Genomics-led Discovery of Antimicrobials From *Scytonema*

Nathan Henry Sprague Byrne

John Innes Centre

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy (PhD)

September 2024

Abstract

Natural products are an important source of antibiotics and other clinically relevant compounds. Unfortunately, the rise of antibiotic resistance has rendered existing antibiotics ineffective in treating multidrug resistant pathogens. This development has coincided with a drop-off in the discovery of novel natural products. The combination of these two factors has created a desperate need for the discovery of novel bioactive molecules.

One barrier to finding novel bioactive molecules is the constant rediscovery of the same compounds which requires lengthy dereplication. By investigating understudied bacterial species, the chance of rediscovering an existing molecule is lower and thus the chance of finding something novel is higher. *Scytonema* are a genus of cyanobacteria of the order *Nostocales*. When the genome sequence for *Scytonema hofmannii* PCC 7110 was published it was the largest bacterial genome sequenced to date. The genome, over 12 MB, is predicted to produce over 30 specialised metabolites; only six molecules have been isolated from this strain. Apart from *Scytonema hofmannii* PCC 7110 very little is known about the rest of the *Scytonema* genus. Many *Scytonema* strains have no published literature about them, let alone investigations into their ability to produce specialised metabolites.

In this thesis, the ability of the *Scytonema* genus to produce novel specialised metabolites was investigated. The genomes of 10 strains of *Scytonema* were sequenced, where none had previously publicly available genomic information. These genomes were then analysed to elucidate their ability to produce novel specialised metabolites. Screening was undertaken to determine if these strains produced bioactive metabolites, and attempts were made to trigger the production of novel specialised metabolites. This work revealed two antimicrobial compounds. The first, a known compound, led to attempts to determine the biosynthetic gene cluster responsible for its production. The second, a novel compound, led to attempts to structurally characterise this molecule.

Acknowledgements

First and foremost, I would like to thank the two greatest post docs a PhD student could hope for, Natalia Miguel-Vior and Javier Santos-Aberturas. Their unwavering support and combined supreme knowledge has guided me through every step of my PhD journey, there is not a chance that I would have completed the PhD without them. I will be eternally grateful for their friendship and mentoring. In a very close second, I must thank my super visor, Andy Truman, for not only putting up with me for the last four years, but for all the support and guidance throughout my PhD journey. I am also grateful for my secondary supervisor, David Lea-Smith, for all the invaluable scientific input. I would also like to thank Jose Moreno Cabezuelo for his cyanobacterial expertise.

I'm grateful to have been in the wonderful Truman group and thankful for many members past and present who have taught me so much. I'd like to thank Alaster Moffat, Jonny Ford, Louis Fong and Liam Chapman for all the great memories. I must also thank my protégé, Kenny Budiman, for his assistance and all the memorable moments. The Molecular Microbiology department has been a fantastic place to work and I have positive memories of almost every member. A few people I am particularly grateful for are Catriona Thompson, Joseph Sallmen and Max Jordan. Your friendship over the years has meant so much.

Finally I'd like to thank Annie for her unending support and love during the final year of my PhD and whilst I was writing my thesis. I must also thank the sport of ultrarunning and the Norfolk Gazelles for keeping me somewhat sane during this PhD Journey.

This work was funded by the UKRI-BBSRC Norwich Research Park Biosciences Doctoral Training Partnership.

Author's Declaration

The research described in this thesis was conducted at the John Innes Centre between October 2020 and September 2024. All data described here are original and were obtained by the author unless otherwise attributed in the text. No part of this thesis has previously been submitted for a degree at this or any other academic institution. Work pertaining to the validation of the scytoscalarol BGC and the third screening attempt were assisted by UEA year in industry student Kenny Budiman.

List of Abbreviations

Abbreviation	Definition
ACP	Acyl carrier protein
ANI	Average nucleotide identity
BGC	Biosynthetic gene cluster
DH	De-hydratase
DMAPP	Dimethylallyl pyrophosphate
DNA	Deoxyribonucleic acid
EIC	Extracted ion chromatogram
ELSD	Evaporative light scattering detector
ER	Enoylreductase
EtOAc	Ethyl acetate
GCF	Gene cluster family
GFPP	Geranylfarnesyl diphosphate
GNPS	Global natural product social molecular networking
HMM	Hidden markov model
HPLC	High-performance liquid chromatography
IPP	Isopentyl pyrophosphate
IPTG	Isopropyl β-d-1-thiogalactopyranoside
KR	Ketoreductase
KS	Ketosynthase
Lan	Lanthionine
LB	Lysogeny broth
LC-MS	Liquid chromatography – mass spectrometry
MB	Megabase
MeLan	Methyllanthionine
MEP	Methyl-D-erythritol phosphate
MiBIG	Minimum information about a biosynthetic gene cluster
MIC	Minimum inhibitory concentration

MRSA	Methicillin resistant *Staphylococcus aureus*
MT	Methyl-transferase
MVA	Mevalonic acid
NCBI	National centre for biotechnology information
NEB	New England biolabs
NMR	Nuclear magnetic resonance
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
OD	Optical density
OSMAC	One strain many available compounds
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PKS	Polyketide synthase
QIB	Quadram Institute of Bioscience
RBS	Ribosome binding site
RiPP	Ribosomally synthesised post-translationally modified peptide
RPM	Revolutions per minute
RRE	RiPP recognition element
rRNA	Ribosomal ribonucleic acid
SDS-PAGE	Sodium dodecyl sulphate poly-acrylamide gel electrophoresis
SNA	Soft nutrient agar
TE	Thioesterase
UEA	University of East Anglia
VRE	Vancomycin resistant *Enterococcus*
WHO	World health organisation

List of Figures

List of Tables

Table of Contents

Chapter 1: Introduction

1.1 The history of antibiotics and their discovery

The discovery and introduction of antibiotics during the twentieth century may have been the largest positive development in human health. It was not until the twentieth century that Salvarsan, an antibiotic used to treat syphilis, was introduced in 1910[1] and was the first antibiotic to be used clinically. Paul Ehrlich, who discovered Salvarsan, referred to his discovery as chemotherapy as he was using a chemical to treat disease. It was almost twenty years later that the next major advancement was made, when Alexander Fleming famously observed the action of penicillin on contaminated petri dishes. However, Fleming was unable to isolate penicillin and was unable to convince the wider scientific community of the importance of his observation. Penicillin was first purified in 1940, by Norman Heatley[2]. The first clinical trial with pure penicillin was carried out by Alexander Fleming in 1943 when he successfully treated a patient suffering from *Streptococcal meningitis*[3]. It was not until 1945 that the structure of penicillin was known. Dorothy Hodgkin confirmed the beta-lactam structure by X-ray crystallography[4], which allowed the creation of synthetic derivatives to overcome resistance[5,6,7] (Figure 1.1).

The first use of the term 'antibiotic' was by Selman Waksman in the late 1930's. Waksman was the first to identify that the *Streptomyces* genus was a prolific producer of antimicrobial natural products[8]. In 1943 Waksman isolated streptomycin (figure 1.1) from two strains of *Streptomyces griseus*. Streptomycin was the first effective treatment for tuberculosis and Waksman was awarded the Nobel prize in physiology and medicine for its discovery in 1952. He later, in 1949, isolated neomycin from *Streptomyces fradiae.* These discoveries paved the way for a period in time, spanning the 1950s and 60s is called 'The golden age' of antibiotic discovery. Over half of the antimicrobials currently in use today were discovered during this time[9].

The mass discovery of antibiotics unsurprisingly led to the mass use of antibiotics. This resulted in the emergence of antibiotic resistance. Random mutation and horizontal gene

transfer are mechanisms by which bacteria are quickly able to develop resistance to antibiotics[10]. During the golden age, new antimicrobials were being discovered fast enough to overcome this developing resistance. However, in the nineteen-eighties antibiotic discovery had slowed down enough to make emerging resistance a real threat. From the nineteen-fifty's through to the eighties most natural product discovery was done by screening bacteria and fungi against pathogenic microbes. Any positive hits would be purified, chemically characterised and tested clinically. The *Streptomyces* genus proved to be particularly fruitful, producing over fifty-five percent of antimicrobials discovered during this time[11]. Many natural product families were discovered this way. From nineteen-seventy up until the twenty-first century screening for natural products continued, however only one, daptomycin[12], obtained clinical approval. One reason for this drop in new antibiotics is that all easily available antimicrobials had been discovered, and more intensive methods of discovery were required. This led to continuous rediscovery of the same compounds meaning dereplication is always required. During this time many semi or wholly synthetic derivatives of existing antimicrobials were developed[13]. By the year two thousand most pharmaceutical companies had stopped screening natural products and instead switched to screening libraries of synthetic compounds[14].

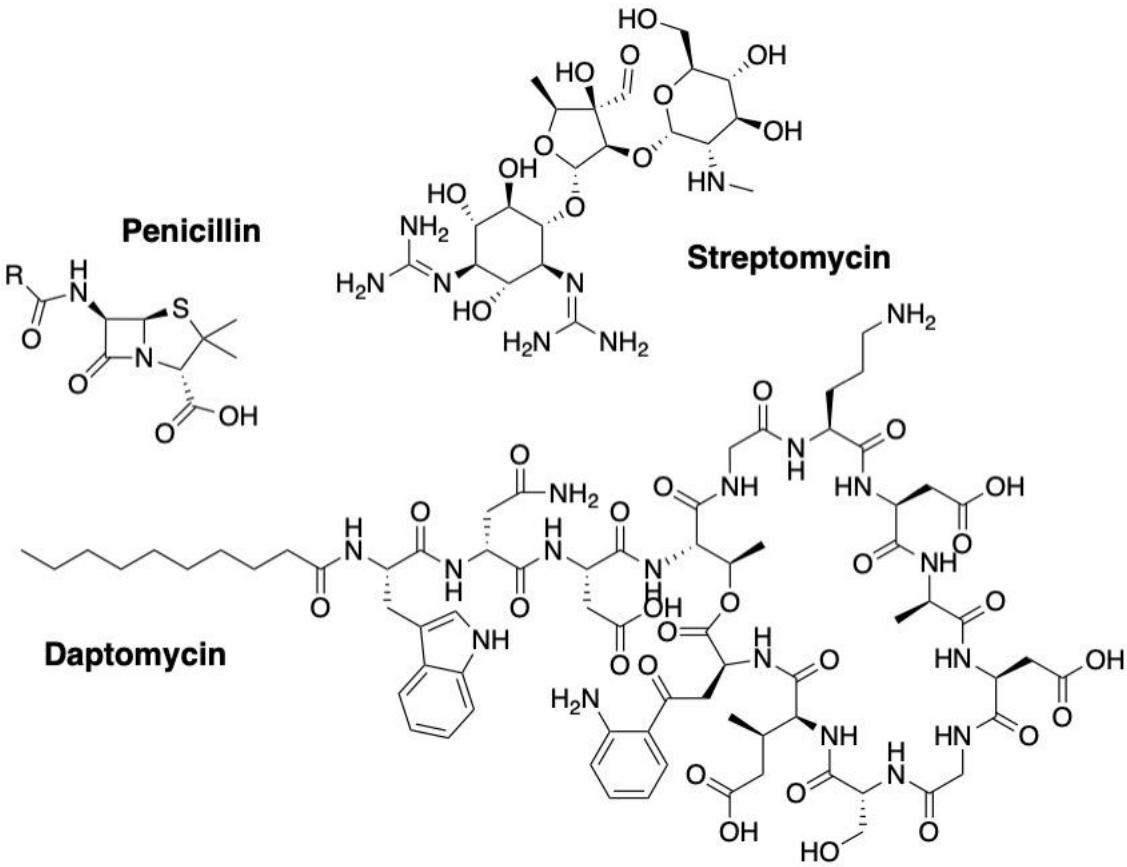

Figure 1.1: The structures of three antimicrobial natural products with historic significance. Penicillin was the first antibiotic discovered, by Alexander Fleming in the late 1930's. The general structure of the penicillin family is shown, the R group is varied. Streptomycin was discovered by Selman Waksman in 1943, Waksman was the first to coin the term antibiotic. Daptomycin is one of the most recent antibiotics discovered, the only new antibiotic to gain clinical approval in the 21st century.

1.2 The use of genomics in natural product discovery

When the first *Streptomyces* genome was published in 2002, it was found to be much larger and contain more secondary metabolite gene clusters than predicted[15]. This genome showed that most bacterial gene clusters are not expressed under standard laboratory conditions. It is thought that less than ten percent of secondary metabolite biosynthetic gene clusters (BGCs) are expressed during routine growth[13]. Now, whole genome sequencing is relatively inexpensive which, combined with advancements in genome analysis, allow researchers to have a full picture of a bacterium's biosynthetic potential. The number, and type of gene clusters is easily determined, which allows researchers to focus on strains with high 'biosynthetic potential'. Gene cluster homology also shows if one or more gene clusters in a strain already have a known product from another strain. This reduces the need for dereplication as certain molecules can be excluded from an early stage in the discovery process. Researchers can now predict and focus in on novel pathways. Advances in DNA synthesis and cloning techniques means that expressing these potentially novel gene clusters is easier than ever. It is also possible to predict the structure of a natural product from genomic information, which allows researchers to link characterised natural products back to its gene cluster. Genomic information helps with understanding the biosynthesis of a natural product and allows for potential modification of its pathway, which could help improve the yields of a natural product or could allow for modifications of the structure.

One of the most impactful genomic tools for natural product discovery is antiSMASH[16], which is a tool that mines microbial genomes for specialised metabolite BGCs. AntiSMASH uses manually curated rules to define which core biosynthetic functions need to be in a genomic region for it to constitute a BGC. To identify biosynthetic function it uses profile hidden Markov models (HMMs) from the protein family and domain databases, Pfam[17], TIGRFAMs[18] and SMART[19]. It also incorporates HMMs from Bagel[20], which is not only a database for RiPP precursor peptide core regions and HMM motifs of associated genes but is also a web server to mine RiPP and bacteriocin gene clusters. AntiSMASH also gives

context on the similarity of a particular BGC to those found in other strains. To do this it uses a ClusterBlast algorithm which produces three outputs, all of which use the same algorithm, just with differing reference datasets. ClusterBlast will show regions from the antiSMASH database which have similarity to the cluster in question. SubClusterBlast will show sub-cluster units with similarity to the query BGC. And KnownClusterBlast will show cluster in the MIBiG database that have similarity. Minimum Information about a Biosynthetic Gene cluster[21] (MIBiG) is a standardised format that describes the minimal required information to uniquely characterise a BGC, it also has a database of characterised BGCs which other genomic tools, such as antiSMASH, can use. By incorporating so many tools and databases into one workflow, antiSMASH represents a quick and easy way to gain a lot of information about the number, type and arrangement of specialised metabolite BGCs. ClusterBlast can act as an easy dereplication step, by checking the BGCs predicted for a particular strain and seeing if any have high homology to a characterised BGC. AntiSMASH produces a .gbk file for each gene cluster region that it identifies. These files can then be inputted into other bioinformatic tools for further analysis. Two such tools are BiG-SCAPE[22] and CORASON. BiG-SCAPE (biosynthetic gene similarity clustering and prospecting engine) builds gene cluster similarity networks from inputted BGCs and can link these to BGCs in the MIBiG database. From this CORASON (core analysis of syntenic orthologues to prioritise natural product gene clusters) elucidates the phylogenetic relationships between these BGCs[22].

Another useful toolkit to aid in the discovery of novel natural products is the Global Natural Products Social Molecular Networking[23] (GNPS). GNPS, on a fundamental level, is a database for LC-MS/MS data which is crowdsourced and open access. The rest of the toolkit is built using this database and allows the comparison of your data to the database. GNPS not only helps with dereplication but can help characterise novel molecules by seeing similarities to the chromatograms of characterised molecules. GNPS also builds molecular networks from MS/MS data which identifies which molecules or fragments are related and builds an interactable network between these molecules.

There are many publicly available databases of natural products, that are not only easily searchable but collate all information about a particular specialised metabolite. Two of the most prolific examples are The Dictionary of Natural Products, which is the most comprehensive source of natural product information available, and The Natural Products Atlas[24]. The newest version of the NP Atlas (2.0) contains over 30,000 compounds alongside full taxonomic descriptions. Version 2.0 of the NP Atlas also integrates the CyanoMet[25] data base which is the largest publicly available dataset of cyanobacterial specialised metabolites. These databases allow for easier dereplication due to the ease of checking whether a particular strain has had existing compounds isolated from it. An example of this can be seen in Chapter 4.

1.3 A brief overview of notable natural product families

1.3.1 Polyketides

Polyketides are a prolific family of structurally diverse natural products and have been a rich source of clinically relevant pharmaceuticals[26,27]. Polyketides are produced by enzymatic assembly lines called polyketide synthases (PKSs). Polyketide synthases are usually multifunctional enzymes which are classified into three types[28]. Type I PKSs are large proteins which are comprised of multiple functional domains. Type I PKSs are further divided into iterative PKSs, where the same set of enzymatic domains catalyse multiple chain elongation steps, and assembly-line PKSs where each catalytic domain acts once[28]. Iterative PKSs are predominantly found in fungi though eukaryotic examples have been found. Type II PKSs are multi-enzyme complexes comprised of monofunctional proteins, and these complexes often work iteratively. Type II PKSs are predominantly found in bacteria, although they are extremely rare in cyanobacteria. Type III PKSs are simpler, chalcone-like synthases which produce a product from a single active site. For type I and II PKSs, the required catalytic PKS domains are: an acyltransferase (AT), a ketosynthase (KS) and a thioesterase (TE). An acyl carrier protein (ACP) is also required. The ACP domain is post-translationally modified with a phosphopantetheinyl arm which allows the sub-unit (usually acetyl-CoA or an analogue) to be loaded, this is catalysed by an AT domain[29]. The

ACP then moves the starter unit to a KS where chain elongation occurs through the addition of extender units (usually malonyl-CoA or methylmalonyl-CoA) by decarboxylative Claisen condensation. This ACP-bound β-ketothioester intermediate can be further modified by multiple optional enzymes such as a ketoreductases (KR), dehydratase (DH), enoylreductase (ER) and methyltransferase (MT)[28]. A KR domain reduces a β-keto group to a β-hydroxy group, and a DH can then remove the hydroxy group leaving an α-β-unsaturated alkene. The ER domain can reduce the α-β double bond to a single bond. These modifications will happen on the sub-unit most recently added to the chain. This cycle is repeated until the desired product is made and then transferred to the final catalytic domain, the TE which hydrolyses the completed polyketide chain from the ACP. Polyketides are structurally diverse and many have potent bioactivities, as a result many polyketides have found clinical use[30]. One example is the antibiotic tetracycline, which has been used to treat a range of infections including malaria, cholera and syphilis. It is currently on the World Health Organisation's list of essential medicines. Tetracyclines are a large family of antibiotics but the first, chlortetracycline was isolated from *Streptomyces aureofaciens* in 1945 by Benjamin Duggar[31].

1.3.2 Non-ribosomal peptides

Nonribosomal peptides are a structurally diverse range of natural products produced by large multienzyme machineries called nonribosomal peptide synthetases (NRPSs). The biosynthesis of NRPs occurs in three main phases. The first domain is an adenylation domain (A domain) which activates an amino acid by turning it into an adenylate, using ATP, then it transfers the activated amino acid to a peptidyl carrier protein (PCP)[32]. The PCP has been covalently modified with a phosphopantetheine cofactor[32], which is attached to the amino acid and peptide substrates through a thioester linkage. Next, condensation domains (C domain) catalyse the amide bond formation between the thioester group on the growing peptide chain from the previous module with the amino acid of the current module, extending the PCP bound peptide chain[33]. Optional domains can modify the growing peptide chain, a condensation-cyclisation domain (Cy domain) can catalyse the cyclisation of amino acid side chains such as serine, threonine or cysteine to form oxazolidines or thiazolidines[33]. Epimerisation domains (E domain) can catalyse the

inversion into a D-amino acid. Finally, after all elongation steps have been completed a thioesterase domain (TE domain) hydrolyses the polypeptide chain from the terminal PCP. Further post-translational modifications can then modify the polypeptide chain such as, halogenation, glycosylation, acylation or hydroxylation. The structural diversity of nonribosomal peptides is vast, especially as there are many possible subunits – not only the twenty proteogenic amino acids but non-proteogenic amino acids can be incorporated as well. As a result, many nonribosomal peptides have found clinical use. Examples include the antibiotics Dactinomycin and vancomycin[34] (Figure 1.2), both on the World Health Organisations list of essential medicines[35].

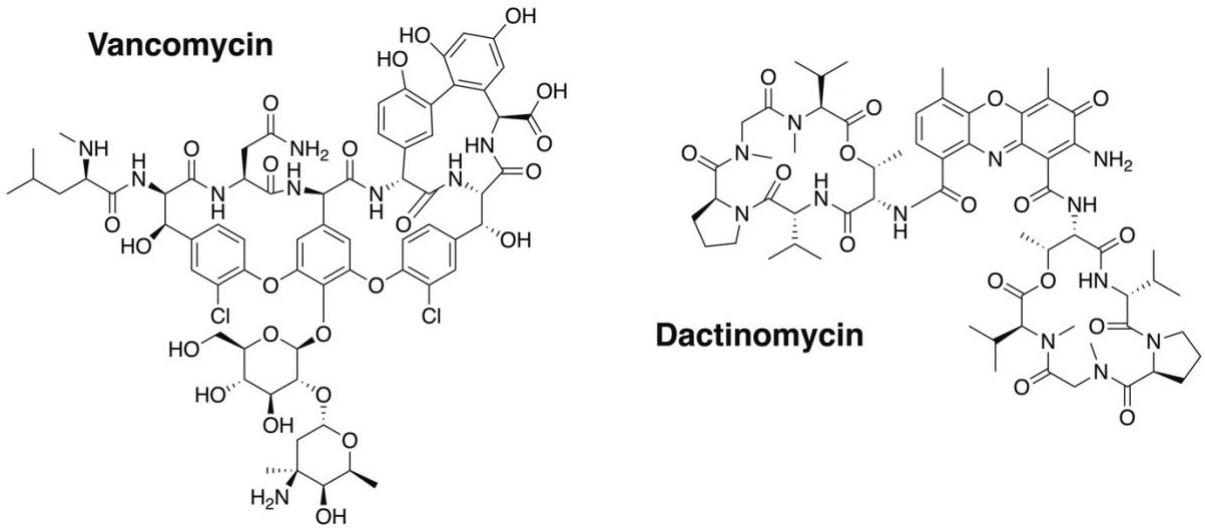

Figure 1.2: The structures of vancomycin and dactinomycin, both are NRPs. Both compounds have found prolific clinical use as antibiotics and both are on the WHOs list of essential medicines.

1.3.3 Terpenes

Terpenes are a large group of natural products, abundant in plants as well as bacteria. Terpenes are classified by the number of carbons in their backbone[36,37]. Monoterpenes have ten carbons, sesquiterpenes have fifteen, diterpenes have twenty, the rarer sesterterpenes have twenty-five carbons and triterpenes have thirty carbons. Terpenes are built up from five carbon isoprene subunits either in the form of dimethylallyl pyrophosphate (DMAPP) or isopentenyl pyrophosphate (IPP). These precursors are structural isomers but are produced by two distinct pathways, these pathways are normally mutually exclusive in organisms. The mevalonate (MVA) pathway is usually found in archaea and eukaryotes and produces IPP. The non-mevalonate, methyl-D-erythritol phosphate, (MEP) pathway is found in most bacteria and produces DMAPP. These five-carbon subunits are then condensed together to produce the ten-carbon molecule, geranyl pyrophosphate which is the precursor to monoterpenes and monoterpenoids. Geranyl pyrophosphate can be converted to farnesyl pyrophosphate (fifteen carbons) to produce sesquiterpenes, or into geranylgeranyl pyrophosphate (twenty carbons) for the production of diterpenes or into geranylfarnesyl pyrophosphate (twenty-five carbons) for the production of sesterterpenes, which is catalysed by a terpene synthase which can also cyclise the terpene. Finally modifying enzymes such as a cytochrome P450 can modify the basic terpene hydrocarbon to produce the final mature terpenoid. Terpenes are used as building blocks for other important molecules, for example steroids are derivatives of the terpene squalene. Terpenes are known to play a role in plant communication and in disease resistance[38]. Terpenes can also have clinical use, such as the diterpenoid paclitaxel, which is used to treat breast cancer[39].

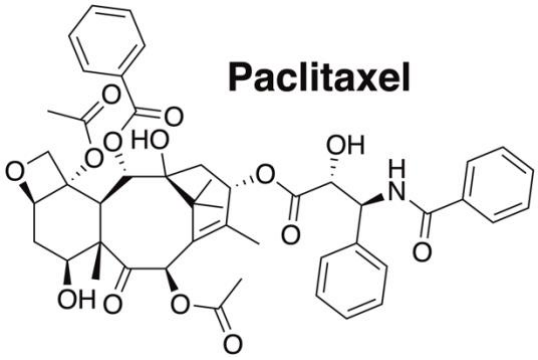

Figure 1.3: The structure of paclitaxel, a diterpene that has clinical use in the treatment of breast cancer.

1.3.4 Ribosomally synthesised post translationally modified peptides

Ribosomally synthesised post translationally modified peptides (RiPPs) are initially ribosomally synthesised as a precursor peptide[40]. The precursor peptide is made up of a sequence conserved N-terminal leader peptide, which is usually between 20 and 110 residues long[41], and a core peptide which have great sequence variability[42]. Following the core peptide some RiPPs have a recognition sequence at the C-terminus. This recognition sequence can be important for subsequent excision and post-translational cyclisation steps. In eukaryotes a signal peptide is sometimes present at the N-terminus which directs the RiPP to a specific compartment of the cell. The core peptide region then undergoes post-translational modifications, which are carried out by tailoring enzymes encoded in the RiPP BGC. Finally, the non-core regions of the peptide are removed by proteolysis to produce the mature peptide. An overview of general RiPP biosynthesis is shown in Figure 4.

In a few rare cases, such as in bottromycins, the leader peptide is not found at the N-terminus but instead at the C-terminus[43]. In these cases it is referred to as a follower peptide to differentiate[44]. A small number of tailoring enzymes are responsible for the majority of post-translational modifications. An example of this is that the genes encoding radical SAM dependant methyltransferases are found in the BGCs of thiopeptides,

proteusins and bottromycins[45]. Many of these enzymes predominantly recognise the leader (or follower) peptide rather than the core peptide that is modified. Though the core peptide requires particular motifs specific to the modifying enzyme. Many RiPP tailoring enzymes recognise and bind to their corresponding precursor peptide using a domain called the RiPP recognition element (RRE)7. These RREs exist either as a discrete protein or fused to a lager protein domain and have been found in over 50% of prokaryotic RiPP classes7. Discrete RREs have been shown to bind to the leader peptide and act as a scaffold for the recruitment of tailoring enzymes46.

