allow to differentiate between the possible arrangements. However, the sample obtained in this work thus far was not pure enough, there were multiple hydrogen environments overlapping in the proton NMR (as seen in Figure 3.30) and thus the structure of target compound could not be elucidated further from the data obtained at this point.

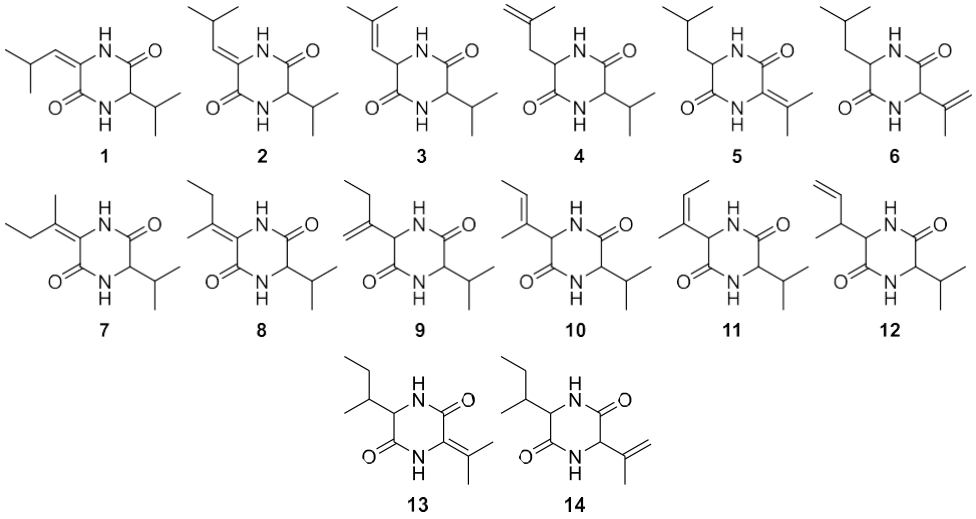


Figure 3.32 Possible structures for the monomer of target compound A (C11H18N2O2); where 1-6 are combinations of valine and leucine, and 7-14 are combinations of valine and isoleucine; 1 and 2, 7 and 8, and 10 and 11 are E/Z stereoisomers; any of valine, leucine or isoleucine may be L- or D-enantiomers.

This compound is inspired by the class of natural products diketopiperazines (DKPs). They are cyclic dipeptides, characterized by a bicyclic structure where two amino acids are linked through a peptide bond, and the resulting cyclic compound includes two carbonyl groups (at the 2 and 5 positions of the cycle). This structure, typically involving amino acid derivatives, makes DKPs versatile molecules with a wide range of biological functions. They are known for their antimicrobial properties, which have been observed in various microorganisms, including bacteria, fungi, and even viruses. DKPs produced by fungi and bacteria often show strong antibacterial effects, particularly against Gram-positive and Gram-negative bacteria, as well as antifungal and antiviral properties. For instance, compounds such as *cyclo-(L-Pro-L-Leu)* and *cyclo-(L-Pro-L-Val)* exhibit substantial antibacterial activities against *Staphylococcus aureus* and *Escherichia coli*, as well as antifungal properties against *Candida albicans*.[196] Several DKPs have been shown to inhibit the proliferation of cancer cells and induce apoptosis (programmed cell death) in a range of cancer types. For example, *cyclo-(L-Pro-L-Tyr)* has demonstrated antifungal activity and cytotoxicity against human breast cancer cells, colon cancer cells, and liver cancer cells.[197]

3.16 Target 421: Increasing culture volume to 6 L

In order to provide enough material for isolation, 6 L of *E. gracilis* were cultured in LB media with *S. glaucescens* over 6 weeks at 25 °C. 5 mL samples taken from each for HPLCMS/MS analysis to confirm the presence of target 421 – which was

observed in 4 of the 5 mL sample HPLC chromatograms through EICs – which indicated the correct RT, m/z and 211 monomer in the MS2. However, all six litre cultures, not just the four where the target was confirmed, were dried down to account for the possibility that target 421 was present in the other two 1 L cultures but absent in the 5 mL samples. This variation was observed as described previously in Chapter 2 in the triplicate 5 mL samples taken from the 1 L cultures of axenic algal cultures.

The 6 L of LB cocultures were then combined and dried down *in vacuo* over the course of 2 weeks, reducing the volume to 500 mL. The resulting mixture was then extracted into ethyl acetate and butanol, and each extract concentrated to a brown liquid (<10 mL) *in vacuo*. A 1 mL sample of each organic extract was taken and analysed through HPLCMS/MS and target 421 was identified in both – in a higher relative abundance in the ethyl acetate extract. Ethyl acetate is more volatile and extracts fewer other compounds, so was selected as the focus for initial extraction.

The ethyl acetate extract was dissolved into five x 10 mL solutions of acetonitrile and water – from 20%, 40%, 60% to 100% acetonitrile. Salt accumulated in each sample, removed through syringe filtering twice removing all solid residue. A 1 mL sample was taken from each and the presence of 421 confirmed through EIC of the HPLCMS/MS chromatograms.

Each solution was then run through semi-prep HPLC on a 10 to 50% acetonitrile gradient over 30 minutes and fractions collected every minute. 1 mL of each fraction was then transferred to HPLC vials and analysed for target 421 using LCMS/MS EIC analysis and was identified in the fractions 14 – 17 corresponding to ~18-20% acetonitrile.

The 421-containing fractions from all 4 HPLC runs were then recombined and run on a longer, shallower HPLC column of 10 to 35% acetonitrile over 60 minutes with fractions collected each minute. 1 mL of each fraction was transferred to HPLC vials

which were then analysed through HPLCMS/MS. The 421 target was identified in a single fraction at 15 minutes corresponding to ~19% acetonitrile.

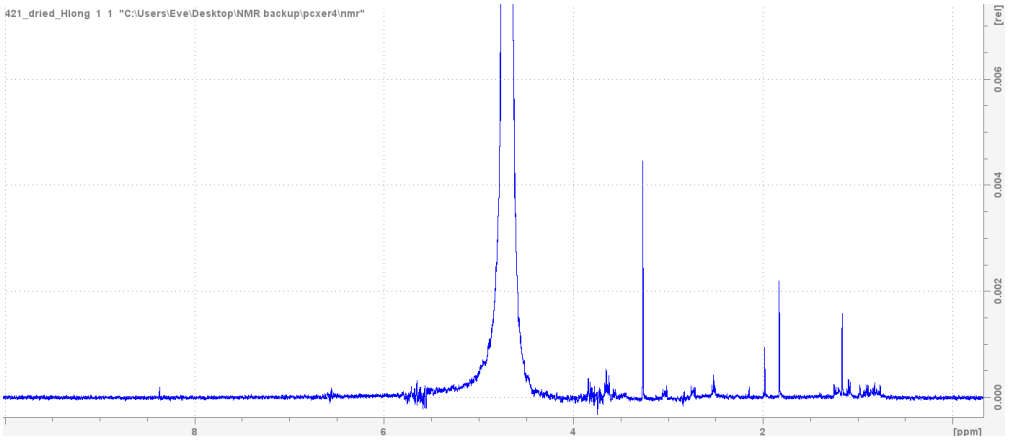


Figure 3.33 H^1 NMR of fraction 15 containing target 421

The TIC of fraction 15 showed that there are many non-target compounds present. NMR was run, and as can be seen in figure 3.33 in which the intensity is incredibly low. There is also very little similarity between the first and second NMR and it is not possible to identify which peaks are a result of target 421. Another purification method through gel permeation LPLC was attempted using a Sephadex LH-20 column and an isocratic 20% acetonitrile mobile phase. 39 x 10 mL UV active fractions were collected and 1 mL samples taken from each fraction and analysed using HPLCMS/MS analysis. Unfortunately, no target compound was observed through this analysis so the growth was repeated once more.

3.17 Target 421: Final purification attempt

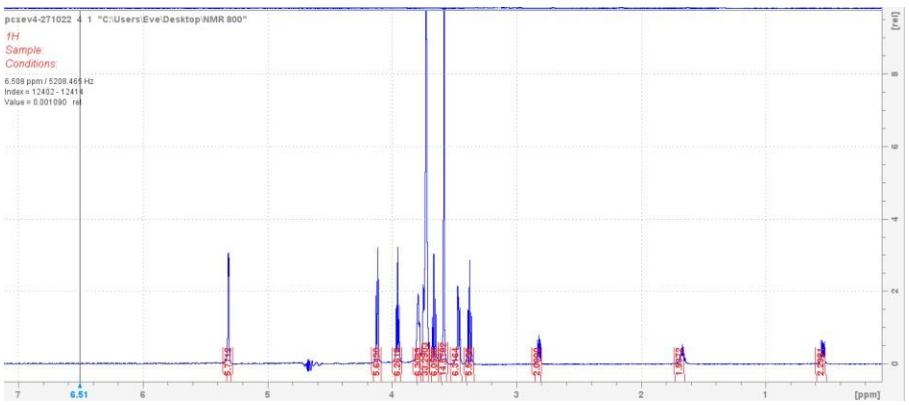


Figure 3.34 H^1 NMR of fraction containing target 421 on 800 MHz NMR

This time 6 L of culture was grown once more but instead of HPLC followed by SE, HPLC was repeated twice, once with the initial gradient and then again with a gradient of 0 to 20% acetonitrile over 20 minutes. The compound was then submitted to NMR analysis. This resulted in a cleaner spectrum as shown in Figure 3.34. However, it also appears dissimilar from the previous NMR spectra produced. This could indicate that different metabolites are produced in each 421-growth stage, which co-elute with the compound in a manner that is hard to predict and therefore difficult to solve. Additionally, this is still not a clear sample. Analyses of the 2D revealed the presence of more than one compound in the sample. In addition, integration of the peaks showed that the compound or compounds present here contained far more hydrogen environments than would be expected from a compound with a mass of 211 m/z. Therefore it is not possible to determine the structure of Target 421 from this data. A search of the predicted compound structure did not reveal any results, indicated that is a novel compound but there is not sufficient evidence to suggest that a compound of this structure produced by Coculturing of *Euglena gracilis* with *Streptomyces glaucescens*.

3.18 Chapter Three Conclusions

In this chapter, the impact of coculturing and culture volumes on the natural product diversity of algae and *Streptomyces* strains has been explored. Coculturing was found to significantly increase the diversity of natural products, highlighting its potential as a strategy to enhance the metabolic output of these organisms. Additionally, altering culture volumes was shown to have a profound effect on the metabolite profiles, with larger culture volumes leading to distinct differences in the compounds produced, possibly due to changes in resource availability and microbial competition.

A possible novel diketopiperazine (DKP) was identified based on its unique fragmentation patterns, though the structure could not be confirmed. The lack of a matching compound in online databases suggests that this compound may be a novel discovery, warranting further investigation into its potential biological activity and structural elucidation.

These findings underscore the complexity of microbial metabolism and the importance of culture conditions in determining the diversity and yield of natural products. The potential identification of novel compounds, particularly from underexplored species like *Euglena*, opens avenues for future research into their pharmacological applications. Further exploration into the impacts of coculturing

and culture volume variations offer an avenue to increase natural product diversity in *Euglena* without the need for genomic tools.

Chapter 4: Investigation of novel phosphonate compound in *Euglena gracilis*

4.1 Relevance of phosphonates as clinically relevant natural products

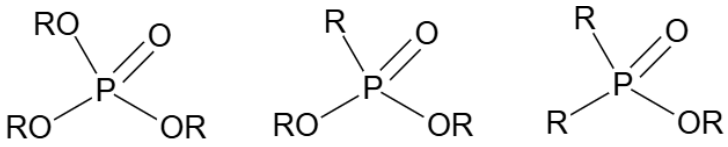


Figure 4.1 Structure of phosphate, phosphonate and phosphinate

Phosphonates are characterized by their direct carbon-phosphorus (C–P) bonds, distinguishing them from other organophosphorus compounds (Figure 4.1). Their biological significance stems from their structural mimicry of phosphate esters, carboxylic acids, and tetrahedral intermediates formed during carbonyl transformations, which are ubiquitous in metabolic and signalling pathways. This mimicry enables phosphonates to function as potent small-molecule inhibitors of various biological processes.[198] Their tetrahedral geometry, combined with their ability to engage in electrostatic interactions, hydrogen bonding, and metal ion complexation, makes them highly effective molecular inhibitors. Another key advantage of phosphonates as drug candidates is their stability in vivo. The enzymatic cleavage of C–P bonds is rare in most mammals, including humans, likely due to the prevalence of phosphate as a phosphorus source. As a result, phosphonate-based drugs exhibit resistance to metabolic degradation, enhancing their therapeutic potential.[199]

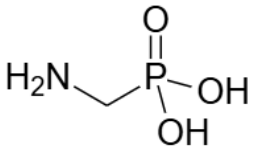


Figure 4.2 Structure of 2-aminoethylphosphonic acid

Phosphonic acids are often considered analogues of carboxylic acids, despite significant differences in geometry (tetrahedral phosphorus vs. planar carbon), acidity (phosphonic acids being more acidic), and steric bulk (phosphorus having a larger atomic radius), phosphonic acids are often being recognized as false substrates or inhibitors by enzymes and receptors. Phosphonic acids, such as α-aminoethylphosphonic acid (Figure 4.2), are generally non-toxic, water-soluble, and stable in vivo. They can form hydrogen bonds with proteins and chelate metal ions at enzyme active sites, further contributing to their inhibitory potential. The tetrahedral geometry of phosphorus also allows phosphonates to mimic the high- energy transition states of ester and amide hydrolysis, a property

that has inspired extensive research into their use as protease inhibitors.[200,201] Enzyme inhibitors targeting metalloproteases, serine proteases, and other hydrolytic enzymes have been developed based on phosphonate chemistry. Although a wealth of phosphonate-derived inhibitors has been investigated, relatively few have been approved for clinical use.[202]

4.2 Biosynthesis of phosphonates

Phosphonate biosynthesis has been extensively studied in bacterial models, with key biosynthetic enzymes characterized.[203] The early stages of phosphonate and phosphinate biosynthesis are highly conserved across species, most known phosphonates originate from the PEP biosynthetic pathway, with the exception of the K-26 family of *Actinomycetes*-derived tripeptides, which exhibit angiotensin-converting enzyme (ACE) inhibitory activity.[204]

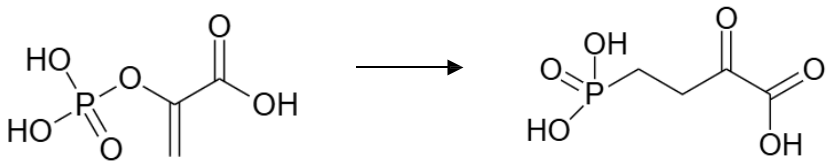


Figure 4.3 PepM catalyses the isomeration of PEP to PnPy

The PEP biosynthetic pathway begins with the isomerization of phosphoenolpyruvate (PE P) to phosphonopyruvate, catalyzed by phosphoenolpyruvate mutase (PepM). Due to its universal role in phosphonate biosynthesis, PepM serves as a reliable genetic marker for identifying phosphonate-producing organisms. Genome mining has revealed that approximately 5% of sequenced bacterial genomes, 5% of actinomycete isolates, and 7% of metagenomic datasets contain PepM homologs.[205]

To date, less than 30 naturally occurring small-molecule phosphonates have been identified.[206] Well-characterized examples include fosfomycin, a broad-spectrum antibiotic from *Streptomyces*;[207] phosphinothricin, a herbicide;[208] FR-900098, an antimalarial agent;[209] and rhizocticin, an antifungal tripeptide from *Bacillus subtilis*.[210]

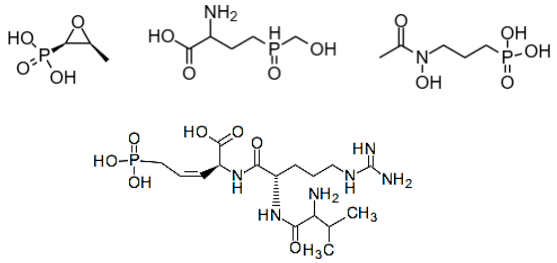


Figure 4.4 Top left to right: Fosfomycin, phosphinothricin, FR-900098. Bottom: rhizocticin

4.3 Phosphonate based drug examples

Several naturally occurring phosphonates have demonstrated significant pharmaceutical potential. The herbicide bialaphos, produced by *Streptomyces viridochromogenes* and *Streptomyces hygroscopicus*, contains the non-proteinogenic amino acid phosphinothricin (see Figure 4.4), which mimics the tetrahedral intermediate of glutamine synthetase, leading to ammonia accumulation and cell death. Although phosphinothricin itself is not efficiently transported across cell membranes, its incorporation into tripeptides facilitates uptake via oligopeptide permeases. Once inside the cell, peptide hydrolysis releases the active phosphinothricin, classifying it as a "Trojan horse" antibiotic. [203]A similar strategy is observed in dehydrophos, a tripeptide from *Streptomyces luridus*, which releases acetylphosphonate methyl ester inside target cells, inhibiting pyruvate oxidase and pyruvate dehydrogenase.[211] Antibiotics plumbemycins and rhizocticins, also lyse within the cell to release of 2-amino-5-phosphono-3-cis-pentenoic acid (APPA) which inhibits threonine synthase.[212]

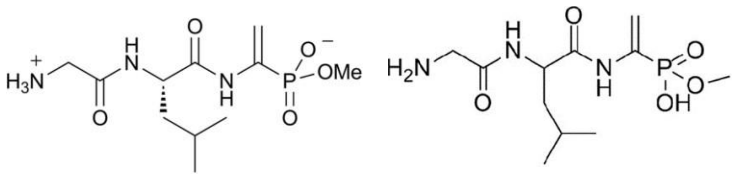


Figure 4.5 Structure of dehydrophos and plumbermycin A

Bisphosphonates, structural analogues of pyrophosphates, selectively bind to hydroxyapatite and inhibit osteoclast-mediated bone resorption. They are widely used in treating osteoporosis, Paget's disease, myeloma, and bone metastases and have also been explored for delivering chemotherapeutic agents to bone tissue.[213, 214]

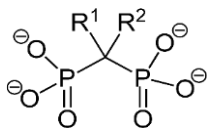


Figure 4.6 Functional group of bisphosphonates

Peptidyl phosphonate diaryl esters serve as covalent inhibitors of serine hydrolases, a key enzyme superfamily involved in metabolic regulation and neurodegenerative

diseases. These compounds act as irreversible inhibitors by phosphorylating the active-site serine residue, and their selectivity can be modulated through peptide modifications. This approach has been particularly successful for metalloproteases, whose active sites contain metal ions that interact with the phosphonate moiety, leading to the development of potent enzyme inhibitors with diverse applications, including antibiotics, anticancer drugs, herbicides, and antivirals.[215]

4.4 Potential for phosphonate production in *Euglena*

While bacterial phosphonate biosynthesis is well-characterized, phosphonate production in eukaryotic microorganisms remains largely unexplored. Genetic sequencing of *Euglena gracilis* strain CCAP1224/5Z by Ellis O'Neill identified two PepM homologs, see Appendix, indicating its potential to synthesize phosphonates.[140] Previously, only one phosphonate, 2-aminophosphonate, has been identified in *Euglena gracilis*.[153] The identification of two *PepM* homologs in *Euglena gracilis* CCAP1224/5Z provides the first genetic evidence that this organism possesses the potential for phosphonate biosynthesis beyond the previously identified 2-aminophosphonate. The presence of a second *PepM* enzyme strongly suggests the existence of an unidentified phosphonate, highlighting a previously unexplored metabolic capability in *E. gracilis*. By attempting to confirm and elucidate the structure of this unknown phosphonate, this Chapter aims contributes to a broader understanding of phosphonate metabolites in *E. gracilis* and explore methods for the isolation and structural determination of a novel phosphonate compound in *Euglena*.

4.5 Screening *Euglena gracilis* for the presence of phosphonate.

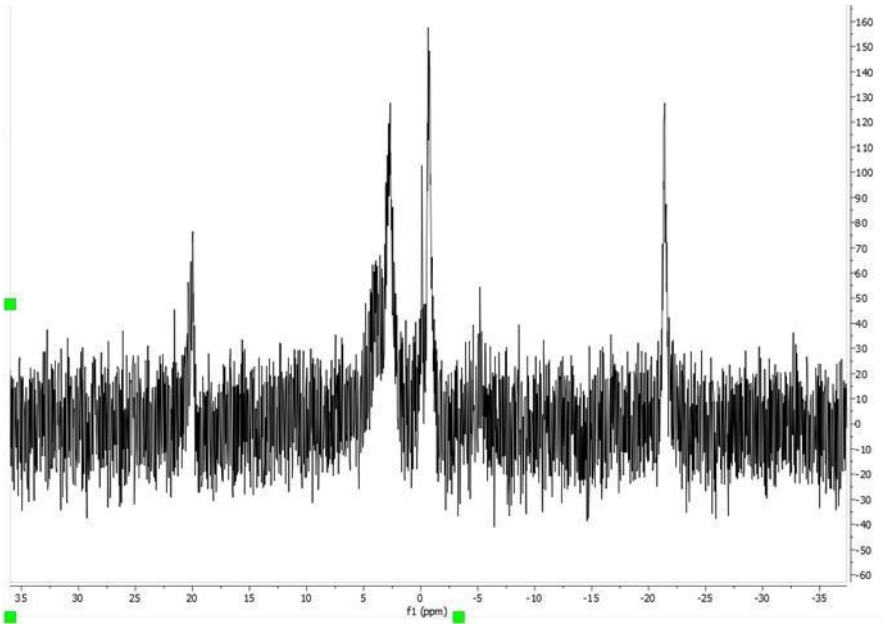


Figure 4.7 ^{31}P NMR spectra of supernatant of *Euglena gracilis* grown in EG:JM and glucose.

Four 1 litres of *Euglena gracilis* were cultured in EG:JM + glucose for 4 weeks, 1 x 5 mL was harvested from one of the litre cultures and centrifuged to remove biomass from the sample. 2-aminoethylphosphonate is contained within the *Euglena* cell,[216] to avoid reidentification of this known phosphonate only the supernatant was assayed for the presence of phosphonate peaks. Screening culture extracts using ^{31}P NMR provides a reliable method for phosphonate identification, as phosphonates exhibit distinct chemical shifts between 5–40 ppm.[217] However, it is crucial to differentiate phosphonates from cyclic phosphate diesters, which may fall within a similar range. Treatment of samples with alkaline phosphatase and phosphodiesterase will degrade these diesters which causes them to shift out of the phosphonate range.

The supernatant was concentrated down *in vacuo* to a brown residue and suspended in 666 ul of D_2O. When this extract was analysed through ^{31}P NMR a peak was observed at 20 ppm (see Figure 4.7). To ascertain whether this peak was the result of a phosphonate or a cyclic phosphate diester the ^{31}P NMR analysis was treated with calf intestinal alkaline phosphatase (CIAP) addition incubation at 37 °C. After treatment the peak at 20 ppm was still observed, excluding its identification as a cyclic phosphate diester in the extract.

4.6 Activated charcoal for the initial purification of *E. gracilis* cultures

After the phosphonate compound was confirmed, attempts were made to isolate the natural product so it could be structurally elucidated. The culture supernatants were initially cleaned through separation down an activated charcoal column. Activated carbon is already established as efficient at removing low concentrations of pollutants from wastewater, including antibiotics.[218] Predominantly, molecules physically interact with the activated charcoal through adsorption, and to a lesser extent Van der Waals forces and hydrophobic interactions. The high number and heterogeneity of pores, comprising of micropores, mesopores, and macropores, alongside the incredibly large surface area of activated charcoal, with one gram of activated carbon possessing a surface area of over 3,000 square meters, makes it suitable for removing the large and diverse number of molecules present in culture supernatant.[219]

The supernatant was then concentrated down *in vacuo* before it was washed down a charcoal column with increasing ethanol concentration from 0 – 100% in 25% increments. The fractions were dried down and reconstituted in D_2O for P^{31} NMR analysis (Figure 4.8). The peak at 20 ppm was observed in the 75% water 25% ethanol fraction, alongside phosphate peaks. H^1 NMR and C^{13} NMR showed the presence of a multitude of peaks, and the fraction was taken forward to further purification.