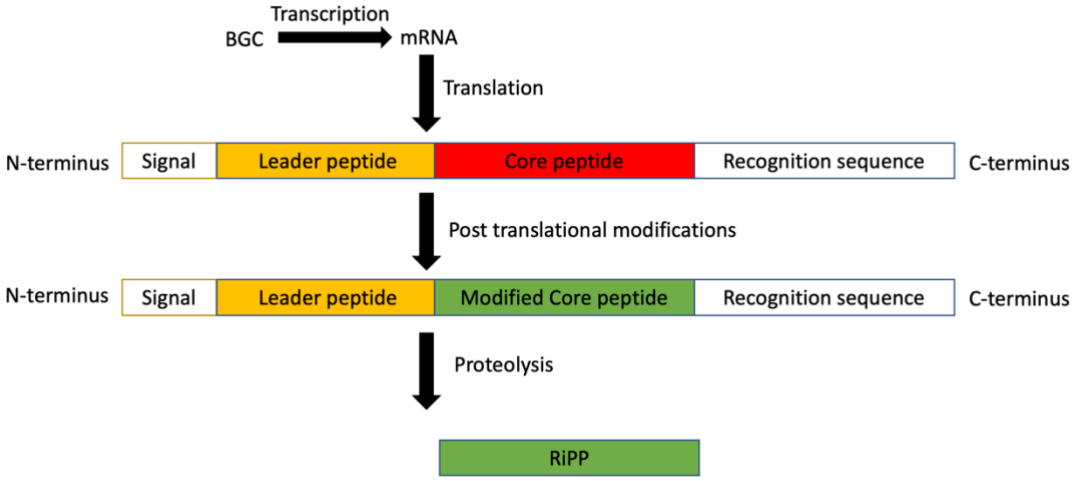

Figure 1.4: Workflow for the biosynthesis of RiPPs. The signal peptide and recognition sequence are optional and not a part of all RiPP biosynthetic pathways. The leader and core peptides are integral for RiPP biosynthesis.

1.4 RiPP sub-families and post-translational modifications

The structural diversity of RiPPs, which gives rise to the huge number of RiPP subfamilies, is created by a sophisticated set of post-translational modifications. Below the post-translational modifications that produce some of the most well studied RiPPs are described Many more classes of RiPP do exist alongside many more possible post translational modifications, leading to the great structural diversity amongst RiPPs[48].

1.4.1 Cyanobactins

Cyanobactins are exclusively produced by cyanobacteria. It was previously thought that macrocyclisation was a defining feature of cyanobactins but recently highly modified linear cyanobactins have been discovered[49]. Despite this, N to C macrocyclisation is extremely common and carried out by subtilisin-like serine proteases[50]. These macrocycles are often further modified by O- or C- prenylation caused by a [3,3]-sigmatropic Claisen rearrangement.[51] Interestingly with cyanobactins the precursor peptide can encode multiple products[44]. Cyanobactins usually contains azol(in)e rings, which are produced by ATP-dependant cyclodehydration of Cys, Thr or Ser residues catalysed by a YcaO-like cyclodehydratase.[50] Trikoramide A (Figure 5) is a cyclic decapeptide cyanobactin that was isolated from the marine cyanobacterium *Symploca hydnoides*[52] It possesses cytotoxicity against the AML2 and MOLT-4 cancer cell lines with IC_{50} values of 8.2 and 4.8 µM respectively[52].

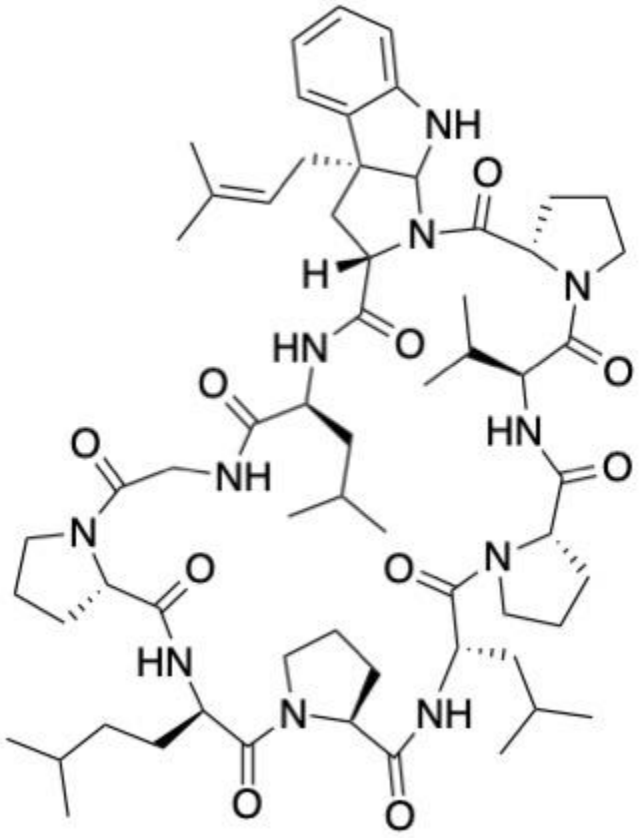

Figure 1.5: The structure of trikoramide A, a cyclic decapeptide cyanobactin isolated from *Symploca hydnoides*

1.4.2 Lanthipeptides

Lanthipeptides are defined as peptides that contain meso-lanthionine (Lan) or 3-methyllanthionine (MeLan) residues. These residues are formed through the 1,4-conjugate addition of a cysteine residue onto a dehydrated amino acid[53]. Addition to a serine will produce a Lan residue and addition to a threonine produces a MeLan residue. Following the 1,4-conjugate addition, the resulting enolate needs to be protonated to produce the final residue[53]. Nisin (Figure 1.5) was the first lanthipeptide reported and when its structure was reported in 1971 it was shown to be one of the longest known RiPPs[54]. A lanthionine residue is made by two alanine residues whose β-carbons are connected vie a thioether linkage[44]. This thioether linkage is produced when Ser/Thr residues in the core peptide are dehydrated to 2,3-didehydroalanine and (Z)-2,3 didehydrobutyrine respectively[50]. After this the unsaturated amino acids undergo a Michael type addition to install a Cys thiol to produce a lanthionine residue[50]. Lanthipeptides that have antibiotic activity are called lantibiotics[55]. Lanthipeptides are divided into four groups depending on the biosynthetic enzymes that install the Lan and MeLan residues. For class I lanthipeptides the dehydration is carried out by a dedicated dehydratase called LanB and the cyclisation is carried out by a LanC cyclase[44]. Class I lanthipeptides are distinct for having two separate enzymes carry out the dehydration and cyclisation as for class II, III and IV lanthipeptides both are carried out by a bifunctional lanthionine synthetase. For class II lanthionine synthetases, designated LanM[44], the N-terminal dehydration domain is distinct to class III and IV synthetases and shows little sequence homology. Class III and IV synthetases have distinct C-terminal cyclisation domains. Class V lanthipeptide gene clusters contain a Hop-A1 like protein and a phosphotransferase and these two enzymes catalyse a key dehydration step7. Class V lanthipeptides are covered in more detail in chapter 5.

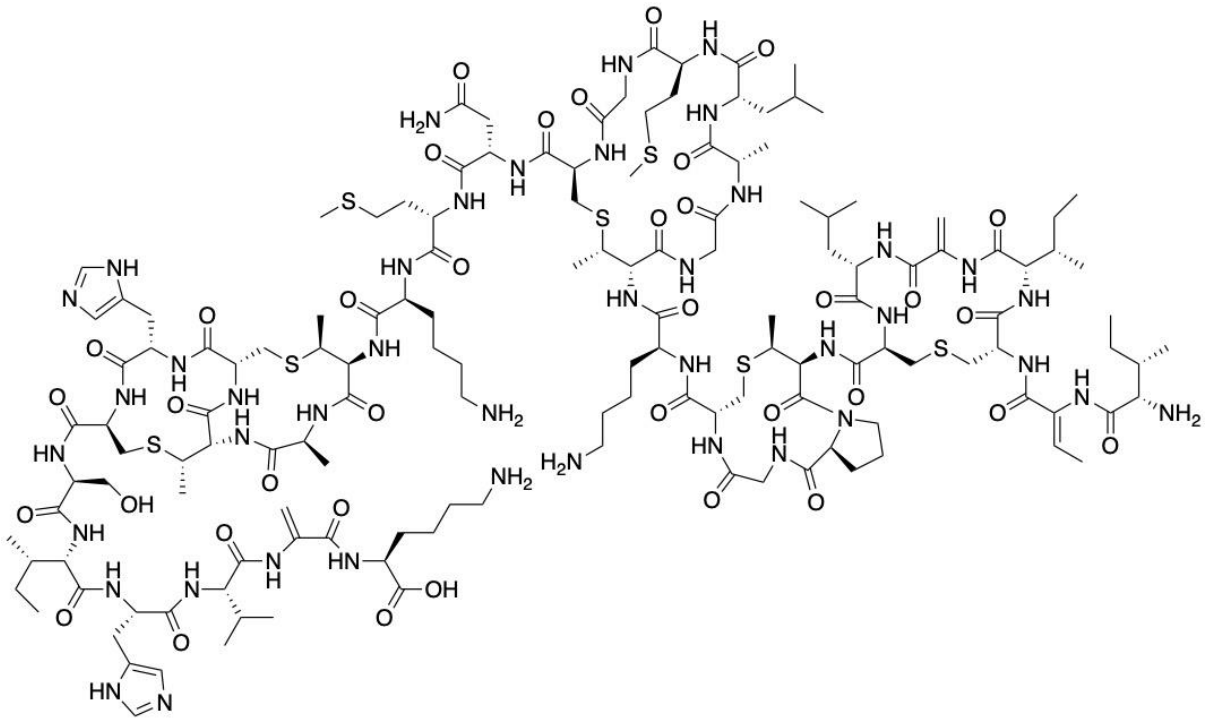

Figure 1.6: The structure of Nisin, the first reported lanthipeptide. It was first isolated from *Lactococcus lactis*

1.4.3 Lasso peptides

Lasso peptides are characterised by an N-terminal macrolactam and a C-terminal peptide tail that threads through the lactam.[50] The C-terminal peptide tail is held in place by bulky side chains so that it does not unthread. The macrolactam is formed through a dehydration reaction between the N-terminal amine and a side chain carboxylic acid group from Asp/Glu residues in the core peptide region. This macrocyclisation is carried out by two lasso maturation enzymes referred to as B and C proteins.[50] The B protein is a cysteine protease that cleaves off the leader peptide which exposes the N-terminal amine. The C protein then activates the carboxylate side chain of an Asp/Glu residue by adenylation[50]. The macrolactam is then formed by the nucleophilic attack of the N-terminal amine[57]. It has been suggested that the B and C proteins form an active complex[44] as their action is dependent on the presence of the other.

1.4.4 Thiopeptides

The defining feature of thiopeptides is a macrocyclic core with a central six-member nitrogenous heterocycle and further thiazoles.[44] They are produced by the dehydration of Ser/Thr residues and cyclodehydration of Cys/Thr/Ser residues in the core peptide[44]. The central six-member heterocycle is produced by an enzymatic Diels-Alder like [4 + 2] cycloaddition[58]. Many thiopeptides also contain post-translational modifications that are not derived from the core peptide sequence[59], which greatly increases the structural diversity of thiopeptides. An example is nosiheptide (Figure 7) which shows antimicrobial activity against methicillin-resistant *Staphylococcus aureus*[60].

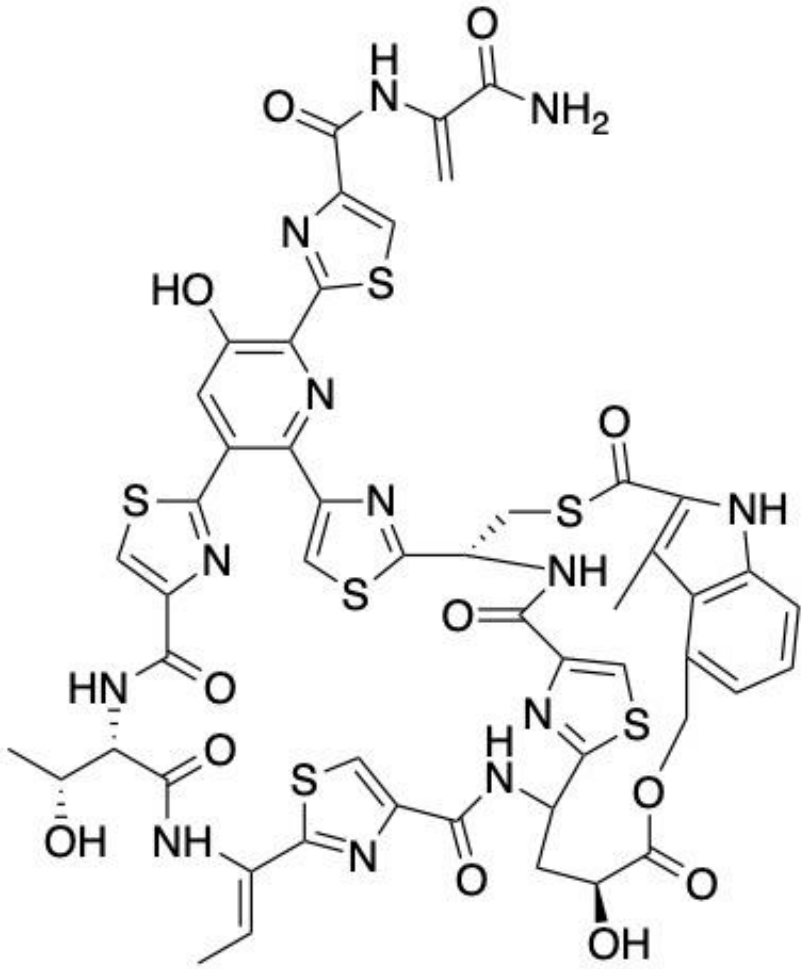

Figure 1.7: The structure of nosiheptide, a thiopeptide that shows antimicrobial activity against Gram-positive bacteria.

1.5 Cyanobacteria

1.5.1 Ecological role of cyanobacteria

Cyanobacteria are a phylum of photosynthetic prokaryotes. Despite being often referred to as blue-green algae, they bear no relation to eukaryotic algae. Cyanobacteria were one of the first organisms on earth, with fossil records of cyanobacteria dating back over three and a half billion years[61]. Cyanobacteria were responsible for the transforming Earth's original carbon dioxide rich atmosphere into the oxygen rich environment we have today. Over the last three and a half billion years, cyanobacteria have been able to evolve an incredible variety across the phylum and fulfil many different biological niches. Cyanobacteria are able to inhabit a diverse range of environments, including both aquatic and terrestrial. Cyanobacteria also have a wide range of morphologies including unicellular, surface attached, filamentous and colony forming[62]. Cyanobacteria have developed interactions with other micro-organisms, one example is cyanolichens where the cyanobacteria have a symbiotic relationship with multiple microorganisms[63]. In cyanolichens cyanobacteria play an integral ecological role, providing the majority of the nitrogen and carbon. It is thought that many cyanobacterial specialised metabolite gene clusters have evolved to interact with other organisms[61].

1.5.2 Cyanobacterial blooms and toxicity

Cyanobacteria are infamously known for their ability to create toxic blooms[64]. Eutrophication is a build-up of nutrients in a salt or freshwater environment and is the main cause of toxic blooms. Eutrophication leads to an increase in the growth of cyanobacteria, sometimes leading to large visual coverings. The increase in cyanobacteria will remove oxygen from the environment as when they die their de-composition by other micro-organisms depletes the oxygen[65]. The removal of oxygen upsets the ecological balance but the most damaging effect is the build-up of toxic metabolites. Many cyanobacteria produce specialised metabolites that are neurotoxins, cytotoxins and hepatoxins[61]. One notorious example is anatoxin-a which is a neurotoxic alkaloid produced by the cyanobacterium

Anabaena flos-aquae7. Herds of livestock have been killed by toxic blooms of *A. flos-aquae* producing anatoxin-a. The prevalence of toxic cyanobacterial blooms is increasing due to an increase in eutrophication mainly caused by agricultural run off[61].

1.6 Cyanobacterial specialised metabolite production

Cyanobacteria are a rich source of structurally diverse bioactive specialised metabolites. It is thought that the complex and competitive communities which cyanobacteria inhabit are one reason for this[61]. The full extent of the capabilities of cyanobacteria to produce specialised metabolites has been poorly understood resulting in it being underestimated. Many cyanobacterial specialised metabolites have been wrongly attributed to higher order species such as marine sponges. One example is dolastatin-10, which was first isolated from a mollusc[67]. Screening of environmental metagenome libraries revealed that a cyanobacteria, *Caldora penicillate*, possessed the gene cluster responsible for producing dolastatin-10. Thus, it is now known that cyanobacteria attached to the mollusc produced dolastatin-10 and not the mollusc itself. In both marine and terrestrial environments cyanobacteria are prone to predation. The production of bioactive specialised metabolites acts as a defence mechanism. It has been shown that typical marine grazers such as crabs and sea urchins do not predate upon *Moorea producens*[61], which is thought to be because of the toxicity of the specialised metabolites it produces. Ypaoamide produced by *Moorea producens* has been shown to repel sea urchins and multiple fish species[68].

Outside of their ecological niches, cyanobacterial secondary metabolites have promising uses as pharmaceuticals and therapeutics. Multiple cyanobacterial specialised metabolites are currently undergoing phase two and three clinical trials. Cryptophycins are a family of depsipeptides that have been isolated from *Nostoc* sp. and show promising anti-tumour activity[69]. Cryptophycins also have anti-fungal activity and it thought that cyanobacteria produce Cryptophycins to outcompete fungi[70]. Cryptophycins attack tubulin microfilaments preventing cell division and have been shown to be effective against multi-drug resistant tumours[70]. Cryptophycin-1 is currently in stage one human clinical trials.

Dolastatins are another large class of cyanobacterial natural products with potential clinical use. Dolastatin-10 is a pentapeptide which contains four unique amino acids, dolavaline, dolaisoleucine, dolaproline and dolaphenine[67]. The 'dola' variant differs from the standard amino acid as a hydrogen atom on the amine group of each amino acid is replaced with a methyl group. Unfortunately dolastatin-10 did not pass phase II trials as it caused peripheral neuropathy in over forty percent of patients[71]. A derivative of dolastatin-10 named TZT-1027 that has the thiazoline ring removed is currently in phase I trials[71,72]. The structures of both molecules are shown in Figure 1.8. Dolastatin-15 is a linear peptide and also has activity against multiple cancer cell lines. An analogue of dolastatin-15, ILX-651, is currently in stage II clinical trials. Another dolastatin derivative is monomethyl auristatin E which has found clinical use as an antibody drug conjugate used to Hodgkin Lymphoma and other cancers[73]. Monomethyl auristatin E is too toxic to be administered on its own, administering it as an antibody drug conjugate allows for the targeting of the compound to the tumour reducing overall toxicity[73]. Despite this, nervous system toxicity is still a big problem when administering auristatin antibody conjugates[74].

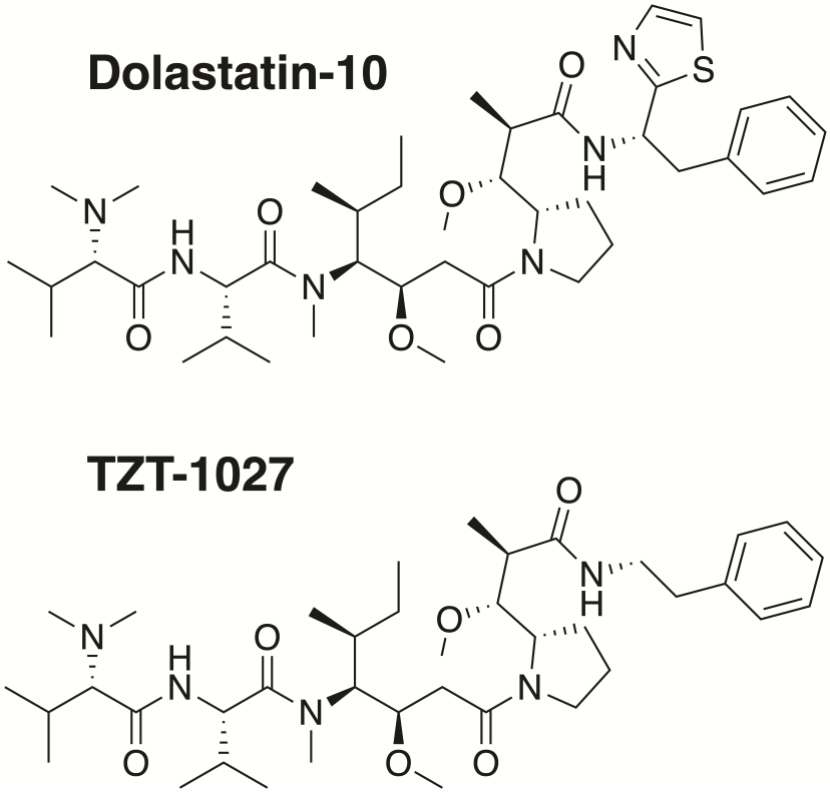

Figure 1.8: The structure of dolastatin-10, a pentapeptide natural product with anti-cancer activity. Dolastatin-10 failed stage II clinical trials. A derivative of dolastatins-10 – TZT-1027 which lacks the thiazoline ring is currently in phase I trials.

Many cyanobacterial specialised metabolites also show antimicrobial activity. Interestingly, Nostoc species are the most fruitful source of cyanobacterial antimicrobials[75]. One example is the carbamidocyclophane family of molecules, which were isolated from *Nostoc* sp. CAVN2. These molecules have shown inhibition of methicillin-resistant *Staphylococcus aureus* at nano molar levels7. Another example of antimicrobial cyanobacterial specialised metabolites are the pitipeptolides. Pitipeptolides C-F were isolated from the marine cyanobacterium *Lyngbya majuscule*[77]*.* They are cyclodepsipeptides and have been shown to have activity against both cancer cell lines and *Mycobacterium tuberculosis.* It was shown that an N-methylation of phenylalanine was integral to the antimicrobial activity of the pitipeptolides[77](Figure 1.9).

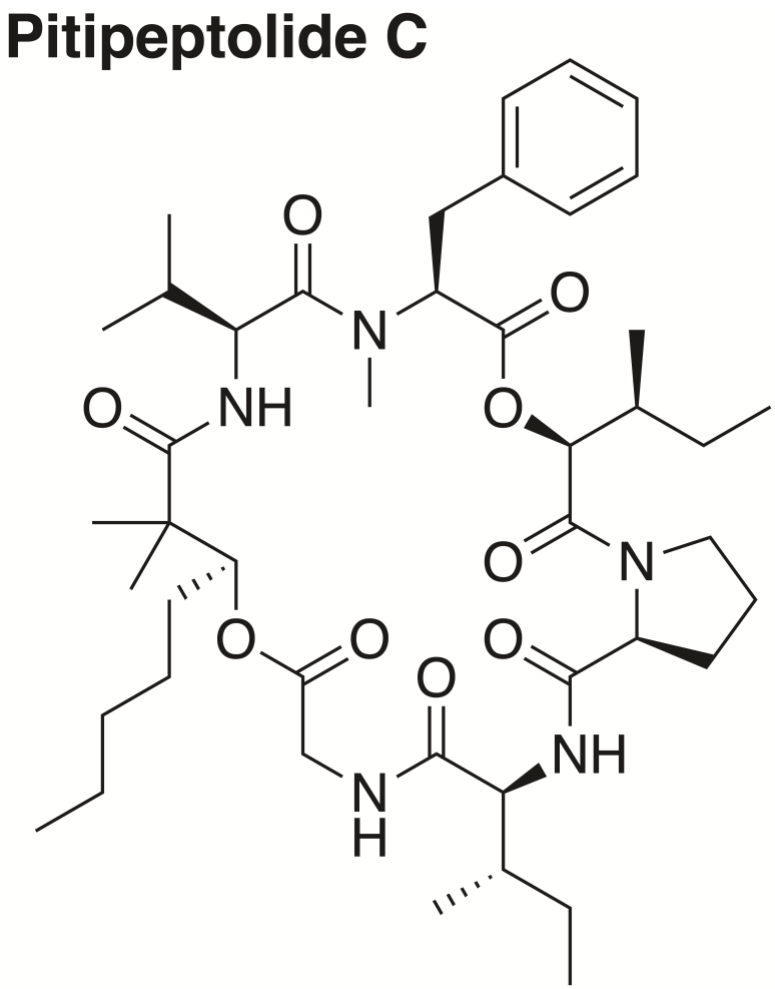

Figure 1.9: The structure of pitipeptolide C, a molecule with activity against cancer cell lines first isolated from *Lyngbya majuscule.*

Despite many cyanobacterial specialised metabolites having strong activity, none have yet passed through clinical trials and made it to the clinic. This lack of progression is usually due to unintended toxicity, however examples such as the dolastatins show that it is possible to synthesise derivatives that minimise unintended effects. Monomethyl auristatin E is a further example that through derivatisation and further targeting of the molecule via an antibody, it is possible to produce clinically relevant pharmaceuticals from toxic cyanobacterial natural products[73]. This is an area that could be explored further with other bioactive cyanobacterial natural products that have been overlooked due to their toxicity.

1.7 The *Scytonema* genus

1.7.1 Overview of *Scytonema*

Scytonema is a genus of filamentous cyanobacteria of the order Nostocales. *Scytonema* form heterocysts, which is a specialised cell that fixes nitrogen[78]. These heterocysts are formed during nitrogen starvation. The nitrogenase responsible is inactivated by O_2 so the heterocyst envelope acts as a barrier to oxygen[79]. *Scytonema* are found in a diverse range of environments and conditions[80,81]. They can grow in both salt and freshwater either free floating or attached to a surface. They have also been found to grow on rocks, trees and multiple other non-aquatic surfaces. They are also found in lichens[82], a symbiosis of cyanobacteria and a fungus. *Scytonema* also form communities with other bacteria. It is unknown if this is a symbiotic relationship or if other bacteria predate on the *Scytonema* due to their ability to fix carbon and nitrogen. *Scytonema* are also known to be a key component of biological soil crust in deserts and dry climates[83,84].

1.7.2 Existing *Scytonema* genomic information

In 2012 the genome of *Scytonema hofmannii* PCC 7110 was published[85]. The genome spanned over 12 mega bases (MB) and contained 12,356 coding sequences, at the time this was the most gene-rich prokaryote that was known[85]. Of these genes almost fifty percent showed no homology to other strains. The number of novel genes increases the likelihood of discovering novel natural products with novel chemistry. The next *Scytonema* genome to be published was *Scytonema millei* VB511283 in 2015[81]. This genome was sequenced using only a Ilumina HiSeq platform resulting in a poor-quality genome with 118 scaffolds. The length of the genome was originally published at 11.6 MB with a GC content of 51%. This GC content is far too high for a *Scytonema* strain which tend to have a GC content in the low forties. A higher quality genome for *Scytonema millei* VB511283 was released in 2020 with only 26 contigs. This genome spans 6 MB and has a GC content of 44.5%. This highlights the need for good quality genome assemblies and also shows that not all *Scytonema* strains have extremely large and gene rich genomes. This trend of poor quality

published genomes continued in 2020 when the genome for *Scytonema* sp. UIC 10036 was published alongside the isolation of scytodecamide from this strain7. *Scytonema* sp. UIC 10036 was sequenced using combined short-read Illumina and long-read PacBio. Despite this combined sequencing approach the resulting genome spans 379 contigs, a vast amount. The result of this large number of contigs is that the total genome size is 14.1 MB, making it by far the largest *Scytonema genome.* However, the large number of contigs likely means the total genome size is either inflated by duplicated DNA, or contains contaminant DNA. Despite these issues, the researchers were still able to attribute a BGC to scytodecamide[86] which shows that even poor quality genomes can yield useful information. Two further *Scytonema* genomes have been uploaded to the antiSMASH database, *Scytonema* sp. NIES 2130 and *Scytonema* sp. NIES 4073. *Scytonema* sp. NIES 4073 is well sequenced, the genome is covered in 5 contigs and is 9.8 MB in size. AntiSMASH predicts NIES 4073 contains 24 gene cluster regions. *Scytonema* sp. NIES 2130 is also referred to as *Scytonema* sp. HK-05. The NIES 2130 genome is covered in 43 contigs and is 9.8 MB in size. AntiSMASH predicts 28 gene cluster regions for NIES 2130.

1.7.3 A timeline of *Scytonema* natural product discovery

Multiple natural products have been isolated from *Scytonema,* although none have found a clinical use. The first natural product isolated from *Scytonema* was cyanobacterin, which was first discovered in 1982[87]. Cyanobacterin inhibits the growth of photosynthetic organisms but does not exhibit activity against non-photosynthetic bacteria[87]. Cyanobacterin was first isolated from *Scytonema hofmannii* UTEX 2349, however this strain has recently been reclassified to *Tolypothrix* sp. PCC 9009. In 1990 tolytoxin (also known as scytophycin B) was found to be produced by the strains, *Scytonema mirabile* BY-8-1, *Scytonema burmanicum* DO-4-1 and *Scytonema ocellatum* DD-8-1, FF-65-1 and FF-66-3[88]. Tolytoxin is a scytophycin and is known to have cytotoxic activity. Scytophycins are a family of polyoxygenated macrolides isolated from *Scytonema*[89]*.* Tolytoxin has been shown to induce apoptosis on human cancer cell lines which is an essential feature of anticancer drugs[90]. In 1997 five diacylated sulfoglycolipids were isolated from *Scytonema* sp. TAU SL-30-1-4[91]. These sulfoglycolipids were found to inhibit the reverse transcriptase of HIV-1, potentially leading to a treatment for HIV. In 2001 two natural products were isolated from

Scytonema hofmanni PCC 7110, scyptolin A and B[92]. Both scyptolin A and B are cyclic depsipeptides featuring a nineteen-member cyclised ring. Interestingly both contain the uncommon residue 3'-chloro-N-methyl-tyrosine. Both molecules were shown to inhibit porcine pancreatic elastase[92]. A further natural product, hofmannolin, was isolated from *Scytonema hofmanni* PCC 7110 in 2003. Hofmannolin is also a cyclic depsipeptide with two characteristic differences to other cyanopeptolins. It has an O-methylated tyrosine residue and a 2-hydroxy-3-methyl-valeric acid residue at the N-terminus[93]. However, unlike the previous scyptolin A and B, hofmannolin has no identifiable bioactivity. Also in 2003 scytovirin, a protein comprised of ninety-five amino acids was isolated from *Scytonema varium*[94]. Scytovirin contains ten cysteines which form five intrachain disulphide bonds. Scytovirin was found to bind to the viral coat proteins gp120, gp160 and gp41 and inhibit HIV-1.

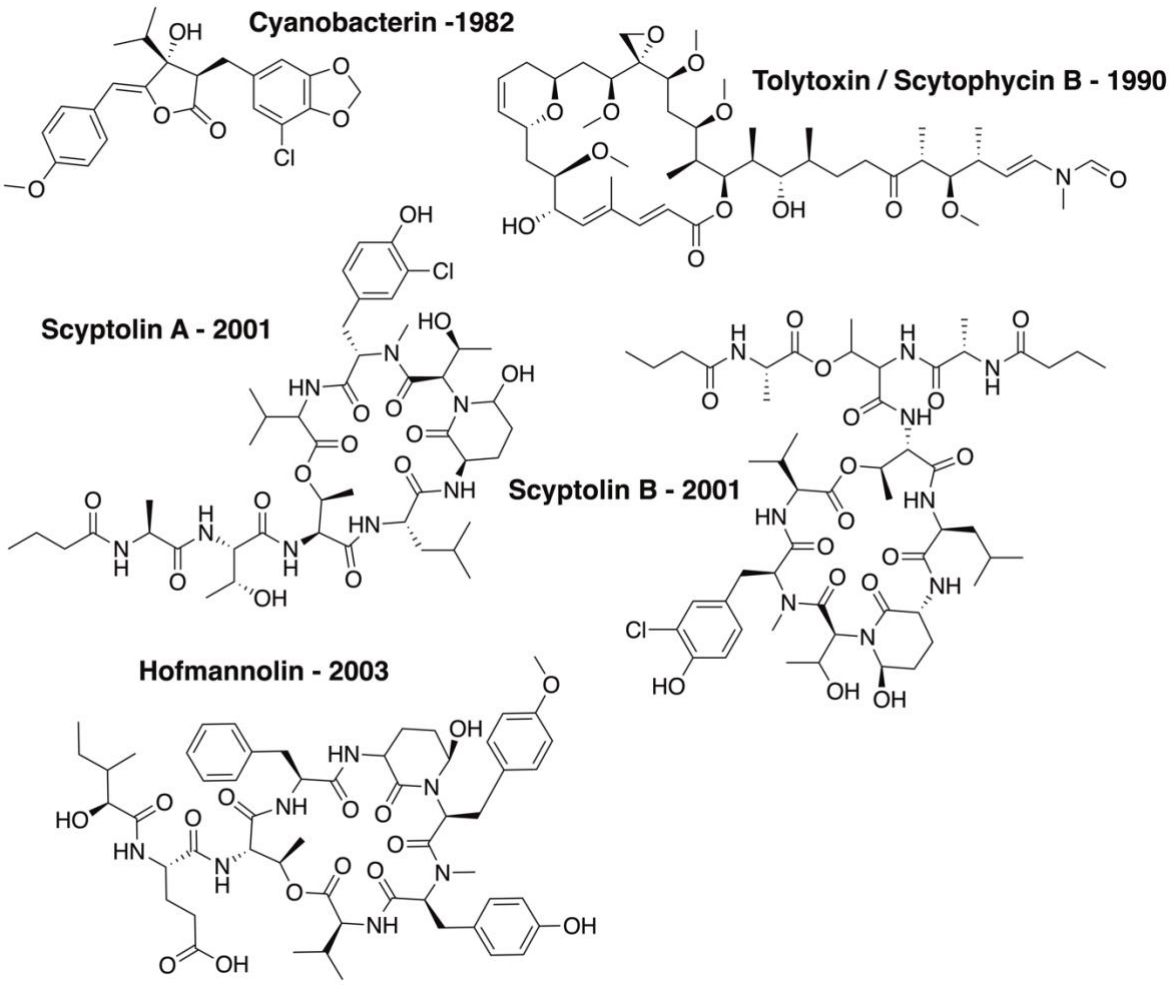

Figure 1.10: The structures of the five natural products isolated from *Scytonema* strains from 1982 – 2003.

In 2008 nostodione A was discovered from a strain of *Scytonema hofmanni* that was isolated from Lake Michigan[95]. Nostodione A is an alkaloid derived from the condensation of tryptophan and phenylpropanoid-derived subunits[95]. Nostodione A had previously been reported as a degradation product of scytonemin[96], a cyanobacterial sunscreen pigment isolated from *Nostoc commune.* Nostodione A was shown to have an IC_{50} of 50 µM in a proteasome inhibitory assay. In 2009 an antimicrobial sesterterpene was isolated from *Scytonema* sp. UTEX 1163[97]. This molecule, called scytoscalarol, contains a guanidino group which had notbeen previously seen in cyanobacterial terpenes, though examples have been found since[98]. Scytoscalarol has been shown to have broad spectrum antimicrobial activity. It was shown to be active against *Bacillus anthracis, Staphylococcus aureus, Escherichia coli, Candida albicans* and *Mycobacterium tuberculosis* with MIC values ranging from 2 to 110 µM. In 2010 two cyclic peptides were isolated from *Scytonema hofmanni* UTEX 1834[99]. Scytonemide A contains an unusual imino linkage and Scytonemide B is a depsipeptide containing a 3-hydroxyoctanoic acid moiety. They were discovered using a bio-assay guided fractionation using a proteasome inhibition assay. However scytonemide B was found to be inactive whilst scytonemide A inhibited the 20S proteasome with an IC_{50} value of 96 nM.

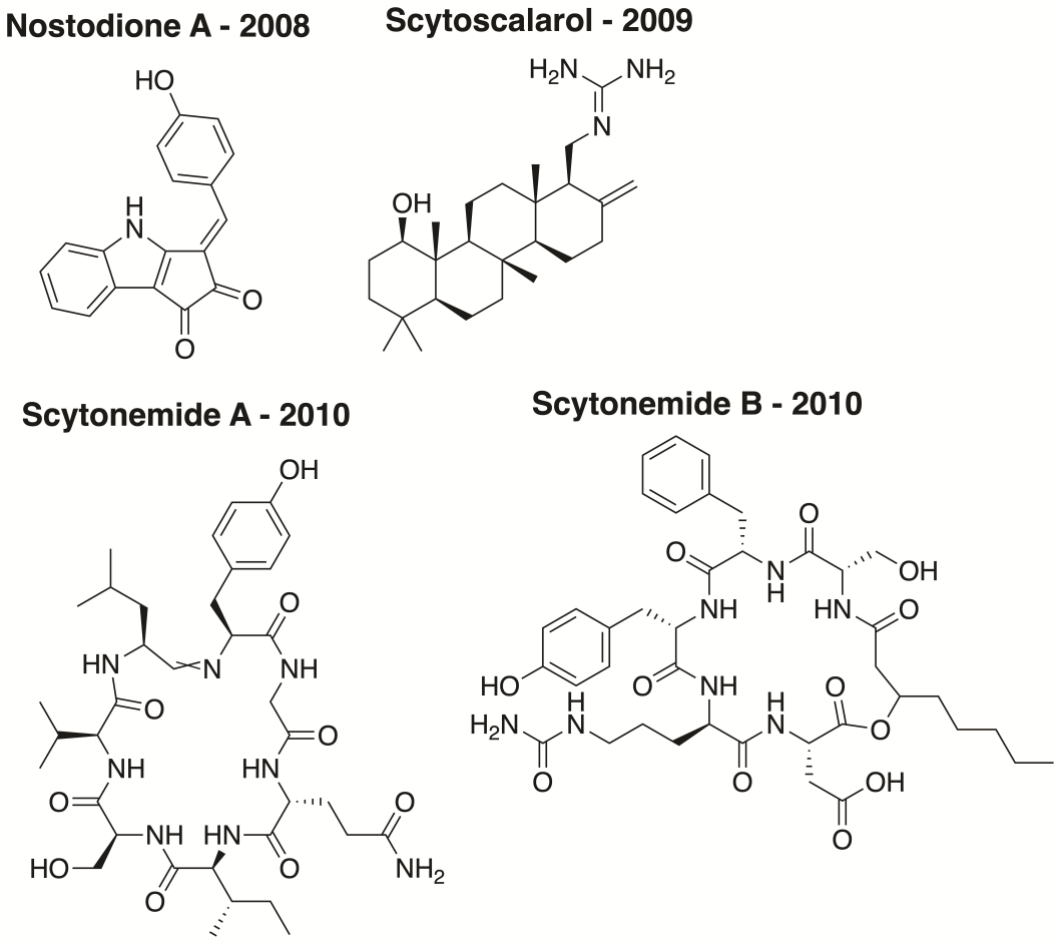

Figure 1.11: The structures of four natural products isolated from *Scytonema* strains from 2008 – 2010.

It was a further ten years later when the next *Scytonema* natural product was reported. In 2020 scytodecamide was isolated from *Scytonema* sp. UIC 100367. Scytodecamide is a cyanobactin which is a family of cyanobacterial RiPPs. Cyanobactins are a varied family, previously being defined as cyclic molecules which contain heterocyclised amino acids. However, there are now cyanobactins that are linear and contain no heterocyclised amino acids. Scytodecamide is a linear decapeptide with an N-terminal N-methylation and a C-terminal adenylation which expands the genetic diversity of cyanobactins7. Scytodecamide was found to have no cytotoxic activity and no further activities were reported7. Later in 2020 three new laxaphycins were isolated from *Scytonema hofmanni* PCC 7110[100]. Laxaphycins are a family of macrocyclic lipopeptides which are produced by numerous cyanobacteria. Laxaphycins contain either eleven or twelve amino acids and are classified as A-type or B-type respectively[100] both types contain either β-aminooctanoic acid or β-aminodecanoic acid. These three new laxaphycins, called scytocyclamides A2, B2 and B3, were all linked to the same hybrid NRPS PKS gene cluster[100]. Interestingly it was found that scytocyclamides A2 and B2 act synergistically to produce anti-fungal activity. On their own neither exhibits activity however when combined they were shown to inhibit the growth of *Aspergillus flavus*[100].

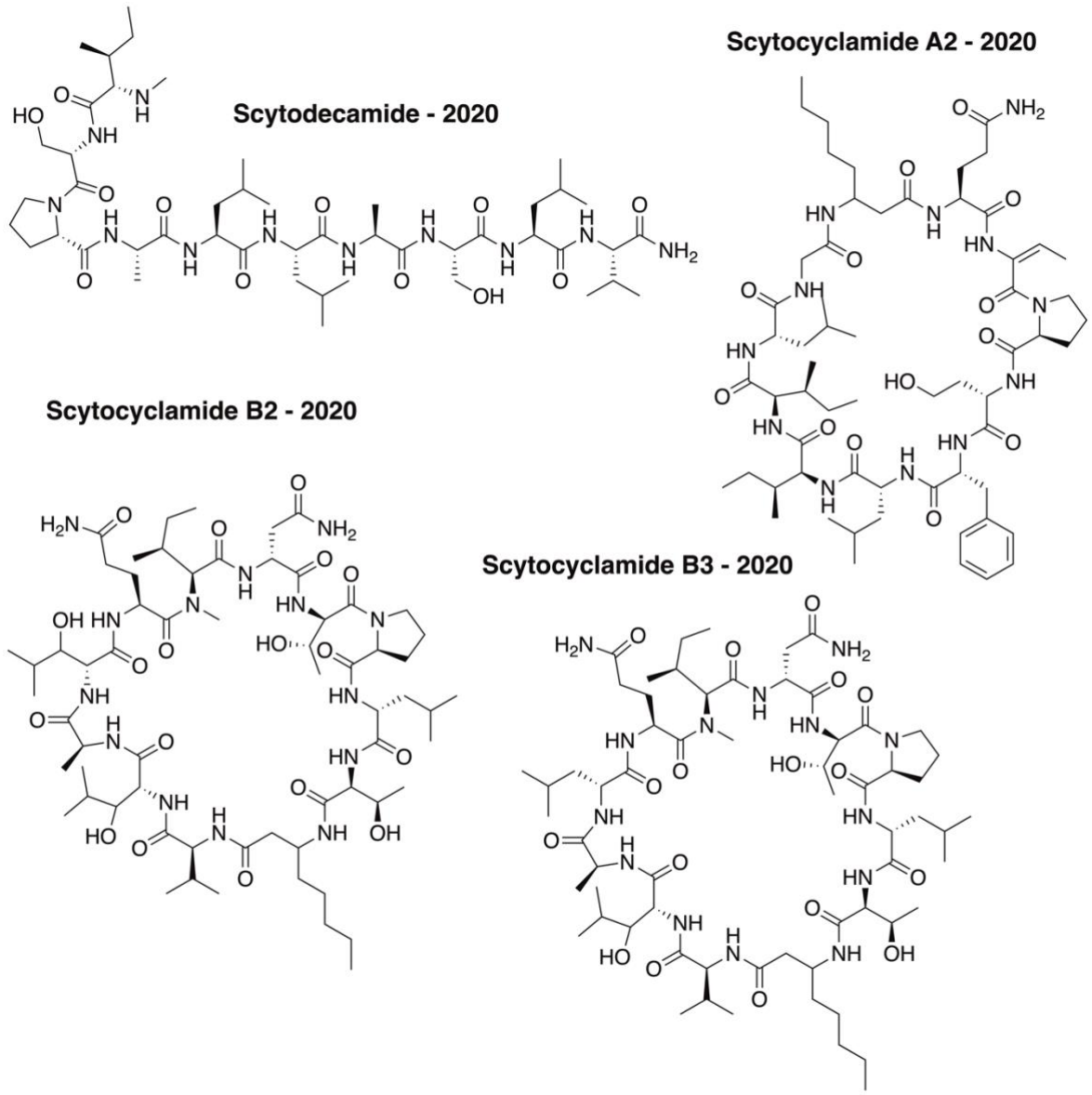

Figure 1.12: The structures of four natural products isolated from *Scytonema* strains in 2020.

As shown above *Scytonema hofmannii* strains are a fruitful source of natural products. The Natural Products Atlas contains 51 *Scytonema* natural products, though this number is inflated by multiple variants of some molecules. Eighteen of these entries in the NP Atlas are attributed to *Scytonema hofmannii strains. Scytonema hofmanni* strains have been shown to have large genomes with many natural product gene clusters. Scytonema, in this regard, show similarity to *Streptomyces*. Comparably very few natural products have been discovered from *Scytonema*. Only six different natural products have been isolated from *Scytonema hofmannii* PCC 7110, even though antiSMASH predicts 28 BGCs for this strain. Despite being the most studied strain of *Scytonema,* less than 25% of its specialised metabolites have been discovered. The antiSMASH shows that *Scytonema* species are understudied despite their high biosynthetic potential, and represent a potential treasure trove of novel specialised metabolites to be discovered.

1.8 Aims of the PhD

The goal of this PhD was to explore the ability of the *Scytonema* genus to produce specialised metabolites. The ultimate goal was to discover novel bioactive specialised metabolites from *Scytonema* species. To do this the following goals were set.

- Research international cyanobacterial strain collections and purchase all available strains of *Scytonema.*
- Attempt to culture all obtained strains of *Scytonema*, and establish a protocol for their propagation in our lab.
- Obtain high-quality whole genome sequences for all obtain strains of *Scytonema*.
- Through the use of computational genomic tools, analyse the *Scytonema* genomes to assess their capabilities to produce specialised metabolites.

After these steps were carried out the project the aim of the project would shift to obtaining novel *Scytonema* specialised metabolites, and then to understand their biosynthesis. The project would then be fluid, following the most promising leads to achieve the ultimate goal of discovering a novel bioactive specialised metabolite from *Scytonema*. The two goals set were:

- Screen the *Scytonema* strains to analyse their ability to produce bioactive metabolites. The culture conditions of the strains could would be altered to try and trigger the production of specialised metabolites
- From the genomic analysis of the *Scytonema* strains, identify particular BGCs of interest and then attempt to express them heterologously.

Chapter 2: Advancing the genomic context of the *Scytonema* genus

2.1 Chapter introduction and aims

The aim of this chapter was to obtain as many strains of diverse *Scytonema* species as possible, sequence them, and then determine their biosynthetic capabilities. The published genome of *Scytonema hofmannii* PCC 7110[85] showed that compared to other cyanobacteria *Scytonema* have large genomes with many specialised metabolite gene clusters. These BGCs have been shown to produce novel natural products[92,93]. However, very few *Scytonema* strains had high quality sequence data available. *Scytonema* have a low GC content and are usually found in co-culture with multiple contaminants. These contaminants can attach to the cyanobacterial cells underneath the sheath that coats the cells which makes it extremely difficult to create axenic cultures[101]. These contaminants complicate genome sequencing resulting in a metagenomic pipeline being required. This chapter also includes the analysis of the metagenomes that were assembled and an overview of the biosynthetic capabilities of *Scytonema.*

2.1.1 Objectives

For this chapter there were three main objectives:

1. Obtain as many strains of *Scytonema* as possible, from worldwide strain collections.
2. Acquire high quality whole-genome sequences for all strains obtained.
3. Analyse the newly sequenced *Scytonema* strains to assess their biosynthetic potential.

2.2 Obtaining strains of *Scytonema*

The initial goal was to obtain as many strains of *Scytonema* as possible, to do this strain collections from around the world were searched and every strain that was available was purchased. *Scytonema* sp. NIES 2130 was obtained one year after the rest of the strains and as a result was not included in the analysis of this chapter. The obtained *Scytonema* strains were isolated from a diverse range of environments from across the globe (Figure 2.1).

Figure 2.1: Map documenting the original location of isolation of obtained strains of Scytonema. The location of isolation was not available for all strains.

Table 2.1: All strains of *Scytonema* obtained, with the strain collection they were identified form and whether it was able to reproducibly culture them in the lab.

Strain	**Strain Collection**	**Culturable?**	**Sequence Available?**	**Culture Media**
Scytonema sp. UTEX 1163	University of Texas	✓	✗	Soilwater: GR+
Scytonema crispum UTEX LB 1556	University of Texas	✓	✗	Modified Bold 3N
Scytonema sp. UTEX LB 2588	University of Texas	✗	✗	BG-11
Scytonema hofmannii UTEX B 1834	University of Texas	✓	✗	Soil Extract Medium
Scytonema hofmannii UTEX B 2349	University of Texas	✓	✓	BG-11
Scytonema sp. UTEX B EE33	University of Texas	✓	✗	BG-11
Scytonema bohnerii SAG 255.80	Göttingen University	✓	✗	ES
Scytonema javanicum SAG 39.90	Göttingen University	✗	✗	Z 45/4
Scytonema lyngbyoides SAG 40.90	Göttingen University	✗	✗	BG-11
Scytonema mirabile SAG 83.79	Göttingen University	✓	✗	ES
Scytonema myochrous SAG 46.87	Göttingen University	✓	✗	ES
Scytonema sp. SAG 67.81	Göttingen University	✓	✗	ES
Scytonema spirulinoides SAG 41.90	Göttingen University	✗	✗	MiEB12

Scytonema sp. NIES 2130	Japanese National Institute	✓	✓	BG-11
Scytonema PCC *7110*	Pasteur Culture Collection	✓	✓	BG-11
Scytonema PCC *7814*	Pasteur Culture Collection	✓	✗	BG-11

2.3 Initial propagation of *Scytonema* strains

Following delivery of the strains, they were subcultured in 175 mL tissue culture flasks containing 50 mL of the media recommended by the strain collections, as well as BG-11 and ES media. They were also subcultured onto solid BG-11 and ES agar plates. The strains were grown in a plant growth room at a temperature of 25 ^{o}C, a light intensity of 1000 lux and a day night cycle of 16:8. Four strains, as indicated in Table 2.1, were unable to be propagated. In addition, other isolates of *Scytonema crispum* have been reported to produce multiple neuro-toxic molecules[102] as a result no further work was done with this strain after it was sequenced, on account of health and safety considerations. All culturable strains, apart from *Scytonema javanicum,* would grow in BG-11 liquid medium. *Scytonema javanicum* would only grow on solid media, which most of the remaining strains would not grow on. Going forward strains that were easily sub-cultured in liquid BG-11 were focused on, with strains being subcultured every four to six weeks.

2.4 Assessing the purity of the acquired *Scytonema* strains

All of the strains we acquired, except for *Scytonema hofmannii* PCC 7110, were xenic cultures due to them being environmental isolates. A xenic culture contains multiple isolates whereas an axenic culture is a monoculture. Visible growth of these contaminants was not seen, the culture medium always remained transparent apart from the clumps of *Scytonema* cells. To get an estimation of the purity of our strains the V4-5 region of the 16s rRNA subunit was amplified for Sanger sequencing. The sequencing chromatograms

showed the number of bases read at each position and therefore provided an indication of how many contaminants were present. It is worth noting this is a very crude approach, however it was quick and cheap, which showed us that our strains each contained multiple contaminants (Figure 2.2). The level of contamination was confirmed by light microscopy as multiple unicellular cells were visible when looking at the cyanobacterial filaments (Figure 2.3).

It is known to be difficult to produce pure cultures of cyanobacteria[101]. Contaminant bacteria form strong attachments to cyanobacterial cells within the sheath that covers the cyanobacterial cell, which affords protection to antibiotics and manual forms of purification. One theory is that *Scytonema* possess so many specialised metabolite gene clusters because they have to control the growth of so many contaminants / symbionts. Thus, by eliminating the contaminants we would potentially stop the production of bioactive specialised metabolites. As a result, it was decided not to attempt any form of purification of the *Scytonema*. We discussed with our sequencing collaborators, Prof. Alison Mather and Dave Baker, whether these impure cultures would negatively affect the quality of our sequencing. Due to their experience in sequencing metagenomes they did not think it would be an issue so it was decided to sequence the *Scytonema* strains as they were.

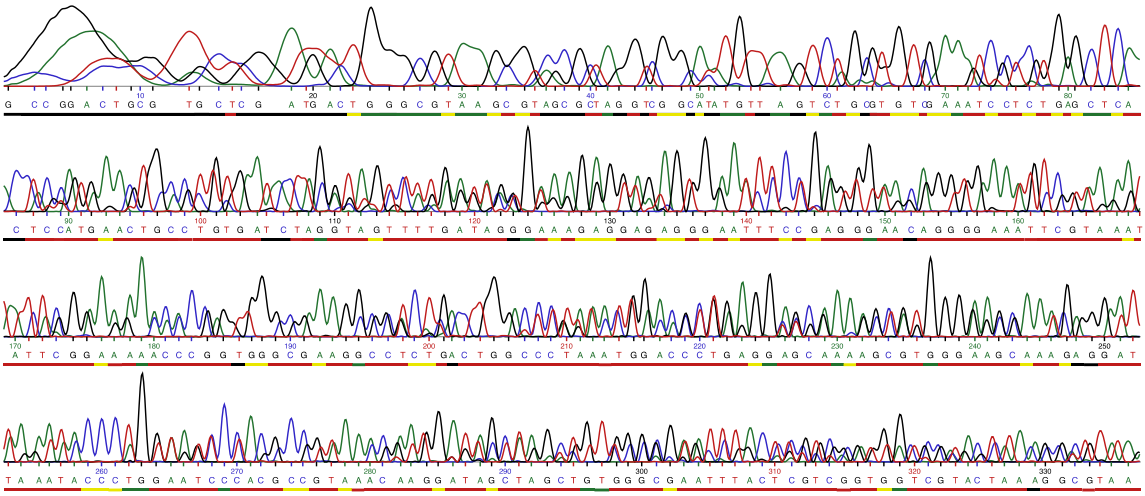

Figure 2.2: Sanger sequencing chromatogram for the 16S V4-5 region of *Scytonema* sp. UTEX EE33. Each coloured line represents a different base. As multiple lines are seen at each position it is clear multiple species are present in this sample.