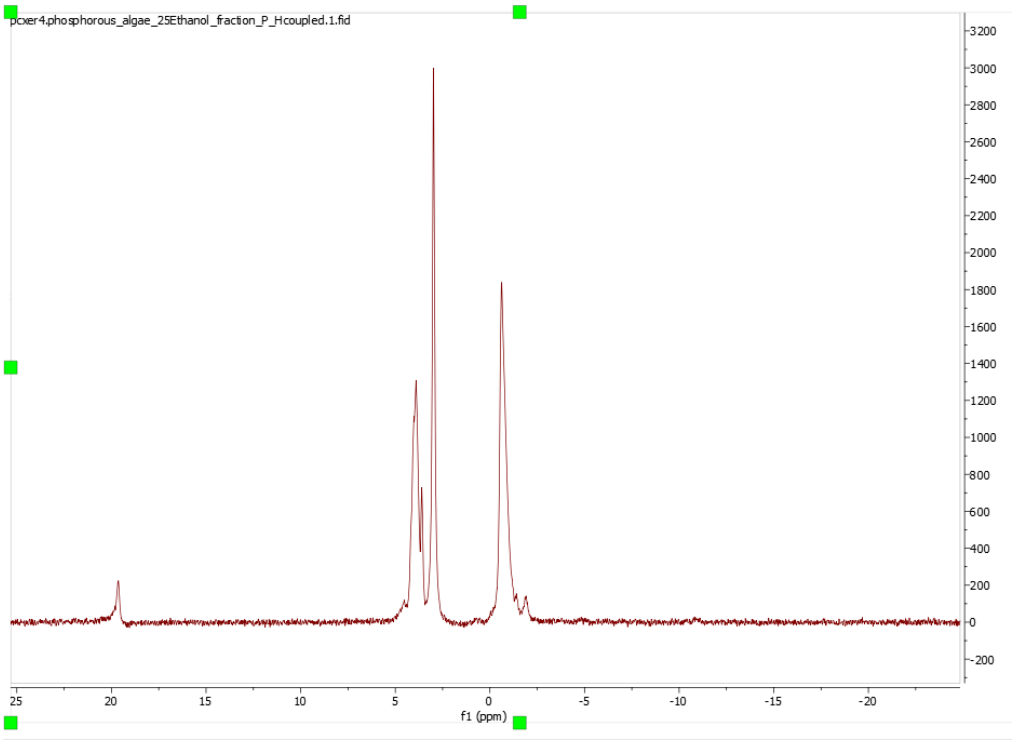


Figure 4.8 ^{31}P NMR spectra of 25% ethanol fraction

4.7 Further purification through ion exchange chromatography

In order to further purify the mixture, with particular focus on the separation of the phosphates from the phosphonates, the 25% ethanol fraction was then split into two halves and run down two 1 mL ion exchange columns: one anion exchange column and one cation exchange column. The anion exchange column was functionalised with a sulphate ion (SP) and the cation exchange column with quaternary amine (Q), both possessing strong charges at a range of pH values.

Due to the presence of an extra oxygen bond in phosphates than phosphonates, they can become more negatively charged so should have a higher affinity to the anion exchange column, and a lesser affinity to the cation exchange column. The columns were both eluted with three column volumes of 10 mM ammonium carbonate solution and 100 mM ammonium carbonate solution. Phosphorous NMRs were then ran on the fractions to assess the success of separation. Both phosphate and phosphonate peaks were observed in the first wash, 10mM ammonium carbonate solution, for both columns indicating no separation had occurred. This was thought to be perhaps due to an oversaturation of the column, as the material was highly concentrated and complex and the column were only 1 mL in volume - a larger column was selected to attempt ion exchange based separation again.

4.71 Further purification through ion exchange chromatography: DEAE chromatography

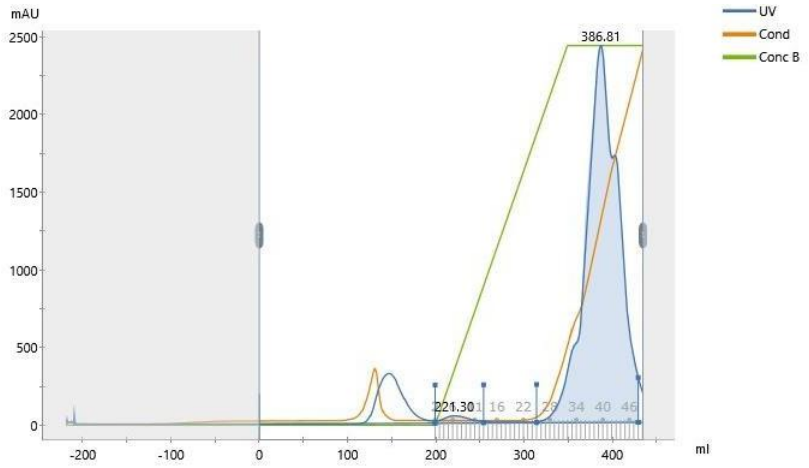


Figure 4.9 AKTA chromatogram of DEAE column

The fractions were injected onto a DEAE column, which possessed a column volume of 620 mL, with a 0 to 50 mM ammonium carbonate mobile phase over 30 minutes. The fractions were collected at 1 minute intervals, the fractions were freeze dried and reconstituted in D_2O and the presence of the phosphonate detected using ^{31}P NMR. The phosphonate was found in 3 fractions: 6, 7, and 8 minutes corresponding

to 9 – 13% 50 mM ammonium carbonate. This was not successful in removing all the phosphate compounds, as can be seen in the NMR, where the other phosphorus-containing compounds remain present at -2-5ppm.

Although there were still phosphate environments present in the phosphonate containing fractions, 6, 7 and 8, there is some reduction in the number observed compared to the fraction from the charcoal column, particularly in fraction 8. Additionally, NMRs of fractions eluting later off the DEAE column (such as fraction 41) revealed the presence of phosphate containing compounds eluting off the column much later, revealing that at least some of the phosphate containing compounds were being retained by the DEAE column and thus removed from the phosphonate.

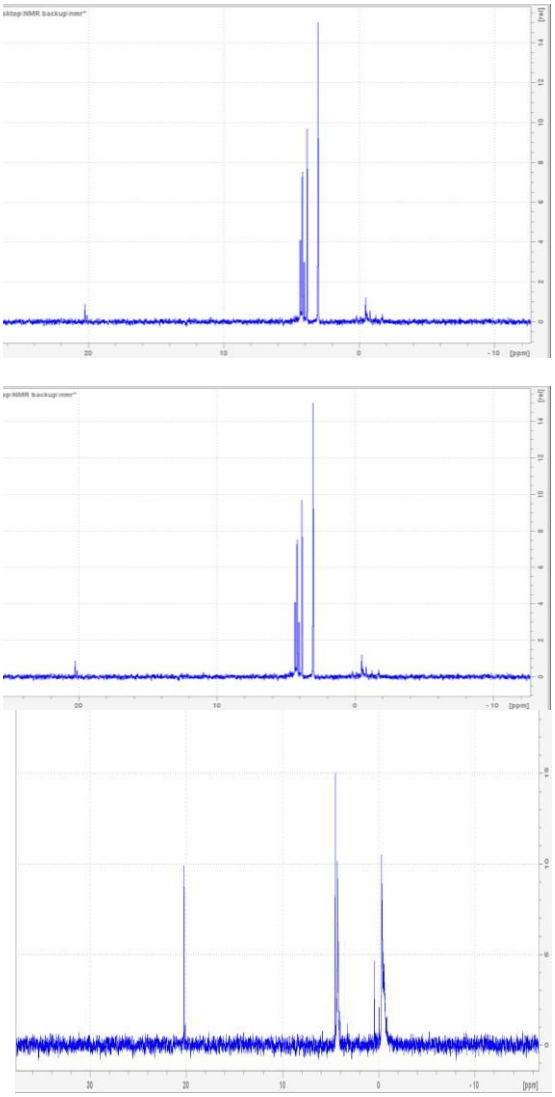


Figure 4.10 ^{31}P NMR spectra of DEAE fractions 6 (top), 7 (middle) and 8 (bottom)

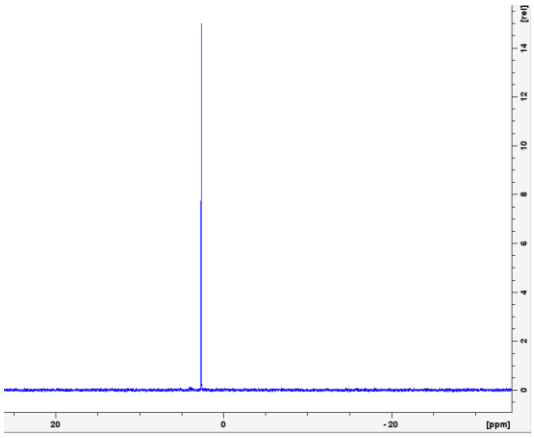


Figure 4.11 ^{31}P NMR spectra of fraction 41.

The phosphonate containing fractions corresponded to the small peak in UV (221 mAU) as can been seen in the chromatogram in Figure 4.12. As neither the P-C or P-O bond absorb UV light at 290 nm the peak is not a consequence of these bonds, nor could it be certain if the absorbance is a result of a chromophore comprising part of the phosphonate containing molecule, or due to unrelated compounds eluting at the same time point. In order to determine any certainty on the structure of this unknown phosphonate, it needed to be purified further. For this purpose, the fractions 6, 7 and 8 were combined, although fraction eight displayed a purer quality of phosphonate, all fractions were combined in order to attempt purification on the highest concentration of phosphonate possible.

Two different methods were attempted for further purification of the mixture. One half was resubmitted onto the DEAE column, but this time the mobile phase was adjusted to from ammonium carbonate (0 to 50mM) to NaCl (0 to 1M) over 30 minutes. This adjustment was based on the higher affinity for Cl- ions to DEAE than carbonate ions, which could possibly improve separation between the phosphates and phosphonates by competing more strongly than the phosphonates for the diethylaminoethyl functional group. However, the high salt concentration preventing locking onto the fractions by the NMR, making validation of this impossible. An attempt was made to remove the salt using a desalting column, but NMR on the samples did not reveal any phosphonate. This could be because the compound was too similar in size to salt and remained in the salt containing fraction, as the column resolution of the desalting column was too high for adequate separation of small molecules, being designed for the desalting of large proteins.

The other half of the phosphonate containing mixture was applied to an Fe column. In order to generate a Fe column, an IMAC Sepharose High Performance column was stripped with EDTA and recharged with Fe by washing it with 0.1 M

$FeCl_3$. In phosphate there is one more oxygen atom available to bond to Fe(III), meaning it should bind more strongly than phosphonate. 0-0.1 M Tris buffer was run down the column in attempt to displace the phosphonates. The three hydroxyls in tris buffer should have higher affinity for Fe(III) than phosphonate in the same manner as the phosphate, but no phosphonate was recovered. It is possible that Tris was not strong enough at 0.1 M to fully elute the phosphonate - although five column volumes were used it may have been better to also use a higher molarity buffer. At this point all phosphonate containing material had been lost, and the growth was repeated, this time at 6L to provide more material for purification.

4.8 Purification through size exclusion chromatography

The same process of culturing and purifying with activated charcoal was used as previously in the Chapter. The fraction containing phosphonate was once again found in the 25% ethanol fraction. As previous attempts at separating the phosphonate from the phosphate using ion exchange had proved ineffective, this time an attempt was made using size exclusion chromatography. For this an LH-20 Sephadex column was used as this had been designed for the separation of natural products and low-weight peptides, with an exclusion limit of 5000 Da and a column volume of 130 mL. An isocratic mobile phase of 20% ethanol was used, and 5 mL fractions were collected. The fractions were lyophilized and phosphorous NMR run on every UV-active peak until the phosphonate was identified in fraction 25, the second UV peak observed.

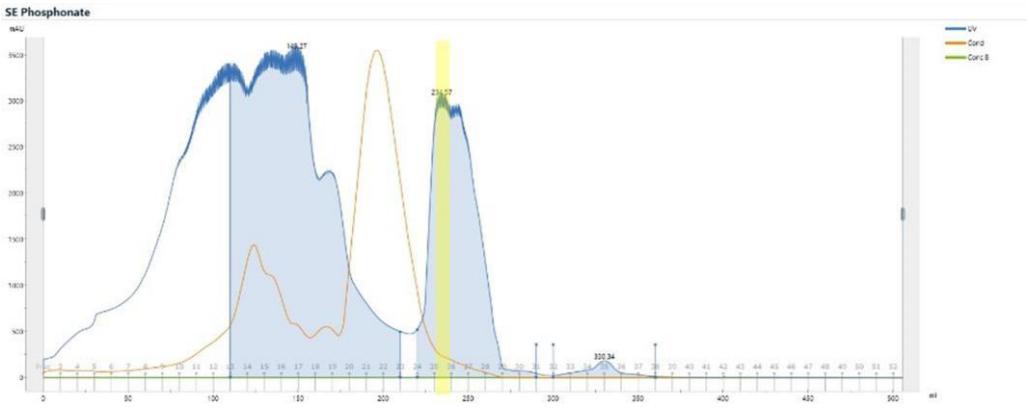


Figure 4.12 Chromatogram of LH-20 SE column with fraction 25 highlighted in yellow

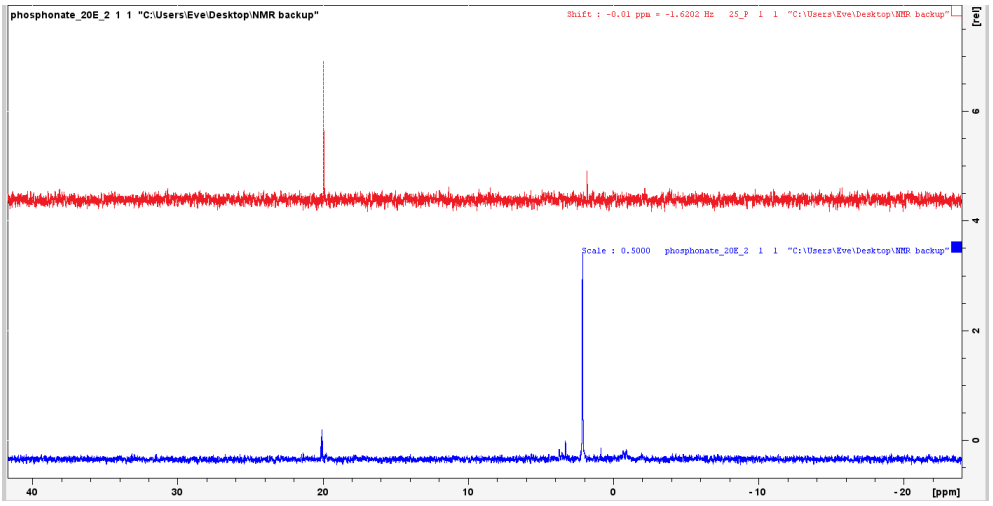


Figure 4.13 Phosphorous NMR of 25% ethanol fraction from charcoal column (bottom) and fraction 25 of SE column (top)

The size exclusion column proved effective at separating the phosphates from the phosphonates, as this was the first time in the isolation process that the phosphonate environments were relatively more intense than the phosphate environments in the NMR (Figure 4.13).

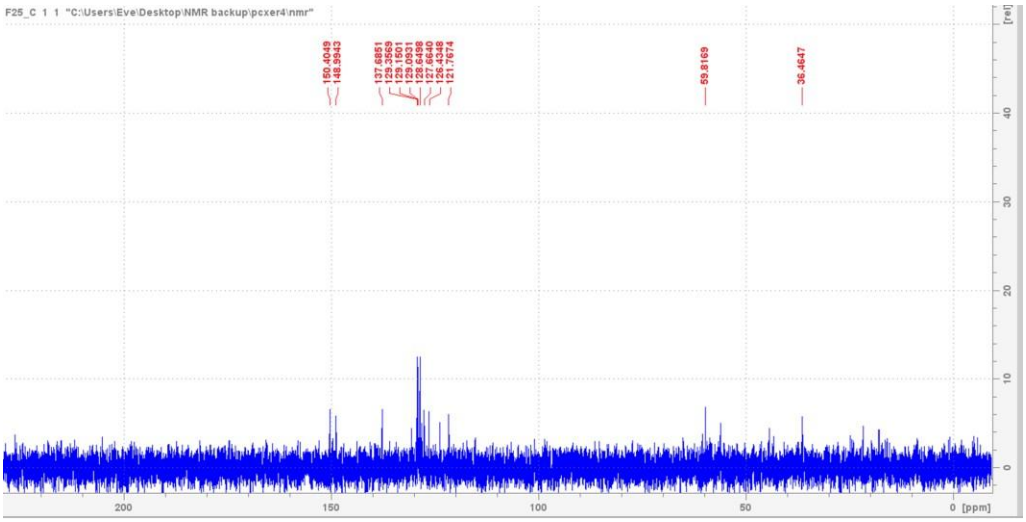


Figure 4.14 Carbon NMR of fraction 25 of SE column

Proton and carbon NMRs were then run on this fraction to assess the level of purity. In the carbon NMR there are twelve environments, ten of which are within the 120 – 150 region indicating they are all sp^2 hybridised, with the six in 121 – 136 most likely part of an aromatic ring, which would be contributing the UV absorbance seen in the chromatogram. The other two are in the sp^3 hybridised region, with the peak at 59.8169ppm possibly due to an alcohol. Although not quantitative, the high noise

to peak ratio in the carbon NMR is a consequence of the low concentration of compound in the sample. Additionally, these peaks could not be a consequence of the phosphonate compound but of a contaminant, with the phosphonate compound being present in too low quantities to be observed at all.

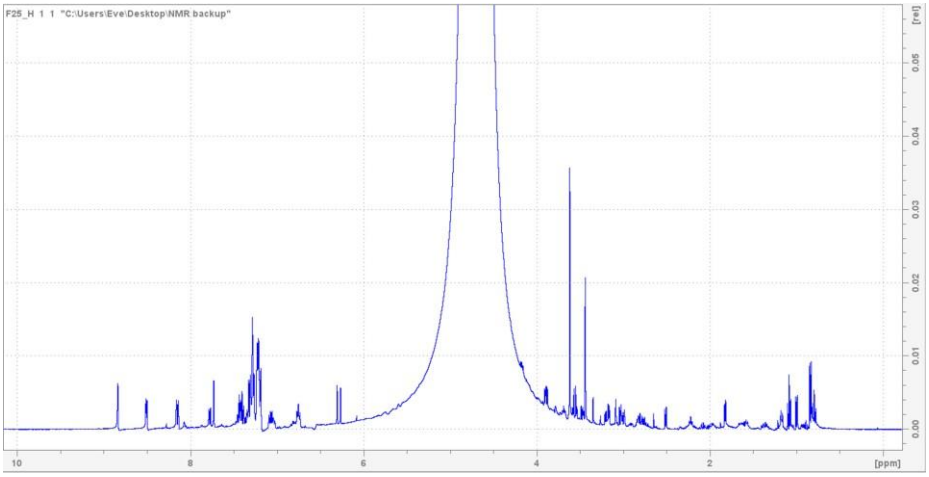


Figure 4.15 Proton NMR of fraction 25 of SE column

The proton NMR contains at least 40 environments (see figure 4.15), not accounting for those masked by the water peak. There are hydrogens in the aromatic range which matches the observations in the carbon NMR, and there are no peaks past 8 ppm indicating the absence of aldehydes or esters in the mix. However, the NMR is not clean enough to allow for structural elucidation and as such further purification steps were taken.

4.81 Further purification through use of a Dowex ion exchange column

In order to remove the phosphate from the phosphonate, anion exchange was selected as the chromatography method this time. However, instead of the 1 mL Q-trap column used previous, a column was made using 40 g Dowex. This anion exchange column has the same functionality as the Q-trap column, a quaternary amine, but it was thought that the larger surface area should improve the separation. However, despite washing with water, 50 mM ammonium carbonate and 500 mM ammonium carbonate the phosphonate was not recovered. This was thought to be possibly due to the ammonium carbonate wash lacking the strength to fully elute the phosphonate from the quaternary amines. An attempt to wash the column with acetic acid buffer in order to promote dissociation of the phosphonate also yielded no phosphonate. The fractions were all combined and dried down to observe if perhaps the phosphonate was simply diluted at too low a concentration across the different fractions to be observed using NMR, but it was not observed again.

4.9 Purification through SE and semi-prep HPLC

The culturing and purification with activated charcoal was repeated once more. The 25% ethanol fraction was submitted to size exclusion chromatography according to the same methods described previously in the chapter.

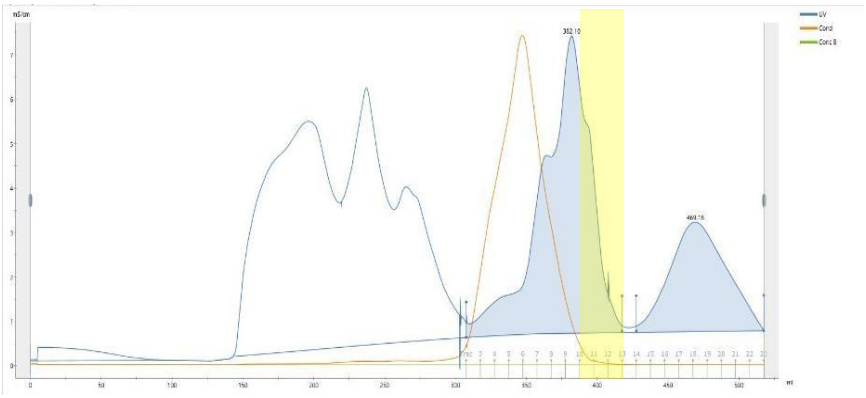


Figure 4.16 Size exclusion chromatogram of 25% ethanol fraction

This time the phosphonate-containing fractions were observed in three fractions coming off the SE column, which could indicate the column was being impacted by residual build up from the complex mixtures being run on it or possibly higher concentration of phosphonate being present.

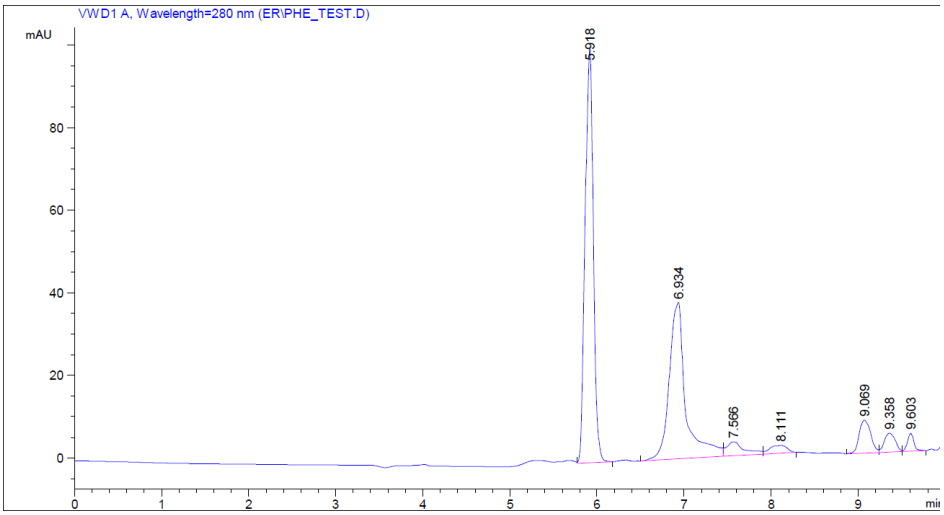


Figure 4.17 Semi-prep HPLC trace of phosphonate containing fractions from SEC

After SE, a method other than ion-exchange was required for secondary purification as this had proved to be ineffective thus far. Therefore, the phosphonate-containing fractions were combined and submitted to purification through reverse-phase HPLC. As phosphonates are polar molecules, a high-water content mobile phase was selected consisting of 0 – 50% B over 10 minutes, where B is 0.1% formic acid in

acetonitrile. This was anticipated to work as a preliminary experiment to assess where the phosphonate eluted and then perform a secondary HPLC purification step with a shallower gradient around this point. The prewash, the runoff and fractions every minute were collected, dried down and ran through phosphorus NMR but the phosphonate was not recovered.

The UV peaks indicates the present of at least seven chromophores in the mixture at this point, showing the presence of at least seven other compounds in this stage of the purification process. These began eluting at six minutes which is equivalent to 30% acetonitrile - it was anticipated the phosphonate containing fraction would be comprised of more water-soluble molecules that would elute much earlier in the gradient. It is possible that they are simply not UV-active. Regardless, this method proved ineffective and culturing was started again.

4.10 Purification through SE and DEAE

The consistent loss of observable phosphonate containing peaks after repeated purification attempts could be attributed to degradation of the phosphonate compound. Therefore, attempts were made to streamline the purification process. In order to negate the lengthy process of eluting through a charcoal column, the supernatant was this time centrifuged with activated charcoal. In this case the phosphonate was not recovered until the 100% ethanol fraction, and a higher volume of ethanol had to be used to dislodge the phosphonate. This likely is due to the higher quantity of activated charcoal used when centrifuging, which was intended to compensate for the reduction in exposure time of the material to the activated carbon caused by the switch from column to centrifuge but too much activated charcoal was used.

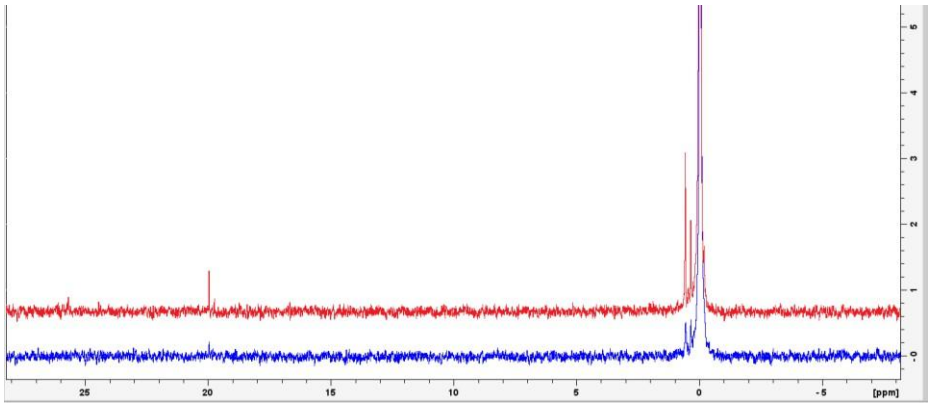


Figure 4.18 100% ethanol fractions, first wash shown in red and second in blue

In order to estimate the concentration of phosphonate produced from 6 L of culture, a standard of methyl phosphonic acid was used. For a molarity of 78 mM, the methylphosphonic integrates relative to the phosphonate peak at a ratio of 1:0.0025, making the concentration of phosphonate 0.2 mM (see Figure 4.19). The structure of methyl phosphonic acid is simple and known, it would be simple to account for that when analysing NMR of an otherwise purified phosphonate compound if it can not be removed.