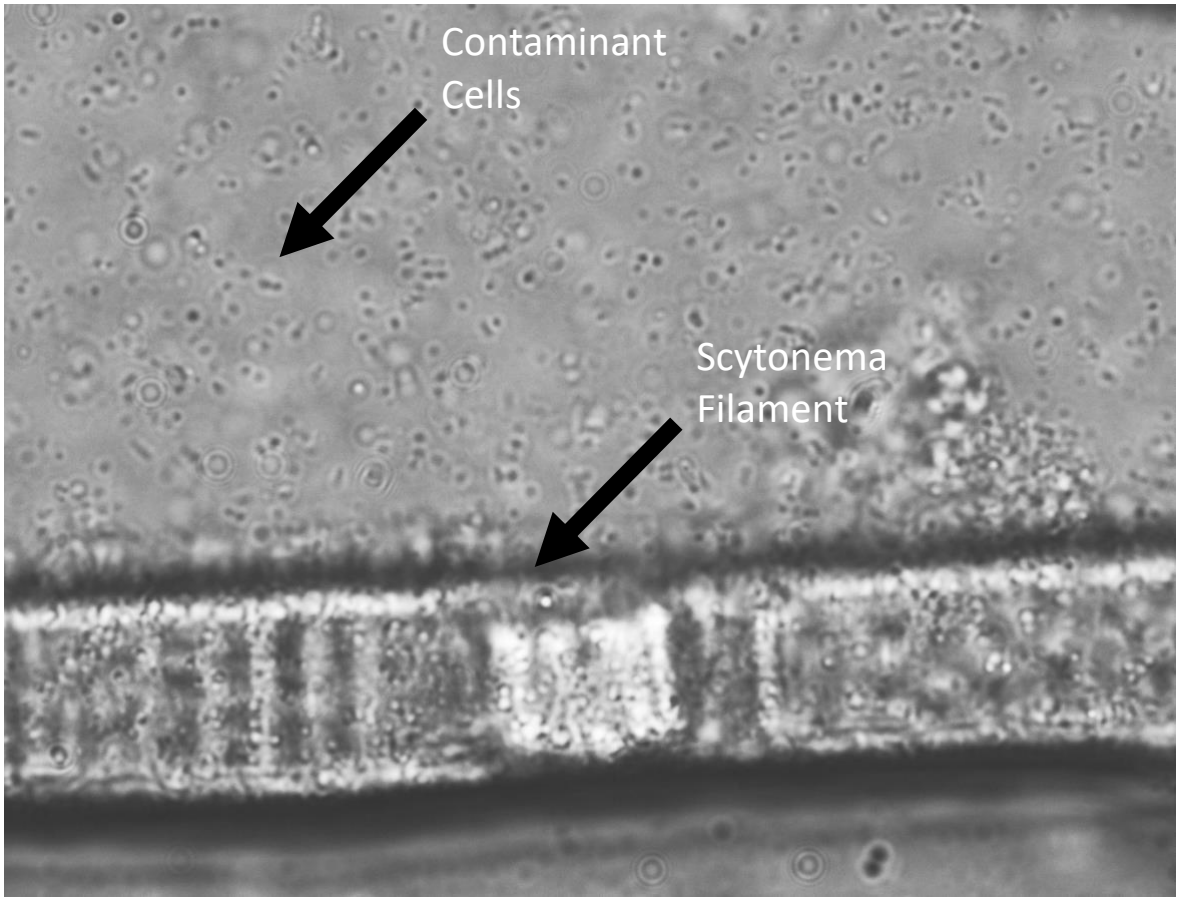

Figure 2.3: Light microscopy of a filament from *Scytonema javanicum* SAG 39.90. Unicellular contaminants are seen both attached to the filament and in the surrounding medium. The *Scytonema* filament runs across the image with small unicellular contaminants clearly seen in the surrounding supernatant.

2.5 Sequencing of *Scytonema* strains

All sequencing was carried out using gDNA that was extracted by myself in our lab. This gDNA was then provided to the various sequencing platforms.

The enhanced sequencing platform at MicrobesNG was initially assessed for sequencing of *Scytonema bohnerii.* This is a hybrid genome sequencing pipeline combining Oxford Nanopore and Illumina sequencing, which we initially believed would provide the read quality needed to assemble the cyanobacterial genomes in the presence of contaminants. However, the returned assemblies and raw reads contained no cyanobacterial DNA. After consultation with MicrobesNG it was decided that another sequencing service would be used as MicrobesNG had no previous cyanobacterial sequencing experience and were unable to correct this issue.

The *Scytonema* strains were sequenced at the Quadram Institute Bioscience (QIB) in collaboration with Prof. Alison Mather, group leader at the QIB, and Dave Baker, head of sequencing at the QIB. Rhiannon Evans (QIB) carried out the library prep and sequencing, and Dr Samuel Bloomfield (QIB) advised on assembling and analysing the genomes, all final genomes were assembled and analysed by myself. We provided gDNA of all culturable *Scytonema* strains, unfortunately *Scytonema hofmannii* UTEX 2349 failed library prep twice and as a result was not sequenced. An existing genome exists for *S. hofmannii* UTEX 2349, however it is low quality. We were provided with long read data produced using an Oxford Nanopore PromethION and 250 base pair paired end Illumina reads. Illumina sequencing is highly accurate, but however the short reads result in more contigs and a more fragmented genome. Oxford Nanopore sequencing is less accurate but provides long reads, up to multiple megabases (MB) long[103]. These long reads act as a scaffold which the Ilumina reads are mapped to. Our goal was to map the *Scytonema* genomes in as few contigs as possible. A contig is a set of overlapping DNA segments that represent a region of DNA. The more contigs that make up a bacteria genome the more fragmented it is and information is lost. If a specialised metabolite gene cluster lies on the edge of a contig then we do not know all

of the genes that make up that gene cluster, which makes cluster analysis and cloning impossible. Biosynthetic gene clusters (BGCs) such as PKSs and NRPSs are large and highly repetitive and thus cannot be effectively sequenced using just short reads.

2.5.1 Genome assembly of *Scytonema* strains

Genome assembly was carried out with assistance from Dr Samuel Bloomfield, a post-doctoral researcher in the Mather Group at QIB. The genomes were assembled using Unicycler[104], a pipeline for the hybrid assembly of bacterial genomes. OPERA-MS[105], an alternate hybrid metagenome assembly pipeline was also trialled. However, OPERA-MS produced higher contig genomes and therefore Unicycler was favoured. As our cultures were not axenic and contained both *Scytonema* and contaminants, a metagenomic pipeline was required to separate the contigs into individual species. This process of separating the contigs is called binning and was done using Metabat27. CheckM[107] was then used to analyse these bins. CheckM looks for single copy genes and tries to assign phylogeny based on this. It will then assign a completeness and a contamination score. The completeness score looks to see what percentage of the expected single copy genes are present. The contamination score looks to see if any of the single copy genes are found more than once. A completeness score greater than ninety-five percent and a contamination score less than five percent indicate a good quality assembly[107].

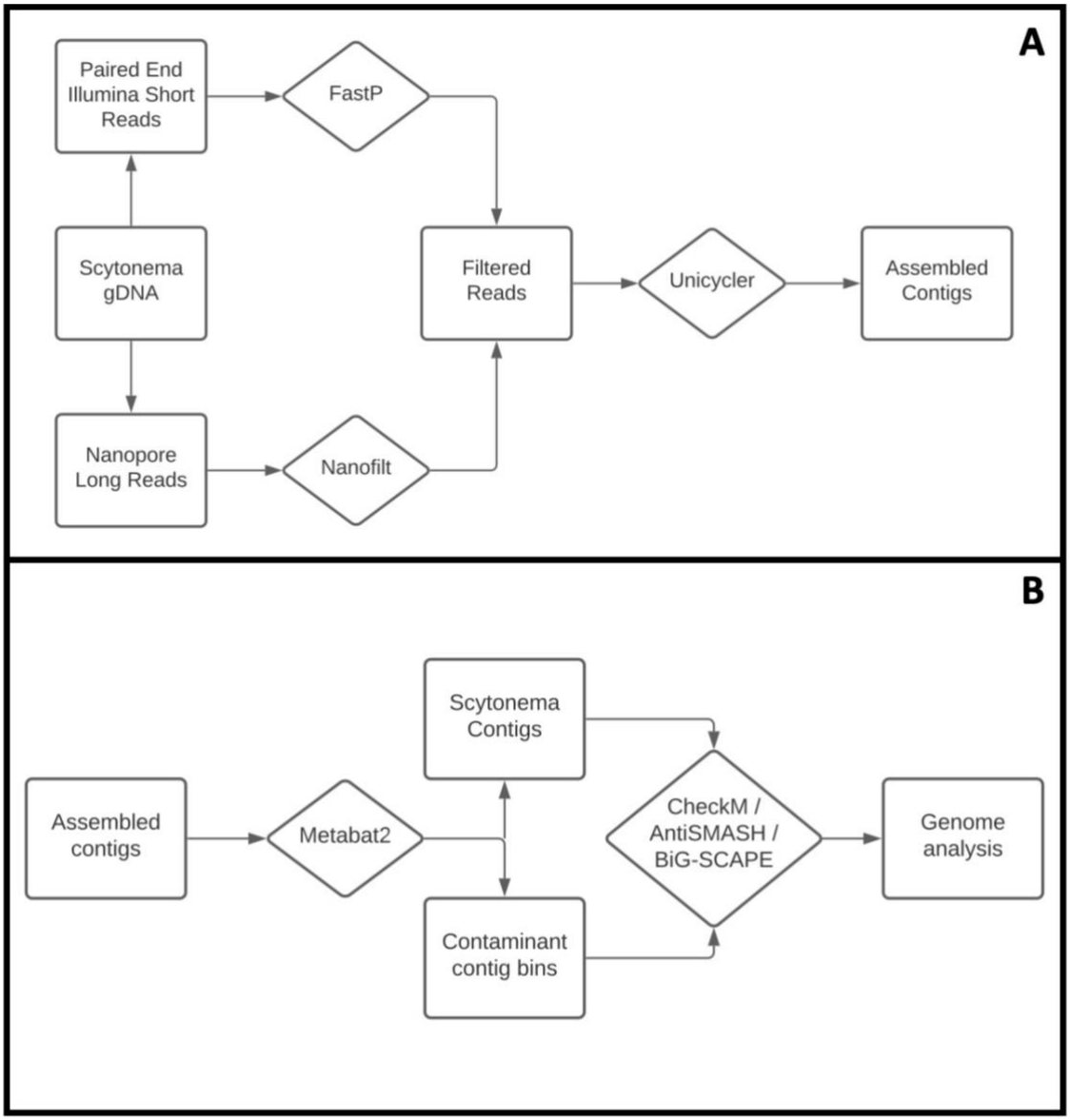

Figure 2.4: A – Workflow for the assembling of the Scytonema genomes showing the computational tools used. B – Workflow for the genomic analysis of the assembled Scytonema genomes.

2.6 Analysis of full genome assembly results

To analyse the quality of our genome assemblies CheckM was used. For each strain multiple bins are acquired, which tells us the number of bacterial contaminants each strain has. CheckM assigns a marker lineage for each bin based on the single copy genes that it finds. CheckM will try to assign the closest lineage possible. The completeness and contamination, as mentioned previously, check for all expected single copy genes and if they appear more than once. The GC percentage is useful as *Scytonema* are low GC, usually around forty percent, which differentiates them from most of the contaminants. The number of contigs, largest contig and N50 value allow the quality of the assembly to be analysed. The N50 value is the size of the largest contig that can bridge the midpoint of the genome, a high N50 value indicates a well assembled genome. For a few bins, CheckM is unable to assign a marker lineage and thus cannot assign a completeness or contamination score. This lack of assignment could be because the bin contains too few single copy genes or the quality of assemblies in that bin is poor. Each strain contains a bin marked Phylum: Cyanobacteria which denotes the *Scytonema* genome. Interestingly, no bin contains a cyanobacterial contaminant.

2.6.1 *Scytonema* myochrous SAG 46.87

The *Scytonema myochrous* genome is assembled into ten contigs with the largest contig spanning seventy-five percent of the genome which indicates a high-quality genome assembly. The completeness and contamination scores are both well within acceptable limits[107]. Two separate *Rhizobiales* contaminants are seen, both of which are well assembled in single digit contigs. Two further bins are found, bins one and two. Bin one contains a genome over nine megabases however no marker lineage could be assigned. This lack of assignment could either be because this bin contains a novel bacterium, or this bin contains DNA from multiple bacteria and is the result of poorly binned contigs. Bin two

has a low completeness score which could be the result of the bacterium being poorly sequenced.

Table 2.2: Sequencing and assembly data for all species found within our isolate of *Scytonema myochrous* SAG 46.87 as reported by CheckM.

Bin	Marker Lineage	Completeness (%)	Contamination (%)	GC%	Genome Size	Contigs	Longest Contig
1	-	-	-	64.2	9,110,892	36	1,522,939
2	Kingdom: Bacteria	18.1	0	60.1	2,297,215	75	197,327
3	Order: Rhizobiales	98.4	5.01	68.1	4,743,908	3	4,625,593
4	Order: Rhizobiales	94.3	2.25	65.6	5,338,152	9	2,208,682
5	Phylum: Cyanobacteria	99.3	1.73	43.7	9,706,163	10	7,676,305

2.6.2 *Scytonema* sp. SAG 67.81

The Scytonema sp. SAG 67.81 genome was assembled well in only seven contigs. The largest contig is not huge at only 2.9 MB but multiple contigs over 2 MB allow the majority of the genome to be covered with minimal gaps. Six contaminants are found with this strain, five of which are well assembled with less than ten contigs. Interestingly the contaminant of the order *Cytophagales* is a single contig assembly with excellent completeness and contamination scores.

Table 2.3: Sequencing and assembly data for all species found within our isolate of *Scytonema* sp. SAG 67.81 as reported by CheckM.

Bin	Marker Lineage	Completeness (%)	Contami-nation (%)	GC%	Genome Size	Contigs	Largest Contig
1	Class: Alphaproteobacteria	100	0.65	67.5	3,127,005	5	1,605,520
2	Phylum: Cyanobacteria	98.4	0.96	42.3	7,654,720	7	2,903,637
3	Order: Actinomycetales	93.4	4.91	69.7	2,847,966	10	709,143
4	Kingdom: Bacteria	91.2	2.20	62.7	4,434,055	2	3,484,014
5	Order: Sphingomonadales	46.4	0.51	64.4	2,025,361	38	318,101
6	Order: Burkholderiales	99.8	1.56	65.9	4,446,141	4	2,811,366
7	Order: Cytophagales	99.1	0.26	48.6	5,966,037	1	5,966,037

2.6.3 *Scytonema* crispum UTEX 1556

Scytonema crispum was the poorest quality genome assembled. The genome was assembled in by far the most contigs, twenty-nine, and the largest contig is only one and a half megabases. Two contaminants did assemble very well, one being a single contig genome and the other a two contig genome. This result shows that there was not an issue with the sequencing or assembly process. Twenty-nine contigs meant that it was still possible to obtain useful information about the genome of *Scytonema crispum*. This is the first *Scytonema crispum* strain to be sequenced.

Table 2.4: Sequencing and assembly data for all species found within our isolate of *Scytonema crispum* UTEX 1556 as reported by CheckM.

Bin	Marker Lineage	Completeness (%)	Contamination (%)	GC (%)	Genome Size	Contigs	Largest Contig
1	Kingdom: Bacteria	2.74	0	64.3	208,445	20	52,473
2	Order: Rhizobiales	98.6	1.26	67.2	5,141,435	21	1,120,877
3	Order: Sphingomonadales	99.4	1.31	64.8	3,597,641	10	1,325,862
4	Kingdom: Bacteria	94.9	1.85	60.7	3,238,804	2	2,728,819
5	Phylum: Cyanobacteria	98.9	2.34	42.6	8,246,590	29	1,557,078
6	Order: Rhizobiales	98.1	0.63	60.3	3,102,143	1	3,102,143
7	Order: Burkholderiales	100	0.86	69.3	3,984,040	1	3,984,040

2.6.4 *Scytonema* sp. UTEX 1163

On first glance the *Scytonema* sp. UTEX 1163 genome assembly looks poor due to its 22 contigs. However, the largest contig covers almost seventy percent of the genome and as a result the majority of the genome is covered with few breaks, allowing for in-depth analysis. Five contaminants are seen for this strain, of which four are well assembled. Interestingly this is one of two strains to have a contaminant from the family *Xanthamonadaceae* which is a family that contains known plant pathogens such as *Xanthomonas oryzae,* which causes rice blight.

Table 2.5: Sequencing and assembly data for all species found within our isolate of *Scytonema* sp. UTEX 1163 as reported by CheckM.

Bin	Marker Lineage	Completeness (%)	Contamination (%)	GC (%)	Genome Size	Contigs	Largest Contig
1	Order: Rhizobiales	99.6	4.96	69.0	5,121,320	11	1,275,436
2	Order: Rhizobiales	31.6	0.00	64.9	2,560,247	73	131,250
3	Family: Xanthomonadaceae	87.3	0.43	67.4	3,248,560	6	1,130,967
4	Order: Sphingomonadales	97.6	1.72	67.1	4,420,310	8	1,712,612
5	Order: Sphingomonadales	99.2	13.2	64.8	3,777,221	3	2,904,559
6	Phylum: Cyanobacteria	99.5	3.82	43.8	9,877,337	22	6,786,929

2.6.5 *Scytonema* sp. PCC 7814

The CheckM analysis for *Scytonema* sp. PCC 7814 indicates that this is one of the less contaminated strains in our collection. Two contaminants are found, however neither is well sequenced. When purchased from the Pasteur Culture Collection the strain was supposed to be axenic. It clearly is not axenic but is less contaminated than many other *Scytonema* isolates. The *Scytonema* genome is assembled to a high quality. The largest contig covers almost eighty percent of the genome and only eight contigs cover the genome.

Table 2.6: Sequencing and assembly data for all species found within our isolate of *Scytonema* sp. PCC 7814 as reported by CheckM.

Bins	Marker Lineage	Completeness (%)	Contamination (%)	GC %	Genome Size	Contigs	Largest Contig
1	Order: Rhizobiales	2.80	0.00	64.1	300,592	12	139,084
2	Order: Sphingomonadales	99.1	7.20	63.6	5,509,590	39	566,428
3	Phylum: Cyanobacteria	99.3	1.73	43.7	9,683,567	8	7,801,628

2.6.6 *Scytonema* javanicum SAG 39.90

The *Scytonema javanicum* genome is one of the best assembled genomes, where the entire genome is covered in only three contigs and two of these are almost four megabases long. Four contaminants were detected via sequencing, although only one (an Actinomycetales strain) is well assembled. It is worth noting that *Scytonema javanicum* was the only strain grown on solid media.

Table 2.7: Sequencing and assembly data for all species found within our isolate of *Scytonema javanicum* SAG 39.90 as reported by CheckM.

Bins	Marker Lineage	Completeness (%)	Contamination (%)	GC %	Genome Size	Contigs	Largest Contig
1	Order: Sphingomonadales	97.3	61.2	64.0	8,070,706	207	303,873
2	Kingdom: Bacteria	52.9	7.26	67.8	2,883,827	112	270,635
3	Order: Actinomycetales	97.7	1.85	68.3	3,984,995	10	1,876,408
4	Order: Rhizobiales	94.9	5.11	65.6	5,434,108	28	1,070,155
5	Phylum: Cyanobacteria	99.8	1.29	42.7	8,037,170	3	4,001,603

2.6.7 *Scytonema* sp. UTEX EE33

Whilst the number of contigs for *Scytonema* sp. UTEX EE33 is high, the overall quality of the assembly is high. This high quality is shown by the largest contig being over six megabases. Two contaminants are found, both of which are well sequenced shown by the low number of contigs.

Table 2.8: Sequencing and assembly data for all species found within our isolate of *Scytonema* sp. UTEX EE33 as reported by CheckM.

Bins	Marker Lineage	Completeness (%)	Contamination (%)	GC %	Genome Size	Contigs	Largest Contig
1	Class: Alphaproteobacteria	99.1	1.74	67.5	3,601,124	5	1,314,857
2	Class: Alphaproteobacteria	100	1.06	69.8	3,015,426	6	1,193,612
3	Phylum: Cyanobacteria	99.3	2.13	43.8	9,876,330	17	6,336,728

2.6.8 *Scytonema* bohnerii SAG 255.80

The *Scytonema bohnerii* genome is very well sequenced in only five contigs and the largest contig spans over eighty percent of the genome. Only one contaminant was detected, which is also well sequenced.

Table 2.9: Sequencing and assembly data for all species found within our isolate of *Scytonema bohnerii* SAG 255.80 as reported by CheckM.

Bin	Marker Lineage	Completeness (%)	Contamination (%)	GC %	Genome Size	Contigs	Largest Contig
1	Order: Actinomycetales	99.0	0.06	70.0	3,205,304	4	1,406,333
2	Phylum: Cyanobacteria	99.3	0.33	41.2	8,814,754	5	7,523,817

2.6.9 *Scytonema* mirabile SAG 83.79

The *Scytonema mirabile* assembly is very high quality being only four contigs. Only one additional bin is assembled. The bin only contains six hundred kilobases of DNA and no marker lineage is able to be assigned. The GC content of this bin, 40.5 percent, is the same as *Scytonema mirabile*, this DNA could belong to the *Scytonema* and may simply be an artefact of the assembly pipeline or a plasmid that is separately binned. As a result, *Scytonema mirabile* appears to be the only axenic strain in our collection, despite the strain collection stating that it was xenic.

Table 2.10: Sequencing and assembly data for all species found within our isolate of *Scytonema mirabile* SAG 83.79 as reported by CheckM.

Bin	**Marker Lineage**	**Completeness (%)**	**Contamination (%)**	**GC %**	**Genome Size**	**Contigs**	**Largest Contig**
1	-	-	-	40.5	696,619	7	207,682
2	Phylum: Cyanobacteria	99.3	0.11	40.5	8,535,723	4	3,879,605

2.6.10 *Scytonema* hofmannii UTEX 1834

The *Scytonema hofmanni* UTEX 1834 genome is the largest of the strains that were sequenced at twelve and a half megabases. Given its size, it is well assembled in fourteen contigs with the largest contig spanning almost fifty percent of the genome. The only contaminant is a single contig assembly of a Xanthamonadaceae.

Table 2.11: Sequencing and assembly data for all species found within our isolate of *Scytonema hofmanni* UTEX 1834 as reported by CheckM.

Bin	Marker Lineage	Completeness (%)	Contamination (%)	GC %	Genome Size	Contigs	Largest Contig
1	Family: Xanthamonadaceae	99.9	0.69	67.2	3,835,248	1	3,835,248
2	Phylum: Cyanobacteria	99.8	1.53	41.9	12,598,546	14	5,965,281

2.6.11 Analysis of assembled *Scytonema* genomes

Table 2.12 shows the assemblies for the *Scytonema* strains. Overall the quality of the bacterial genome assemblies is high. The assembly for *Scytonema crispum* is the poorest. The majority of the sequencing reads map to contaminants and not to the *Scytonema crispum* assembly, leading me to believe that this is the most contaminated strain and as result the assembly has more contigs. The rest of the genomes have an L50 value of two or less, which shows that the majority of the assemblies are very contiguous and therefore reduces the chances of a BGC lying on a contig boundary. In some cases, the number of contigs is not directly proportional to how fragmented an assembly is. Seven of the assemblies contain circular plasmids, which increase the number of contigs without making the assembly more fragmented. Two of the assemblies contain a circular main chromosome. The fact that a single contig was able to cover the whole of the main chromosome speaks to the quality of the assemblies. The sequencing shows that *Scytonema* do generally possess large genomes, where the smallest is over seven and a half megabases with most between eight and ten megabases. Interestingly, *Scytonema*

hofmannii UTEX 1834 has a genome larger than *Scytonema hofmannii* PCC 7110 at over twelve and a half megabases.

Table 2.12: Assessment of the quality of sequencing and assembly of the Scytonema strains. Data was reported by CheckM. L50 is the number of contigs required to cover fifty percent of the genome. L75 is the number of contigs required to cover seventy-five percent of the genome. Predicted BGCs (Biosynthetic Gene Clusters) is the number of gene cluster regions identified by AntiSMASH 7.0.

Strain	Contigs	Total Size (bp)	Largest contig (bp)	Circular contigs	GC %	L50	Predicted BGC's
Scytonema bohnerii SAG 255.80	5	8,814,753	7,523,817	3	41.2	1	15
Scytonema crispum UTEX 1556	29	8,246,890	1,557,078	0	42.6	4	14
Scytonema mirabile SAG 83.79	4	8,535,723	3,879,605	0	40.5	2	18
Scytonema hofmannii UTEX 1834	14	12,598,546	5,965,281	8	41.9	2	28
Scytonema javanicum SAG 39.90	3	8,037,170	4,001,603	1	42.7	2	18
Scytonema sp. PCC 7814	8	9,683,567	7,801,628 (circular)	3	43.7	1	29
Scytonema myochrous SAG 46.87	10	9,706,163	7,676,305 (circular)	5	43.7	1	31
Scytonema sp. UTEX EE33	17	9,876,330	6,336,728	6	43.8	1	24
Scytonema sp. SAG 67.81	7	7,654,720	2,903,637	0	42.3	2	22
Scytonema sp. UTEX 1163	22	9,877,337	6,786,929	6	43.8	1	30

2.6.12 Conclusions drawn from assembly analysis

The CheckM data shows that overall the *Scytonema* assemblies are of a high quality. Somewhat surprisingly, the majority of the contaminants are also well assembled, with multiple being single contig assemblies, which validates our sequencing and assembly pipelines. There does seem to be a link between the purity of the culture and the quality of the assembly. *S. bohnerii*, *S. mirabile* and *S. hofmanni* UTEX 1834 all have one or zero contaminants and are all very well assembled. Conversely, *Scytonema crispum* is the worst assembled strain and also the most contaminated.

A summary of the contaminant strains is shown in Figure 2.5. The most common contaminants are of the orders Rhizobiales and Sphingomonadales. Rhizobiales can fix nitrogen and are known to have a positive symbiotic relationship with both plants and cyanobacteria[108]. Rhizobiales are also found as a third member in cyanolichens, alongside fungi and cyanobacteria[108]. Sphingomonadales and the rest of the contaminants are all known cyanobacterial contaminants[109]. The role of these co-associated bacteria is unknown.

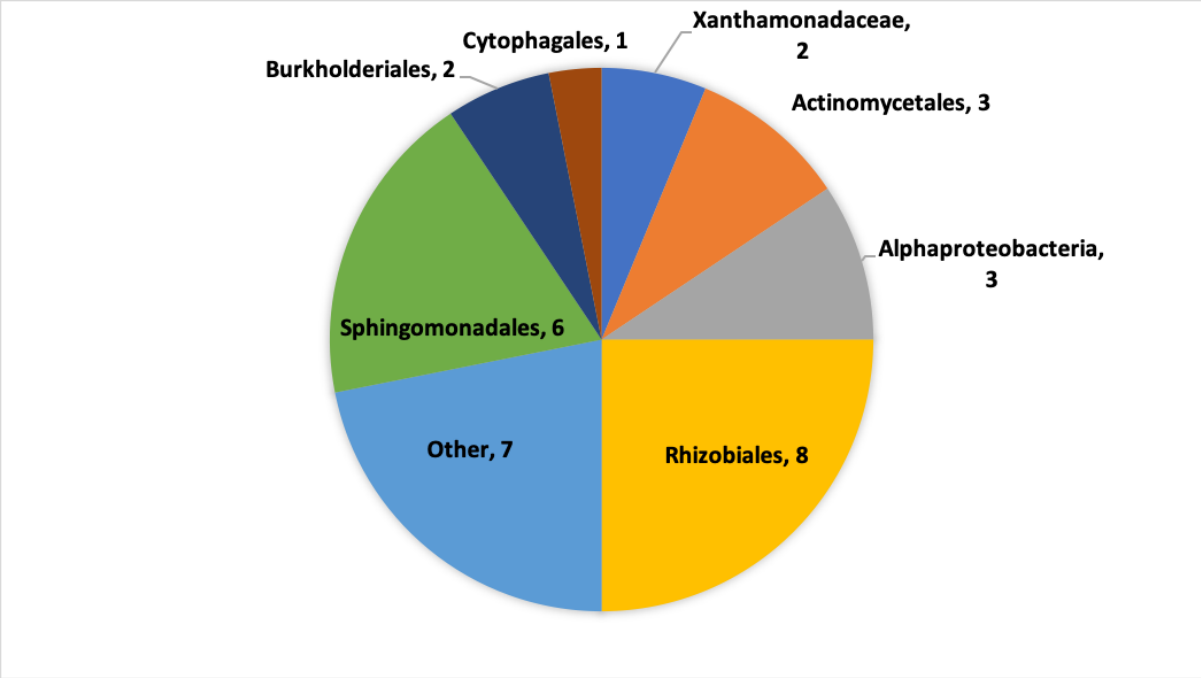

Figure 2.5: Pie chart representing the proportion of each type of contaminant found within all of the sequenced *Scytonema* strains. Data is based on marker lineage as reported by CheckM.

2.7 Analysing the biosynthetic potential of *Scytonema*

2.7.1 Analysing the phylogeny of sequenced *Scytonema* strains

The classification of cyanobacterial strains is known to be inconsistent with strains often being reclassified[110]. This inconsistent classification is due to strains being originally classified based on morphology rather than genomic data. In order to confirm that all of our strains are indeed *Scytonema* a phylogenetic tree was produced using AutoMLST[111]. AutoMLST stands for automated multi-locus species tree. AutoMLST maintains a database of the top ten highest quality genomes for each species based on reference genomes obtained from the NCBI Refseq database[112]. The average nucleotide identity (ANI) of input strains is compared to this database and the nearest reference organisms from NCBI RefSeq are selected. Single copy genes with dN/dS values less than one are then identified. These single copy genes are then aligned and trimmed before the tree is produced. The phylogenetic tree, shown in Figure 2.6, can be split into three distinct clades based upon the closely related strains. The Tolypothrix clade contained two of our *Scytonema* strains, *Scytonema crispum* UTEX 1556 and *Scytonema hofmannii* UTEX 2349. *Scytonema hofmannii* UTEX 2349 has since been reclassified to *Tolypothrix* sp. PCC 9009. *Scytonema crispum* UTEX 1556 grew in a single large clump unlike other *Scytonema*, which grow in multiple small balls. It is also worth noting that other strains of *Scytonema crispum* have been reclassified to the *Heteroscytonema* genus[102]. Our results comply with this and suggest *Scytonema crispum* species are distinct from *Scytonema. Scytonema mirabile* SAG 83.79 and *Scytonema bohnerii* SAG 255.80 both clade with *Nostoc* strains. Both of these strains form biofilms which is typical of *Nostoc* but not common for *Scytonema.* Seven of our strains are in the *Scytonema* clade. Three strains not previously classified as *Scytonema* are also in this clade, which highlights the poor classification of Cyanobacteria.

Figure 2.6: Phylogenetic tree containing all of our sequenced *Scytonema* strains produced by AutoMLST. Highlighted strains are our sequenced strains that I inputted to AutoMLST, the rest of the strains are the closest related strains as defined by AutoMLST. Three clades are identified based upon the closest related matches as well as the morphology of the strain. The *Tolypothrix* clade is shown in yellow. The *Nostoc* clade is shown in red. The *Scytonema* clade is shown in green.

2.7.2 AntiSMASH analysis of sequenced *Scytonema* strains

AntiSMASH[16] is an online tool that searches for specialised metabolite gene clusters in a given input sequence. All of our sequenced strains were analysed using AntiSMASH 7.0. The AntiSMASH results showed that strains within our *Scytonema* clade contained between eighteen and thirty-one BGCs. The three strains that were sequenced but fell outside the *Scytonema* clade *S. mirabile*, *S. bohnerii*, and *S. crispum* were predicted to have fifteen, fifteen and eighteen BGCs respectively. The fact that the three strains with the fewest BGCs fell outside of the *Scytonema* clade, lends credence to our hypothesis that true *Scytonema* are richer in specialised metabolite gene clusters, compared to closely related genera. Figure 2.7 shows the number of AntiSMASH gene cluster regions predicted for *Scytonema* and closely related strains. Eighty percent of strains in the *Scytonema* clade are predicted to contain twenty or more BGCs. Whereas, only ten percent of strains in Figure 2.7 outside of the *Scytonema* clade contain twenty or more BGCs, which clearly shows that *Scytonema* are enriched in BGCs compared to closely related strains. Five of our sequenced strains contain over twenty-five clusters. One notable feature is that many of the hybrid gene cluster regions contain multiple discrete gene clusters, each potentially capable of producing a separate natural product. One characterised example is found in *S. hofmannii* PCC 7110 (shown in Figure 2.8), where one AntiSMASH gene cluster region contains the BGCs that produce Scyptolin[92] A, B and Hofmannolin, as well as a further putative microviridin BGC. These populated regions mean that the total number of natural products that these *Scytonema* strains can produce is likely to be higher than the total number of AntiSMASH gene cluster regions. Based on this analysis, *Scytonema* are rich in RiPPs, NRPSs and terpenes, which constitute 65% of the total BGCs (Figure 2.9).

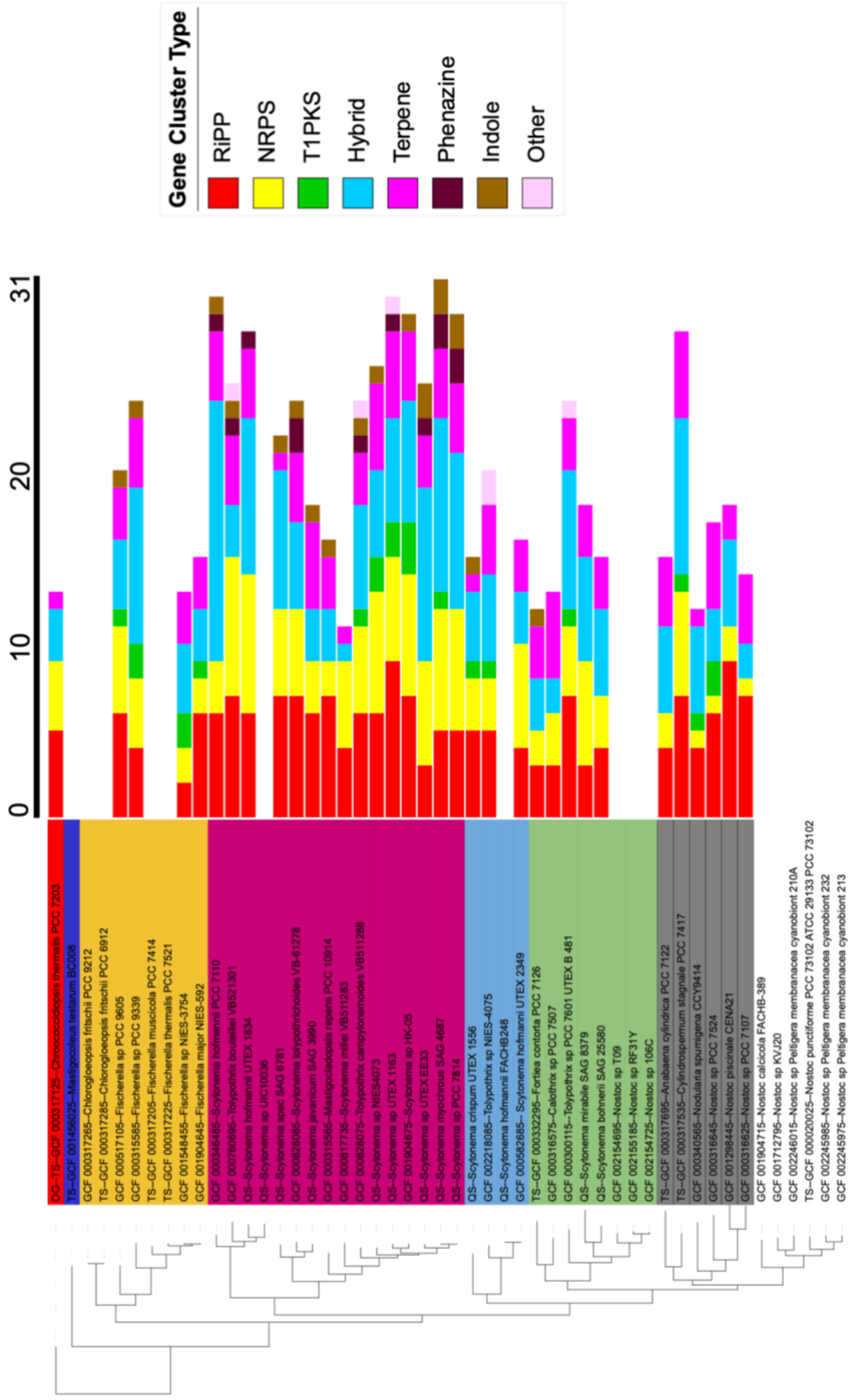

Figure 2.7: A phylogenetic tree containing the *Scytonema* clade, which is shown in plum, in addition to closely related cyanobacterial strains. Strains are overlaid with the number of predicted AntiSMASH gene cluster regions, colour coded to indicate the type of gene cluster. Strains which do not have the number of gene clusters shown have assemblies that are too fragmented to get reliable gene cluster data.

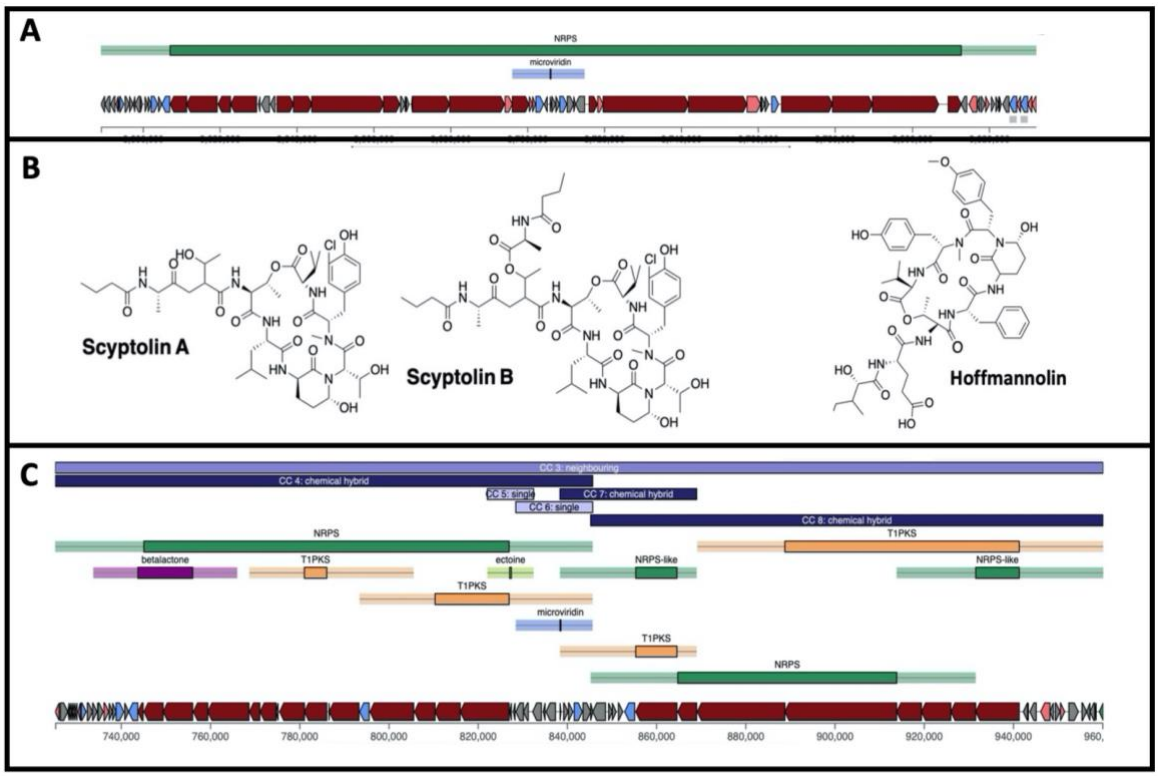

Figure 2.8: A - An AntiSMASH gene cluster region from *S. hofmannii* PCC 7110. this region contains the BGCs that produce Scyptolin A, B and Hofmannolin. The structures of these molecule are shown in B. A gene cluster from *S. hofmannii* UTEX 1834 is shown in C. Like the cluster from PCC 7110 this AntiSMASH gene cluster region is predicted to contain multiple individual BGCs. This region is predicted to produce three NRPS/PKS hybrids, a microviridin and an ectoine. No known products have been linked to this mega cluster from UTEX 1834.

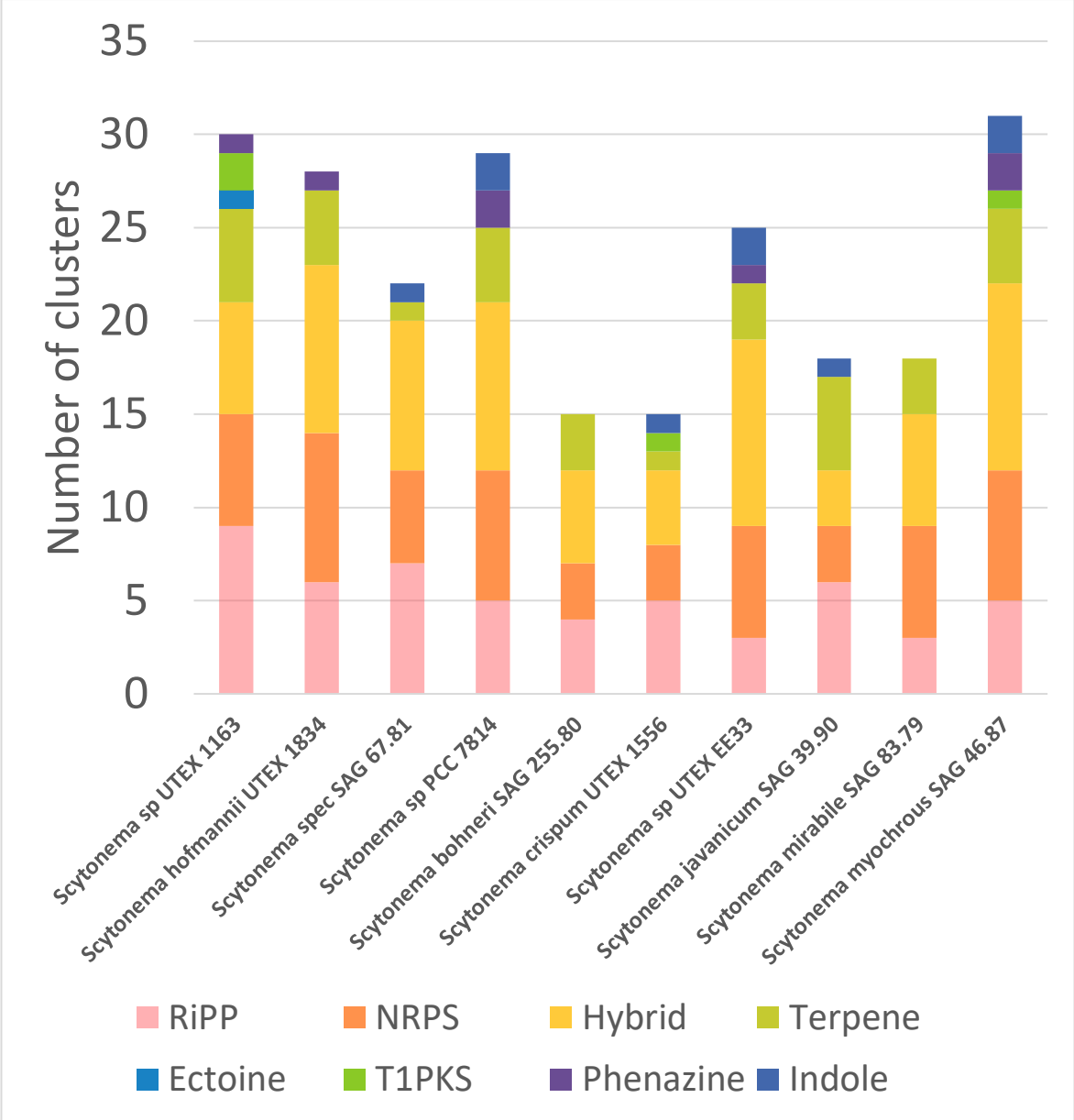

Figure 2.9: Graph showing the number of predicted gene cluster regions for each sequenced strain of *Scytonema*. The number of clusters is predicted using AntiSMASH and each is a single gene cluster region. The class of each gene cluster region is shown by colour, if multiple biosynthetic gene clusters are predicted it is shown as a hybrid cluster alongside true hybrid clusters such as NRPS-PKS hybrids.

2.7.3 Exploring the diversity between *Scytonema* gene clusters

To determine how much similarity there was between the specialised metabolite gene clusters amongst the *Scytonema* strains a program called BiG-SCAPE[22] was used. BiG-SCAPE takes antiSMASH gene clusters alongside MIBiG annotated clusters as input to build BGC similarity networks and identify gene cluster families (GCFs). These sequences are then translated into a string of Pfam domains. The percentage of shared domains is then calculated alongside the percentage of pairs of adjacent domains as well as the overall sequence identity between the protein domains. BiG-SCAPE then calculates a distance matrix based on these values and creates links between the BGCs based upon a distance matrix cut off. A lower value for the cut off creates tighter groups of more similar compounds, a higher value creates broader families. All sequenced *Scytonema* genomes were included as input alongside already sequenced *Scytonema* genomes that were found within the *Scytonema* clade as shown in Figure 2.7. The default matrix cut-off is 0.3 however this was raised to 0.5 to try and increase the connectivity. Despite this permissive setting the results, as shown in Figure 2.10, show very few links being made, seventy percent of the gene clusters inputted form no links. This is much less than expected and less than most other bacterial genera, which could be due to including not enough strains and introducing more strains into the analysis could produce more links. When comparing our results to similar studies, such as an exploration of *Rhodococcus* NRPS distribution[113], our lack of connectivity is clear. In the *Rhodococcus* study many large clusters are formed, and there are very few singletons, less than 1% of inputted BGCs. The matrix cut off they used was 0.6, this is similar to our value of 0.5 but slightly more permissive. The more permissive cut off is not enough to cause such a contrast in results. They also used many more inputted BGCs, as many more *Rhodococcus* strains have been sequenced when compared to *Scytonema.* It is worth noting that our dataset is relatively small and it is assumed more connections would form with further genomes and BGCs added. Despite these caveats it is clear that *Scytonema* are an extremely diverse genus.

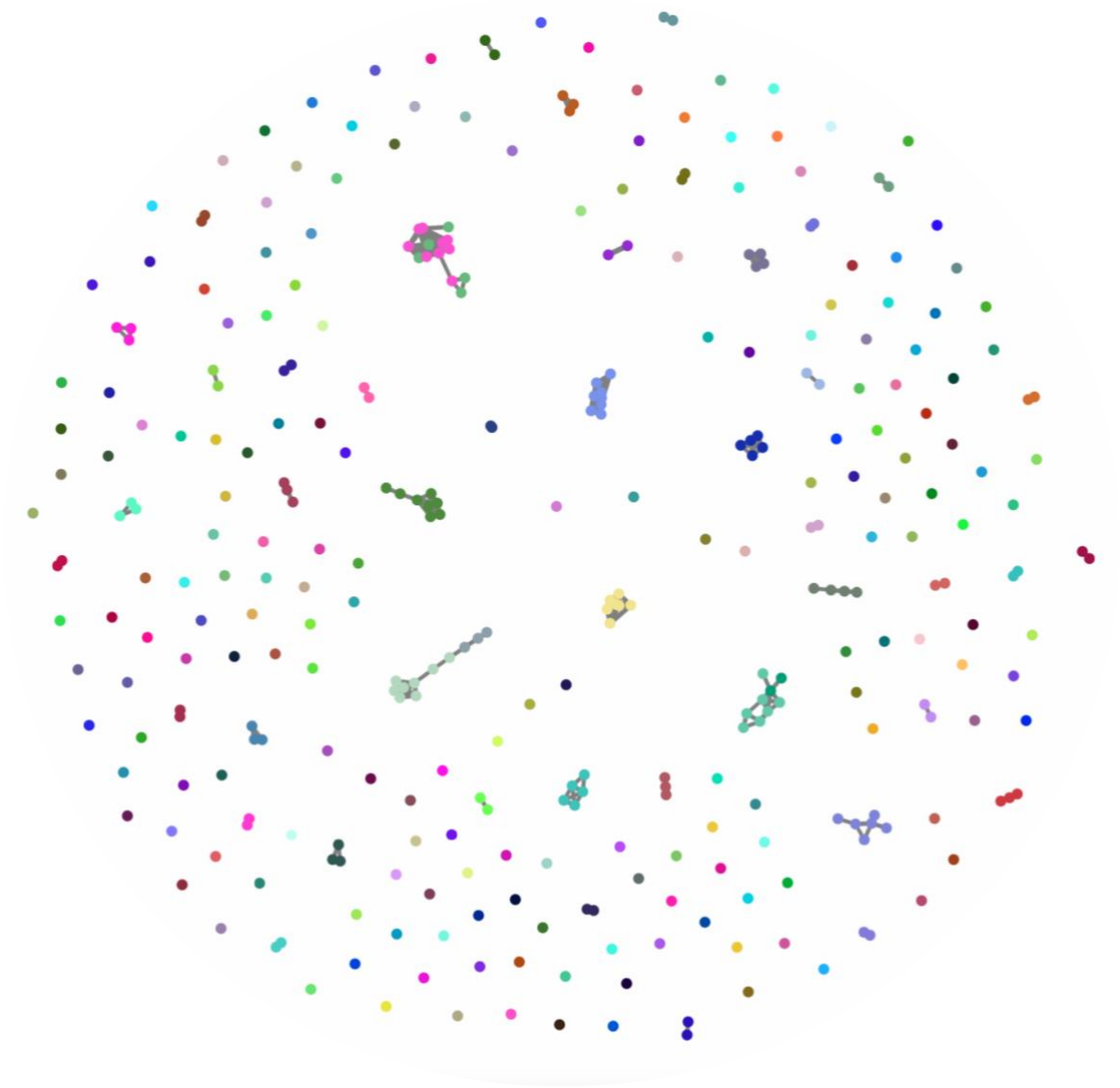

Figure 2.10: Gene cluster homology network created by BiG-SCAPE. Each dot represents a gene cluster from the input data set which is all strains in the *Scytonema* clade as defined by the AutoMLST phylogenetic tree. A matrix cut off of 0.5 was used. Colours are assigned arbitrarily by BiG-SCAPE and all connected gene clusters will share the same colour. Few clusters are seen, the majority of the gene clusters do not form connections.

2.7.4 Exploring the diversity of *Scytonema* genes

The BiG-SCAPE analysis confirmed that there is little homology between *Scytonema* gene clusters. To see if this lack of homology extended to all genes, Roary[114] was used to carry out a pangenome analysis. Roary extracts all coding sequences from the inputted strains and performs an all against all comparison using BlastP. The genomes of all strains in the *Scytonema* clade were used as an input. The output from Roary shows how many strains have a homologue of a particular gene, a seventy percent identity cut off was used. This identity cut off was used as it is the lowest value recommended by Roary and represents an identity cut-off often used for core genes from the same genus[115]. Even with this lowest recommended identity cut off sixty-five percent of genes are only found in a single strain, which shows that *Scytonema* strains are very distinct from one another. This lack of homology could be because of the diverse environments they are found in (Figure 2.1) and the diverse communities of symbionts that they exist with. It also shows that we are only scratching the surface of *Scytonema* as we would expect to eventually hit the pan-genome limit where no new families are found. We are clearly far from that limit currently.

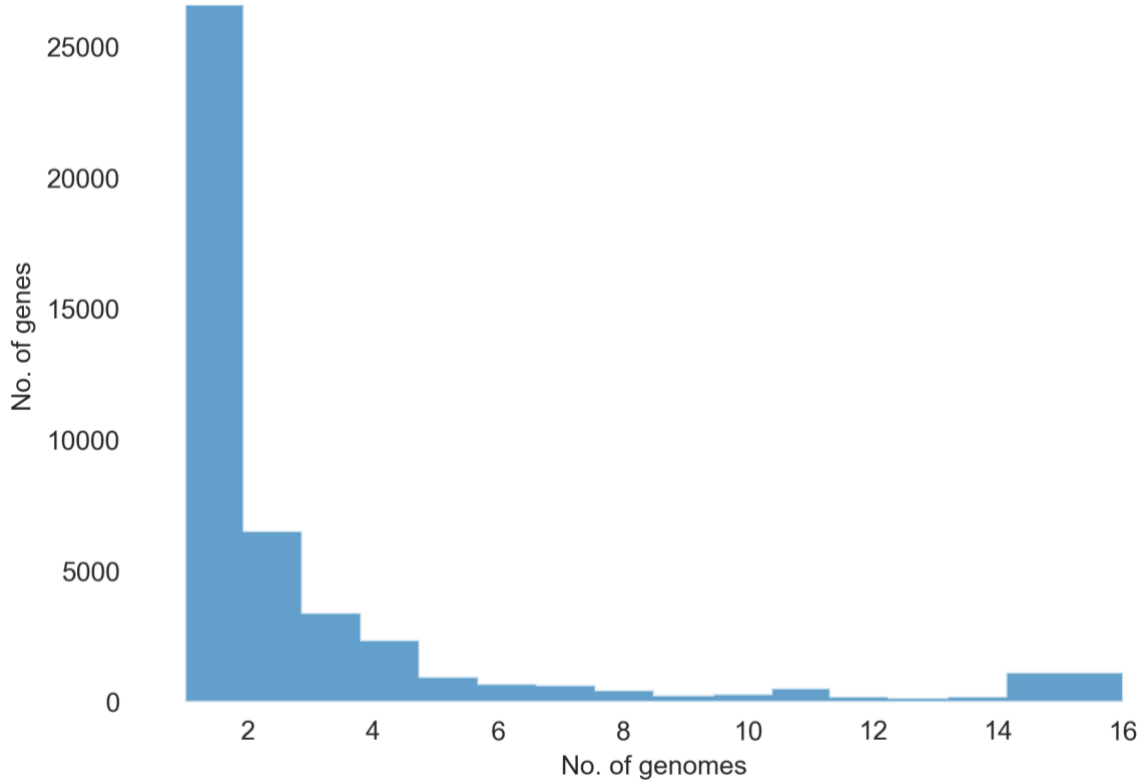

Figure 2.11: Graph showing the results of a pan genome analysis carried out using Roary. The graph shows how many strains within the *Scytonema* clade, as defined by the AutoMLST phylogenetic tree, share a particular gene (with seventy percent or higher identity).

2.8 Homology of *Scytonema myochrous* 46.87 and *Scytonema* sp. PCC 7814

Scytonema sp. PCC 7814 is predicted to contain twenty-nine specialised metabolite gene cluster regions. It was observed in the BiG-SCAPE networking that these twenty-nine gene clusters each formed a link with a BGC from *S. myochrous*. It was found the genes in each pair of linked clusters were one-hundred percent homologous. *Scytonema myochrous* contains two further gene clusters. The phylogenetic tree (Figure 2.6) shows that these strains are closely related. The genome assemblies for both of these strains have circular chromosomes meaning that the assembly of the chromosome is complete and full information is known. The chromosomes are of different sizes which confirms that these are distinct strains. To assess for synteny between these strains, their genomes were aligned using Mauve7. Coloured regions indicate homology with straight red lines indicating contig boundaries. It clearly shows that the main chromosome is not only rearranged but contains different portions of the DNA. This rearrangement shows that these genomes have undergone some rearrangement, which differentiates them along with the additional BGCs in *S. myochrous*. These two strains initially had distinct phenotypes despite their sequence similarity. PCC 7814 grows extremely well and forms dense cultures, whereas *S. myochrous* initially grew extremely poorly. However, after three years of continuous subculture *S. myochrous* has begun to grow well with a similar growth phenotype to PCC 7814. It is unknown what has caused this shift as the culture conditions have not changed. It is possible that this has been caused by a gene mutation or the loss of a plasmid, which could be confirmed by re-sequencing the genome.