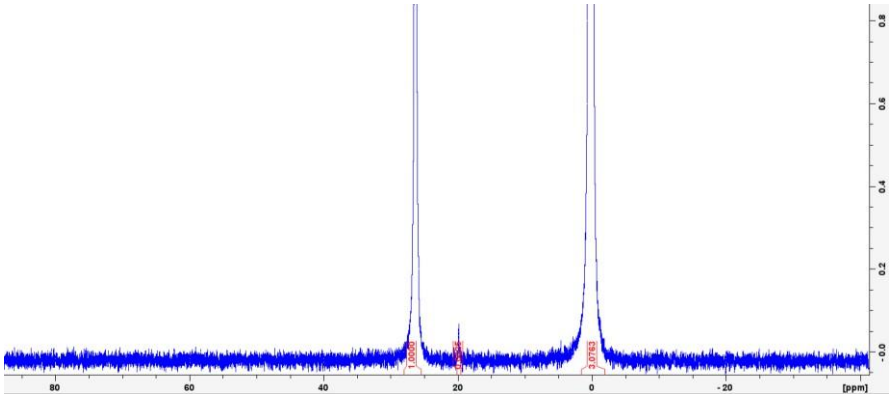


Figure 4.19 Phosphorous NMR with standard of methyl phosphonic acid shown at 26 ppm

This highlights one of the main difficulties in isolating this phosphonate, it is produced in very small amounts, and is difficult to separate from the chemically similar phosphate which are present in much higher amounts, a concentration of 2893 mM in this sample.

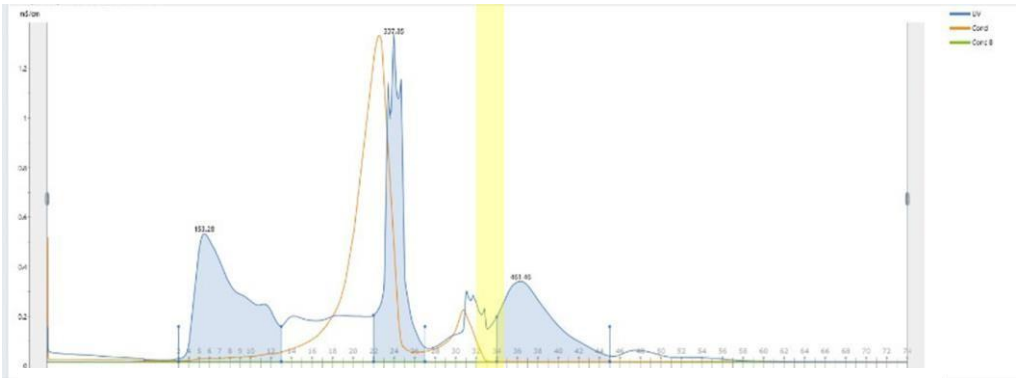


Figure 4.20 Size exclusion chromatogram of 100% ethanol fraction with fraction 33 highlighted in yellow

The 100% ethanol fraction was then dried down and submitted to SE, in this case it eluted in fractions 32, and 33. The ratio of methyl phosphonic acid to phosphonate decreased from 400 times as much to 14 times as much in fraction 33. However, it has not been separated from the phosphonate-containing compounds, which remained at 40 time as much.

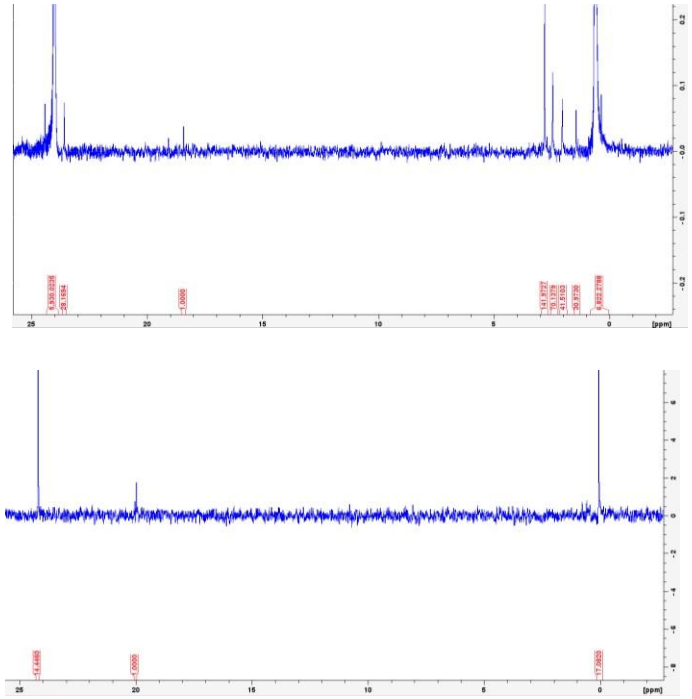


Figure 4.21 Phosphorous NMR of fraction 32 (top) and 33 (bottom).

4.11 Mass spectrometry analyses

Due to previous secondary purification attempts resulting in the loss of the phosphonate-containing compound, combined with the already low concentration of phosphonate in the sample, further purification did not appear to be a viable next step. Instead, molecular networking was explored as an alternative approach to aid in structural determination. By constructing a molecular network using HPLC-MS/MS spectra from Fosfomycin and methyl phosphonic acid standards, alongside fractions 32 and 33, it was hypothesized that clustering and fragmentation patterns could provide insights into the molecular properties of the unknown phosphonate.

To prepare for analysis, the fractions were dried down and reconstituted in methanol. Over the following days, a method was optimized for detecting Fosfomycin in negative mode. When complete, the fractions were submitted for phosphorus NMR to confirm retention of the phosphonate peak; however, it was found to be absent. The repeated “loss” of the phosphonate peak throughout the purification process initially appeared to be due to the challenges of handling trace concentrations of phosphonate within complex mixtures. However, this supports evidence of a degradation of the phosphonate compound.

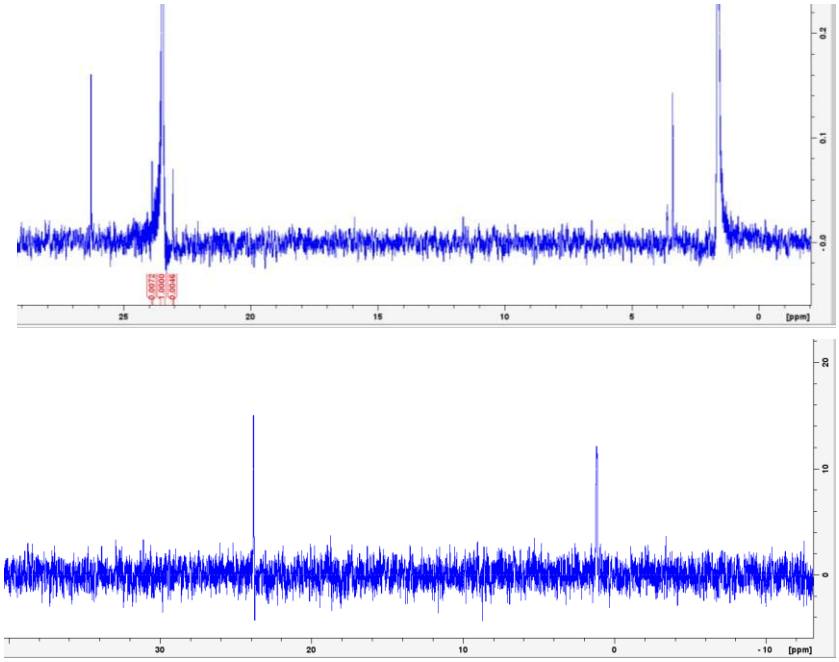


Figure 4.22 Phosphorous NMR of fraction 32 (top) and 33 (bottom). After 5 days spent at 4 °C in methanol. The peak at 20ppm present in Figure 4.26 is no longer observed.

4.12 NMR analysis of impure phosphonate containing fractions

To gain structural insights into the phosphonate compound, the phosphonate-containing fraction was analysed using high-field 800 MHz NMR following the charcoal purification step. Phosphorus-decoupling experiments were employed to help identify peaks associated with the phosphonate moiety. Additionally, a modification of growth conditions—culturing for two weeks in EG:JM with 30 mM glutamic acid followed by two weeks in EG:JM supplemented with glucose—resulted in a significant increase in biomass yield, from less than 1 g/L to over 2 g/L. This increase correlated with an enhanced production of phosphonate, enabling the isolation of a fraction with a higher relative concentration of the phosphonate compound (Figure 4.23).

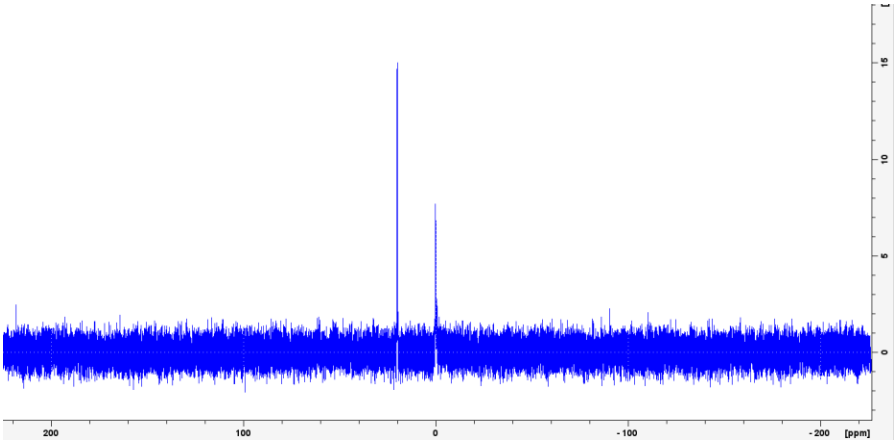


Figure 4.23 Phosphorous NMR of 25% ethanol fraction.

Despite the increased phosphonate concentration, the presence of multiple overlapping signals in the NMR spectra made peak assignment challenging. However, comparison of the ^{13}C-^{31}P coupled and decoupled 1D NMR experiments revealed two splitting patterns that may be indicative of a C–P bond.

For phosphonates, the expected J-values of P–C coupling can vary significantly depending on the nature of the bond, hybridization state, and distance between phosphorus and carbon atoms. The strongest P–C coupling occurs when the phosphorus is directly bonded to a carbon atom, resulting in a one-bond coupling constant (^{1}J_PC). In phosphonates, these values typically range between 100–150 Hz, with variations depending on the electronic environment. For example, alkyl phosphonates generally exhibit ^{1}J_PC values near 120–140 Hz, whereas aryl phosphonates tend to fall within the 110–130 Hz range. The strength of the coupling is influenced by the hybridization of the carbon; sp²-hybridized carbons in aromatic systems may show slightly lower J-values than sp³-hybridized carbons in alkyl groups.[220]

Two-Bond P–C Coupling (^{2}J_PC) and Three-Bond P–C Coupling (^{3}J_PC) When the phosphorus and carbon atoms are separated by one or more intervening bonds, the coupling constants decrease significantly. Two-bond (^{2}J_PC) couplings, where phosphorus and carbon are separated by one non-hydrogen atom (e.g., P–O–C or P–C–C), typically fall within the 5–20 Hz range. The precise value depends on bond angles, electronic effects, and conformational flexibility. Three-bond (^{3}J_PC) couplings, where phosphorus and carbon are separated by two intervening atoms, generally exhibit smaller values, often between 2–10 Hz.

The presence of electronegative substituents (e.g., hydroxyl or ester groups) can influence J_PC values by altering electron density and bond polarization. For instance, in phosphonates containing a P–C–O linkage, the coupling constants may be slightly lower than in purely alkyl phosphonates due to electron-withdrawing effects. Additionally, hydrogen bonding and solvation effects can further modulate the observed J-values in solution-state NMR.[221]

The first splitting pattern identified is a peak at 29 ppm, which indicates that this carbon is in a predominantly alkyl environment (Figure 4.24). This peak exhibited doublet splitting in the ^{13}C-^{31}P coupled spectrum, which was absent in the decoupled spectrum. This suggests that the peak corresponds to a carbon directly influenced by phosphorus coupling. The J-value of these peaks was calculated at 4.62 Hz, the low J-value could indicate that this carbon is three bonds away from the phosphorous.

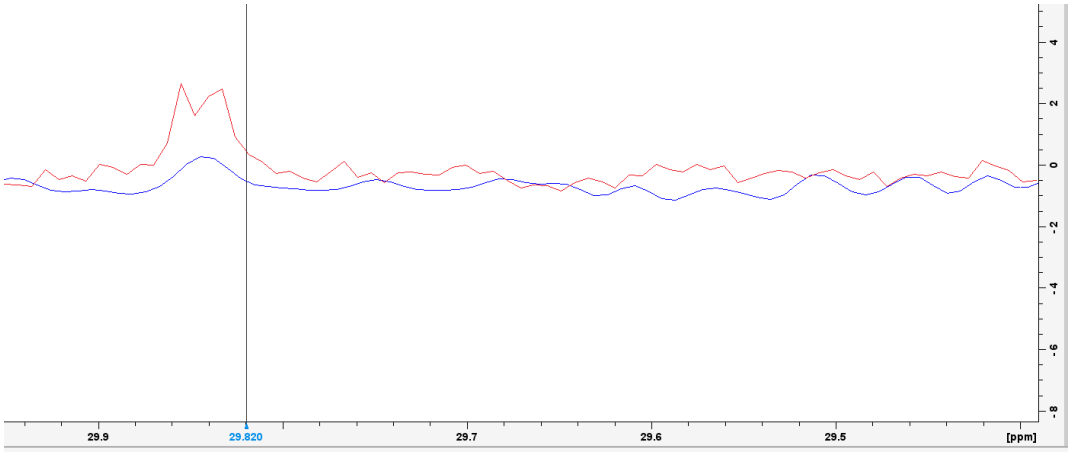


Figure 4.24 Section (29.5-29.9) of overlaid C^{13} NMR of carbon-phosphorous decoupled (blue) and carbon-phosphorous coupled (red).

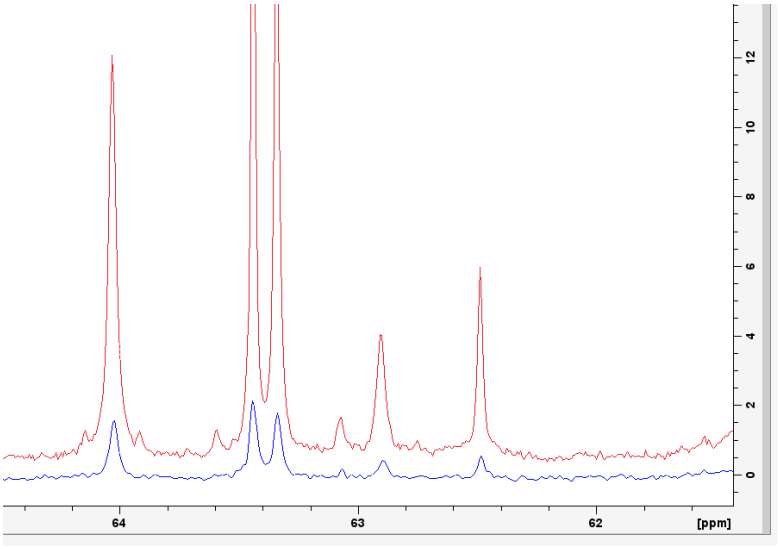


Figure 4.25 Section (62-64) of overlaid C^{13} NMR of carbon-phosphorous decoupled (blue) and carbon-phosphorous coupled (red).

There is also the presence of peaks observed in the coupled P-C experiment that are not observed in the decoupled experiment at 64 ppm (Figure 4.25). The larger singlet present at 64 ppm is unlikely to be part of the same molecule that produces the two small peaks either side of it in the coupled experiment due to the difference in integration. This could indicate a singlet peak from the phosphonate compound is hidden by the larger peak, only becoming visible in the coupled experiment when it splits into a doublet. The coupling value between these peaks is 80 Hz which is close to the 100Hz lower limit range for coupling 2 bonds away from phosphorous. The reduced J-value could be explained by the presence of electronegative oxygen bound to this carbon. C-OH peaks are expected around 70 ppm in carbon NMR. Therefore, this suggests the presence of a hydroxy group in the phosphonate molecule which is closer to the phosphorous atom than the alkyl at 28.8 ppm. No further conclusions could be drawn from the HMBC, HSQC or TOCSY experiments as the mixture is too contaminated. No peaks were observed in the 2D ^{13}C-^{31}P, which is most likely due to the low concentration of phosphonate in this sample.

4.13 Conclusions of Chapter Four

This study confirmed the presence of a novel phosphonate compound in *Euglena gracilis*, however, the compound was observed to be produced in exceptionally low amounts, which, combined with its apparent instability, posed significant challenges to its purification. Multiple purification strategies were pursued, yet complete isolation remained unachievable due to these limitations. The instability of the compound, particularly during the purification process, suggests that degradation may be influenced by environmental factors such as pH.

Future work will focus on conducting time-course stability studies of phosphonate-containing fractions across a range of pH levels to determine the conditions under which degradation occurs most slowly. Identifying an optimal pH could inform adjustments to the purification process, improving yield and stability. If a stable fraction can be obtained, a molecular network could be developed according to the parameters outlined in Section 4.11, potentially facilitating structural elucidation through comparative fragmentation analysis.

In addition to optimizing purification conditions, genetic approaches could be explored to enhance the production of the phosphonate compound. One strategy would be to increase the native expression of PepM genes in *Euglena gracilis* through targeted gene regulation techniques such as promoter engineering to drive higher transcriptional activity of PepM. Additionally, CRISPR knock outs of PepM genes could be analysed to see if the phosphonate compound can no longer be observed, providing evidence of a direct link between these genes and the unknown phosphonate.

.

Chapter Five: Are Mycosporine-like amino acids (MAAs) produced by *Euglena gracilis?*

5.1 Mycosporine-Like Amino Acids: Structure, Biosynthesis, and Applications

Mycosporine-like amino acids (MAAs) are small, secondary metabolites characterized by their unique cyclohexenone or cyclohexenimine chromophore conjugated to an amino acid or amino alcohol group.[222] These compounds serve as natural sunscreens due to their strong absorption of ultraviolet (UV) radiation in the range of 310–365 nm, effectively protecting cells from UV-induced damage. In addition to their photoprotective function, MAAs exhibit antioxidant, anti-inflammatory, and anti-carcinogenic properties, making them attractive candidates for biotechnological and pharmaceutical applications.[223]

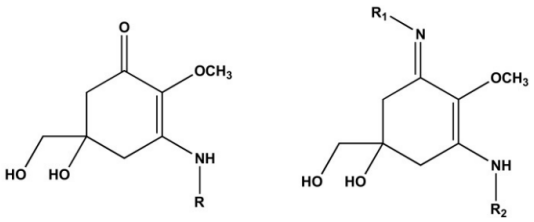


Figure 5.1 Basic structure of mycosporine (left) and mycosporine-like amino acid (right)

The core structure of MAAs consists of a cyclohexenone or cyclohexenimine ring system, which enables efficient UV absorption. The substitution pattern of this core structure varies, with different MAAs possessing additional amino acids or amino alcohol groups, leading to differences in spectral properties and biological activity.[224] Examples of common MAAs include mycosporine-glycine, shinorine, porphyra-334, palythine, and asterina-330 (Figure 5.2). The precise structure determines their stability and capacity to scavenge reactive oxygen species (ROS), contributing to their biological significance beyond UV protection.

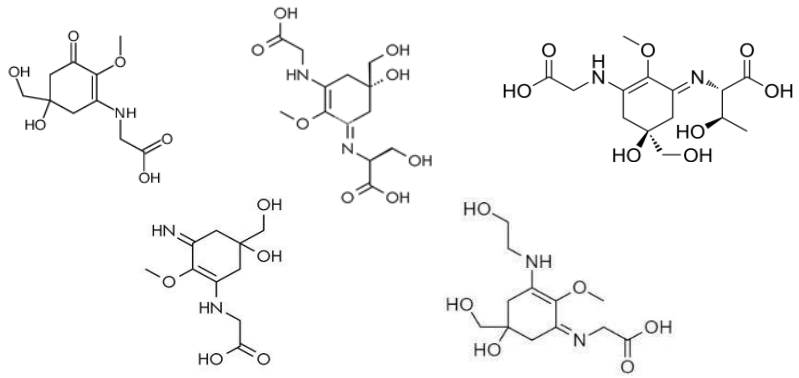


Figure 5.2 Top, left to right: mycosporine-glycine, shinorine, porphyra-334.
Bottom: palythine, and asterina-330

MAAs are synthesized via the shikimate pathway, an essential metabolic route responsible for the production of aromatic amino acids and secondary metabolites in microorganisms, fungi, and plants. The key precursor in MAA biosynthesis is 3-dehydroquinate, which undergoes enzymatic modifications, including cyclization and amination, to form the cyclohexenone or cyclohexenimine core. In cyanobacteria and algae, the biosynthetic pathway involves an O-methyltransferase and a non-ribosomal peptide synthetase (NRPS) system, which incorporates different amino acid moieties to diversify MAA structures.[225]

Genomic studies have identified key genes involved in MAA biosynthesis, such as those encoding dehydroquinate synthase (DHQS) and O-methyltransferases, which regulate the enzymatic steps required for MAA formation. The presence of these genes in various marine organisms suggests that horizontal gene transfer may have played a role in the evolutionary distribution of MAAs across different taxa.[226]

5.2 Distribution of Mycosporine-Like Amino Acids in Nature

MAAs are widely distributed in marine and freshwater organisms, particularly in cyanobacteria, red algae, dinoflagellates, and some fungi. Their presence is strongly correlated with exposure to high UV radiation, indicating an adaptive response to protect cells from photodamage. In marine ecosystems, MAAs are abundant in coral reef organisms, including symbiotic dinoflagellates (Symbiodinium spp.), which transfer these compounds to their coral hosts. Additionally, various microalgae and macroalgae accumulate MAAs to shield themselves from UV stress, particularly in intertidal and shallow-water habitats.[222]

Among algal species, red algae (Rhodophyta) are particularly rich sources of MAAs. Species such as Porphyra, Gracilaria, and Palmaria produce high concentrations of MAAs, including porphyra-334 and shinorine. Green algae (Chlorophyta) and brown algae (Phaeophyceae) also contain MAAs, though at lower concentrations. The production of MAAs in algae is often regulated by environmental factors such as light intensity, UV exposure, and nutrient availability. Studies have shown that algal cells increase MAA synthesis in response to elevated UV radiation, supporting their role in photoprotection.[227]

5.3 Identification and Isolation of Mycosporine-Like Amino Acids from Algae

The extraction and identification of MAAs from algae typically involve aqueous or methanolic extraction followed by chromatographic separation. The most common analytical techniques include high-performance liquid chromatography (HPLC) coupled with photodiode array detection (PDA) and mass spectrometry (MS). These methods allow for the precise quantification and structural elucidation of MAAs based on their characteristic UV absorption maxima (310–365 nm) and mass-to-charge ratios.

Nuclear magnetic resonance (NMR) spectroscopy is also employed for structural characterization, particularly for novel MAAs with unique substitution patterns.[228]

5.4 MAAs as natural products

Due to their potent UV-absorbing and antioxidant properties, MAAs are increasingly utilized in skincare formulations as natural photoprotective agents. Unlike synthetic UV filters, which may cause environmental and health concerns, MAAs are biodegradable, non-toxic, and provide broad-spectrum UV protection. Several commercial sunscreen products now incorporate algal-derived MAAs to enhance photoprotection while reducing oxidative stress and skin aging.[229]

Beyond photoprotection, MAAs exhibit promising anti-cancer properties due to their ability to neutralize ROS and modulate cellular stress responses. Oxidative stress is a major contributor to DNA damage and carcinogenesis, and MAAs have been shown to mitigate these effects by acting as efficient antioxidants. Recent studies suggest that MAAs can inhibit UV-induced apoptosis in keratinocytes, reducing the risk of skin cancer development.[230]

Furthermore, some MAAs demonstrate anti-proliferative effects on cancer cells by modulating key signalling pathways involved in cell cycle regulation and apoptosis. The potential for MAAs to enhance DNA repair mechanisms and protect against UV-induced mutations highlights their therapeutic value in cancer prevention and treatment.[231] At present, there have been no MAAs identified from *Euglena* although their ability to produce these compounds seems likely due to the prevalence of MAA production across algal species. The following chapter explores MAA production in *Euglena gracilis*.

5.5 Analysis of the biomass of *E. gracilis* through UV-spectroscopy

Mycosporine-like amino acids (MAAs) are intracellularly retained metabolites therefore unlike in previous analyses, extraction efforts in this chapter focused on the algal biomass rather than the supernatant. Previous studies have successfully extracted MAAs from *Rhodophyta* by refluxing biomass in 10-25% methanol or ethanol in water.[232] Accordingly, a 6 L culture of *Euglena gracilis* was centrifuged at 4500 rpm, and the resulting biomass was frozen and lyophilized to yield 1.30 g of dry material. This material was suspended in 250 mL of 10% methanol in RO water and heated to reflux overnight. The resulting green solution was filtered under pressure to remove solids and subsequently concentrated to 10 mL in vacuo.