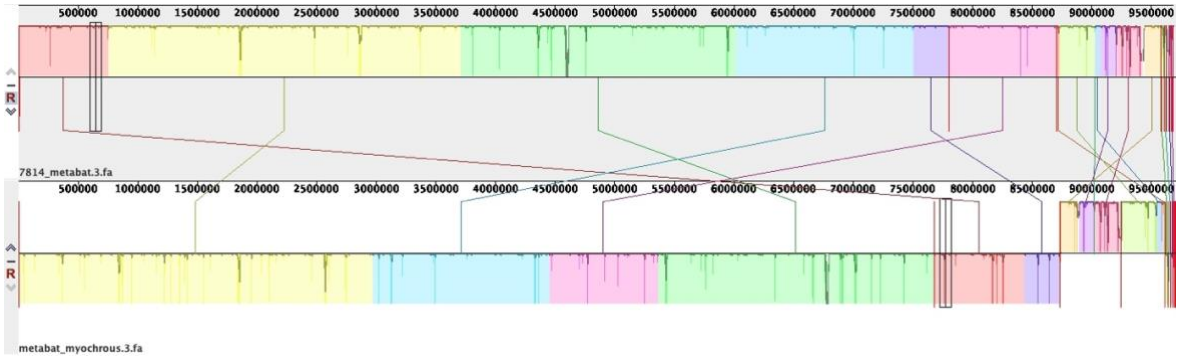

Figure 2.12: This figure shows the homology between the genomes of *Scytonema myochrous* SAG 46.87 and *Scytonema* sp. PCC 7814. Homologous regions are highlighted in the same colour and linked by a line. Contigs in the assembly are shown by straight red lines between highlighted regions.

Chapter 3: OSMAC based screening of *Scytonema* strains to trigger the production of novel natural products

3.1 Chapter introduction and aims

3.1.1 One strain many available compounds

A common approach towards activating cryptic gene clusters is the 'one strain many compounds' (OSMAC) approach[117]. By culturing the bacteria in a variety of conditions different stresses and selection pressures will be exerted on the bacteria. This approach has been shown to activate different cryptic gene clusters[118,119]. OSMAC strategies include: co-cultivating with other strains, adding elicitors, changing the media composition, or altering the culture conditions[120]. Carbon catabolite repression is known to have a large effect on the control of transcription[121], so altering carbon source can trigger the production of specialised metabolites. For example, glycerol suppresses the production of carbapenems in *Erwinia carotovora*[121] so altering the carbon source can improve their production. Other key nutrients, such as nitrogen, sulphur and phosphorus can also be altered to affect the specialised metabolite profile of a strain[122]. Co-culturing with other strains can also trigger the production of specialised metabolites due to the antagonism between strains[122]. Co-culturing of bacteria can also lead to horizontal gene transfer between strains leading to the production of novel natural products. One example of this is the horizontal gene transfer from *Streptomyces padanus* to *Rhodococcus fascians* when they were co-cultured[123]. This gene transfer resulted in the production of Rhodostreptomycin A and B which are streptomycin-like compounds with broad spectrum antimicrobial activity[123].

3.1.2 Applying the OSMAC approach to *Scytonema*

AntiSMASH analysis revealed the prolific potential of *Scytonema* to produce numerous specialised metabolites (section 2.7.2). By undertaking an OSMAC based screen using our *Scytonema* strains, it may be possible to turn on some of these gene clusters that are not active under standard lab conditions. As a result, new specialised metabolites could be discovered. Also, due to the fact that these strains are understudied and some have zero published papers regarding them, it may be possible to discover novel natural products that are produced even under lab conditions. *Scytonema* strains were grown in BG-11 medium, where variations could be made to this medium to limit key nutrients. Experiments in this chapter focused on variants of BG-11 medium rather than trialling alternate mediums, which is due to the self-sufficient nature of cyanobacteria. *Scytonema* are not reliant on external carbon and nitrogen sources able to fix their own carbon and nitrogen. As our *Scytonema* cultures are xenic, the limitation of carbon and nitrogen controls contaminant growth resulting in the cyanobacteria not being outcompeted. As a result, most cyanobacterial media are very similar concoctions of trace metals alongside some key nutrients such as phosphate and iron (Section 7.1). As most cyanobacterial media are so similar it was thought that growing the *Scytonema* in each would have little effect on the metabolic profile of the *Scytonema.* It was decided to focus on BG-11 medium and then alter key nutrients. BG-11 contains no carbon source, thus this could not be altered. A carbon source could be added, however due to the *Scytonema* strains being mixed cultures this could trigger symbionts / contaminants to outcompete the *Scytonema.* This increased growth of co-occurring bacteria could also be a positive, as co-culturing is an effective OSMAC approach[124,125]. Other key growth factors could also be altered, including growth vessel, light intensity and the length of the day – night light cycle.

3.1.3 Aims of this chapter

The goal of this work was to grow the *Scytonema* strains under a variety of conditions to trigger the production of novel metabolites. Strains would be grown under a variety of conditions, then the cultures would then be extracted and analysed. The extracts would be analysed by LC-MS to look for differences in the metabolite profiles. Extracts would also be screened for antibacterial and antifungal activity against a selection of clinically relevant indicator strains. If bioactive metabolites were found by these bioassays, dereplication can be carried out using the LC-MS data. If the bioactivity cannot be explained by already known metabolites, then the compound would be isolated and characterised.

3.1.4 Objectives

For this chapter there were three main objectives:

1. Carry out screening of *Scytonema* strains using varied growth conditions and media compositions.
2. Extract and analyse metabolites through a combination of LC-MS and bioassays.
3. Identify a novel bioactive compound and then purify and characterise the compound.

3.2 First screen of *Scytonema* strains

3.2.1 Parameters of first *Scytonema* screen

Normally, the first nutrient to be altered for fermentation experiments would be the carbon source. However, *Scytonema* are photosynthetic and are grown without a carbon source. Frequently, the next nutrient to be altered is the nitrogen source. Therefore, for

this first screen, *Scytonema* were grown in BG-11 media with three different nitrogen conditions, standard BG-11, which utilises sodium nitrate ($NaNO_3$), no sodium nitrate but an equimolar amount of ammonium chloride (NH_4Cl), and also nitrogen free (no $NaNO_3$). Cyanobacteria can use ammonium chloride as a source of nitrogen but it is known to inhibit their growth compared to sodium nitrate7. *Scytonema* can fix their own nitrogen so we hypothesised that by limiting the available nitrogen they would grow slower as they would have to pay a metabolic cost to fix their own nitrogen, especially as they generate specialised heterocysts for nitrogen fixation[79]. This switch to fixing their own nitrogen will result in a large change in metabolism and physiology, which could cause novel specialised metabolites to be produced. Cultures were grown in 50 mL flasks for 5 weeks, in duplicate.

The following ten strains were included in the first screen: *Scytonema* sp. UTEX 1163, *Scytonema hofmannii* UTEX 1834, *Scytonema hofmannii* UTEX 2349, *Scytonema* sp. UTEX EE33, *Scytonema bohnerii* SAG 255.80, *Scytonema mirabile* SAG 83.79, *Scytonema* sp. SAG 67.81, *Scytonema* sp. NIES 2130, *Scytonema* PCC 7110 and *Scytonema* PCC 7814. The strains were grown in two light conditions: twenty-four hour constant light and on a sixteen : eight day night cycle. There is a large metabolic shift when cyanobacteria are photosynthesising compared to when respiration is happening at night[127]. The cultures were sampled at two time points: after three and five weeks, which was done to see if the metabolite profile changes over time, 5 mL of culture was taken for each extraction. In addition, three variants of BG-11 were tested, each with a different nitrogen source as stated above. Therefore, each strain, with each nitrogen source was grown under the two light cycles, 24 hour constant light and a 16:8 day night cycle. All of these conditions were tested in duplicate for a total of 120 samples. As each sample was extracted from two time points, the total number of extracts was 240. For the extraction the 5 mL sample was mixed with an equal volume of ethyl acetate, shaken, and the ethyl acetate fraction was taken and dried down. The resulting powder was then resuspended in 1 mL of 50% acetonitrile.

3.2.2 Results of first *Scytonema* screening experiment

All extracts were screened against the following panel of indicator strains: B17.06226 *Enterococcus faecium* (VRE), *Staphylococcus aureus* (MRSA), *Pseudomonas aeruginosa* PA01, *E. coli* NR698 and *Candida albicans.* Only one strain, *Scytonema* sp. UTEX 1163 showed antimicrobial activity, and this activity was seen against all indicator strains and across a range of conditions. An antimicrobial terpenoid, scytoscalarol[97], is known to be produced by this strain and this is covered in more detail in chapter 4. Figure 3.1 shows an example of a spot-on lawn assay of UTEX 1163 extracts against MRSA. Activity is seen across all nitrogen conditions with similar levels of inhibition, which shows that the nitrogen source had little effect on the production of this antimicrobial. Also, there is poor reproducibility. Only one set of nitrogen source duplicates, out of the three, had the same result. It is unknown whether this is because of an issue with the extraction procedure or if the molecule was only produced in one of the duplicates. Due to the lack of results from the first screen, changes needed to be made to improve the results from a second screen. Of course, it was possible that the *Scytonema* strains simply do not produce any antimicrobial NPs. However, due to the number of BGCs present in the *Scytonema* strains and the complex communities in which they dwell, this was thought to be unlikely. Due to the lack of results from the screen it was decided not to do LC-MS analysis of the samples (apart from UTEX 1163, covered in chapter 4). This was decided as we believed lots of improvements could be made to the screening protocol based on things learnt during this first screen and due to the lack of observed bioactivity and inconsistent results between duplicates. As a result, it would be a better use of time and money to push ahead with a second screen and then do LC-MS analysis on those results.

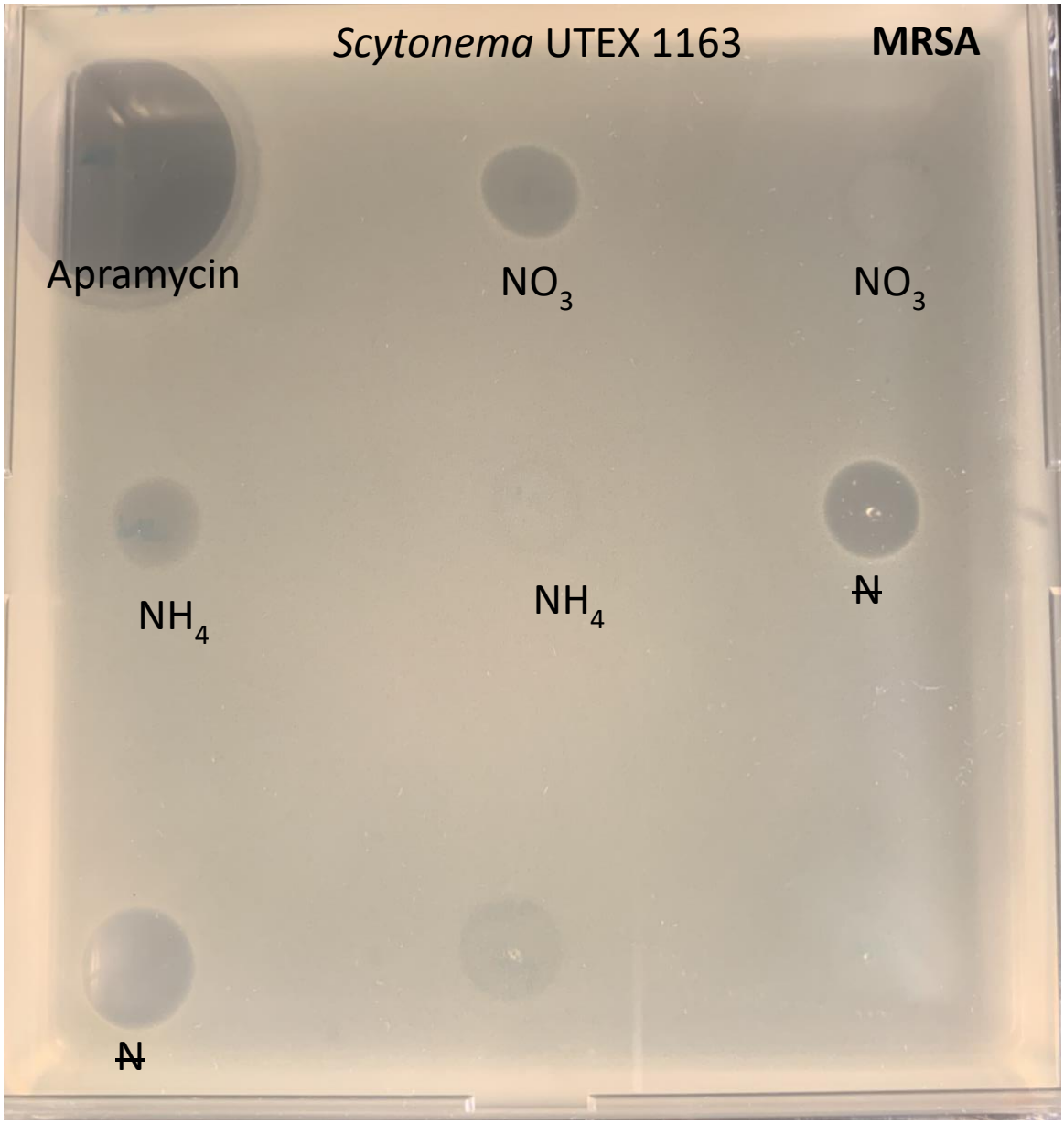

Figure 3.1: Spot-on-lawn bio-assay against methicillin resistant *Staphylococcus aureus* (MRSA). All extracts were from cultures of *Scytonema* sp. UTEX 1163, grown on a 16:8 day night cycle and the nitrogen source for each spot is marked on the plate. A control of apramycin, 50 µL/ml is shown in the top left corner. Zones of inhibition indicate that molecules with antimicrobial activity are found within that extract.

3.2.3 Effects of nitrogen source on the growth of *Scytonema*

Whilst changing the nitrogen source had no differential effect on the production of bioactive metabolites, it did have a large effect on the growth rate of the *Scytonema* strains. Prior to the screen it was hypothesised that by growing the strains with a less bio-available nitrogen source (ammonium chloride) or even no nitrogen source would result in poorer growth. While *Scytonema* can fix nitrogen via heterocysts, this is a metabolically costly process[128] and as a result would negatively impact growth. However, the inverse was seen (Figure 3.2). Cultures grown with ammonium chloride grew the poorest, which was expected as it is known that ammonium chloride negatively impacts cyanobacterial growth[126] . Surprisingly, cultures grown without nitrogen grew very well and many strains also produced a red pigment when grown without nitrogen. The nature of this pigment is unknown. However, leghaemoglobin is a red-pigmented heme containing protein that is found in Rhizobia[129] and found to be crucial in nitrogen fixation in legume root nodules[129]. Whilst rhizobia are a common contaminant in our *Scytonema* strains (Figure 1.3) the red pigment is seen in strains for which no rhizobia are found. Some cyanobacteria strains produce leghaemoglobin[130], whilst others, including *Nostoc commune* UTEX 584, produce an alternate heme containing protein, cyanoglobin[131]. It is unknown whether *Scytonema* produce leghaemoglobin or cyanoglobin but *Scytonema* are closely related to *Nostoc commune.* As a result, this red pigment could be due to the presence of the nitrogen fixing heme containing protein cyanoglobin. Though this would need to be confirmed.

Sodium Nitrate Ammonium Chloride No Added Nitrogen

Figure 3.2: Three cultures of *Scytonema* sp. UTEX 1163, each grown with a different nitrogen source. The leftmost culture is standard BG-11 which includes sodium nitrate, the culture is well grown and the media is notably a light red colour. In the middle is BG-11 with ammonium chloride instead of sodium nitrate, there is clearly less biomass and no red colour in the media. Finally, the rightmost culture is grown without a nitrogen source. The cultures is clearly well grown and the media is a deep blood red colour.

3.3 Second screen of *Scytonema* strains

3.3.1 Changes to screening methodology based on results of the first screen

One possible issue with the first screen was that bioactive specialised metabolites could be produced and extracted, but just in too low of a concentration to be detected. In that screen, extracts of 5 mL were taken after three and five weeks. Samples were taken at two time points to see how the metabolite profile changes over time. However, this is not possible if metabolites are too low of a concentration to be detected via bioassays. It could be possible to detect metabolite production by LC-MS but the large number of samples would render this prohibitively expensive. For future screens, it was decided to extract the whole 50 mL culture after 4 weeks This would improve the quantity of extracted metabolites ten-fold. The volume of 50 % acetonitrile that the final extract was resuspended in was also reduced ten-fold, from 1 mL to 100 µL. These two changes would increase the concentration of extracted metabolites 100 times, maximising the chance of detecting a bioactive metabolite if one is produced. As a result of these changes the experimental set-up for the second screen was as follows: cultures were grown in 50 mL flasks for four weeks. The 50 mL cultures were then frozen and lyophilised, The resulting powder was then resuspended in 100 µL of acetonitrile.

It was also decided to do all further screens under one day – night cycle, the standard 16:8 cycle. This standardisation of the light periodicity was done so that more media compositions could be screened as the total number of samples being screened was fixed. Given the poor results from the ammonium chloride growth, strains would be grown with nitrogen (sodium nitrate) and without added nitrogen. The initial screen

showed that the strains grew well under both conditions and removing the nitrogen had an effect on metabolism, as seen by the red pigment produced.

3.3.2 Parameters of second *Scytonema* screen

Once again strains would be grown in variations of BG-11 media. In order to impart stress on the strains, two of the BG-11 variants would limit key nutrients, phosphate and iron. For low phosphate the concentration of added K_2HPO_4 was lowered from 0.22 mM to 0.022 mM. For low iron the final concentration of ferric ammonium citrate was lowered from 20 µM to 2 µM. The third BG-11 variant was to add 30 mM acetate to the cultures. The aim of this was to increase the growth rate of co-occurring bacteria found in the *Scytonema* cultures. Co-culturing strains is a powerful OSMAC method to obtain new compounds[122]. It is possible that *Scytonema* possess so many BGCs because of the complex communities that they exist in, and they produce compounds to ensure that they are not outcompeted as non-motile organisms. Standard BG-11 contains no carbon source as the *Scytonema* are able to produce all the carbon they need through photosynthesis. This lack of carbon also limits the growth of contaminants, which are not photosynthetic and cannot produce their own carbon. Acetate was chosen as it is less bio-available than other carbon sources[132] and it is important that the *Scytonema* does not get fully out-competed. Also, since each *Scytonema* strain has a unique community of co-occurring strains the effect of the acetate could be different between strains. The final BG-11 variant included the antibiotic spectinomycin at a concentration of 50 µg/mL. This concentration was intended to be a sub-inhibitory level to act as an elicitor via the triggering of stress responses. The addition of sub-inhibitory levels of antibiotics has been shown to trigger natural product production in Actinobacteria[133,134]. The ten strains screened in the first experiment were all used in this second screen. They were all grown

in the four conditions stipulated above both with added nitrogen ($NaNO_3$) and without nitrogen, resulting in eight separate BG-11 variations. These were all, once again, grown in duplicate for a total of 160 extracts.

3.3.3 Results of second *Scytonema* screen – impact of conditions on growth

Unfortunately, it became apparent very quickly that the concentration of spectinomycin was too high, which resulted in the quick death of the *Scytonema.* Similarly, cultures which contained both acetate and sodium nitrate had very fast growth of co-occurring bacteria. When grown in standard BG-11, growth of co-occurring bacteria is not visible, however in the cultures which contained both acetate and sodium nitrate, the media became cloudy within 24 hours. This fast and visible growth of co-occurring bacteria out competed the cyanobacteria, which was likely due to the acetate concentration being too high. This contaminant growth shows how the lack of carbon and nitrogen in the standard culturing of *Scytonema* in the lab keeps contaminants at bay, as almost all are not photosynthetic or able to fix nitrogen. No other cyanobacteria are found as contaminants so none of the contaminants are photosynthetic. A common contaminant are Rhizobiales (Figure 1.3) which are able to fix nitrogen[108]. None of the other contaminants are known to fix nitrogen. Interestingly, cultures which contained acetate but with no added nitrogen source saw less growth of co-occurring bacteria, which meant that the *Scytonema* was not outcompeted and allowed it to grow alongside the co-occurring bacteria. Cyanobacterial growth was lower in this condition when compared to standard BG-11, as fewer clumps were seen.. This slowing in growth is likely due to the increased competition for resources, also the co-occurring bacteria are likely using the *Scytonema* as a nitrogen source[82]. *Scytonema* grown in both the low phosphate, and low iron conditions grew very well, as if they were grown in standard BG-11. This lack of change in growth could suggest that the amount of phosphate and iron in standard BG-11 is in such

a large excess, that limiting it to a tenth still does not limit the nutrients enough to impart stress.

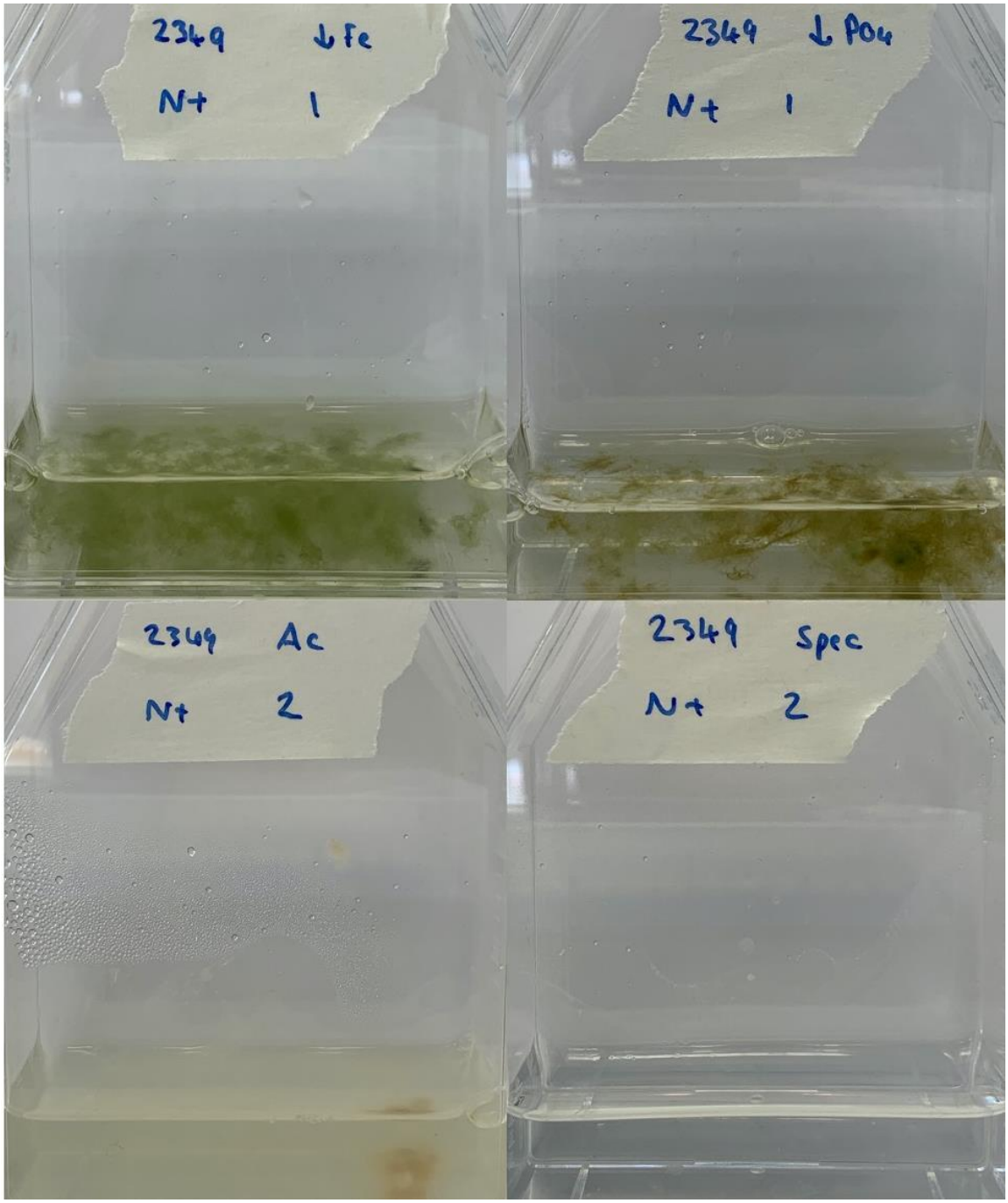

Figure 3.3: Four cultures of Scytonema sp. UTEX 2349, all grown with nitrogen added in the four variations trialled in the second screen. Top left: 1/10th iron culture, the culture has grown well with vibrant green cells. Top right: 1/10th phosphate culture, the culture has grown well but the cells are brown in colour. Bottom left: culture supplemented with 30 mM sodium acetate, the media is cloudy showing strong growth of co-occurring bacteria. The Scytonema has clearly been outcompeted as there is minimal growth. Bottom right: culture supplemented with 50 µM spectinomycin, the amount of antibiotic was clearly too much as no cyanobacterial growth is seen.

3.3.4 Bio-assay results from second *Scytonema* screen

As with the first screen, the whole culture was extracted with an equal volume of 50% acetonitrile, concentrated and final samples were prepared in 100 µL of 50% acetonitrile. This final solvent allowed the final samples to be directly used for both spot-on-lawn bioassays and LC-MS analysis. These final samples were 100 times as concentrated as the first screening experiment. Spot on lawn assays were carried out against the following indicator strains: B17.06226 *Enterococcus faecium* (VRE), *Staphylococcus aureus* (MRSA), *Pseudomonas aeruginosa* PA01, *E. coli* NR698 and *Candida albicans.*

Once again, *Scytonema* sp. UTEX 1163 showed strong activity against all indicator strains tested. This activity is attributed to the production of scytoscalarol and is covered in more detail in chapter 4. Excitingly, two more strains, *Scytonema* sp. NIES 2130 and *Scytonema* sp. UTEX 1834 showed bioactivity against *Pseudomonas aeruginosa* PA01. For both strains, activity was seen across multiple conditions. The zones of inhibitions, shown in Figure 3.4, were faint, especially compared to the zones of inhibition produced by UTEX 1163. Nonetheless, it was decided to pursue this bioactivity.

1. Acetate + NO_3
2. Acetate + NO_3
3. Low PO_4
4. Low PO_4
5. Acetate
6. Acetate
7. Low PO_4
8. Low PO_4
9. Low Iron
10. Low Iron
11. Acetate + NO_3
12. Acetate + NO_3
13. Acetate
14. Acetate

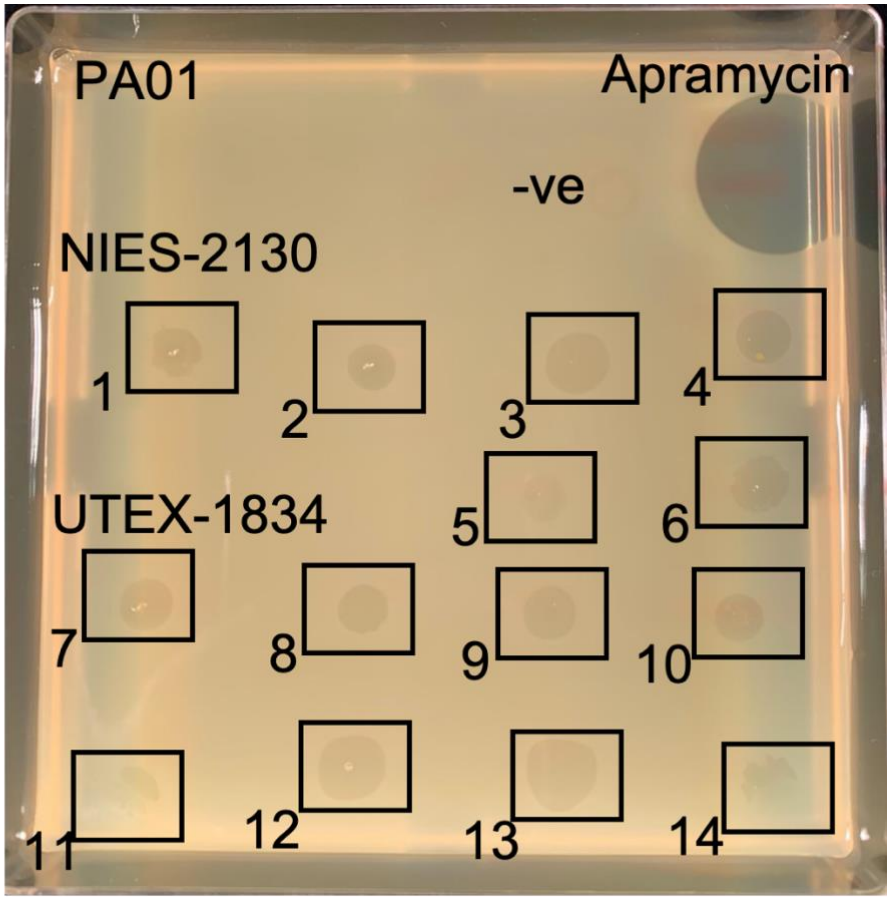

Figure 3.4: A spot-on-lawn assay against *Pseudomonas aeruginosa* PA01 using extracts from the second screening experiment. Activity is seen from extracts of two strains, *Scytonema* sp. NIES-2130 and *Scytonema* sp. UTEX 1834. Activity is seen across a variety of conditions, shown on the left. At the top apramycin is used as a positive control and a negative control of 50% acetonitrile. Spots 1-6 were extracts from *Scytonema* sp. NIES-2130 and spots 7 – 14 were from extracts of *Scytonema* sp. UTEX 1834.

3.3.5 Pursuing activity against *Pseudomonas aeruginosa* PAO1

In order to pursue this bioactivity further growth of both strains, UTEX 1834 and NIES 2130, was scaled up to 1 L. Low phosphate BG-11 was used as the medium, as both strains produced activity against PA01 in this condition. All growth conditions were the same as the second screening experiment except instead of two 50 mL flasks of each being grown, twenty were used. Once grown for four weeks the cultures were harvested using the same protocol as the second screen. The final extract was dissolved in 1 mL of 50 % acetonitrile, which should be 2 times as concentrated as the extract from the second screen. Unfortunately, when the spot-on-lawn assays against *Pseudomonas aeruginosa* PA01 were repeated, no activity was seen. The experiment was repeated again and once again it was not possible to reproduce the PAO1 bioactivity. Due to the lack of reproducibility, it was decided to continue on to a third screening attempt.

3.4 Third screen of *Scytonema* strains

3.4.1 Parameters of third screening attempt

The third screening attempt was assisted by Kenny Budiman, a UEA placement student under my supervision. The parameters used in the second screening attempt had produced some visual differentiation in the growth of the *Scytonema* strains (Figure 3.3). However, some of the conditions clearly did not work, or could be improved upon. Firstly, it was decided not to add spectinomycin to any cultures, instead standard BG-11 was used, to act as a control. Whilst cultures with acetate and no nitrogen did yield an interesting growth phenotype, the *Scytonema* in cultures with both acetate and nitrogen were quickly outcompeted. To try and improve upon this, the acetate concentration was lowered from 30 mM to 10 mM. The goal here was to balance the growth of *Scytonema* and co-occurring bacteria in both nitrogen containing and nitrogen free cultures. Finally, whilst there was a visual difference between low phosphate and low iron cultures, the *Scytonema* grew extremely well in all instances, which suggests that the nutrients were not limiting. To improve upon this the phosphate and iron levels were lowered to one-

hundredth of the level present in standard BG-11. This meant the ferric ammonium citrate concentration was lowered to 0.2 µM and the K_2HPO_4 concentration was lowered to 0.0022 µM. Once again, these four conditions were assessed with and without adding sodium nitrate, which provided a total of eight fermentation conditions. As in the second screen, these conditions were all conducted in duplicate for 160 total cultures. Extractions were carried out in the same manner as the second screen, the resulting extract were also analysed by LC-MS.

3.4.2 Results of third *Scytonema* screen – impact of conditions on growth

The tweaks to the growth conditions were a success as each growth condition produced a different phenotype with clear cyanobacterial growth. The reduction to 10 mM sodium acetate was effective as the *Scytonema* did not get out-competed in cultures that contained both acetate and sodium nitrate. The reduction to 1 / 100^{th} of phosphate and iron seemed to be effective in imparting stress on the *Scytonema* through nutrient starvation. This nutrient starvation is seen by both of these conditions having a different phenotype to *Scytonema* grown in standard BG-11, this can be seen (Figure 3.5).

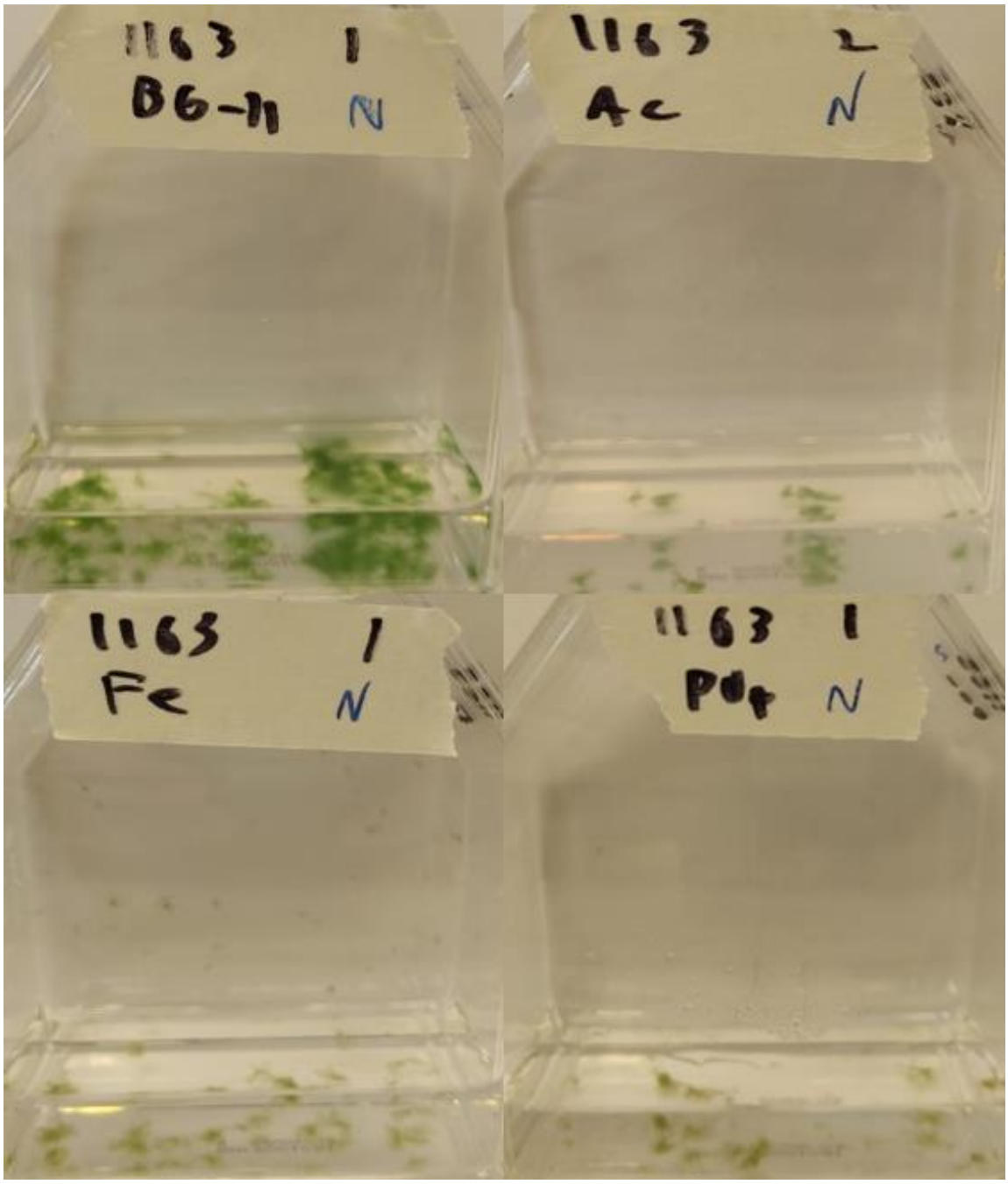

Figure 3.5: Four cultures of *Scytonema* sp. UTEX 1163, all grown with nitrogen added in the four variations trialled in the third screen. Top left: standard BG-11 culture, the culture has grown well with vibrant green cells. Top right: 10 mM acetate culture, the culture has grown and has not been outcompeted, as in the second screen. The amount of growth is clearly a lot less than the standard BG-11 however the media is still translucent indicating the contaminants have not grown strongly.. Bottom left: culture with $1/100^{th}$ iron,, the culture has grown less than the standard BG-11 though more than the acetate supplemented culture. The cells are brown in colour indicating nutrient depletion.. Bottom right: $1/100^{th}$ phosphate culture, similar to the $1/100^{th}$ iron culture in that the cells have grown, though less well than standard BG-11 and the cells are brown in colour.

3.4.3 Results of bio-assays from third *Scytonema* screen

Extracts were prepared in the same way as the previous screening attempt and screened against the same panel of indicator strains, as well as *Bacillus subtilis*. These showed that *Scytonema* sp. UTEX 1834 had activity against *E. coli* NR698[135] (which has a permeabilised outer membrane) and *Bacillus subtilis,* bioassay results against *E. coli* NR698 can be seen in Figure 3.7. Interestingly, these UTEX 1834 extracts did not show any activity against *Pseudomonas aeruginosa* PA01, which was seen in the previous screen. Whilst the previous screen did not show that UTEX 1834 had any activity against *E. coli* NR698. This difference in activity indicates a shift in metabolite profile and that the differing activities seen between screens is due to different molecules being produced.

3.5 Investigating the antimicrobial activity from *Scytonema* sp. UTEX 1834

3.5.1 Confirming the bioactivity from *Scytonema* sp. UTEX 1834 was reproducible

To confirm that this activity was reproducible, twenty 50 mL cultures of *Scytonema* sp. UTEX 1834 were grown in BG-11 containing 1/100th iron. Initially one culture was extracted in order to confirm bioactivity was still present. These cultures were extracted by freezing and lyophilising the cultures after four weeks. The resulting powder was then resuspended in 100 µL of acetonitrile and all samples were then pooled. This extract showed activity against both *E.coli* NR698 and *Bacillus subtilis,* the same as the extracts from the third screen. The remaining cultures were harvested and antimicrobial activity was confirmed by bioassay. At this point it became clear that more material would be required for purification and characterisation of the molecule and a scale up would be needed.

3.5.2 Scaling up production of *Scytonema* sp. UTEX 1834

It was decided to initially scale up production to 10 L of culture. The main limiting factor was the size of the plant-growth cabinet used to grow the *Scytonema* strains. The second limiting factor was the logistics of extracting from high volumes of culture. Previously, cultures were frozen, then freeze dried to dry down the cultures. This procedure was not logistically possible with 10 L of culture as the freeze drying alone would take approximately 50 days, based on the time taken for previous extractions. As a result, new extraction methods would need to be trialled and validated. Firstly, the 10 L of *Scytonema* sp. UTEX 1834 was grown in 1/100th iron BG-11 without nitrogen. In the third screen multiple media compositions saw bioactivity from UTEX 1834, however low iron without nitrogen was the most reproducible. These cultures were grown in 175 mL tissue culture flasks containing either 50 mL or 100 mL of the media. 100 mL cultures were trialled to try and maximise space in the plant growth cabinet. Previous attempts to scale up culture volume had resulted in a change in phenotype so it was unknown whether UTEX 1834 would still produce the bioactive molecule in 100 mL cultures. These cultures were initially grown for three weeks as this was when growth visually plateaued. At this point a 50 mL and a 100 mL culture were freeze dried and extracted to confirm the antimicrobial activity was still seen – which it was. This activity gave us the green light to trial alternate extraction procedures. The first extraction procedure trialled was a liquid : liquid extraction with ethyl acetate (EtOAc). Initially the culture was extracted from three times with equal volumes of EtOAc. The EtOAc was then dried down and the resulting extract was resuspended in 50% acetonitrile. A spot-on-lawn assay confirmed that all three sequential extracts had activity against *B. subtilis* and *E. coli* NR698. The remainder of the 10 L of culture was then extracted in the same way.

3.5.3 Purification and mass determination of antimicrobial molecule from *Scytonema* sp. UTEX 1834

In order to determine which molecule produced by UTEX 1834 was bioactive and to work towards purifying it, fractionation was carried out by semi-preparative HPLC, taking thirty second fractions. The extract from the 10 L of culture was used for this, keeping a small amount behind as a control. After fractionation, a spot-on-lawn assay was carried out against *Bacillus subtilis* to determine which fractions had bioactivity. The results are shown in Figure 3.6 and show that only one fraction, fraction 7, had bioactivity.

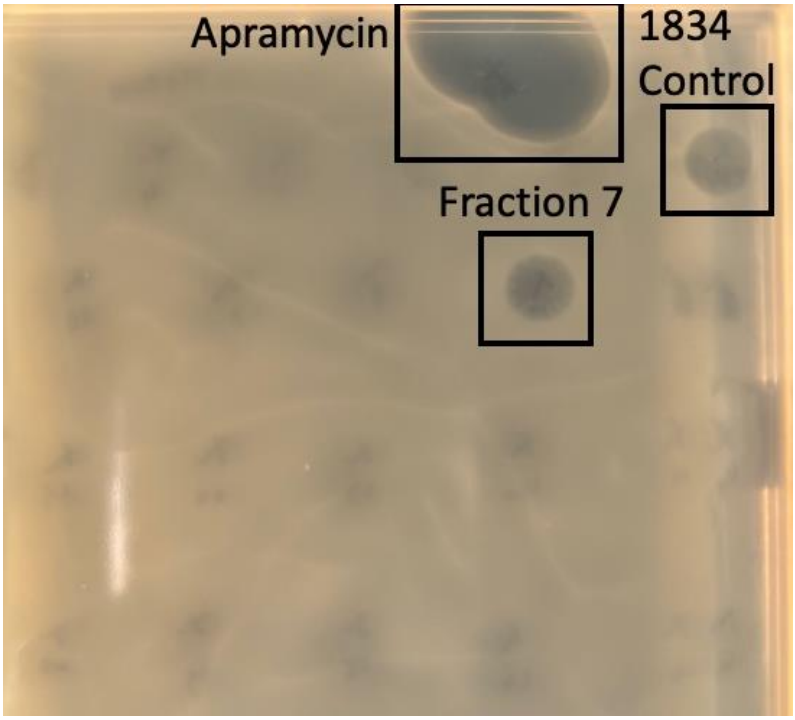

Figure 3.6: A spot-on-lawn assay against *Bacillus subtilis* using fractions of the extract from *Scytonema* sp. UTEX 1834. A positive control of apramycin was used, a control of UTEX 1834 extract before fractionation was also used to show the activity. Only one fraction, fraction 7, had bioactivity. All other fractions are spotted on the plate however no zones of inhibition can be seen.

After confirming bioactivity was only seen in fraction 7, all fractions were analysed by LC-MS to look for differential peaks. As bioactivity was only seen in fraction seven, any peak seen in the LC-MS chromatogram for fraction seven that was absent from the others could be the bioactive molecule of interest. When comparing the LC-MS chromatograms of the fractions (Figure 3.7) a peak is clearly seen in fraction seven as well as the crude UTEX 1834 extract, but is absent in all other fractions. The molecule responsible for this peak had a protonated mass of 1597.8352 Da. This was determined to be a protonated mass as a doubly protonated mass was also seen. It was not possible to assign a predicted formula as multiple plausible formulae were predicted. To ensure that this molecule was novel and had not been previously reported, dereplication was carried out. Firstly, the following databases were consulted for molecules with the same mass: NP Atlas[24], Dictionary of Natural Products, and ChemSpider. No molecules were found with the same, or a related mass. A literature search was also carried out to look for reported NPs with similar masses that had been isolated from *Scytonema* or *Nostocales* strains, but no such molecule was found. It is important to note that there are two reported specialised metabolite from *Scytonema* sp. UTEX 1834, scytonemide A and B[99]. These molecules have no reported antimicrobial activity and the masses of these molecules are not found in our extracts.

The putative bioactive molecule was then purified by two further rounds of semi-preparative HPLC, the final round taking six-second slices. Each fraction was analysed by LC-MS and pure fractions which contained only the molecule of interest were combined and dried completely. Before being dried, a sample was taken and a spot-on-lawn bioassay was carried out. Activity was still seen, this assay combined with the LC-MS chromatogram of the sample showing that it only contained the molecule with a mass of 1597.8352 Da, which strongly suggests that this is the bioactive molecule. The total dry mass obtained was approximately 0.2 mg. Due to the very low amount obtained, a further 25 L of UTEX 1834 was grown, over two cycles. These cultures were grown in

flasks containing 50 mL of media for four weeks. Extraction was carried out as previously stated for the third screen.

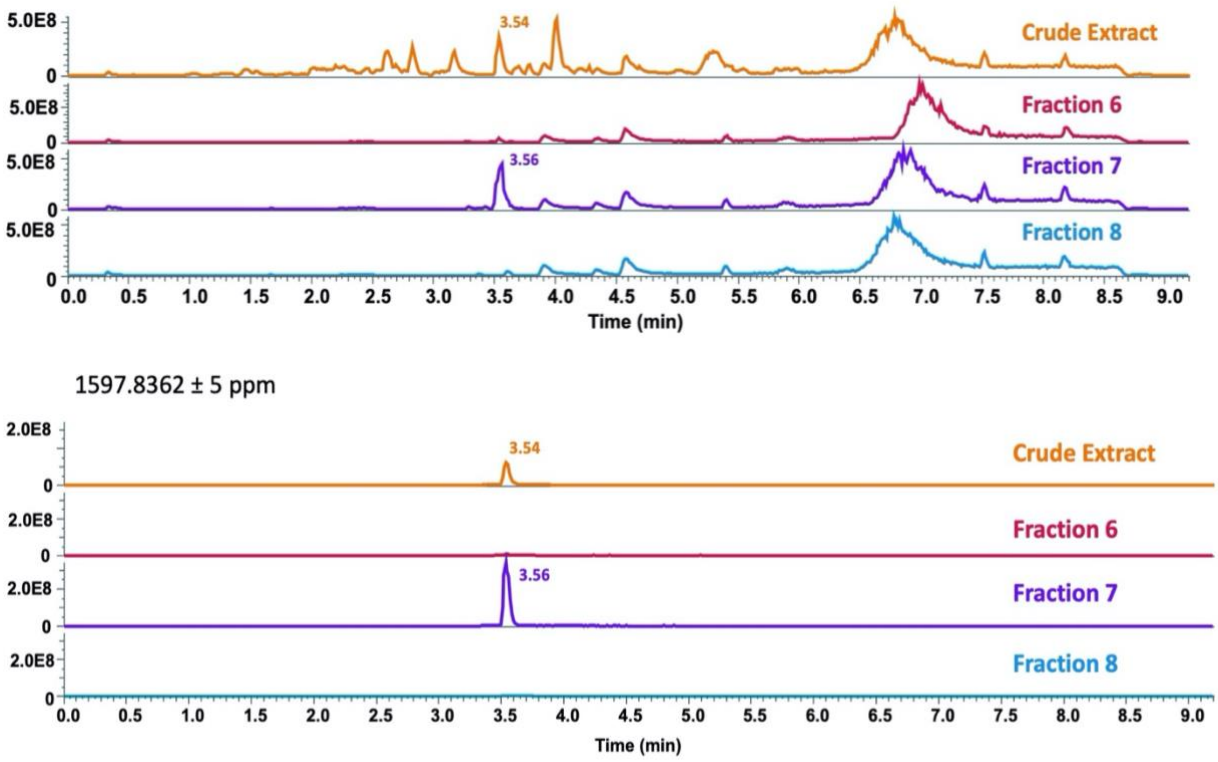

Figure 3.7: Top: An LC-MS base peak chromatograms of fractions of *Scytonema* sp. UTEX 1834 extracts alongside a sample of the crude extract before fractionation. A differential peak is seen in fraction 7 at 3.56 minutes, fraction seven was the only fraction to have antimicrobial activity. Bottom: An LC-MS extracted ion chromatogram of the extracts selecting for the mass of the compound seen at 3.56 minutes in fraction seven. This mass is not seen in any other fraction.

3.5.4 Further optimisation of extraction procedure of antimicrobial molecule from *Scytonema* sp. UTEX 1834

The total amount of compound extracted, 0.2 mg from 10 L of culture, is low and indicates that the extraction procedure is not fully optimised or that productivity is low (as described in section 7.5.2). Now that there was a mass to follow, extracts could be analysed by LC-MS to see the amount of compound being extracted with each method. Multiple extraction methods were tested, each using one 50 mL culture of UTEX 1834 that had been grown for three weeks. Firstly, the method used to extract from the 10 L of culture was trialled. A 1 mL pre-extraction sample was taken from the culture and was extracted three times with an equal volume of EtOAc. After extraction, a 1 mL sample of the remaining culture was taken. These pre and post extraction samples were taken to see what proportion of the total amount of compound was extracted in these three extractions. These extracts were processed as before and the final samples analysed by LC-MS.

The extracted ion chromatograms (EICs) of the pre and post extraction samples, looking for the mass of the bioactive compound (*m/z* 1597.8352) are shown in Figure 3.8. Three notable things were apparent from the chromatograms. Firstly, there appeared to be very little to none of the compound in the pre-extraction sample. Secondly, each extract contained the same amount of the compound, this indicates that not all of the compound is being extracted as a decrease in the amount of compound extraction with each subsequent fraction is expected. Lastly, there is still a large amount of the compound present in the post-extraction sample, thus this extraction method is not effective. These results were unexpected, especially the lack of compound in the pre extract sample, so the extractions were repeated the same as before to confirm the results. All of the results were exactly the same (data not shown). As the pre extract sample was just taken from the supernatant, it was hypothesised that the compound could be attached to the cyanobacterial cells, or not efficiently exported.

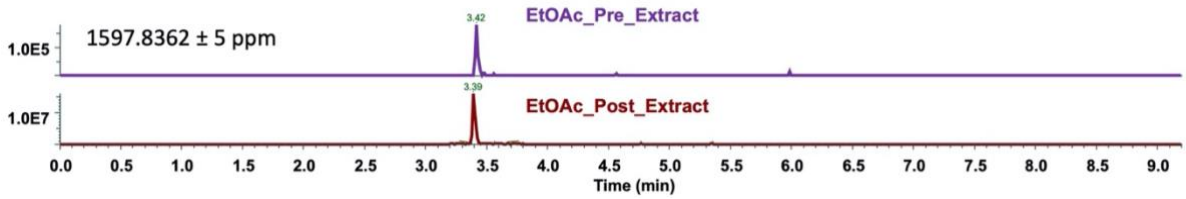

Figure 3.8: Extracted ion chromatograms, selecting for the mass of the bioactive compound. The top chromatogram is a pre-extraction sample of the culture medium. The bottom chromatogram is a post-extraction sample of the culture medium. 100 times as much of the bioactive compound is found in the post extraction sample compared to the pre-extraction sample. It is thought that the compound is retained within the cells thus the pre-extract sample having little of the compound.

Two further extraction methods were trialled to try and answer the questions arisen from the previous extractions. Firstly, the previous extraction procedure was repeated, but with 11 consecutive extractions with an equal volume of EtOAc. This large number was chosen to make sure all of the compound was extracted, and the total number of consecutive extractions required to extract all of the compound. Once again, a pre and post extract was taken. The 50 mL of culture was extracted with an equal volume of EtoAc, shaken, and the ethyl acetate fraction as then taken and dried down. The resulting powder was then resuspended on 100 µL of acetonitrile. The extraction was then repeated ten further times. The EIC chromatograms are shown in Figure 3.9. Surprisingly, it took nine consecutive extractions for the amount of compound in each extract to start to reduce, and even after eleven extractions there is still compound left in the post-extraction extract. These extractions showed that ethyl acetate is not an effective solvent to extract this molecule. To extract all of the compound from the 25 L of culture, 250 L of ethyl acetate would be required, which is an unworkable amount in a standard laboratory.

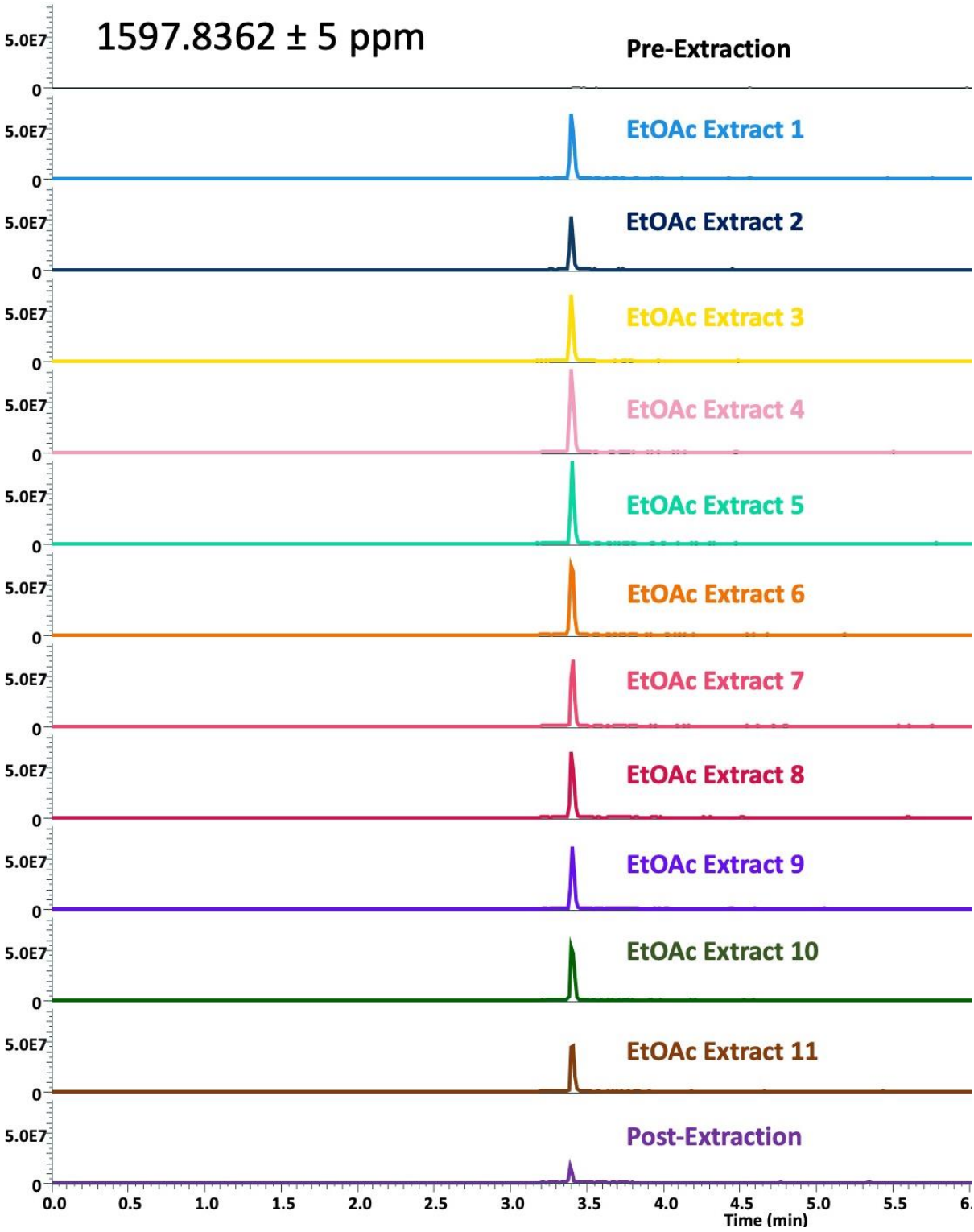

Figure 3.9: Extracted ion chromatograms of ethyl acetate extracts from *Scytonema* sp. UTEX 1834. The culture was extracted with an equal volume of ethyl acetate sequentially 11 times. Total amount of the bioactive compound extracted doesn't decrease until the tenth extraction. Even after eleven sequential extractions, there is still compound left in the post-extraction sample.