The extract was further purified using preparative high-performance liquid chromatography (prep-HPLC) with a reverse-phase C18 column. The mobile phase consisted of MQ water (0.1% formic acid) as solvent A and acetonitrile (0.1% formic acid) as solvent B. A gradient similar to that used for MAA isolation from *Rhodophyta* was

applied, ranging from 0-5% B over 1 minute, increasing to 5-7% B over 5 minutes, followed by 7-9% B over another 2 minutes, and finally progressing from 9% to 100% B over 10 minutes (total runtime: 20 minutes). MAAs have been identified to elute from a C18 column in 98-90% water, therefore compounds eluting in 10% B or less and absorbing between 200-400 nm were considered indicative of MAA presence.

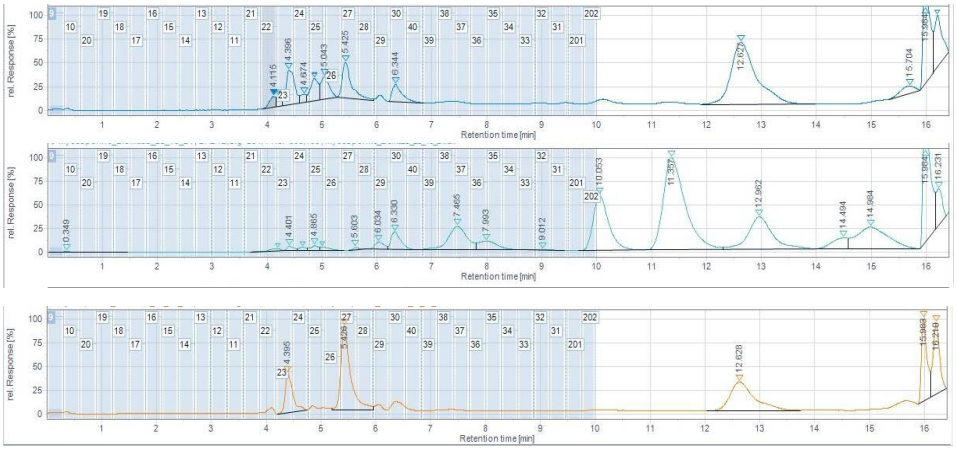


Figure 5.3 Chromatographic spectra from prep-HPLC of *E. gracilis* biomass extract showing absorption at 333 nm (top), 254 nm (middle), and 372 (bottom)

Eight peaks exhibiting UV absorbance at 333 nm were detected between 4 and 7 minutes, corresponding to 7–9% acetonitrile in water. To distinguish endogenous *Euglena*-derived compounds from media-derived artifacts, sterile media was subjected to identical chromatographic conditions. The media blank revealed four peaks in the same retention time range, indicating that at least two *Euglena* components exhibited UV activity in this region, potentially more if compounds co-elute with media components. Based on these findings, fractions 22, 24, 28, 29, and 30 were selected for secondary purification, while fractions 23 and 27, which aligned with regions of high UV absorbance in the EGJM media spectra, were excluded.

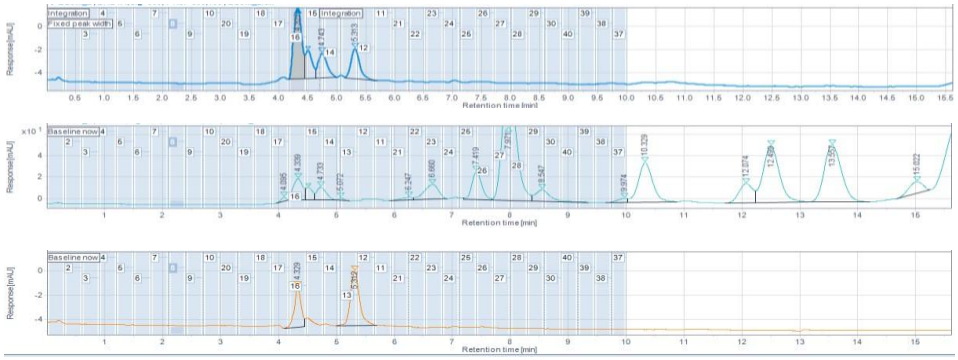


Figure 5.4 Chromatographic spectra from prep-HPLC of EGJM media at 333 nm (top), 254 nm (middle), and 372 (bottom)

5.6 Secondary purification of UV-active fractions through additional prep-HPLC

The selected fractions were reinjected onto a 0–10% acetonitrile gradient over 10 minutes, enabling resolution of discrete UV-absorbing compounds across the gradient, as shown in Figure 5.5 Fractions containing UV-active peaks were lyophilized, reconstituted in D2O, and analysed using nuclear magnetic resonance (NMR) spectroscopy on a Bruker 400 MHz spectrometer to determine functional groups. However, weak signal-to-noise ratios precluded the identification of any peaks due to the large water signal.

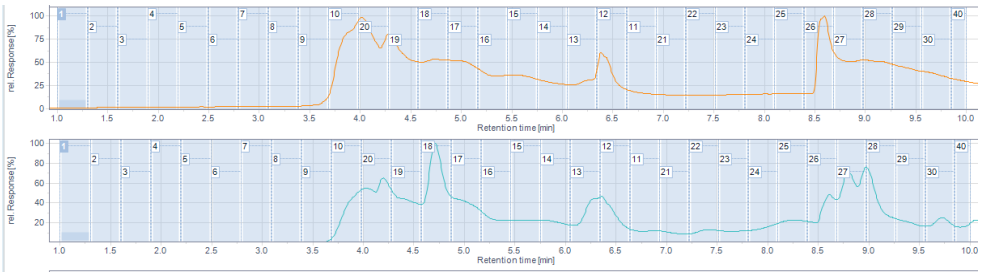


Figure 5.5 Chromatographic spectra from prep-HPLC of fraction 22 on a 10-minute gradient from 0 – 10% acetonitrile in water. Absorption at 254 nm (top) and 333 nm (bottom).

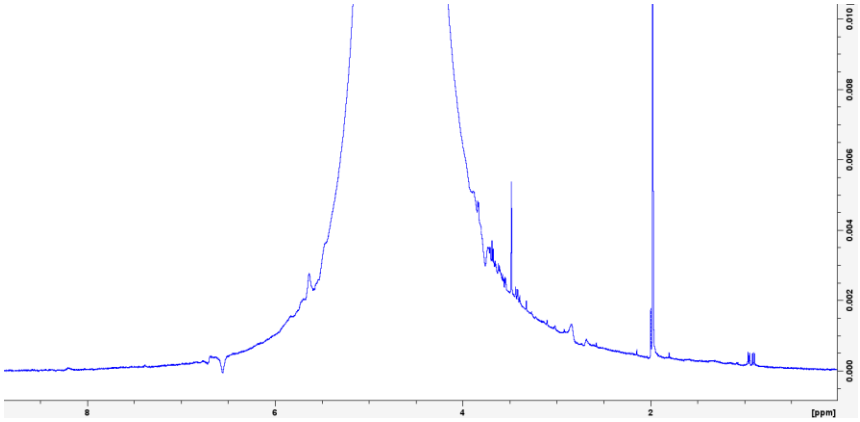


Figure 5.6 1H NMR of fraction 19 from Section 5.6

5.7 MS analyses of UV active fractions from secondary HPLC purification

An alternative route to compound identification within UV-active fractions is through MS, in order to identify compounds present in the UV active fractions which eluted from the secondary HPLC analyses in section 5.6, a mass spectrometry-based approach was employed. Each UV-active fraction from the secondary prep-HPLC purification was subjected to HPLC-MS/MS analyses using the same mobile phase conditions as employed in section 5.5. Despite the presence of a diode array detector (DAD), its sensitivity was insufficient to resolve UV absorption in the analysed fractions. Given the lack of UV data, MS analysis was focused on the retention time region corresponding to previously identified UV-active peaks in section 5.5, with an additional one-minute buffer to account

for potential retention time drift due to minor variations in column volume and injected analyte composition.

The spectra were screened for m/z values within the typical range for MAAs (270-380 Da) in the 4–7 minute retention time window, along with characteristic MAA fragmentation patterns. Expected MS2 fragmentation included neutral losses corresponding to water (18 Da), carbon dioxide (42 Da), formaldehyde (30 Da), hydroxyalkyl chains such as C_2H_4O (44 Da), and aminoalkyl groups such as $C_2H_4NH_2$ (60 Da). No fragmentation indicative of known MAA fragmentation patterns were observed in the raw spectra.[228,233] To facilitate MAA identification, a molecular network was constructed to identify commonalities among UV-active fractions and differentiate *Euglena*-derived metabolites from media components.

Figure 5.7 Molecular network of UV-active compounds from section 5.6 and EGJM media blanks

The molecular network analysis identified 12 nodes meeting the defined parameters (m/z range of 200-400 Da, retention time of 2-5 minutes). However, none of these nodes corresponded to distinct MS1 peaks in the raw spectra. Moreover, the anticipated presence of shared MAA-related nodes across multiple fractions was not observed. Instead, a significantly higher number of nodes (38) fitting these criteria were identified in the EGJM media blank. Given the inconclusive results, the extraction and initial HPLC purification processes were repeated to generate a more data-rich molecular network.

Number of spectra	m/z	RT/ min
12	294.988	4.2
7	329.256	3.9
6	344.074	2.4
5	267.157	2.6
4	362.085	2.3
4	348.069	2.7
4	224.453	4.8
3	301.142	2.3
3	269.113	2.4
3	366.057	2.5
3	338.342	4.9
2	382.837	2.2
2	372.925	2.2

Table 5.0 Non-control nodes with a m/z range of 200-400 Da, retention time of 2-5 minutes

5.8 MS analyses of all fractions eluted from HPLC purification of *Euglena* biomass Growth, extraction and purification was repeated as described in Section 5.5. Subsequent HPLCMS/MS analysis and molecular networking of all fractions eluted from this HPLC purification revealed a possible MAA cluster. Notably, compounds in this cluster were only detected in extreme polarities from the HPLC column—eluting at 0-7% acetonitrile and at 100% acetonitrile, but absent in intermediate fractions. This unusual distribution suggested the potential association of MAAs with residual cell membrane components not fully removed during extraction. A literature search did not identify any previously reported MAAs matching the observed m/z values in this molecular network. However, putative MAA structures were assigned to the possible MAAs present within this network to assist with future structural elucidation.

m/z	RT/ min
225.865	3.8
235.459	3.9
238.8	4.0
238.999	4.0
240.007	4.0
244.125	4.0
252.979	4.2
281.475	4.7
282.23	4.7
286.292	4.8
298.704	5.0
300.107	5.0
305.162	5.0

Table 5.1 Nodes with possible MAA character. Nodes in cluster highlighted in yellow

As well as the nodes highlighted in yellow in Table 5.1, the cluster also contains a node with a mass of 490.278 m/z that elutes at 3.7 minutes. This could be indicative of a large MAA compound. Possible MAA structures were generated for almost all the m/z values listed in the table.

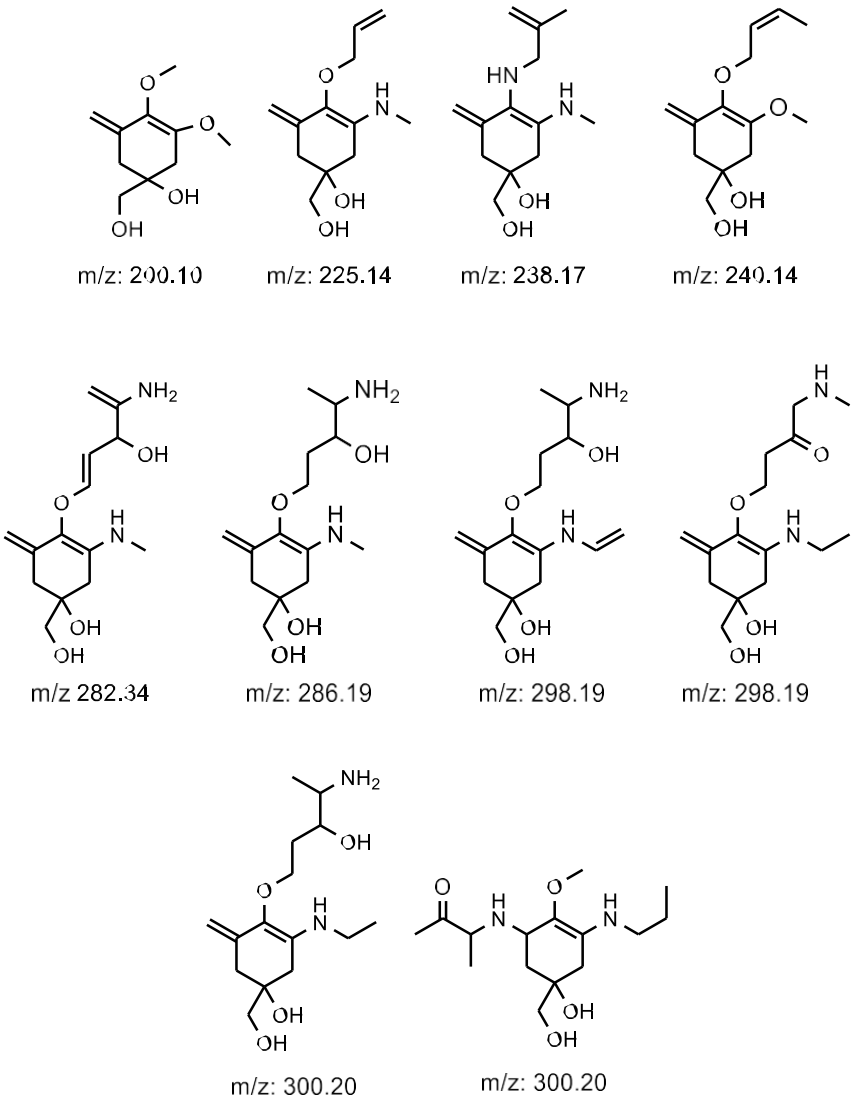


Figure 5.8 Suggested structures for possible MAAs present in extract

Some fractions displayed particularly clean spectra as can be observed in Figure 5.9.

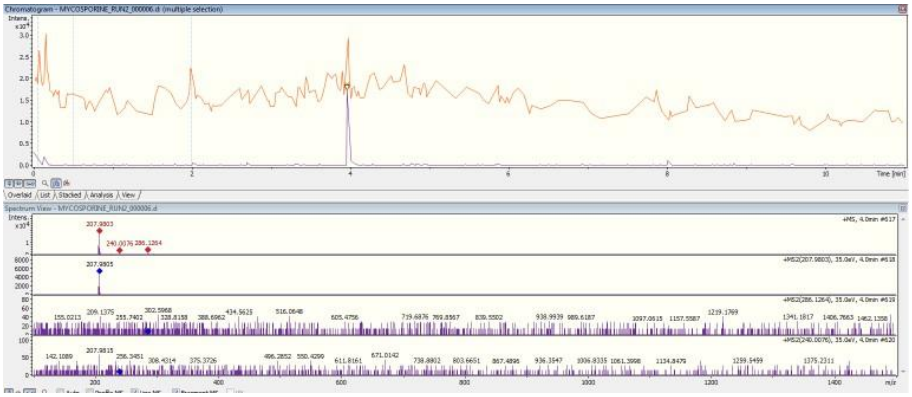


Figure 5.8 TIC shown in purple of fraction 6 from HPLC purification, compared to EGJM blank media control (shown in orange)

This fraction contained 3 compounds thought to be indicative of MAAs in the extract, and appeared clean in the chromatogram. Therefore, it was dried down and submitted for NMR on the 800 MHz spectrometer.

5.9 NMR study of fraction six

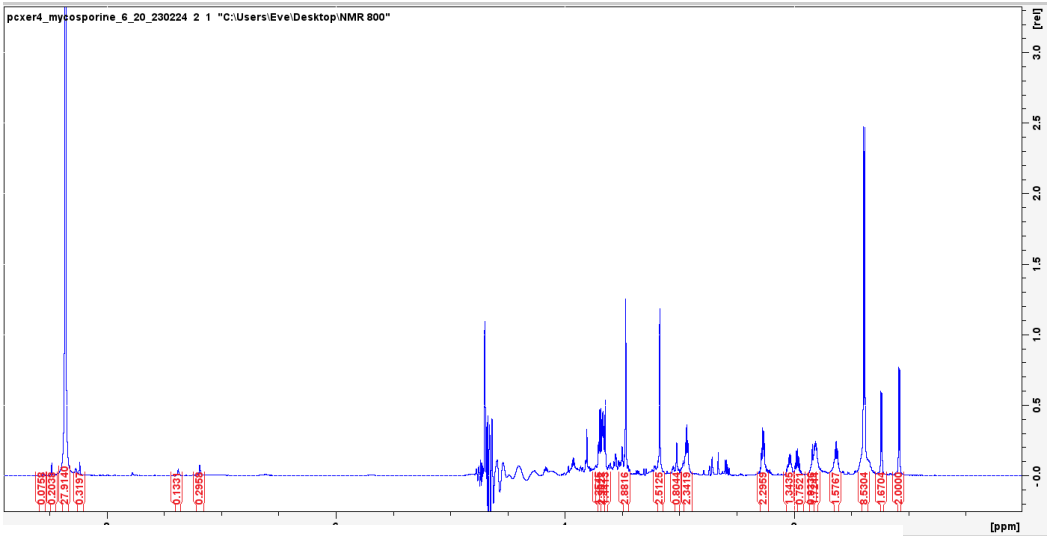


Figure 5.9 ^{1}H NMR of fraction 6

The presence of two very small peaks between 7 and 8 ppm in the ^{1}H NMR spectrum suggests the presence of either aromatic protons or deshielded exchangeable protons (e.g., amides, phenols, or enols). Given the UV absorbance at 333 nm, the sample likely contains conjugated chromophores, which could be mycosporine-like amino acids (MAAs). Alternatively the peaks in this region could correspond to aromatic protons in a conjugated system, possibly from a pyrrole, indole, or other heteroaromatic compounds, which are commonly found in algal metabolites. Alternatively, these peaks might arise from amide (-NH) or enol (-OH) protons that are hydrogen-bonded and deshielded, shifting them into the 7–8 ppm range. The fact that the peaks are small suggests that

these protons may be present in low concentration, weakly hydrogen-bonded, or subject to exchange broadening, depending on the solvent and sample conditions.

The ^{1}H NMR peaks at 6.4 and 6.6 ppm, corresponding to ^{13}C peaks at 103 and 117 ppm, respectively in the HSQC, suggest the presence of conjugated or heteroaromatic systems within the algal metabolite mixture. The ^{13}C shift at 103 ppm is consistent with a C=C or C-N hybridized carbon, often seen in pyrrole or indole derivatives, while the carbon at 117 ppm suggests a more deshielded sp^2-hybridized carbon, potentially part of an extended conjugation system. The ^{1}H chemical shifts in the 6.4–6.6 ppm range indicate protons on electron-rich double bonds, which could correspond to vinylic protons in α,β-unsaturated carbonyl systems or heterocyclic environments. These assignments align with known MAA structures, where similar shifts have been reported for pyrimidine, imine, or enamine functionalities. The absence of peaks in the fully aromatic region (7–8.5 ppm) further supports the idea that these protons belong to electron-rich, partially conjugated systems rather than fully delocalized benzene-like rings.

No peaks in the 6 – 7ppm range were observed in the DIPSI. The absence of the 6.4 and 6.6 ppm peaks in the DIPSI spectrum suggests that these protons do not exhibit strong scalar (J-coupling) interactions with other protons in the molecule. One possibility is that they are isolated protons, such as those attached to quaternary carbons or part of an extended conjugated system, which would prevent them from showing cross-peaks in DIPSI. Another explanation is that they correspond to exchangeable protons, such as hydroxyl (-OH) or amine (-NH) groups, which undergo rapid solvent exchange, causing broadening or suppression of their DIPSI cross-peaks. Additionally, if these protons are part of a highly conjugated system, such as vinylic or imine protons in mycosporine-like amino acids (MAAs), pyrimidine derivatives, or enamine structures, their weak dipolar interactions may further limit their detectability in DIPSI. Solvent effects or fast relaxation processes could also contribute to the absence of signals, particularly if these protons are subject to broadening due to exchange dynamics or electronic delocalization. The lack of DIPSI cross-peaks strongly indicates that these protons do not engage in significant scalar coupling, making them likely candidates for conjugated, weakly coupled, or exchangeable environments.

If these small peaks are indicative of the presence of MAAs, they also reveal that the MAA fraction is heavily contaminated with co-eluting compounds present in significantly higher concentrations. To enhance the separation of the UV-active species from these contaminants, the growth and extraction processes were repeated with modifications to the purification strategy.

5.10 Modifying growth and purification techniques

The culture conditions for *Euglena gracilis* were adjusted to maximize biomass yield. Initial growth was conducted in six 1 L cultures of EG:JM at pH 4.5 with 30 mM glutamic acid for two weeks. Subsequently, 50 mL from each 1 L culture was transferred to fresh EG:JM medium supplemented with 15 g glucose to stimulate further growth. This modification resulted in a substantial increase in biomass production, yielding 12 g of dry weight from 6 L of culture. The same extraction protocol was employed as previous, followed by preparative HPLC using the same gradients but a modified mobile phase composition: A: MQ water with 5% methanol (0.1% formic acid, 10 mM ammonium formate) and B: 100% methanol (0.1% formic acid, 10 mM ammonium formate). By increasing the polarity of the mobile phase, it was expected that separation of more polar compounds would improve due to a reduced polarity difference between water and methanol compared to water and acetonitrile.

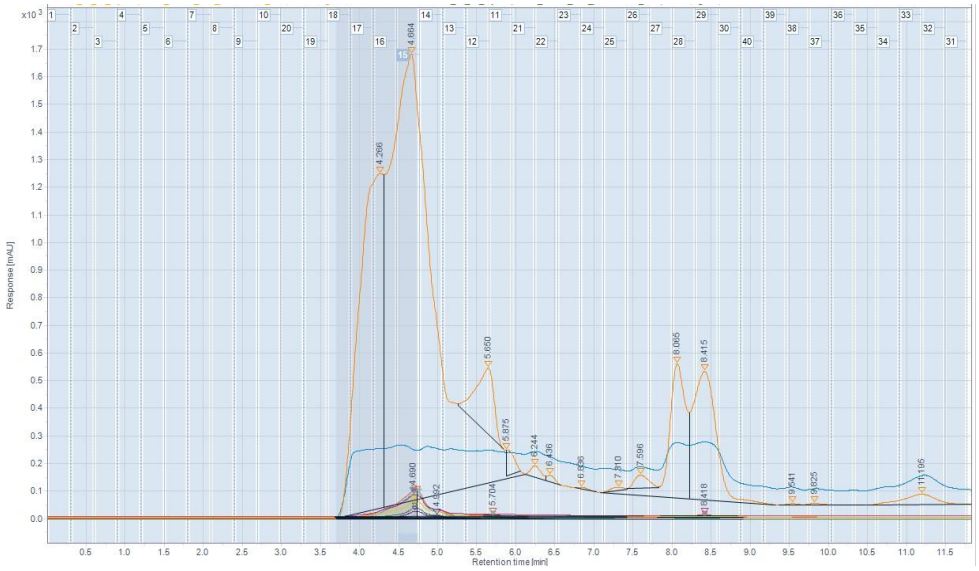


Figure 5.10 Chromatograms from second HPLC purification, key absorbance at 275nm (yellow) and 333 nm (turquoise)

To minimize material loss due to potential degradation, only the most intense 333 nm-absorbing peak observed after the second HPLC purification was selected for additional purification. Before the additional purification, the fraction containing the highest 333 nm absorbance was dried under vacuum and submitted to 800 MHz NMR. The peaks at 6.4 and 6.6 ppm were observed again, but their integration remained significantly lower than the peaks in the alkyl region, which were themselves overlapping and unresolved. The reduced purity of this fraction compared to previous samples, despite undergoing two HPLC purification steps, is likely a consequence of the modified culture conditions and increased biomass used for extraction.

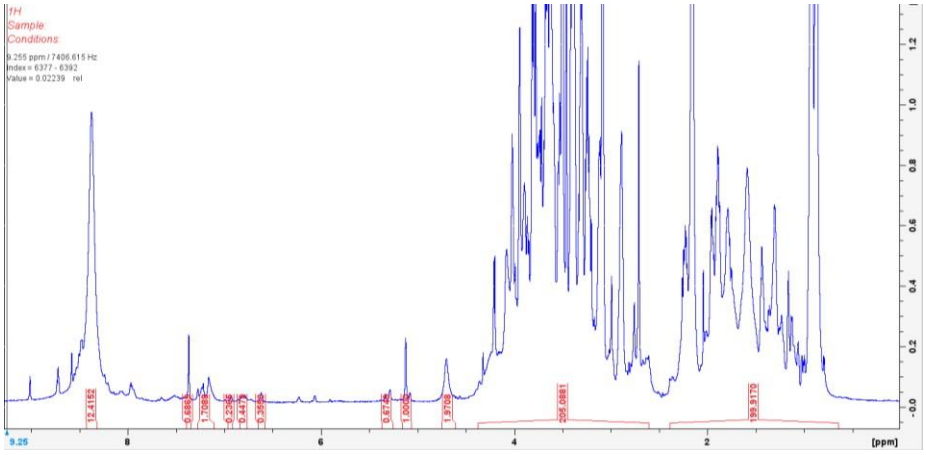


Figure 5.11 ^{1}H NMR of fraction 15

The initial prep-HPLC chromatogram revealed additional UV-active peaks, suggesting that a greater number of metabolites were extracted in this batch, which may have contributed to decreased separation efficiency due to column overloading. Additionally, selecting the peak with the highest UV absorbance at 333 nm may not be solely indicative of an MAA, as co-elution with multiple UV-absorbing compounds could have contributed to the signal.

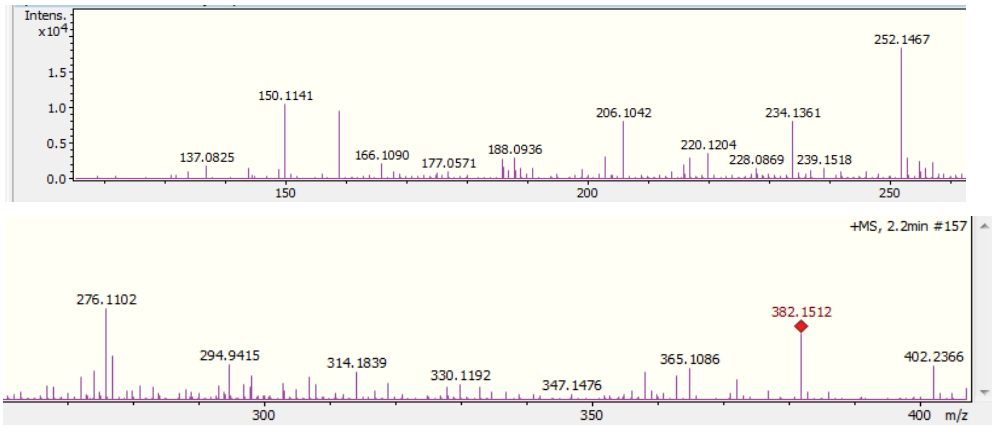


Figure 5.12 MS2 spectrum of 382.1512 m/z

This fraction was subsequently analysed using HPLC-MS/MS, which revealed a possible mycosporine-like amino acid (MAA) peak with a fragmentation pattern that could correspond to the schematic in Figure 5.13. Molecular networking of fraction 15 with the EGJM media control confirmed that no media-derived compounds were present, yet the molecular network and NMR spectra indicated high levels of contamination, necessitating further purification.

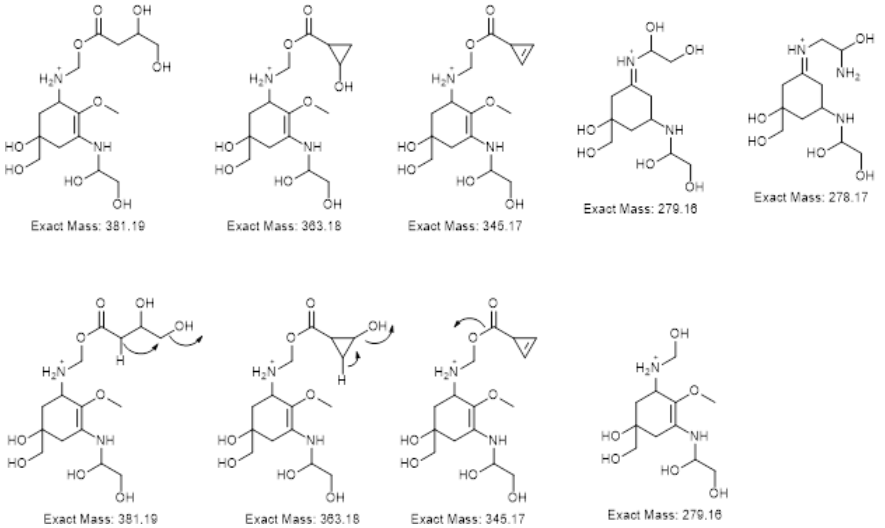


Figure 5.13 Possible fragmentation pattern for 382.1512 m/z

5.11 Pursuing additional purification

The fraction was then subjected to hydrophilic interaction liquid chromatography (HILIC) using a mobile phase of A: 95% acetonitrile, 5% water (5 mM ammonium formate, 0.1% formic acid) and B: 50% acetonitrile, 50% water (5 mM ammonium formate, 0.1% formic acid). The fraction was dissolved in 500 ul of the starting mobile phase composition, sonicated, and injected onto the HILIC column. This process was repeated four times, with the starting composition of A decreasing by 10% with each iteration, ensuring complete solubilization of the material while maintaining an appropriate water percentage for the column's efficiency. The remaining residue was dissolved in methanol and subjected to C18 chromatography under the mobile phase conditions outlined in Section 5.5. No 333 nm-absorbing peaks were detected in this fraction, suggesting that all UV-active compounds had been effectively purified in the HILIC step.

The fractions exhibiting 333 nm absorbance were analysed using 800 MHz NMR, but the resulting spectra displayed only weak signals. Even in the most concentrated fractions, no peaks indicative of conjugated environments were observed, possibly indicating either the degradation of the MAA-like compounds or insufficient analyte concentration. These results suggest that while the purification strategy successfully eliminated a significant

proportion of contaminants, the final MAA fraction may still be too dilute or structurally unstable to allow for definitive spectroscopic characterization.

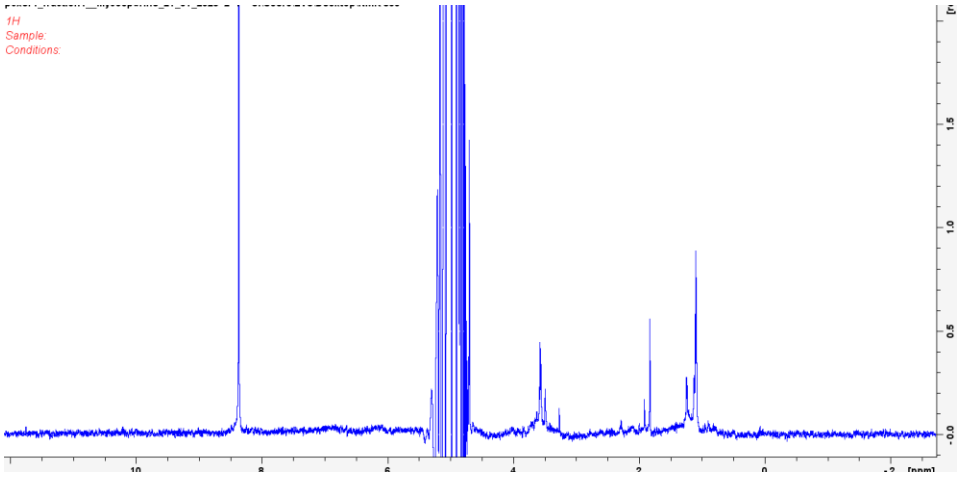


Figure 5.14 Proton NMR for 333nm active fraction from HILIC column

5.12 Conclusions from Chapter Five

This chapter has provided evidence for the presence of MAAs in *Euglena gracilis* the identification of unknown molecules with UV-absorption in the 300 – 400 nm range, with the appropriate molecular weight for MAAs, and in one case a possible fragmentation pattern of MAA fragmentation, as well as the expected polarity. However, despite multiple purification strategies, including preparative high-performance liquid chromatography (HPLC) and hydrophilic interaction liquid chromatography (HILIC), a fully purified sample suitable for definitive structural characterization was not obtained. The presence of co-eluting metabolites and the complexity of the extract likely hindered the isolation of a single, well-resolved MAA fraction. Additionally, limitations in sample stability and degradation during purification may have contributed to the inability to confirm the exact Molecular structure of the detected UV-absorbing compounds.

Future work should focus on refining purification protocols to improve yield and purity. The use of alkaline hydrolysis and enzymatic digestion could be particularly valuable in selectively breaking down non-MAA UV-absorbing contaminants, thereby facilitating the differentiation of MAAs from structurally similar metabolites. Alkaline hydrolysis and enzymatic digestion provide a targeted approach for differentiating Mycosporine-like amino acids (MAAs) from structurally similar UV-absorbing contaminants by exploiting their differential chemical stability. Alkaline hydrolysis involves treating the sample with a strong base, such as NaOH or KOH, at elevated temperatures (40–100°C), which selectively degrades labile UV-absorbing metabolites, including phenolic compounds, flavonoids, and glycosylated derivatives. These compounds undergo base-catalysed cleavage, breaking down into non-UV-absorbing fragments, whereas MAAs, due to their highly stable imine-based core structures, remain intact. This process effectively removes ester- and amide-linked conjugates that might co-

elute with MAAs in chromatographic analyses. Following hydrolysis, enzymatic digestion further enhances purification by selectively degrading protein-bound or conjugated contaminants. Proteases such as trypsin and proteinase K hydrolyse protein-associated UV-absorbing impurities, while glycosidases target glycosylated metabolites, breaking them into non-UV-absorbing components.[234] Esterases and lipases can also be employed to hydrolyse ester-linked aromatic contaminants. Since MAAs are small, non-peptidic, and non-glycosylated molecules, they remain unaffected by enzymatic digestion, ensuring that only unwanted impurities are degraded. These approaches have been successfully applied in previous studies to distinguish MAAs from other UV-protective compounds in algal and microbial extracts.[235]

Future research could explore strategies to enhance the production of Mycosporine-like amino acids (MAAs) in *Euglena gracilis* by targeting key genes, enzymes, and signalling pathways involved in their biosynthesis. Since MAAs are synthesized via the shikimate pathway, a primary focus could be on upregulating genes encoding enzymes such as dehydroquinate synthase and O-methyltransferase, which are essential for the formation of key intermediates like 4-deoxygadusol. While homologs of these enzymes have been identified in *E. gracilis* for other metabolic processes,[236] their roles in MAA biosynthesis remains to be explored. *E. gracilis* contains genes encoding 5-enolpyruvylshikimate-3-phosphate synthase, an enzyme associated with the shikimate pathway,[237] which could be targeted for overexpression to increase the metabolic flux towards MAA production. Additionally, regulatory elements controlling these biosynthetic genes could be characterized and modified to enhance expression under optimal conditions.

Another approach to increase MAA yield involves manipulating environmental factors, particularly light exposure, which has been shown to influence the production of MAAs.[238] Future experiments should investigate the effects of varying light intensities and wavelengths on MAA production in *Euglena*, potentially identifying conditions that maximize biosynthetic output. Furthermore, metabolomic and transcriptomic analyses could provide deeper insights into how MAA synthesis is regulated in response to environmental cues, allowing for the development of metabolic engineering strategies that optimize production. By integrating genetic, biochemical, and environmental approaches, this research could provide a robust framework for enhancing MAA biosynthesis in *Euglena* and expanding its potential applications in biotechnology. Working in tandem, alternative culture conditions may enhance MAA production, while metabolomics-based comparative analyses between *Euglena* and known MAA-producing species could provide additional evidence for the structure of MAAs in *Euglena*.

Chapter Six: Conclusions and Future Directions

This thesis has highlighted a vast quantity of unidentified compounds within the *Euglena* metabolic profile, confirming its relevance as a source from which to isolate novel natural products. From 33 *Euglena* strains, a total of 1017 unknown compounds were reported from Cytoscape Network One, and a total of 21 unknown compounds were reported to be uniquely produced from *E. gracilis* in Cytoscape Network Two, not found in either other algal species or in the five *Stretpmyces* strains. In addition, different coculturing organisms and modified media conditions resulted in the production of unique natural products from *Euglena*. Coculturing *Chlorella vulgaris*, cyanobacteria *Synechococcus* sp.7002. and *Euglena gracilis* with five *Streptomyces strains:, S. coelicolor A3, S. glaucescens, S. limosus, S. griseus, and S. netropsis* was proven to produce compounds not produced in monocultures of either strains with 789 compounds unique to cocultures. Modifying culturing conditions is a faster and more accessible approach to increasing natural product diversity than targeting specific genes and can also be used in the case of organisms such as *Euglena* in which the genome is lacking in sufficient annotation and for which limited genetic tools have been developed. Additionally, the presence of a novel phosphonate compound produced by *Euglena gracilis* was confirmed and there is also strong evidence from mass spectrometry and UV data for the presence of MAAs in *E. gracilis*.

Although no compound was purified enough to make confident annotation in its structure, the number of different purification methods employed within this thesis are expansive enough to suggest that it is not the method of purification that is holding back isolating a novel natural product from *Euglena* but the concentration by which *Euglena* produces natural products. This is highlighted particularly in the case of the novel phosphonate compound, which was seen to only be produced at 0.2 mM concentration. Additionally, from 6L of culture and undergoing just two purification steps, activated charcoal followed by size exclusion chromatography, the peaks indicative of phosphorous splitting are only just above the noise level after 24 hours of scans on a high field 800 MHz NMR which supports a very limited amount of phosphonate compound is harvested from *E. gracilis* cultures. Previous isolation of natural products from *Euglena* have been obtained not from the supernatant but from the biomass, even in the case of the toxin Euglenophycin. Although one might assume that a toxin would be transported extracellularly in order to target non-*Euglena* cells, the majority of Euglenophycin produced is maintained within the cell, [160] presumably to deter predation by digestion.

It might not be possible to culture *Euglena* in sufficient volumes within the laboratory to obtain enough material to isolate novel compounds. If you take the phosphonate compounds as an example, and assume linear concentration increase with increasing media volume, in order to begin with a concentration of just 1 mM phosphonate the culture volume would have to be multiplied by 5 – which brings the total volume to 30 L. In order to circumvent this issue, working in partnership with companies that currently culture *Euglena* in large volumes for the study of biofuels, supplements, or as an alternative food source might offer a route to achieving larger culture volumes.[239] These companies typically utilize the biomass, not the supernatant, for their research, making the supernatant a "waste byproduct" that would have no use for and as such could be inclined to donate it to academic research. They may be less inclined to modify culture media or introduce additional microbes for coculture experiments, as this could disrupt their established production systems. Additionally, these companies are unlikely to have methods for concentrating large volumes of supernatant into manageable amounts for laboratory work.

To address this, the use of an industrial evaporator would be required, similar to those found in wastewater treatment facilities. Such equipment would be essential to concentrate large volumes of supernatant efficiently. However, this process would also concentrate media components in larger amounts than would be removable on laboratory scales necessitating industrial filtration methods, such as large scale activated charcoal adsorption, to remove unwanted substances. Outsourcing the growth of *Euglena* to an external company with the capacity to maximize culture volumes would therefore need to be coupled with access to industrial machinery for downstream processing. Only after these steps could the concentrated material be made available for detailed laboratory work.

It is, however, possible to utilise genetic techniques to increase the yield of natural products *Euglena* produces. Polyketide and non-ribosomal peptide genes have been identified and annotated within *Euglena*.[177] These genes activate biosynthetic pathways that are pivotal for natural product production of two classes, and possible hybrid classes, of natural products. As they have already been identified, targeting these genes for increased expression could increase the concentration of some natural products from *Euglena*.

One of the most effective methods for upregulating gene expression in *Euglena* is promoter engineering. Native promoters such as those from the psbD (photosystem II protein D) or hsp70A (heat shock protein 70A) genes are widely used due to their strong, light-inducible characteristics.[240] These promoters could be exploited to drive

the expression of polyketide and non-ribosomal peptide genes to increase NP production in *Euglena.*

Alternatively, these genes could be transfected into another host. However, splicing large non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) genes from *Euglena gracilis* presents significant challenges due to the complexity of its genome, transcriptome, and post-transcriptional modifications. Unlike bacterial systems where PKS and NRPS gene clusters are typically co-localised and transcriptionally linked, *Euglena* possesses a genome with extensive RNA editing, trans-splicing, and a mix of nuclear and plastid-encoded biosynthetic pathways. These factors make it difficult to identify, assemble, and functionally express full-length PKS and NRPS genes outside their native environment. In particular, *Euglena*'s PKS genes are likely fragmented across multiple loci, interspersed with non-canonical introns or requiring trans-splicing events for full transcript assembly. The modular nature of PKS systems, where multiple enzymatic domains are encoded within a single large open reading frame (ORF), further complicates full-length PCR amplification and cloning, as repetitive sequences and high GC content can cause polymerase slippage and incomplete assembly. Additionally, the functional expression of *Euglena* PKS enzymes may require specific acyl carrier proteins (ACPs), redox partners, or chaperones, which are either unknown or absent in common heterologous hosts like *Escherichia coli* or *Saccharomyces cerevisiae*. Moreover, codon bias differences between *Euglena* and bacterial or fungal expression systems may lead to inefficient translation or misfolded proteins when attempting heterologous expression.[140,177]

The NRPS systems in *Euglena* face similar hurdles, particularly due to the presence of non-standard adenylation and condensation domains that may differ structurally and functionally from well-characterised bacterial NRPS systems. The possibility of PKS-NRPS hybrid pathways in *Euglena* further complicates functional reconstruction, as interactions between these biosynthetic modules may depend on specific regulatory elements, subcellular localisation signals, or auxiliary proteins that are not present in heterologous hosts. Additionally, *Euglena*'s plastid involvement in secondary metabolism suggests that full NRPS-PKS system expression may require engineering a dual-localisation strategy within an alternative algal or plant-based expression host. Current barriers to cloning and expression include the lack of fully characterised NRPS and PKS gene clusters, uncertainty regarding their regulation, and the absence of a functional recombinant expression system for complex *Euglena* biosynthetic enzymes. Future research should prioritise long-read sequencing and transcriptomic validation to map complete PKS-NRPS gene clusters, followed by synthetic biology approaches such as domain-by-domain expression in engineered microbial hosts, plastid-targeted

constructs for algal expression, or modular assembly in specialised Streptomyces strains. Overcoming these barriers will be crucial for unlocking the biotechnological potential of *Euglena* in producing novel polyketide and peptide-based bioactive compounds.[61]

It might be feasible to sidestep the need for a producing organism in its entirety. Currently being trailed by software such as GNPS and Sirius are prediction algorithms that assign structures of unknown compounds based on MS/MS data of known compounds. Should these improve to the point where they can accurately describe the structure of an unknown compound, perhaps with assistive help from developing AI, there would be no need for purification or structural elucidation. In this scenario, the compound could be synthesized in large amounts and assayed for activity. The compound, having been effectively "pre-screened" by millions of years of evolution, would most likely possess bioactivity. This method could function as a recipe book from which synthetic chemists can readily select new structures to synthesize for drug development.[170]

Alternatively, if the prediction methods are not entirely accurate, providing only suggestions of functional groups, this information would still be valuable for improving isolation. Isolation techniques could be tailored to the known properties of the compound, significantly reducing the time required for purification optimization. With fewer steps in the purification process, less material would be lost, ultimately increasing the amount available for subsequent analyses such as NMR. Moreover, the integration of predictive software with machine learning could progressively refine these methods, providing researchers with increasingly reliable structural insights over time.

Additionally, a hybrid approach could emerge, where partial structural predictions are combined with traditional analytical techniques to iteratively identify and refine compound structures. This could create a feedback loop, where experimental data informs software models, and software predictions streamline experimental workflows. Such a system would reduce reliance on large-scale culture systems and open up new avenues for drug discovery, allowing researchers to explore a wider array of potential bioactive compounds with fewer resources. The incorporation of these technologies into existing workflows represents a promising step forward in the efficiency and scalability of natural product research.

This could be further supported by constructing comprehensive molecular networks of *Euglena gracilis* grown under a diverse set of environmental conditions to generate a highly data-rich metabolite profile. By culturing *Euglena* under varying pH levels, researchers could identify pH-dependent metabolic shifts, particularly in secondary

metabolite production, which may reveal novel bioactive compounds. Additionally, introducing heavy metal contamination (e.g., cadmium, lead, or arsenic) could induce stress responses leading to the upregulation of metal-binding metabolites, antioxidant compounds, or chelating agents, which could have pharmaceutical or environmental applications. Another promising avenue involves co-cultivation with bioactive compound-producing fungi, such as *Aspergillus nidulans* or *Trichoderma harzianum*, both of which have been shown to produce polyketides, alkaloids, and antimicrobial peptides.[178] Such microbial interactions may trigger cryptic biosynthetic pathways in *Euglena*, leading to the discovery of previously uncharacterised natural products.

Chapter Seven: Materials and Methods

All materials purchased from Merck unless otherwise stated.

7.1 General Methods

7.11 Media Preparation

YPD agar plates were made from YPD powder.LB broth and LB agar media were made from premade LB powder and LB agar powder, respectively. EG, AF6, Jaworski's Medium (JM), Soil Extraction (SE), and Soil, Water Biphasic (SWBi) media were prepared according to the NIES collection guidelines (https://mcc.nies.go.jp/medium/en/media_web_e.html). JM:SE is a 1:1 ratio of JM in half concentration and SE, while EG:JM is a 1:1 ratio of EG and JM, both in half concentrations. EG:JM plus glucose includes an additional 15 g glucose per L. Solid EG:JM was prepared by adding 25 g of agar to 1 L of EG:JM. EG:JM pH 4.4 plus 30 mM glutamic acid was pH adjusted with 12 M HCl.

7.12 Small-Scale Cell Culturing

Cultures were grown in 10 mL volumes in 50 mL T-flasks at 100 rpm and maintained at 25 °C under a 16 h light/8 h dark cycle. Cultures were harvested after six weeks.

7.13 Large-Scale Cell Culturing

For scale-up from 10 mL cultures, 1 mL of small-scale Culture was inoculated into 1 L of liquid media in a 2 L conical flask and stirred by hand. A portion (1 mL) was returned to the small-scale culture for maintenance. Cultures were 100 rpm and maintained at 25 °C under a 16 h light/8 h dark cycle. and harvested after six weeks. For maintenance of large scale Cultures 250 mL of media in a 500 mL conical flask was inoculated with 10 mL of the relevant culture before extraction from the 6 L cultures. This was then used to inoculate the large scale cell Culturing in future regrowths.

7.14 Small-Scale Extraction

5 mL of each culture was extracted with ethyl acetate (2 × 5 mL), followed by back-extraction with 1 mL deionized water. The remaining aqueous layer was extracted with butan-1-ol (2 × 5 mL) and back-extracted with 1 mL deionized water. The ethyl acetate and butanol fractions were dried under a fume hood and vacuum evaporation, respectively, to leave a residue. Lyophilization was performed to remove residual water. Extracts were reconstituted in 2 mL methanol, vortex mixed, centrifuged (13,000 g, 2 min), and supernatants (1.5 mL) were transferred into HPLC vials.

7.15 Large-Scale Extraction

Identical to small-scale extraction, with proportional adjustments for larger volumes.

7.2 High-Performance Liquid Chromatography (HPLC) and Mass Spectrometry (MS)

7.21 HPLC-MS/MS

Sample fractions were analysed by HPLC-ESIQ/TOF-MS/MS using a Dionex ultimate 3000 UHPLC (Thermo Scientific) coupled to a Bruker Daltonics Impact II Q-ToF high-resolution mass spectrometer. HPLC was performed using a Kinetex reverse-phase C18 column (100 mm × 2.1 mm, 3 µm) with 0.5 mL/min flow rate. MS1 range: 300–1500 m/z. MS2: untargeted, 3 precursors per cycle, fixed collision energy: 10 eV, source gas temperature: 250 °C, gas flow: 8 L/min.

7.22 GNPS and Molecular Networking

HPLC-MS/MS files were converted to mzXML using Trans Proteomic Pipeline software and uploaded to GNPS for Molecular networking under GNPS-recommended parameters. Data were visualized in Cytoscape.

7.3 Semi-Preparative High-Pressure Liquid Chromatography (HPLC)

Sample fractionation was performed on a Kinetex® C18 column (100 Å, 100 mm × 21.2 mm, 5 µm) using a Waters 1525 Binary HPLC pump coupled to a Waters 2489 UV/visible detector. Flow rate: 5 mL/min.

7.4 Compound Detection and Isolation

Fractions containing target compounds were identified based on m/z values, fragmentation patterns, and retention times in LCMS-MS. Solvent was removed in vacuo, leaving a residue.

7.5 Nuclear Magnetic Resonance (NMR) Spectroscopy

Fractions were prepared in either 100% D_2O or 9:1 D_2O:H_2O, up to 666 ul volumes and ^1H NMR spectra were recorded on either a Bruker Ascent 400 MHz spectrometer or Bruker Ascent 800 MHz spectrometer. Acquisition time for ^1H NMR was 2.5 mins on the 400 mHz and 16 hours on the 800 MHz spectrometer.

7.6 Chapter One: Screening of Euglenoid Strains

7.61 Small-Scale Cell Culturing

Thirty-two Euglenoid strains (Culture Collection of Algae and Protozoa) were inoculated in 10 mL of AF6, JM:SE, and SWBi media. Cultures were maintained for six weeks.

7.62 Large-Scale Cell Culturing

Cultures were scaled up to 1 L volumes under identical conditions.

7.63 Extraction

Small-scale and large-scale extractions were performed as described in the General Methods. The litre Cultures were evaporated down *in vacuo* to <25% of their original

volume. They were then washed twice with an equal volume of the extracting solvent, the organic layer was then washed with DI H_2O and the organic layers recombined and removed *in vacuo* to leave a white residue.

7.7 Semi-prep HPLC

Sample fractionation was performed on a Kinetex® C18 column (100 Å, 100 mm × 21.2 mm, 5 µm) using a Waters 1525 Binary HPLC pump coupled to a Waters 2489 UV/visible detector. Flow rate: 5 mL/min.

Mobile phases A: 0.1% formic acid in MQ H_2O and B: 0.1% formic acid in acetonitrile. An initial elution gradient of 10 – 100% B over 15 minutes with a 5-minute hold at 100 %B was used for each extract before specialised shallower gradients were applied.

Target 2.1: 50 – 70% B over 15 minutes, 70 – 100% B over 1 minute, 5 minute hold at 100%B.

Target 2.2: 60 – 90% B over 15 minutes, 90 – 100% B over 1 minute, 5 minute hold at 100%B.

Target 2.3: 50 – 70% B over 15 minutes, 70 – 100% B over 1 minute, 5 minute hold at 100%B.

Target 2.4 60 – 90% B over 15 minutes, 90 – 100% B over 1 minute, 5 minute hold at 100%B.

7.8 NMR

The fractions containing 2.1, 2.3, were dried in vacuo on the rotary evaporator and then dissolved in 100% DI H_2O. ^1H NMR spectra were obtained using the 800 MHz NMR.

7.9 Disc Diffusion Assays

Fractions were dried to residue. Separate LB agar plates were inoculated with *E. coli* and *K. rhizophila.* YPD agar plates were inoculated with *S. cerevisiae*. The inoculant was spread evenly over the plate using a spreader. Ampicillin (0.10 mg/mL) was the positive control. The following extracts were then pipetted onto autoclaved paper disks which were evenly transferred onto each plates.

Target 1.1 containing fraction 30 of the first purification of CCAP1261.6.

Target 2.1 containing fractions after two HPLC purification steps of CCAP1204.9, pooled, dried and redissolved in DI H_2O (5 ul).

Target 2.3 all fractions from first purification step of CCAP1216.3C, pooled, dried and redissolved in DI H_2O (5 ul).

Target 2.4 Fractions 1-29 of first purification step of CCAP1211.3 pooled, dried and redissolved in DI H_2O (5 ul). Target 2.4 containing fraction 30 of previous purification step after purification through second HPLC pooled, dried and redissolved in DI H_2O (5 ul).

Target 2.5 containing fraction 30 of the first purification of CCAP1204.20A.

The plates were then incubated at 37 °C overnight.

7.10 Chapter Two: Coculture and Streptomyces Strain Analysis

7.11 Small-Scale Culturing

The *Streptomyces* strains were obtained from a private collection within the University of Nottingham. Colonies were transferred from LB agar plates via inoculation loops into 10 mL of culture within 50 mL T-flasks and shaken at 100 rpm and maintained at a temperature of 25 ˚C, in a 16 h light 8 h dark cycle. Cultures were harvested after 6 weeks.

7.12 Small Scale Extraction

Identical to Small Scale Extraction for Chapter One, with the exception that the residue left behind was brown in colour.

7.13 Large-Scale Extraction

Identical to Small Scale Extraction for Chapter One, with the exception that the residue left behind was brown in colour.

7.14 HPLC Fractionation and Compound Isolation

HPLC runs employed a Phenomenex Luna C18 Prep HPLC column, with elution gradients optimized per sample type. Target fractions were analysed via LCMS and further purified as necessary.

7.15 HPLC mobile phase used in section 3.141

421 containing fractions from acetonitrile solutions were pooled, dried down in vacuo and dissolved in 10 mL of 10% acetontrile. An initial elution gradient of 10 – 100% B over 15 minutes with a 5-minute hold at 100 %B. Fractions were assayed for the presence of target compounds using HPLCMS/MS. Fractions containing 421 were combined, dried and redissolved in 1 mL of 100% MQ water. They were then submitted purification via a gradient of 0 – 30% acetonitrile over 30 minutes.

7.16 HPLC mobile phase used in section 3.16

Ethyl acetate extract was dried and dissolved in 10 mL solution of 10% acetontriile in MQ water. The mixture was syringe filtered and then run through semi-prep HPLC on a 10 to 50% acetonitrile gradient over 30 minutes and fractions collected every minute. The 421-containing fractions were then recombined, dried, dissolved in 1 mL 10% acetonitrile in water and run on a longer, shallower HPLC column of 10 to 35% acetonitrile over 60 minutes with fractions collected each minute.

7.17 SE chromatography

Fraction 15 was injected onto a Sephadex LH-20 column in an isocratic 20% acetonitrile mobile phase. A 0.25 mL/min flow rate was used and 10 mL fractions were collected.

7.18 Chapter Three: Phosphonate Compound Isolation

7.19 Culturing and Biomass Processing

10 mL of CCAP1224/5z was inoculated in 1 L of EG:JM + glucose and grown in standard temperature and light conditions for four weeks. The Cultures were centrifuged at 4500 x g for 30 mins at 4 °C to separate the biomass from the supernatant. The supernatant was concentrated down to 500 mL on the rotary evaporator.

7.20 Charcoal Column

After growth the Cultures transferred into 50 mL falcons and centrifuged at 4500rpm for 30 minutes at 4 °C. The supernatant was removed and the total volume *in vacuo* to 500 mL and ran through an activated charcoal column (30 cm x 1cm) swollen with 100% DI H_2O. The column was then washed with two column volumes of DI H_2O, 75% DI H_2O 25% ethanol then 50% DI H_2O 50% ethanol and finally 100% ethanol. The fractions were dried to a brown residue *in vacuo* and reconstituted in 1 mL 100% D_2O. 666 ul was then transferred to an NMR tube for ^{31}P NMR analysis. The presence of the C-P peaked was confirmed through identification of the peak at 20 ppm.

7.21 Charcoal Centrifugation I

After growth the Cultures transferred into 50 mL falcons and centrifuged at 4500rpm for 30 minutes at 4 °C. The supernatant was removed and the total volume *in vacuo* to 500 mL and transferred to two 500 mL centrifugation tubs, containing 20 grams of activated charcoal column swollen with 100% DI H_2O. The tubs were vortexed and then centrifuged at 4500rpm for 30 minutes at 4 °C. The supernatant was removed and the process

repeated with solutions of 75% DI H_2O 25% ethanol then 50% DI H_2O 50% ethanol and finally 100% ethanol. The fractions were dried to a brown residue *in vacuo* and reconstituted in 1 mL 100% D_2O. 666 ul was then transferred to an NMR tube for ^{31}P NMR analysis. The presence of the C-P peaked was confirmed through identification of the peak at 20 ppm.

7.22 Charcoal Centrifugation II

After growth the Cultures transferred into 50 mL falcons and centrifuged at 4500 rpm for 30 minutes at 4 °C. The supernatant was removed and the total volume *in vacuo* to 500 mL and transferred to 50 mL falcon tubes, containing 2 grams of activated charcoal column swollen with 100% DI H_2O. The tubes were vortexed and then centrifuged at 4500rpm for 30 minutes at 4 °C. The supernatant was removed and the process repeated with solutions of 75% DI H_2O 25% ethanol then 50% DI H_2O 50% ethanol and finally 100% ethanol. The fractions were dried to a brown residue *in vacuo* and reconstituted in 1 mL 100% D_2O. 666 ul was then transferred to an NMR tube for ^{31}P NMR analysis. The presence of the C-P peaked was confirmed through identification of the peak at 20 ppm.

7.23 Charcoal Centrifugation III

Identical to Charcoal Centrifugation II with the exception of increasing the activated charcoal to 20 grams.

7.24 Charcoal Centrifugation IIII

Identical to Charcoal Centrifugation II with the exception of decreasing the activated charcoal to 0.5 grams, then increasing by 0.5 grams until better separation was observed via ^{31}P NMR, which was at 1.5 grams.

7.25 Strong Ion Exchange columns – Q-trap I and SP column

Q-trap and SP-columns were primed with 10 mM ammonium bicarbonate solution. The phosphonate fractions were dried down and reconstituted into DI H2O. 1mL was loaded onto a 1 mL Q-Trap column and 1 mL onto 1 mL SP-column simultaneously. Both columns were washed with three column volumes of ammonium carbonate solution at 10 mM, 3CV of 50 mM and 3CV of 1 M NaCl. Absence of phosphonate in each fraction confirmed using ^{31}P NMR.

7.26 Q-trap II

Q-trap primed with 10 mM ammonium bicarbonate solution. The phosphonate fractions were dried down and reconstituted into DI H2O. 1mL was loaded onto a 1 mL Q-Trap column. Column was washed with three column volumes of ammonium carbonate

solution at 10 mM, 3CV of 50 mM and 3CV at 500mM. Absence of phosphonate in each fraction confirmed using ^{31}P NMR.

7.27 Weak Ion Exchange Column -DEAE

15 mL of phosphonate containing mixture from charcoal purification I was injected onto a DEAE column with an increasing gradient of 10 – 100% 0.1 M ammonium carbonate mobile phase over 30 minutes at 1 mL/min. 1 mL fractions were collected every minute. These fractions were freeze dried and analysed through ^{31}P NMR to find the phosphonate compound. It was found in the initial run off. It was dried down and run on a DEAE column with an increasing gradient of 10 – 100% 1 M NaCl mobile phase over 30 minutes at 1 mL/min. 1 mL fractions were collected every minute. These fractions were freeze dried and analysed through ^{31}P NMR to find the phosphonate compound. The phosphonate was found in 4 fractions: 6, 7, 8 and 9 minutes.

7.28 Fe column

A nickel-NTA agarose column was washed with 3 column volumes of 50 mM EDTA, then 6 column volumes of H_2O. Then 3 column volumes of Fe(III)Cl_3. Fractions 6 – 9 from the weak ion exchange column were recombined and injected onto the column and washed with three column volumes of 0.1M Tris buffer at a flow rate of 1mL/min. ^{31}P NMR analysis on all fractions revealed no C-P peaks present.

7.29 Size Exclusion Sephadex – LH20 column

LH20 Sephadex resin was swelled with 20% ethanol in DI H_2O overnight. It was then poured into an empty column to make up a 130 mL SE column. After charcoal purification, phosphonate containing fractions were concentrated down to <10 mL *in vacuo* and injected onto the column at a flow rate of 0.25 ml/min in isocratic 20% ethanol until UV returned to baseline (typically three column volumes). 7.5 mL fractions were collected and lyophilised. Presence of phosphonate in each fraction confirmed using ^{31}P NMR.

7.30 Dowex column

40 g DOWEX (1 x 2 200-400 mesh Cl form) was swollen with DI H_2O and added to a 100 mL syringe. It was washed with 3CV of DI H_2O until the flowthrough was pH neutral. The phosphonate containing fractions were dried down and reconstituted in DI H_2O (3 mL) and added to the column. The column was washed with 3 CV of DI H_2O, then 50 mM ammonium bicarbonate solution, then 100 mM ammonium bicarbonate solution. The three fractions were dried down, then reconstituted in 666ul D_2O. Absence of phosphonate in each fraction confirmed using ^{31}P NMR.

7.31 Prep HPLC

A Phenomenex Luna C18 Prep HPLC column (5 µm, 100 × 21.2 mm) was used in conjunction with the Dionex ultiMate 3000 UHPLC. 1 mL of phosphonate containing mixture was injected onto the column. The mobile phase comprised of A; DI H_2O (0.1% FA) and B: acetonitrile (0.1% FA), the gradient ran 0 – 100%B over 10 minutes. Flow rate was set to 5 mL/min and fractions were collected every minute. Absence of phosphonate in each fraction confirmed using ^{31}P NMR.

7.32 Chapter Four: Algal Culturing and MAA Isolation

7.33 Biomass Extraction

Extraction 1 and 2

The cultures transferred into 50 mL falcons and centrifuged at 4500rpm for 30 minutes at 4 °C. The supernatant was removed and the biomass dried via lyophilisation. The biomass obtained 6 L over 2 extractions was: 1.30 g and 1.76 g. The biomass was then heated to reflux in 10% MeOH overnight. The resulting mixture was concentrated down in vacuo and filtered through a 0.2 µm syringe filter to remove particulates.

Extraction 1 and 3

Initial growth was conducted in six 1 L cultures of EG:JM at pH 4.5 with 30 mM glutamic acid for two weeks. Subsequently, 50 mL from each 1 L culture was transferred to fresh EG:JM medium supplemented with 15 g glucose to stimulate further growth. This modification resulted in a substantial increase in biomass production, yielding 12 g of dry weight from 6 L of culture.

7.34 Semi-Preparative High-Pressure Liquid Chromatography – C18 column

Sample extract fractionation was performed on a Kinetex® C18 column (100 Å, 100 mm x 21.2 mm, 5 µm; Phenomenex, CA, USA) using a Agilent 1260 Infinity II HPLC with DAD. Spectra files were acquired to Agilent InfinityLab Software. Flow rate was set to 5 mL/min. UV absorbance chromatograms were observed for wavelengths of 185 nm, 210 nm, 225 nm, 290 nm, 333 nm 372 nm, 400 nm, and 600 nm.

At low volume levels (< 10 mL), the extract was poorly soluble therefore the following runs were each repeated 5 times with 2 mL, and the fractions containing peaks at 333 nm were combined at the end of each run.

Mobile phase I: A: MQ Water(0.1 % FA) B: Acetonitrile (0.1% FA)

First Run: 0-5% B over 1 minute, increasing to 5-7% B over 5 minutes (6-minute total), followed by 7-9% B over 2 minutes (8-minute total), and finally from 9% to 100% B over 2 minutes (20-minute total). The largest 333nm observed was eluted at 6 minutes with 7% B.

Second Run: A gradient of 0-10% B over 10 minutes was used. A peak at 333nm believed to correspond to the target MAA eluted at 4 minutes with 4% B.

Mobile phase III A: 5% MeOH in MQ water(10mM ammonium formate, 0.1% FA)/MeOH (10mM ammonium formate, 0.1% FA)

First Run: 0-5% B over 1 minute, increasing to 5-7% B over 5 minutes (6-minute total), followed by 7-9% B over 2 minutes (8-minute total), and finally from 9% to 100% B over 2 minutes (20-minute total). The target compound (MAA) eluted at 6 minutes with 7% B.

Second Run: A gradient of 0-10% B over 10 minutes was used. A peak at 333nm believed to correspond to the target MAA eluted at 4 minutes with 4% B.

7.35 Semi-Preparative High-Pressure Liquid Chromatography – HILIC column

Sample extract fractionation was performed on Atlantis Silica HILIC Column, (100Å, 3 µm, 2.1 mm X 50 mm) using a Agilent 1260 Infinity II HPLC with DAD. Spectra files were acquired to Agilent InfinityLab Software. Flow rate was set to 5 mL/min. UV absorbance chromatograms were observed for wavelengths of 185 nm, 210 nm, 225 nm, 290 nm, 333 nm 372 nm, 400 nm, and 600 nm.

Mobile phase A: 9:1 acetonitrile: MQ H_2O (5mM ammonium formate) B: 1:1 acetonitrile: water (5mM ammonium formate)

333nm absorbing fractions from **Mobile Phase I** was dissolved in decreasing ratios of A:B mobile phase (500 ul) and centrifuged. The supernatant was then run on the mobile phase with the starting composition corresponding to the ratio of the supernatant mixture. 0.25 mL fractions were collected over 15 minute runs with the following gradients.

90% A: 90% A to 100% B over 15 minutes.

80% A: 90% A to 100% B over 15 minutes.

70% A: 90% A to 100% B over 15 minutes.

60% A: 90% A to 100% B over 15 minutes.

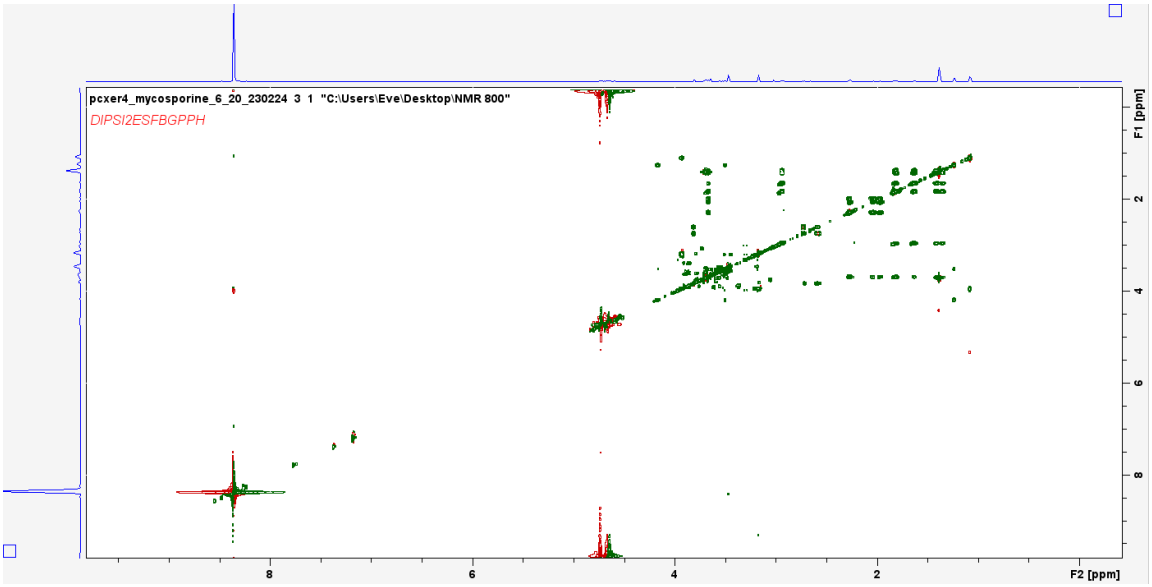
Fractions containing peaks with 333nm absorbance were submitted to NMR.

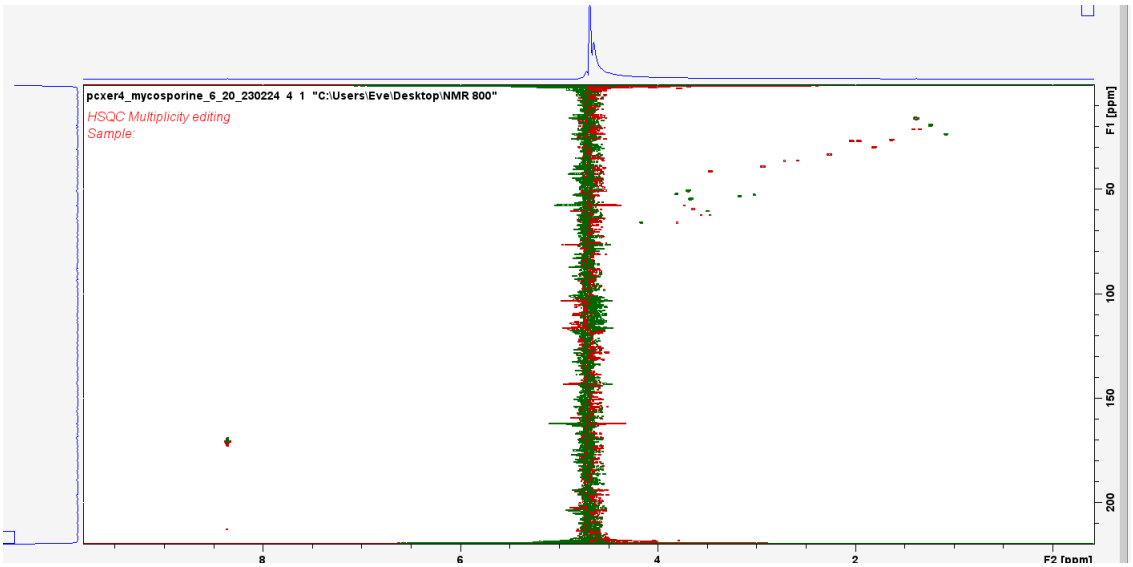
7.36 MAA HPLC-MS/MS

HPLC fractions from Mobile Phase I were analysed by HPLC-ESIQ/TOF-MS/MS in positive ionization mode using a Dionex Ultimate 3000 UHPLC (Thermo Scientific) coupled to a Bruker Daltonics Impact II Q-ToF ultra- high resolution mass spectrometer equipped with a diode array detector (DAD). HPLC was achieved using a Kinetex reverse-phase C18 column (100 mm x 2.1 mm, 3 µm) with mobile phases A: 5% MeOH in MQ water(10mM ammonium formate) and B: MeOH (10mM ammonium formate). The elution gradient was as follows: 0-5% B over 1 minute, increasing to 5-7% B over 5 minutes (6-minute total), followed by 7-9% B over 2 minutes (8-minute total), and finally from 9% to 100% B over 2 minutes (20-minute total). The column oven was set to 30˚C, the flow rate to 0.1ml/min. The MS1 was set to 300-1500 m/z, and the MS2 was untargeted, with a detection limit of 3 precursors per cycle. The fixed collision energy was set to 10 eV, source gas temperature 250 °C, gas flow 8 L/ min.

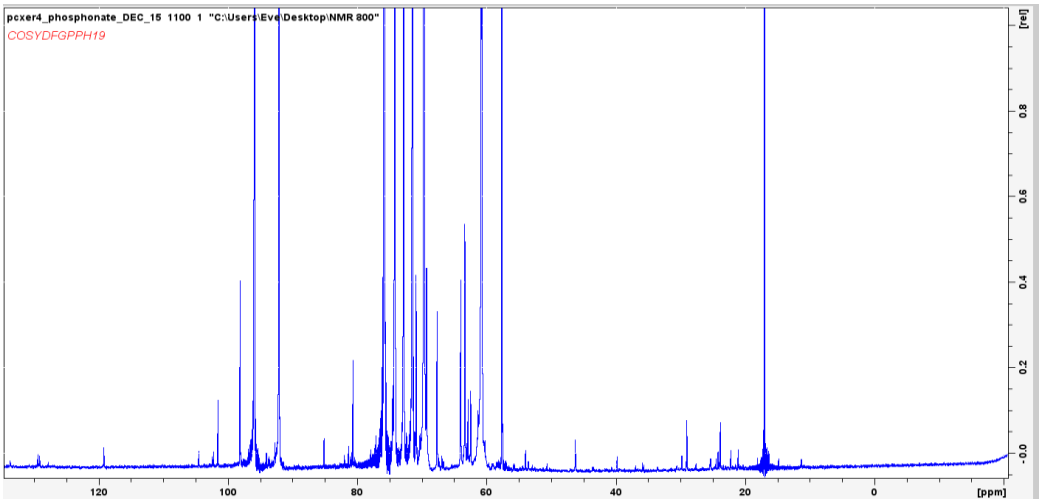
CCAP Code	Algal Species Present
1204.3	*Astasia pertyi*
1204.9	*Astasia ocellata var provasolii*
1204.12	*Astasia hallii*
1204.15	*Astasia klebsii*
1204.17B	*Astasia longa*
1204.20A	*Khawkinea quartana*
1211.3	*Colacium vesiculosum*
1216.3C	*Distigma proteus*
1224.4E	*Euglena geniculata*
1224.5Z	*Euglena gracilis*
1224.7A	*Euglena gracilis var saccharophila*
1224.9C	*Euglena mutabilis*
1224.17P	*Euglena schmitzii*
1224.27	*Euglena clara*
1224.31	*Euglena laciniata*
1224.32B	*Euglena cuneata*
1224.33	*Euglena cantabrica*
1224.35	*Euglena communis*
1224.44	*Euglena deses var intermedia*
1224.48	*Euglena mutabilis*
1224.5	*Euglena van-goori*
1261.6	*Phacus pusillus*
1261.8	*Phacus triqueter*
1261.9	*Monomorphina aenigmaticus*
1271.1	*Rhabdomonas costata*
1271.4	*Rhabdomonas incurva var majo*
1283.2	*Trachelomonas hispida var coronata*
1283.4B	*Trachelomonas volvocina*
1283.8	*Trachelomonas hispida*
1283.13	*Trachelomonas pertyi*
2149	*Euglena viridis*
ESNC	*Euglena sanguiniera*

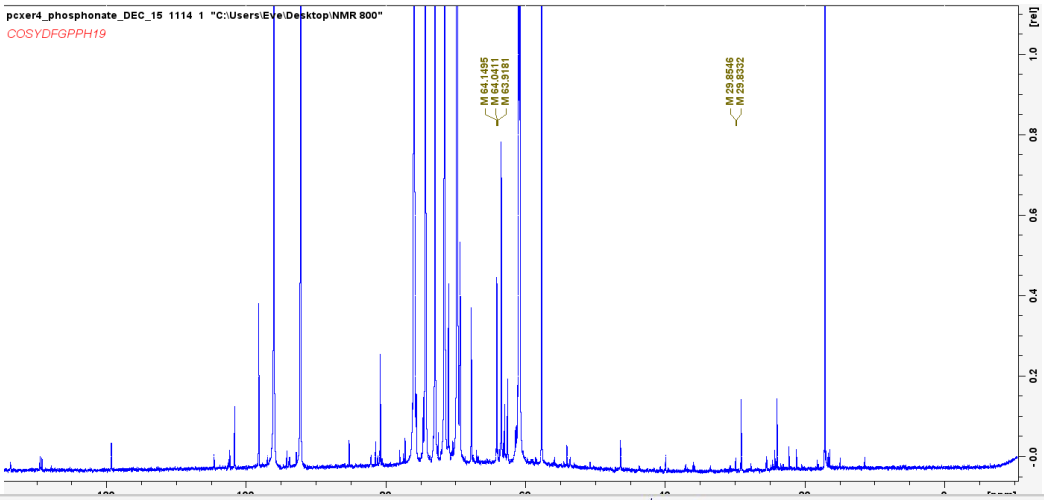
Mycosporine fraction 6 NMR

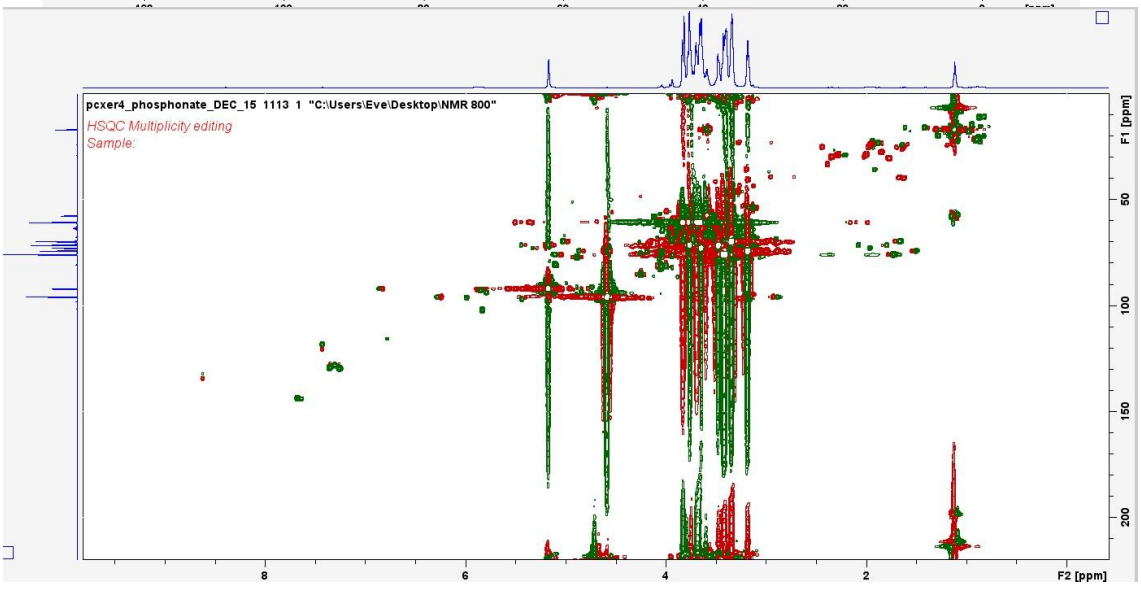




Phosphonate containing fraction NMR with P-C decoupling (top) and coupling (second from top) and HSQC (second from bottom) and DIPSI (bottom)







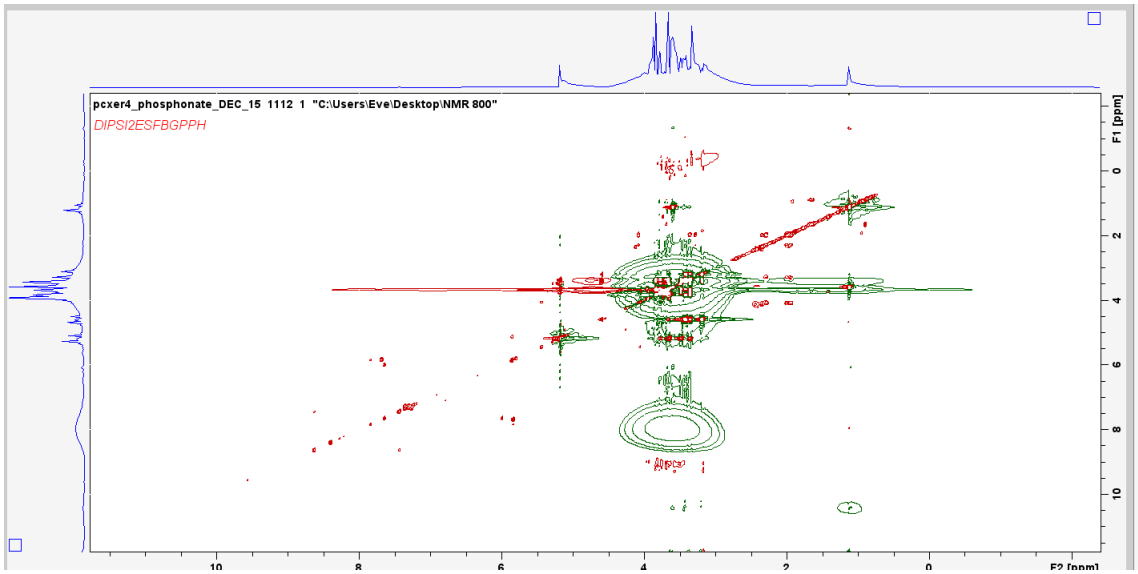


Table of known MAAs used for reference in M/S analysis in Chapter Five

Name	Type	mass	m/z	Max absorbance/nm	Producinig organisms
3-Dehydroquinate	Precursor	174.13	175.05	N/A	*Escherichia coli, Saccharomyces cerevisiae*
4-Deoxygadusol	Precursor	178.14	179.06	~268	*Gadus morhua, Nostoc punctiforme*
Mycosporine-glycine	MAA	198.18	199.07	310	*Anabaena variabilis, Chondrus crispus*
Gadusol	Precursor	210.16	211.06	~270	*Gadus morhua, Nostoc punctiforme*
Mycosporine-serine	MAA	214.2	215.08	310	*Anabaena variabilis, Bangia atropurpurea*
Mycosporine-taurine	MAA	242.22	243.07	309	*Anabaena variabilis, Porphyra umbilicalis*
Mycosporine-methylamine-serine	MAA	244.24	245.1	327	*Various marine organisms*
Mycosporine-glutamine	MAA	255.23	256.1	310	*Anabaena variabilis, Porphyra umbilicalis*

Mycosporine-methylamine-threonine	MAA	258.26	259.11	327	*Various marine organisms*
Mycosporine-2-glycine	MAA	260.24	261.1	334	*Various marine organisms*
Euhalothece-362	MAA	272.27	273.1	362	*Euhalothece sp.*
Asterina-330	MAA	288.3	289.11	330	*Asterina pectinifera, Porphyra yezoensis*
Palythine	MAA	288.3	289.11	320	*Palythoa tuberculosa, Aphanizomenon flos-aquae*
Usujirene	MAA	298.3	299.11	357	*Various marine organisms*
Palythinol	MAA	303.32	304.12	332	*Palythoa tuberculosa, Porphyra yezoensis*
Mycosporine-glycine-valine	MAA	310.32	311.14	335	*Various marine organisms*
Palythene	MAA	317.34	318.13	360	*Palythoa tuberculosa, Porphyra yezoensis*
Palythine-serine	MAA	318.32	319.13	320	*Various marine organisms*
Mycosporine-glutamic acid-glycine	MAA	324.3	325.13	330	*Various marine organisms*
Shinorine	MAA	332.31	333.12	334	*Porphyra yezoensis, Anabaena variabilis*
Palythenic acid	MAA	333.32	334.11	337	*Various marine organisms*
Porphyra-334	MAA	346.33	347.12	334	*Porphyra umbilicalis, Aphanizomenon flos-aquae*
Palythine-serine-sulfate	MAA	398.32	399.1	320	*Various marine organisms*

References

1. Salam, M. A. *et al.* Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. *Healthcare (Basel, Switzerland) 11*(13), (2023).

2. Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov. 20*(200–216), (2021).

3. Hubert, J., Nuzillard, J.-M. & Renault, J.-H. Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem. Rev. 16*(55–95), (2017).

4. Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery*. *20*(200–216), (2021).

5. Baltz, R. H. Genome mining for drug discovery: progress at the front end. *J. Ind. Microbiol. Biotechnol.* 48(9-10), (2021).

6. Zhang, L. *et al.* Advances in Metagenomics and Its Application in Environmental Microorganisms. *Front. Microbiol.* 12, (2021).

7. Alam, K., Abbasi, M. N., Hao, J., Zhang, Y. & Li, A. Strategies for Natural Products Discovery from Uncultured Microorganisms. *Molecules* 26, (2021).

8. Shukla, R. *et al.* An antibiotic from an uncultured bacterium binds to an immutable target. *Cell* 186, 4059-4073.e27 (2023).

9. Demarque, D. P. *et al.* Mass spectrometry-based metabolomics approach in the isolation of bioactive natural products. *Sci. Rep.* 10, 1–9 (2020).

10. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA. Cancer J. Clin.* 70, 7–30 (2020).

11. Saloni Dattani, Fiona Spooner, Hannah Ritchie, and Max Roser (2023) - “Causes of Death” Published online at OurWorldinData.org. Retrieved from: 'https://ourworldindata.org/causes-of-death' [Online Resource].

12. Saima May Sidik. "How COVID has deepened inequality — in six stark graphics," Nature, Nature, vol. 606(7915), pgs 638-639, (2022).

13. Blundell, *et al*. Inequality and the COVID-19 Crisis in the United Kingdom. Annual Review of Economics. 14 (2022)

14. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* 162, 1239–1249 (2011).

15. Barnes, E. C., Kumar, R. & Davis, R. A. The use of isolated natural products as scaffolds for the generation of chemically diverse screening libraries for drug discovery. *Nat. Prod. Rep.* 33, 372–381 (2016).

16. Pan, R., Bai, X., Chen, J., Zhang, H. & Wang, H. Exploring structural diversity of microbe secondary metabolites using OSMAC strategy: A literature review. *Frontiers in Microbiology* vol. 10 294

17. Kilcoyne, et al. 'Natural products and herbal medicines', *Pharmaceutical Medicine*, Oxford Specialist Handbooks (2013)

18. Vaou, N. *et al.* Interactions between Medical Plant-Derived Bioactive Compounds: Focus on Antimicrobial Combination Effects. *Antibiot. (Basel, Switzerland)* 11, (2022).

19. Hashem, S. *et al.* Targeting cancer signaling pathways by natural products: Exploring promising anti-cancer agents. *Biomed. Pharmacother.* 150, 113054 (2022).

20. Choudhari, A. S., Mandave, P. C., Deshpande, M., Ranjekar, P. & Prakash, O. Phytochemicals in Cancer Treatment: From Preclinical Studies to Clinical Practice. *Front. Pharmacol.* **10**, 1614 (2019).

21. Weaver, B. A. How Taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell* **25**, 2677–2681 (2014).

22. Prestinaci, F., Pezzotti, P. & Pantosti, A. Antimicrobial resistance: a global multifaceted phenomenon. *Pathog. Glob. Health* **109**, 309–318 (2015).

23. Muteeb, G., Rehman, M. T., Shahwan, M. & Aatif, M. Origin of Antibiotics and Antibiotic Resistance, and Their Impacts on Drug Development: A Narrative Review. *Pharmaceuticals (Basel).* **16**, (2023).

24. Stefanoudakis, D. *et al.* The Potential Revolution of Cancer Treatment with CRISPR Technology. *Cancers (Basel).* **15**, (2023).

25. Maryam, L. & Khan, A. U. A Mechanism of Synergistic Effect of Streptomycin and Cefotaxime on CTX-M-15 Type β-lactamase Producing Strain of E. cloacae: A First Report. *Front. Microbiol.* **7**, 2007 (2016).

26. Sullivan, G. J., Delgado, N. N., Maharjan, R. & Cain, A. K. How antibiotics work together: molecular mechanisms behind combination therapy. *Curr. Opin. Microbiol.* **57**, 31–40 (2020).

27. Demain, A. L. Antibiotics: Natural products essential to human health. *Med. Res. Rev.* **29**, 821–842 (2009).

28. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).

29. Wilken, R., Veena, M. S., Wang, M. B. & Srivatsan, E. S. Curcumin: A review of anti-cancer properties and therapeutic activity in head and neck squamous cell carcinoma. *Mol. Cancer* **10**, 12 (2011).

30. Moudi, M., Go, R., Yien, C. Y. S. & Nazre, M. Vinca alkaloids. *Int. J. Prev. Med.* **4**, 1231–1235 (2013).

31. Sankhuan, D., Niramolyanun, G., Kangwanrangsan, N., Nakano, M. & Supaibulwatana, K. Variation in terpenoids in leaves of Artemisia annua grown under different LED spectra resulting in diverse antimalarial activities against Plasmodium falciparum. *BMC Plant Biol.* **22**, 128 (2022).

32. Mills-Robertson, F. C., Tay, S. C. K., Duker-Eshun, G., Walana, W. & Badu, K. In vitro antimicrobial activity of ethanolic fractions of Cryptolepis sanguinolenta. *Ann. Clin. Microbiol. Antimicrob.* **11**, 16 (2012).

33. Tong, Y. *et al.* Berberine reverses multidrug resistance in Candida albicans by hijacking the drug efflux pump Mdr1p. *Sci. Bull.* **66**, 1895–1905 (2021).

34. Soares, G. M. S. *et al.* Mechanisms of action of systemic antibiotics used in periodontal treatment and mechanisms of bacterial resistance to these drugs. *J. Appl. Oral Sci.* **20**, 295–309 (2012).

35. Medema, M. H., de Rond, T. & Moore, B. S. Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.* (2021)

36. Sun, D., Gao, W., Hu, H. & Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* **12**, 3049–3062 (2022).

37. Naeem, A. *et al.* Natural Products as Anticancer Agents: Current Status and Future Perspectives. *Molecules* vol. 27 at (2022).

38. de la Torre, B. G. & Albericio, F. The Pharmaceutical Industry in 2019. An Analysis of FDA Drug Approvals from the Perspective of Molecules. *Molecules* **25**, (2020).

39. Achan, J. *et al.* Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. *Malar. J.* **10**, 144 (2011).

40. Carlin, M. G., Dean, J. R. & Ames, J. M. Opium Alkaloids in Harvested and Thermally Processed Poppy Seeds . *Frontiers in Chemistry* vol. 8 (2020).

41. Christensen, S. B. Drugs That Changed Society: History and Current Status of the Early Antibiotics: Salvarsan, Sulfonamides, and β-Lactams. *Molecules* **26**, (2021).

42. Tambo, E., Khater, E. I. M., Chen, J.-H., Bergquist, R. & Zhou, X.-N. Nobel prize for the artemisinin and ivermectin discoveries: a great boost towards elimination of the global infectious diseases of poverty. *Infectious diseases of poverty* vol. 4 58 at (2015).

43. Parra, J. *et al.* Antibiotics from rare actinomycetes, beyond the genus Streptomyces. *Curr. Opin. Microbiol.* **76**, 102385 (2023).

44. Matlock, A., Garcia, J. A., Moussavi, K., Long, B. & Liang, S. Y.-T. Advances in novel antibiotics to treat multidrug-resistant gram-negative bacterial infections. *Intern. Emerg. Med.* **16**, 2231–2241 (2021).

45. Kirst, H. A. Developing new antibacterials through natural product research. *Expert Opin. Drug Discov.* **8**, 479–493 (2013).

46. Zampaloni, C. *et al.* A novel antibiotic class targeting the lipopolysaccharide transporter. *Nature* **625**, 566–571 (2024).

47. Wilson, Z. E. & Brimble, M. A. Molecules derived from the extremes of life: a decade later. *Nat. Prod. Rep.* **38**, 24–82 (2021).

48. Theodoridis, S., Drakou, E. G., Hickler, T., Thines, M. & Nogues-Bravo, D. Evaluating natural medicinal resources and their exposure to global change. *Lancet. Planet. Heal.* **7**, e155–e163 (2023).

49. Tanaka, Y. *et al.* Marine-derived microbes and molecules for drug discovery. *Inflamm. Regen.* **42**, 18 (2022).

50. Lianza, M. *et al.* The Three Pillars of Natural Product Dereplication. Alkaloids from the Bulbs of Urceolina peruviana (C. Presl) J.F. Macbr. as a Preliminary Test Case. *Molecules* **26**, (2021).

51. Ito, T. & Masubuchi, M. Dereplication of microbial extracts and related analytical technologies. *J. Antibiot. (Tokyo).* **67**, 353–360 (2014).

52. Thomas, T., Gilbert, J. & Meyer, F. Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* **2**, 3 (2012).

53. Romano, S., Jackson, S. A., Patry, S. & Dobson, A. D. W. Extending the "one strain many compounds" (OSMAC) principle to marine microorganisms. *Marine Drugs* vol. 16, 244 (2018).

54. Wei, H., Lin, Z., Li, D., Gu, Q. & Zhu, T. [OSMAC (one strain many compounds) approach in the research of microbial metabolites--a review]. *Wei Sheng Wu Xue Bao* **50**, 701–709 (2010).

55. Barderas, M. G. *et al.* Metabolomic Profiling for Identification of Novel Potential Biomarkers in Cardiovascular Diseases. *J. Biomed. Biotechnol.* **2011**, 790132 (2011).

56. Ramsden, J. J. Metabolomics and Metabonomics - *Bioinformatics: An Introduction*. pg 1–9 (2009).

57. Ryan, D. & Robards, K. Metabolomics: The Greatest Omics of Them All? *Anal. Chem.* **78**, 7954–7958 (2006).

58. Strehmel, N., Böttcher, C., Schmidt, S. & Scheel, D. Profiling of secondary metabolites in root exudates of Arabidopsis thaliana. *Phytochemistry* **108**, 35–46 (2014).

59. Kleigrewe, K. *et al.* Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. *J. Nat. Prod.* **78**, 1671–1682 (2015).

60. Lowe, R., Shirley, N., Bleackley, M., Dolan, S. & Shafee, T. Transcriptomics technologies. *PLoS Comput. Biol.* **13**, e1005457 (2017).

61. Gao, W. *et al.* Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in Salvia miltiorrhiza. *BMC Genomics* **15**, 73 (2014).

62. Iqbal, H. A., Low-Beinart, L., Obiajulu, J. U. & Brady, S. F. Natural Product Discovery through Improved Functional Metagenomics in Streptomyces. *J. Am. Chem. Soc.* **138**, 9341–9344 (2016).

63. Kampa, A. *et al.* Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proc. Natl. Acad. Sci.* **110**, E3129–E3137 (2013).

64. Chang, F.-Y., Ternei, M. A., Calle, P. Y. & Brady, S. F. Targeted Metagenomics: Finding Rare Tryptophan Dimer Natural Products in the Environment. *J. Am. Chem. Soc.* **137**, 6044–6052 (2015).

65. Xu, J. & Hagler, A. Chemoinformatics and Drug Discovery. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry* vol. 7 566–600 (2002).

66. Medina-Franco, J. L. & Saldívar-González, F. I. Cheminformatics to Characterize Pharmacologically Active Natural Products. *Biomolecules* **10**, (2020).

67. Chen, Y. & Kirchmair, J. Cheminformatics in Natural Product-based Drug Discovery. *Mol. Inform.* **39**, 2000171 (2020).

68. Aron, A. T. *et al.* Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **15**, 1954–1991 (2020).

69. Augustin, R. *et al.* Lipopeptides as rhizosphere public goods for microbial cooperation. *Microbiol. Spectr.* **12**, e03106-23 (2023).

70. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* **449**, 804–810 (2007).

71. Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V, Jarmusch, A. K. & Dorrestein, P. C. Mass spectrometry-based metabolomics in microbiome investigations. *Nat. Rev. Microbiol.* **20**, 143–160 (2022).

72. Khan, A. K. *et al.* An Insight into the Algal Evolution and Genomics. *Biomolecules* **10**, 1524 (2020).

73. Chapman, R. L. Algae: the world's most important "plants"—an introduction. *Mitig. Adapt. Strateg. Glob. Chang.* **18**, 5–12 (2013).

74. Zuorro, A. *et al.* Natural Antimicrobial Agents from Algae: Current Advances and Future Directions. *International Journal of Molecular Sciences* vol. 25 (2024).

75. O'Neill, E. Mining Natural Product Biosynthesis in Eukaryotic Algae. *Mar. Drugs* **18**, 90 (2020).

76. Guiry, M. D. How many species of algae are there? *J. Phycol.* **48**, 1057–1063 (2012).

77. Wodniok, S. *et al.* Origin of land plants: Do conjugating green algae hold the key? *BMC Evol. Biol.* **11**, 104 (2011).

78. Onyeaka, H. *et al.* Minimizing carbon footprint via microalgae as a biological capture. *Carbon Capture Sci. Technol.* **1**, 100007 (2021).

79. Shukla, P. S., Borza, T., Critchley, A. T. & Prithiviraj, B. Carrageenans from Red Seaweeds As Promoters of Growth and Elicitors of Defense Response in Plants. *Front. Mar. Sci.* **3**, (2016).

80. Maia, J. L. da *et al.* Microalgae starch: A promising raw material for the bioethanol production. *Int. J. Biol. Macromol.* **165**, 2739–2749 (2020).

81. Christenhusz, M. & Byng, J. The number of known plant species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).

82. Derelle, E. *et al.* Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11647–11652 (2006).

83. Tolay, M., Amade, P., Cottalorda, J. M., Meinesz, A. & Jean, D. *Toxicologic Effects of the Invasive Seaweed Caulerpa taxifolia in the Mediterranean*. (1999).

84. Featherston, J. *et al.* The 4-Celled Tetrabaena socialis Nuclear Genome Reveals the Essential Components for Genetic Control of Cell Number at the Origin of Multicellularity in the Volvocine Lineage. *Mol. Biol. Evol.* **35**, 855–870 (2018).

85. Tymon, T. M. *et al.* Some aspects of the iodine metabolism of the giant kelp Macrocystis pyrifera (phaeophyceae). *J. Inorg. Biochem.* **177**, 82–88 (2017).

86. Baweja, P. & Sahoo, D. Classification of Algae BT - The Algae World. in (eds. Sahoo, D. &

Seckbach, J.) 31–55 (2015).

87. Nabout, J. C., da Silva Rocha, B., Carneiro, F. M. & Sant'Anna, C. L. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. *Biodivers. Conserv.* **22**, 2907–2918 (2013).

88. Mehdizadeh Allaf, M. & Peerhossaini, H. Cyanobacteria: Model Microorganisms and Beyond. *Microorganisms* **10**, (2022).

89. Carmichael, W. A world overview — One-hundred-twenty-seven years of research on toxic cyanobacteria — Where do we go from here? BT - Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs. 105–125 (2008).

90. Li, T. *et al.* Advances in investigating microcystin-induced liver toxicity and underlying mechanisms. *Sci. Total Environ.* **905**, 167167 (2023).

91. Raja, R., Hemaiswarya, S., Ganesan, V. & Carvalho, I. S. Recent developments in therapeutic applications of Cyanobacteria. *Crit. Rev. Microbiol.* **42**, 394–405 (2016).

92. Colas, S. *et al.* Anatoxin-a: Overview on a harmful cyanobacterial neurotoxin from the environmental scale to the molecular target. *Environ. Res.* **193**, 110590 (2021).

93. Gao, G., Wang, Y., Hua, H., Li, D. & Tang, C. Marine Antitumor Peptide Dolastatin 10: Biological Activity, Structural Modification and Synthetic Chemistry. *Marine Drugs* 19 (2021).

94. Singh, R. K., Tiwari, S. P., Rai, A. K. & Mohapatra, T. M. Cyanobacteria: an emerging source for drug discovery. *J. Antibiot. (Tokyo).* **64**, 401–412 (2011).

95. Cotas, J., Leandro, A., Pacheco, D., Gonçalves, A. M. M. & Pereira, L. A Comprehensive Review of the Nutraceutical and Therapeutic Applications of Red Seaweeds (Rhodophyta). *Life* vol. 10 (2020).

96. Al-Adilah, H. *et al.* Halogens in Seaweeds: Biological and Environmental Significance. *Phycology* vol. 2 132–171 at (2022).

97. Arulkumar, A., Rosemary, T., Paramasivam, S. & Rajendran, R. B. Phytochemical composition, in vitro antioxidant, antibacterial potential and GC-MS analysis of red seaweeds (Gracilaria corticata and Gracilaria edulis) from Palk Bay, India. *Biocatal. Agric. Biotechnol.* **15**, 63–71 (2018).

98. Zubia, M., Fabre, M.-S., Kerjean, V. & Deslandes, E. Antioxidant and cytotoxic activities of some red algae (Rhodophyta) from Brittany coasts (France). **52**, 268–277 (2009).

99. Abidizadegan, M., Peltomaa, E. & Blomster, J. The Potential of Cryptophyte Algae in Biomedical and Pharmaceutical Applications. *Front. Pharmacol.* **11**, 618836 (2020).

100. Hackett, J. D., Anderson, D. M., Erdner, D. L. & Bhattacharya, D. Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot.* **91**, 1523–1534 (2004).

101. Orefice, I., Balzano, S., Romano, G. & Sardo, A. Amphidinium spp. as a Source of Antimicrobial, Antifungal, and Anticancer Compounds. *Life* **13**, (2023).

102. Agarwal, T. Chrysophyaceae an Important Class of the Heterokontophyta. *J. Pharmacogn. Phytochem.* **6**, 1–4 (2018).

103. Weston, E. J., Eglit, Y. & Simpson, A. G. B. Kaonashia insperata gen. et sp. nov., a

eukaryotrophic flagellate, represents a novel major lineage of heterotrophic stramenopiles. *J. Eukaryot. Microbiol.* **71**, e13003 (2024).

104. Kai, A., Yoshii, Y., Nakayama, T. & Inouye, I. Aurearenophyceae classis nova, a New Class of Heterokontophyta Based on a New Marine Unicellular Alga Aurearena cruciata gen. et sp. nov. Inhabiting Sandy Beaches. *Protist* **159**, 435–457 (2008).

105. Aydın, B. Antibacterial Activities of Methanolic Extracts of Different Seaweeds from Iskenderun Bay, Turkey TT - Antibacterial Activities of Methanolic Extracts of Different Seaweeds from Iskenderun Bay, Turkey. *Int. J. Second. Metab.* **8**, 120–126 (2021).

106. McFadden, G. I., Gilson, P. R. & Hofmann, C. J. B. Division Chlorarachniophyta BT - Origins of Algae and their Plastids. in (ed. Bhattacharya, D.) 175–185 (1997).

107. Guiry, M. D. *et al.* AlgaeBase: An On-line Resource for Algae. *Cryptogam. Algol.* **35**, 105–115 (2014).

108. Fang, L., Leliaert, F., Zhang, Z. H., Penny, D. & Zhong, B. J. Evolution of the Chlorophyta: Insights from chloroplast phylogenomic analyses. *J. Syst. Evol.* **55**, 322–332 (2017).

109. Shah, S. A. A. *et al.* Chemically Diverse and Biologically Active Secondary Metabolites from Marine Phylum chlorophyta. *Mar. Drugs* **18**, (2020).

110. Novoveská, L., Zapata, A. K. M., Zabolotney, J. B., Atwood, M. C. & Sundstrom, E. R. Optimizing microalgae cultivation and wastewater treatment in large-scale offshore photobioreactors. *Algal Res.* **18**, 86–94 (2016).

111. Hudek, K., Davis, L. C., Ibbini, J. & Erickson, L. Commercial Products from Algae BT - Algal Biorefineries: Volume 1: Cultivation of Cells and Products. in (eds. Bajpai, R., Prokop, A. & Zappi, M.) 275–295 (2014)

112. Loke Show, P. Global market and economic analysis of microalgae technology: Status and perspectives. *Bioresour. Technol.* **357**, 127329 (2022).

113. Grossman, A. R. In the grip of algal genomics. *Adv. Exp. Med. Biol.* **616**, 54–76 (2007).

114. Kumar, G. *et al.* Bioengineering of Microalgae: Recent Advances, Perspectives, and Regulatory Challenges for Industrial Application. *Front. Bioeng. Biotechnol.* **8**, (2020).

115. Armbrust, E. V. *et al.* The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism. *Science (80-.).* **306**, 79–86 (2004).

116. Merchant, S. S. *et al.* The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **318**, 245–250 (2007).

117. Read, B. A. *et al.* Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature* **499**, 209–213 (2013).

118. Hanschen, E. R. & Starkenburg, S. R. The state of algal genome quality and diversity. *Algal Res.* **50**, 101968 (2020).

119. Stavridou, E. *et al.* Landscape of microalgae omics and metabolic engineering research for strain improvement: An overview. *Aquaculture* **587**, 740803 (2024).

120. Keeling, P. J. *et al.* The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through

transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).

121. Nelson, D. R. *et al.* Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. *Cell Host Microbe* **29**, 250-266.e8 (2021).

122. Ball, S. G. Eukaryotic microalgae genomics. The essence of being a plant. *Plant Physiol.* **137**, 397–398 (2005).

123. O'Neill, E. C. *et al.* The transcriptome of Euglena gracilis reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol. Biosyst.* **11**, 2808–2820 (2015).

124. Cagney, M. H. & O'Neill, E. C. Strategies for producing high value small molecules in microalgae. *Plant Physiol. Biochem.* **214**, 108942 (2024).

125. Jeong, B., Jang, J. & Jin, E. Genome engineering via gene editing technologies in microalgae. *Bioresour. Technol.* **373**, 128701 (2023).

126. Serif, M., Lepetit, B., Weißert, K., Kroth, P. G. & Rio Bartulos, C. A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom Phaeodactylum tricornutum. *Algal Res.* **23**, 186–195 (2017).

127. Weyman, P. D. *et al.* Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* **13**, 460–470 (2015).

128. Shin, Y. S. *et al.* Targeted knockout of phospholipase A2 to increase lipid productivity in Chlamydomonas reinhardtii for biodiesel production. *Bioresour. Technol.* **271**, 368–374 (2019).

129. Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M. & Weeks, D. P. Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryot. Cell* **13**, 1465–1469 (2014).

130. Greiner, A. *et al.* Targeting of Photoreceptor Genes in Chlamydomonas reinhardtii via Zinc-Finger Nucleases and CRISPR/Cas9. *Plant Cell* **29**, 2498–2518 (2017).

131. Hopes, A., Nekrasov, V., Kamoun, S. & Mock, T. Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. *Plant Methods* **12**, 49 (2016).

132. Zhang, Y.-T. *et al.* Application of the CRISPR/Cas system for genome editing in microalgae. *Appl. Microbiol. Biotechnol.* **103**, 3239–3248 (2019).

133. Schwartzbach, S. & Shigeoka, S. *Euglena: Biochemistry, Cell and Molecular Biology*. (2017)

134. Kostygov, A. Y. *et al.* Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* **11**, 200407 (2021).

135. Häder, D.-P. & Hemmersbach, R. Euglena, a Gravitactic Flagellate of Multiple Usages. *Life (Basel, Switzerland)* **12**, (2022).

136. Wijesinghe, P. Light-deformable microrobots shape up for the biological obstacle course. *Light Sci. Appl.* **13**, 103 (2024).

137. Bicudo, C. E. de M. & Menezes, M. Phylogeny and Classification of Euglenophyceae: A Brief Review. *Front. Ecol. Evol.* **4**, (2016).

138. Bedard, S., Roxborough, E., O'Neill, E. & Mangal, V. The biomolecules of Euglena

gracilis: Harnessing biology for natural solutions to future problems. *Protist* **175**, 126044 (2024).

139. Ebenezer, T. E. *et al.* Transcriptome, proteome and draft genome of Euglena gracilis. *BMC Biol.* **17**, 11 (2019).

140. Dooijes, D. *et al.* Base J originally found in Kinetoplastida is also a minor constituent of nuclear DNA of Euglena gracilis. *Nucleic Acids Res.* **28**, 3017–3021 (2000).

141. Ebenezer, T. E. *et al.* Euglena International Network (EIN): Driving euglenoid biotechnology for the benefit of a challenged world. *Biol. Open* **11**, (2022).

142. Fields, O., Hammond, M. J., Xu, X. & O'Neill, E. C. Advances in euglenoid genomics: unravelling the fascinating biology of a complex clade. *Trends Genet.* (2024)

143. O'Neill, E. C., Trick, M., Henrissat, B. & Field, R. A. Euglena in time: Evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspect. Sci.* **6**, 84–93 (2015).

144. Hallick, R. B. *et al.* Complete sequence of Euglena gracilis chloroplast DNA. *Nucleic Acids Res.* **21**, 3537–3544 (1993).

145. Maruyama, S., Suzaki, T., Weber, A. P. M., Archibald, J. M. & Nozaki, H. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evol. Biol.* **11**, 105 (2011).

146. Zheng, X., Wang, Y., Yang, T., He, Z. & Yan, Q. Size-fractioned aggregates within phycosphere define functional bacterial communities related to Microcystis aeruginosa and Euglena sanguinea blooms. *Aquat. Ecol.* **54**, 609–623 (2020).

147. Toyama, T. *et al.* Enhanced production of biomass and lipids by Euglena gracilis via co-culturing with a microalga growth-promoting bacterium, Emticicia sp. EG3. *Biotechnol. Biofuels* **12**, 205 (2019).

148. Shtratnikova, V. Y. *et al.* Complete genome assembly data of paenibacillus sp. RUD330, a hypothetical symbiont of euglena gracilis. *Data Br.* **32**, 106070 (2020).

149. Jasso-Chávez, R. *et al.* Microaerophilia enhances heavy metal biosorption and internal binding by polyphosphates in photosynthetic Euglena gracilis. *Algal Res.* **58**, 102384 (2021).

150. Mendoza-Cózatl, D. G. *et al.* Phytochelatin–cadmium–sulfide high-molecular-mass complexes of Euglena gracilis. *FEBS J.* **273**, 5703–5713 (2006).

151. Kaszecki, E. *et al.* Euglena mutabilis exists in a FAB consortium with microbes that enhance cadmium tolerance. *Int. Microbiol.* **27**, 1249–1268 (2024).

152. O'Neill, E. C. *et al.* Exploring the Glycans of Euglena gracilis. *Biology (Basel).* **6**, (2017).

153. Muchut, R. J. *et al.* Elucidating carbohydrate metabolism in Euglena gracilis: Reverse genetics-based evaluation of genes coding for enzymes linked to paramylon accumulation. *Biochimie* **184**, 125–131 (2021).

154. Kottuparambil, S., Thankamony, R. L. & Agusti, S. Euglena as a potential natural source of value-added metabolites. A review. *Algal Res.* **37**, 154–159 (2019).

155. Gissibl, A., Sun, A., Care, A., Nevalainen, H. & Sunna, A. Bioproducts From Euglena

gracilis: Synthesis and Applications . *Frontiers in Bioengineering and Biotechnology* vol. 7 (2019).

156. Kim, S. *et al.* Biofuel production from Euglena: Current status and techno-economic perspectives. *Bioresour. Technol.* **371**, 128582 (2023).

157. Xie, Y. *et al.* Paramylon from Euglena gracilis Prevents Lipopolysaccharide-Induced Acute Liver Injury. *Front. Immunol.* **12**, 797096 (2021).

158. Takeyama, H. *et al.* Production of antioxidant vitamins, beta-carotene, vitamin C, and vitamin E, by two-step culture of Euglena gracilis Z. *Biotechnol. Bioeng.* **53**, 185–190 (1997).

159. Zimba, P. V *et al.* Euglenophycin is produced in at least six species of euglenoid algae and six of seven strains of Euglena sanguinea. *Harmful Algae* **63**, 79–84 (2017).

160. Sorokina, M. & Steinbeck, C. Review on natural products databases: where to find data in 2020. *J. Cheminform.* **12**, 20 (2020).

161. Bouslimani, A., Sanchez, L. M., Garg, N. & Dorrestein, P. C. Mass spectrometry of natural products: Current, emerging and future technologies. *Natural Product Reports* vol. 31 718–729 (2014).

162. National Library of Medicine. *ChemIDplus Advanced*. National Library of Medicine. (1999)

163. Davis & Vasanthi, Seaweed Metabolite Database *Bioinformation* 5(8): 361-364 (2011).

164. MassBank | MassBank Europe Mass Spectral DataBase. https://massbank.eu/MassBank/.

165. MassBank of North America. https://mona.fiehnlab.ucdavis.edu/spectra/statistics?tab=0.

166. Xue, J., Guijas, C., Benton, H. P., Warth, B. & Siuzdak, G. METLIN MS2 molecular standards database: a broad chemical and biological resource. *Nature Methods* vol. 17 953–954 (2020).

167. Vincenti, F. *et al.* Molecular Networking: A Useful Tool for the Identification of New Psychoactive Substances in Seizures by LC–HRMS. *Front. Chem.* **8**, 572952 (2020).

168. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* vol. 34 828–837 at (2016).

169. Kind, T. & Fiehn, O. Strategies for dereplication of natural compounds using high-resolution tandem mass spectrometry. *Phytochemistry Letters* vol. 21 313–319 (2017).

170. Wibowo, J. T. *et al.* Biotechnological Potential of Bacteria Isolated from the Sea Cucumber Holothuria leucospilota and Stichopus vastus from Lampung, Indonesia. *Marine Drugs* vol. 17 (2019).

171. Chen, J. *et al.* Caffeoylquinic acid derivatives isolated from the aerial parts of Gynura divaricata and their yeast α-glucosidase and PTP1B inhibitory activity. *Fitoterapia* **99**, 1–6 (2014).

172. Naumoska, K., Jug, U., Metličar, V. & Vovk, I. Oleamide, a Bioactive Compound,

Unwittingly Introduced into the Human Body through Some Plastic Food/Beverages and Medicine Containers. *Foods* vol. 9 at (2020).

173. Neupane, R. P. *et al.* Cytotoxic Sesquiterpenoid Quinones and Quinols, and an 11-Membered Heterocycle, Kauamide, from the Hawaiian Marine Sponge Dactylospongia elegans. *Marine Drugs* 17 (2019).

174. Park, S. *et al.* Smenospongidine suppresses the proliferation of multiple myeloma cells by promoting CCAAT/enhancer-binding protein homologous protein-mediated β-catenin degradation. *Arch. Pharm. Res.* **40**, 592–600 (2017).

175. Lin, H.-Y. & Lin, H.-J. Polyamines in Microalgae: Something Borrowed, Something New. *Mar. Drugs* **17**, 1 (2018).

176. O'Neill, E. C. *et al.* The transcriptome of Euglena gracilis reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol. Biosyst.* **11**, 2808–2820 (2015).

177. de Vries, R. P. *et al.* Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. *Genome Biol.* **18**, 28 (2017).

178. Changi, S., Matzger, A. J. & Savage, P. E. Kinetics and pathways for an algal phospholipid (1,2-dioleoyl-sn-glycero-3-phosphocholine) in high-temperature (175–350 °C) water. *Green Chem.* **14**, 2856–2867 (2012).

179. Lauritano, C. *et al.* Lysophosphatidylcholines and Chlorophyll-Derived Molecules from the Diatom Cylindrotheca closterium with Anti-Inflammatory Activity. *Mar. Drugs* **18**, 166 (2020).

180. Meesil, W. *et al.* Genome mining reveals novel biosynthetic gene clusters in entomopathogenic bacteria. *Sci. Rep.* **13**, 20764 (2023).

181. Bind, S., Bind, S., Sharma, A. K. & Chaturvedi, P. Epigenetic Modification: A Key Tool for Secondary Metabolite Production in Microorganisms. *Front. Microbiol.* **13**, (2022).

182. Hassan, S. S. U. *et al.* Stress Driven Discovery of Natural Products From Actinobacteria with Anti-Oxidant and Cytotoxic Activities Including Docking and ADMET Properties. *Int. J. Mol. Sci.* **22**, (2021).

183. He, J. *et al.* Metabolic Responses of a Model Green Microalga Euglena gracilis to Different Environmental Stresses. *Front. Bioeng. Biotechnol.* **9**, 662655 (2021).

184. Peng, X.-Y. *et al.* Co-culture: stimulate the metabolic potential and explore the molecular diversity of natural products from microorganisms. *Mar. life Sci. Technol.* **3**, 363–374 (2021).

185. Han, J. *et al.* Co-culturing bacteria and microalgae in organic carbon containing medium. *J. Biol. Res. (Thessalonike, Greece)* **23**, (2016).

186. Bertrand, S. *et al.* Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery. *Biotechnol. Adv.* **32**, 1180–1204 (2014).

187. Kim, J. Y., Oh, J.-J., Jeon, M. S., Kim, G.-H. & Choi, Y.-E. Improvement of Euglena gracilis Paramylon Production through a Cocultivation Strategy with the Indole-3-Acetic Acid-Producing Bacterium Vibrio natriegens. *Appl. Environ. Microbiol.* **85**, (2019).

188. Yuan, B. *et al.* Identification of the Neoaspergillic Acid Biosynthesis Gene Cluster by Establishing an In Vitro CRISPR-Ribonucleoprotein Genetic System in Aspergillus melleus. *ACS omega* **8**, 16713–16721 (2023).

189. Cueto, M. *et al.* Pestalone, a new antibiotic produced by a marine fungus in response to bacterial challenge. *J. Nat. Prod.* **64**, 1444–1446 (2001).

190. Oh, D. C., Kauffman, C. A., Jensen, P. R. & Fenical, W. Induced production of emericellamides A and B from the marine-derived fungus Emericella sp. in competing co-culture. *J. Nat. Prod.* **70**, 515–520 (2007).

191. Labeda, D. P. *et al.* Phylogenetic study of the species within the family Streptomycetaceae. *Antonie Van Leeuwenhoek* **101**, 73–104 (2012).

192. Ridgway, N. D. The role of phosphatidylcholine and choline metabolites to cell proliferation and survival. *Crit. Rev. Biochem. Mol. Biol.* **48**, 20–38 (2013).

193. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem* **3**, 619–627 (2002).

194. Dittmann, E., Fewer, D. P. & Neilan, B. A. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. *FEMS Microbiol. Rev.* **37**, 23–43 (2013).

195. Rhee, K.-H. Cyclic dipeptides exhibit synergistic, broad spectrum antimicrobial effects and have anti-mutagenic properties. *Int. J. Antimicrob. Agents* **24**, 423–427 (2004).

196. Vela-Corcia, D. *et al.* Cyclo(Pro-Tyr) elicits conserved cellular damage in fungi by targeting the [H+]ATPase Pma1 in plasma membrane domains. *Commun. Biol.* **7**, 1253 (2024).

197. Galezowska, J. & Gumienna-Kontecka, E. Phosphonates, their complexes and bio-applications: A spectrum of surprising diversity. *Coord. Chem. Rev.* **256**, 105–124 (2012).

198. Turhanen, P. A., Demadis, K. D. & Kafarski, P. Editorial: Phosphonate Chemistry in Drug Design and Development. *Frontiers in chemistry* 9 (2021).

199. Burchacka, E. *et al.* Development and binding characteristics of phosphonate inhibitors of SplA protease from Staphylococcus aureus. *Protein Sci.* **23**, 179 (2014).

200. Oleksyszyn, J. & Powers, J. C. B. T.-M. in E. [30] Amino acid and peptide phosphonate derivatives as specific inhibitors of serine peptidases. in *Proteolytic Enzymes: Serine and Cysteine Peptidases* vol. 244 423–441 (Academic Press, 1994).

201. Allen, K. N. & Dunaway-Mariano, D. Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci.* **29**, 495–503 (2004).

202. Metcalf, W. W. & Van Der Donk, W. A. Biosynthesis of Phosphonic and Phosphinic Acid Natural Products. **78**, 65–94 (2009).

203. Bown, L. *et al.* A Novel Pathway for Biosynthesis of the Herbicidal Phosphonate Natural Product Phosphonothrixin Is Widespread in Actinobacteria. *J. Bacteriol.* **205**, (2023).

204. Yu, X. *et al.* Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20759–20764 (2013).

205. Ju, K. S., Doroghazi, J. R. & Metcalf, W. W. Genomics-Enabled Discovery of Phosphonate Natural Products and their Biosynthetic Pathways. *J. Ind. Microbiol. Biotechnol.* **41**, 345 (2014).

206. Cao, Y., Peng, Q., Li, S., Deng, Z. & Gao, J. The intriguing biology and chemistry of fosfomycin: the only marketed phosphonate antibiotic. *RSC Adv.* **9**, 42204–42218 (2019).

207. Hoerlein, G. Glufosinate (phosphinothricin), a natural amino acid with unexpected herbicidal properties. *Rev. Environ. Contam. Toxicol.* **138**, 73–145 (1994).

208. Wiesner, J. *et al.* FR-900098, an antimalarial development candidate that inhibits the non-mevalonate isoprenoid biosynthesis pathway, shows no evidence of acute toxicity and genotoxicity. *Virulence* **7**, 718–728 (2016).

209. Borisova, S. A., Circello, B. T., Zhang, J. K., van der Donk, W. A. & Metcalf, W. W. Biosynthesis of rhizocticins, antifungal phosphonate oligopeptides produced by Bacillus subtilis ATCC6633. *Chem. Biol.* **17**, 28 (2010).

210. Circello, B. T., Miller, C. G., Lee, J. H., Van Der Donk, W. A. & Metcalf, W. W. The antibiotic dehydrophos is converted to a toxic pyruvate analog by peptide bond cleavage in Salmonella enterica. *Antimicrob. Agents Chemother.* **55**, 3357–3362 (2011).

211. Gahungu, M. *et al.* Synthesis and biological evaluation of potential threonine synthase inhibitors: Rhizocticin A and Plumbemycin A. *Bioorg. Med. Chem.* **21**, 4958–4967 (2013).

212. Gatti, D. & Adami, S. New Bisphosphonates in the Treatment of Bone Diseases. *Drugs Aging* **15**, 285–296 (1999).

213. Wolanczyk, M. J., Fakhrian, K. & Adamietz, I. A. Radiotherapy, Bisphosphonates and Surgical Stabilization of Complete or Impending Pathologic Fractures in Patients with Metastatic Bone Disease. *J. Cancer* **7**, 121 (2016).

214. Grzywa, R. & Sieńczyk, M. Phosphonic esters and their application of protease control. *Curr. Pharm. Des.* **19**, 1154–1178 (2013).

215. O'Neill, E. C. *et al.* Exploring the Glycans of Euglena gracilis. *Biology (Basel).* **6**, (2017).

216. Feng, W. *et al.* Forms and Lability of Phosphorus in Algae and Aquatic Macrophytes Characterized by Solution 31P NMR Coupled with Enzymatic Hydrolysis. *Sci. Reports 2016 61* **6**, 1–10 (2016).

217. Lee, S.-M., Kim, J.-G., Jeong, W.-G., Alessi, D. S. & Baek, K. Adsorption of antibiotics onto low-grade charcoal in the presence of organic matter: Batch and column tests. *Chemosphere* **346**, 140564 (2024).

218. Derlet, R. Activated Charcoal—Past, Present and Future. *West. J. Med.* **145**, 493–496 (1986).

219. Neeser, J.-R., Tronchet, J. & Charollais, E. Structural analysis of 3-C-phosphonates, -phosphinates, and -phosphine oxides of branched-chain sugars. *Can. J. Chem.* **61**, 2112–2120 (2011).

220. Ernst, L. 13C n.m.r. spectroscopy of diethyl alkyl- and benzyl-phosphonates. A study of phosphorus–carbon spin–spin coupling constants over one to seven bonds. *Org. Magn. Reson.* **9**, 35–43 (1977).

221. Rosic, N. N. & Dove, S. Mycosporine-like amino acids from coral dinoflagellates. *Appl.*

Environ. Microbiol. **77**, 8478–8486 (2011).

222. Punchakara, A., Prajapat, G., Bairwa, H. K., Jain, S. & Agrawal, A. Applications of mycosporine-like amino acids beyond photoprotection. *Appl. Environ. Microbiol.* **89**, e0074023 (2023).

223. Kurita, T. *et al.* Efficient and multiplexable genome editing using Platinum TALENs in oleaginous microalga, Nannochloropsis oceanica NIES-2145. *Genes Cells* **25**, 695–702 (2020).

224. Shick, J. M. & Dunlap, W. C. Mycosporine-like amino acids and related Gadusols: biosynthesis, acumulation, and UV-protective functions in aquatic organisms. *Annu. Rev. Physiol.* **64**, 223–262 (2002).

225. Chen, M., Rubin, G. M., Jiang, G., Raad, Z. & Ding, Y. Biosynthesis and Heterologous Production of Mycosporine-Like Amino Acid Palythines. *J. Org. Chem.* **86**, 11160–11168 (2021).

226. Sun, Y. *et al.* Distribution, Contents, and Types of Mycosporine-Like Amino Acids (MAAs) in Marine Macroalgae and a Database for MAAs Based on These Characteristics. *Mar. Drugs* **18**, (2020).

227. Parailloux, M., Godin, S., Fernandes, S. C. M. & Lobinski, R. Untargeted Analysis for Mycosporines and Mycosporine-Like Amino Acids by Hydrophilic Interaction Liquid Chromatography (HILIC)-Electrospray Orbitrap MS(2)/MS(3). *Antioxidants (Basel, Switzerland)* **9**, (2020).

228. Chrapusta, E. *et al.* Mycosporine-Like Amino Acids: Potential Health and Beauty Ingredients. *Mar. Drugs* **15**, (2017).

229. Suh, S.-S. *et al.* Anti-inflammation activities of mycosporine-like amino acids (MAAs) in response to UV radiation suggest potential anti-skin aging activity. *Mar. Drugs* **12**, 5174–5187 (2014).

230. Gacesa, R. *et al.* P 011 - Mycosporine-like amino acid activation of the Keap1-Nrf2 pathway. *Free Radic. Biol. Med.* **108**, S21 (2017).

231. Zwerger, M., Schwaiger, S. & Ganzera, M. Efficient Isolation of Mycosporine-Like Amino Acids from Marine Red Algae by Fast Centrifugal Partition Chromatography. *Mar. Drugs* **20**, (2022).

232. Cardozo, K. H. M., Carvalho, V. M., Pinto, E. & Colepicolo, P. Fragmentation of mycosporine-like amino acids by hydrogen/deuterium exchange and electrospray ionisation tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**, 253–258 (2006).

233. Bjarnadóttir, M. *et al.* Palmaria palmata as an alternative protein source: enzymatic protein extraction, amino acid composition, and nitrogen-to-protein conversion factor. *J. Appl. Phycol.* **30**, 2061–2070 (2018).

234. Pereira, D. T. *et al.* Optimizing the Extraction of Bioactive Compounds from Porphyra linearis (Rhodophyta): Evaluating Alkaline and Enzymatic Hydrolysis for Nutraceutical Applications. *Mar. Drugs* **22**, (2024).

235. Del Mondo, A., Sansone, C. & Brunet, C. Insights into the biosynthesis pathway of phenolic compounds in microalgae. *Comput. Struct. Biotechnol. J.* **20**, 1901–1913 (2022).

236. Reinbothe, C., Ortel, B., Parthier, B. & Reinbothe, S. Cytosolic and plastid forms of 5-enolpyruvylshikimate-3-phosphate synthase in Euglena gracilis are differentially expressed during light-induced chloroplast development. *Mol. Gen. Genet. MGG* **245**, 616–622 (1994).

237. Llewellyn, C. A. *et al.* Mycosporine-like amino acid and aromatic amino acid transcriptome response to UV and far-red light in the cyanobacterium Chlorogloeopsis fritschii PCC 6912. *Sci. Rep.* **10**, 20638 (2020).

238. Iwasaki, K. *et al.* Visualizing wax ester fermentation in single Euglena gracilis cells by Raman microspectroscopy and multivariate curve resolution analysis. *Biotechnol. Biofuels* **12**, 128 (2019).

239. Lihanová, D. *et al.* Versatile biotechnological applications of Euglena gracilis. *World J. Microbiol. Biotechnol.* **39**, 133 (2023).