After these results, extraction trials were carried out using methanol. Methanol is able to penetrate the cells and extract molecules trapped within. The fact that there was no compound present in the pre-extract supernatant sample indicates that the molecule could be retained by the cells so methanol could be an effective solvent. To test this a 50 mL culture of UTEX 1834 was filtered to remove the cells from the supernatant. The cells were then washed with methanol whilst the supernatant was dried down and resuspended in methanol. These methanolic extracts were then processed and analysed by LC-MS. The EIC chromatograms are shown in Figure 3.10 and show 500 times as much of the compound was in the cell extract compared to the supernatant extract. As only 1 mL of the supernatant was used for the extraction, this result indicates that the cell extract contained fifty times more compound than the supernatant. Also, extracting directly from the cells with methanol extracted approximately ten times more compound than the EtOAc extraction method. This result confirmed the hypothesis that the bioactive compound is either bound to the cells or retained within them. It was decided to extract the remaining 25 L of UTEX 1834 cultures using this method, filtering off the cells then washing them with methanol. The total volume of cells was combined and washed three times sequentially with methanol.

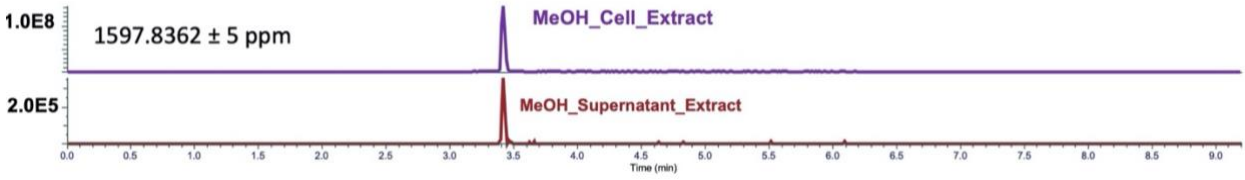

Figure 3.10: Extracted ion chromatograms of methanolic extracts from *Scytonema* sp. UTEX-1834. The top chromatogram shows the amount of compound extracted from the cyanobacterial cells. The bottom chromatogram shows the amount of compound extracted from the supernatant. Approximately 500 times more compound is extracted from the cells. Since only 1 mL of the culture medium was extracted from, overall approximately fifty times more compound is in the cells of a culture compared to the supernatant.

3.6 Attempts to structurally characterise antibiotic compound from *Scytonema* sp. UTEX 1834.

3.6.1 LC-MS analysis of antimicrobial compound

The cells from 25 L of *Scytonema* sp. UTEX 1834 grown in low iron, no nitrogen BG-11 were extracted from using methanol. This extract was purified through two rounds of HPLC, using LC-MS to screen fractions for the compound of interest. Fractions were also screened using bioassays to make sure the activity was not lost. The dried weight of the bioactive compound obtained from 25 L of cultures was 2.5 mg. The goal was to obtain enough compound to characterise the molecule using a combination of LC-MS/MS and nuclear magnetic resonance spectroscopy (NMR).

First, an LC-MS chromatogram was obtained using the pure compound, as well as MS/MS fragmentation data. It was hoped that by analysing the MS/MS data fragments could be identified that could give more information about the overall structure of the molecule. Due to the large size of the molecule the initial hypothesis was that the compound was a peptide or hybrid peptide polyketide, which is because most large, > 1000 MW, natural products are peptidic7. By looking at the fragmentation data it should be possible to see fragments or mass losses that could be attributed to amino acids or short peptides. When analysing the mass losses between fragments, multiple losses of 57.0214 were seen, which can be attributed to the loss of glycine (Figure 3.11). As well as this a loss of 101.0476 was seen which can be attributed to the loss of threonine. As well as these mass losses some of the smaller fragments seen can be attributed to protonated mono or di-peptides. Unfortunately, none of the larger masses could be confidently attributed to a polypeptide chain, which could be due to modifications to the peptides or the inclusion of non-proteogenic amino acids. Despite this finding, multiple masses corresponding to amino acids in the fragmentation data lends credence to our theory that the molecule is peptidic.

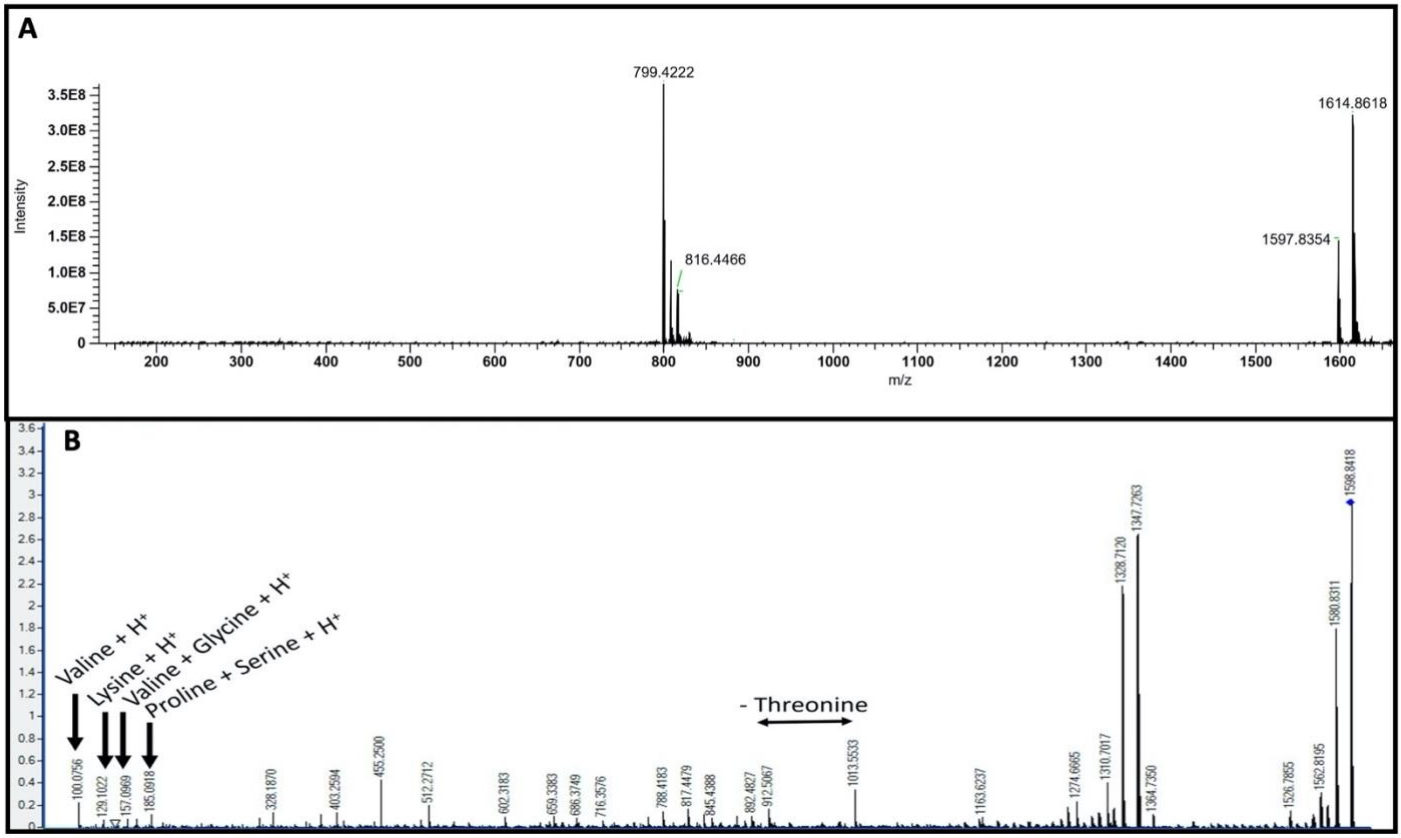

Figure 3.11: A: LC-MS chromatogram of bioactive molecule isolated from *Scytonema* sp. UTEX 1834. The protonated mass of the molecule is seen as 1597.8354. An ammonia adduct can also be seen at 1614.8618. Further to this, an $M+2H^+$ peak is also seen at 799.4222. B: an annotated MS/MS spectra for the bioactive molecule. A loss of the mass of threonine can be seen between peaks at 1013.5533 and 912.5067. Four of the smallest fragments match with the protonated mass of an amino acid or di-peptide. A protonated valine at 100.0756, a protonated lysine at 129.1022, a valine and glycine di-peptide at 157.0969 and a proline and serine di-peptide at 185.0918.

3.6.2 Attempts to use NMR to structurally characterise the antibiotic compound from *Scytonema* sp. UTEX 1834

NMR experiments were carried out by Dr Sergey Nepogodiev, head of the JIC NMR platform. He was provided with purified, dried compound that was weighed at 2.5 mg using a lab scale. An initial ^{1}H proton NMR carried out by Dr Nepogodiev had a low peak intensity, which resulted in Dr Nepogodiev hypothesising that the true amount of compound in our sample was approximately 0.1 mg. This experiment was carried out using deuterated chloroform as the NMR solvent, according to Dr Nepogodiev the compound had poor solubility in this solvent. It is possible that by switching to an alternate solvent, such as deuterated di-methyl sulfoxide (DMSO), that the solubility would increase. As a result, the intensity of the NMR peaks would be increased. Despite this, it was decided that more compound would be required in order to obtain spectra good enough to fully characterise the molecule. Unfortunately, this work was carried out during the final moments of the PhD and time did not allow for further material to be obtained. Further analysis of existing NMR data was delegated to other lab members. The crude NMR spectra that were obtained can be found in the appendix.

3.6.3 Attempts to link antibiotic compound isolated from *Scytonema* sp. UTEX 1834 to its biosynthetic gene cluster

It was hoped that the structural information gleaned from LC-MS, MS/MS and NMR would allow for the identification of the BGC responsible for producing the antimicrobial molecule. The high molecular weight of the molecule indicates that the molecule would be peptidic, which was backed up by the MS/MS chromatogram (Figure 3.11) which shows both the loss of amino acids but also shows fragments with mass equal to protonated amino acids, which would narrow down the BGC to either a RiPP, NRPS or a hybrid NRPS-PKS. AntiSMASH predicts that *Scytonema* sp. UTEX 1834 contains 28 gene cluster regions, including some regions that contain multiple BGCs, such as one region that is predicted to contain 6 discrete BGCs (Figure 2.8). Unfortunately, selecting for

NRPSs, RiPPs and hybrid NRPS-PKS clusters only excludes 7 of the 28 gene cluster regions. The large size of the molecule, 1597.8280 Da, means that 10+ NRPS modules would be required. This requirement excludes a further 5 NRPS containing gene cluster regions with too few NRPS modules. This leaves 16 gene cluster regions. It was predicted that the MS/MS fragmentation data would guide the identification of the gene cluster by identifying key structural motifs. Unfortunately no key motifs were identified, though some amino acids predicted to be in the compound were identified – valine, lysine, glycine, serine and threonine. Not all of the NRPS modules identified by antiSMASH have predictions for the amino acid that they incorporate so it is not possible to rule out any further NRPS or hybrid NRPS-PKS clusters. The same can be said for the RiPP clusters, antiSMASH prediction of RiPP precursor peptides is not good enough to rule out clusters based on knowing these amino acids. As a result, more structural information is required to further eliminate these 16 gene cluster regions.

3.7 Discussion of results from this chapter

The need for novel antimicrobial natural products is well known[10], and cyanobacteria have shown to be a promising source of novel bioactive natural products with examples such as the pitipeptolide family[77] that has activity against *Mycobacterium tuberculosis* and the dolastatin family[67] which has wide ranging cytotoxic activity[137]. *Scytonema* are a promising genus to study due to their large genomes[85], despite this they are understudied with comparatively little published literature. What is known about *Scytonema* proves them to be promising, an example being the discovery of the broad spectrum antimicrobial molecule scytoscalarol[97] from *Scytonema* sp. UTEX 1163. This chapter has shown that further antimicrobial molecules are produced by *Scytonema* and that the OSMAC[122] can successfully trigger their production.

Overall, the results of the screening work were successful, as a potentially novel antimicrobial molecule was detected and isolated. The yield of this molecule was low, requiring further scale up and extractions to isolate enough of the material for structural characterisation. Despite the low yield, bioactivity was easily detectable, showing that the molecule has potent antimicrobial properties. This low yield was attributed to the bio-active compound being retained within the cell, as a result the extraction method needed to be altered. Testing of extractions methods was successful and resulted in an order of magnitude more of the compound being extracted (Figure 3.9), this was achieved by switching the extraction solvent to methanol as it can penetrate and extract molecules from within the cell. Alternatively the cells could have been split opened, for example by sonication. Though this method was not tested this its efficacy cannot be determined.

Apart from this the three screening experiments yielded fewer results than expected. LC-MS analysis for the screens looking for differential expression between culture conditions did not provide any significant results. Only one other bioactive molecule could be reproducibly seen in any of the three screens, scytoscalarol. This is despite ten different strains being tested over three separate screens with a wide range of conditions which resulted in a wide range of phenotypes. A wider selection of indicator strains or screens in general could have been used to test for other activities rather than just anti-bacterial. *Candida albicans* was screened against, but no other fungi. Since *Scytonema* are often found in symbiosis with fungi[63,82], a wider range of fungi could have been screened against. Other cyanobacteria could have been screened against to test for anti-cyanobacterial activity. A screen for cytotoxic activity was carried out in collaboration with Dr Stuart Rushworth from The University of East Anglia, however the results were inconclusive. This experiment could have been repeated to look for toxicity, which is common for cyanobacterial natural products[102,138,139].

3.8 Future work

The further work required to complete this work would be to continue with scaling up growth of *Scytonema* sp. UTEX 1834 in order to purify more of the bioactive compound, which would allow for structural characterisation of the molecule by NMR. Once the structure has been characterised it should be possible to elucidate the BGC responsible for producing it. This would be done by checking RiPP precursor peptides and NRPS module domains to match up to the amino acids that make up the structure of the molecule. Knowing the structural motifs in the molecule and identifying if any post-translational modifications are present would narrow down the RiPP clusters by being able to rule out some RiPP subfamilies. Further bioassays would then be required to find out the full range of activity, and obtain MIC values.

Chapter 4: Exploring the biosynthesis of scytoscalarol

4.1 Background on scytoscalarol and terpene biosynthesis

4.1.1 Previous work on scytoscalarol – discovery and bioactivity

Scytoscalarol is a sesterterpene (C_{25}) first isolated from *Scytonema* sp. UTEX 1163[97]. Scytoscalarol has been structurally characterised and includes a guanidino group. Scytoscalarol was found to have activity against *Bacillus anthracis* (MIC 6 μM), *Staphylococcus aureus* (MIC 2 μM), *Escherichia coli* (MIC 30 μM), *Candida albicans* (MIC 4 μM) and *Mycobacterium tuberculosis* (MIC 110 μM)[97]. Scytoscalarol was found to have cytotoxic activity in a Vero cell assay[140] with an IC_{50} of 135 μM. A Vero cells are a cell line derived from African green monkey kidney epithelial cells[140]. When scytoscalarol was first reported it was the only sesterterpene isolated from cyanobacteria, it was also the first sesterterpene discovered that contains a guanidino group. However, many cyanotoxins contain a guanidino group[141]. Examples include cylindrospermopsin, a cyclic sulphated guanidine alkaloid isolated from *Cylindrospermopsis raciborskii*[142]. Cylindrospermopsin is cytotoxic, hepatotoxic and neurotoxic, it was found to inhibit glutathione and protein synthesis, as well as cytochrome P450s and directly interacts with DNA[139]. The guanidino group was found to be essential for its bioactivity[139]. Cylindrospermopsin is an alkaloid produced by a combined NRPS and PKS BGC which uses guanidinoacetic acid as a starter unit[143]. Two cyanobacterial sesterterpenes which contain a guanidinium group were discovered after scytoscalarol, cybastacine A and B[98]. These molecules were isolated from *Nostoc* sp. BEA-0956 and shown to have antimicrobial activity against *Mycobacterium* spp., *Enterococcus* spp. And *Staphylococcus* spp.[98]. Like scytoscalarol these molecules were first identified by screening for bioactivity, but there is similarly no publicly available genome for the producing strain.

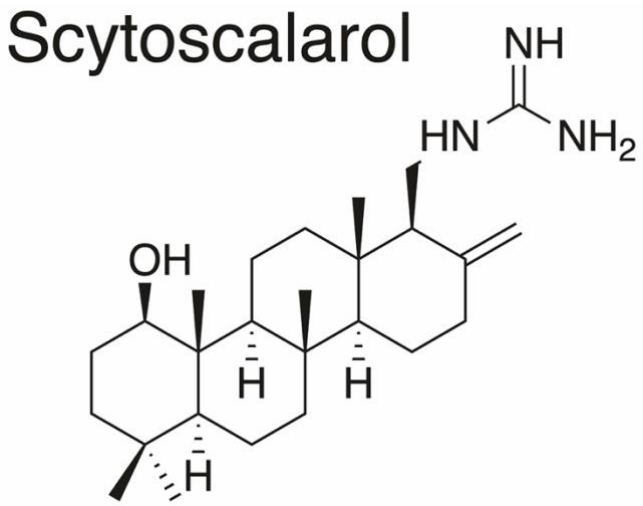

Figure 4.1: The structure of scytoscalarol, a guanidino group containing sesterterpene produced by *Scytonema* sp. UTEX 1163.

4.1.2 Terpene biosynthesis and heterologous expression of sesterterpenes

An overview of terpene biosynthesis can be found in section 1.2.3. A sesterterpene contains a carbon backbone of twenty-five carbon atoms. The final terpenoid molecule is production through a cyclisation of a terpene precursor, catalysed by a terpene synthase. The C_{25} precursor for sesterterpenes is geranylfarnesyl pyrophosphate (GFPP). This GFPP

C_{25} precursor is itself produced through the condensation of five carbon isoprene subunits either in the form of dimethylallyl pyrophosphate (DMAPP) or isopentenyl pyrophosphate (IPP). In order to achieve successful heterologous expression of a sesterterpene, supply of the precursors needs to maximised to ensure enough production of the target molecule so it can be detected. The heterologous host may also lack the machinery to produce the required precursors. Another issue is the heterologous expression of Cytochrome P450s, a common tailoring enzyme for terpenes. P450s are dependent on heme and are difficult to heterologously express for this reason[144].

4.1.3 Hypothesis

The only things known about scytoscalarol previous to this work were its structure and its antimicrobial activity. The BGC responsible for producing scytoscalarol is unknown and the lack of a publicly available genome sequence has prevented genome mining attempts. The mode of action for scytoscalarol's antimicrobial activity is also unknown. By analysing our newly acquired genome of *Scytonema* sp. UTEX 1163 using a combination of antiSMASH and manual Blast search of genes it should be possible to narrow down candidate gene clusters. These possible gene clusters could then be heterologously expressed. The gene cluster could then be heterologously expressed in *E. coli*, especially as plasmids already exist that will boost the precursor supply.

4.1.4 Objectives

For this chapter there were three main objectives:

1. Analyse the *Scytonema* sp. UTEX 1163 genome and identify possible scytoscalarol producing BGC(s).

2. Clone candidate scytoscalarol BGCs and heterologously express them alongside C_{25} terpene precursor.
3. Analyse extracts from heterologous expression and look for evidence of cyclisation of the C_{25} terpene precursor.

4.2 Results

4.2.1 Antimicrobial activity from *Scytonema* sp. UTEX 1163

Results of the first screen (chapter 3.2.2) of acquired *Scytonema* strains showed that *Scytonema* sp. UTEX 1163 grown in no added nitrogen conditions had activity against vancomycin resistant *Enterococcus* (VRE), *Candida albicans,* methicillin resistant *Staphylococcus aureus* (MRSA) and *E. coli.* This matched the known bioactivity profile for the strain suggesting that the bioactive molecule produced was scytoscalarol. This result needed to be confirmed by LC-MS. An extract from *Scytonema* sp. UTEX 1163 grown without nitrogen was analysed by positive mode LC-MS to detect the presence or absence of scytoscalarol.

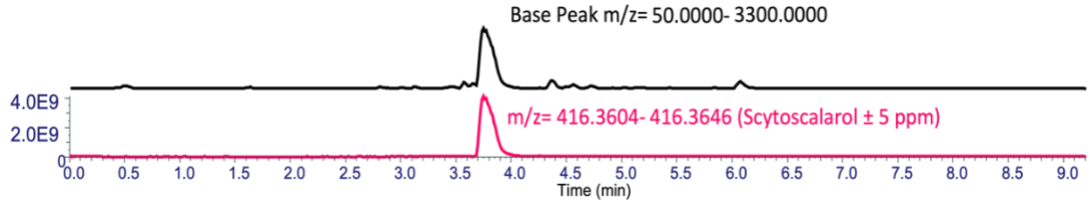

Figure 4.2: LC-MS chromatogram of an extract of *Scytonema* sp. UTEX 1163. The top chromatogram, in black, is a base peak chromatogram showing all ions detected in the sample. The bottom chromatogram, in pink, shows the extracted ion chromatogram of the same sample, selecting for the mass of the protonated scytoscalarol adduct (417.3634). Both traces are the same scale and have equivalent MS/MS fragments.

Figure 4.2 shows the base peak chromatogram and the extracted ion chromatogram corresponding to the mass of scytoscalarol (theoretical $[M+H]^+$ = 417.3634, observed m/z= 417.3625), for an extract from UTEX 1163. These chromatograms show two things, firstly scytoscalarol is clearly found in our extracts from *Scytonema* sp. UTEX 1163. Secondly,

scytoscalarol clearly dominates the rest of the compounds found in the sample. The MS peak area of a molecule is influenced by two main factors, the abundance of the compound and its ability to be ionised. Additionally, the strong bioactivity seen from extracts of UTEX 1163 suggests strong production of scytoscalarol. The large peak area observed in LC-MS is likely due to a combination of these factors, although proper quantification experiments with a pure standard would be required to verify this, but there is currently no commercial standard available for this molecule.

Despite scytoscalarol being a known molecule, there are still interesting questions to be answered. As stated previously the BGC responsible for producing scytoscalarol is unknown, the biosynthetic pathway is unknown and the mechanism of action for its antimicrobial activity is unknown. These questions combined with its rare guanidino-containing terpene structure made it an interesting project. The following work was done in collaboration with Prof. Jeroen Dickschat from The University of Bonn in Germany, an expert in terpene biosynthesis.

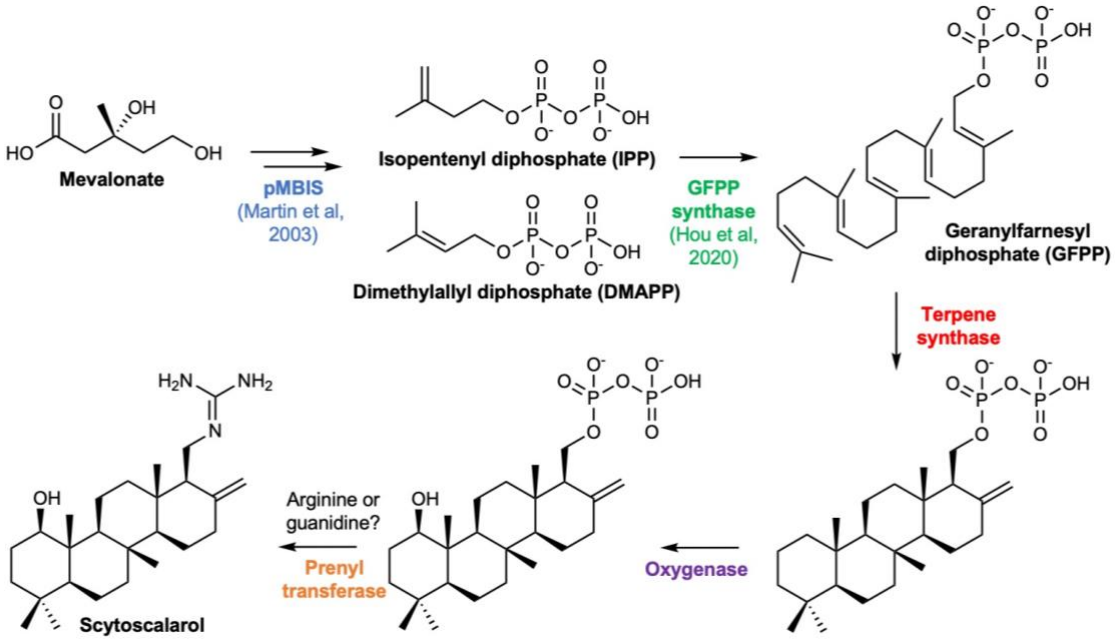

Figure 4.3: Proposed biosynthetic pathway for the biosynthesis of scytoscalarol. First the C^5 subunits are produced, catalysed by genes encoded on the pMBIS plasmid. Secondly these C^5 IPP and DMAPP subunits are then turned into C^{25} GFPP catalysed by genes encoded on the pYE vector. The GFPP is then folded and cyclised by a terpene synthase, an oxygenase, potentially a cytochrome P450 then installs the hydroxyl group. Finally a prenyl transferase transfers an arginine or guanidine followed by cleavage to produce the final molecule.

4.2.2 Analysing the genome of *Scytonema* sp. UTEX 1163

Analysis of *Scytonema* sp. UTEX 1163 using AntiSMASH predicts the presence of twenty-two BGC regions in its genome. Of these twenty-two regions, six contain terpene biosynthesis genes. In order to produce scytoscalarol Prof. Dickschatt hypothesised that two terpene synthase / cyclase genes would be required. The first would cyclise the C_{25} GFPP precursor and the second would act as a prenyltransferase to install the guanidino group. In addition to these two terpene synthase/cyclase genes an enzyme would be required to install the hydroxyl group. A proposed biosynthetic pathway can be seen in Figure 4.3.

The gene cluster shown (antiSMASH region 3.6) in Figure 4.4 was determined to be the most likely cluster to produce scytoscalarol. The hypothetical protein, shown in pink, has a Pfam domain hit for a terpene synthase (PF03927.16). The squalene – hopene cyclase gene (PF13234.6), is usually found in BGCs that produce C_{30} triterpenes, however they are known to be promiscuous and can also produce sesterterpenes[145] like scytoscalarol. This gene cluster fits the hypothetical biosynthetic pathway where two terpene synthases or cyclases would be required. Additionally, cytochrome P450s are known to catalyse a range of modifications including hydroxylations7. The cytochrome P450 in this cluster (PF00067.22) is also annotates as a pentalene oxygenase. Whilst scytoscalarol does not contain pentalene this enzyme could have promiscuity to hydroxylate the fused cyclohexane rings of scytoscalarol.

A multigene Blast search of these three genes revealed that this BGC is well conserved amongst *Scytonema.* Figure 4.5 shows that these genes are also conserved in *Scytonema* sp. HK-05 which is an alternate name for *Scytonema* sp. NIES 2130, a strain we also obtained in our lab. We confirmed by LC-MS that this strain did not produce scytoscalarol in any of the screens undertaken.

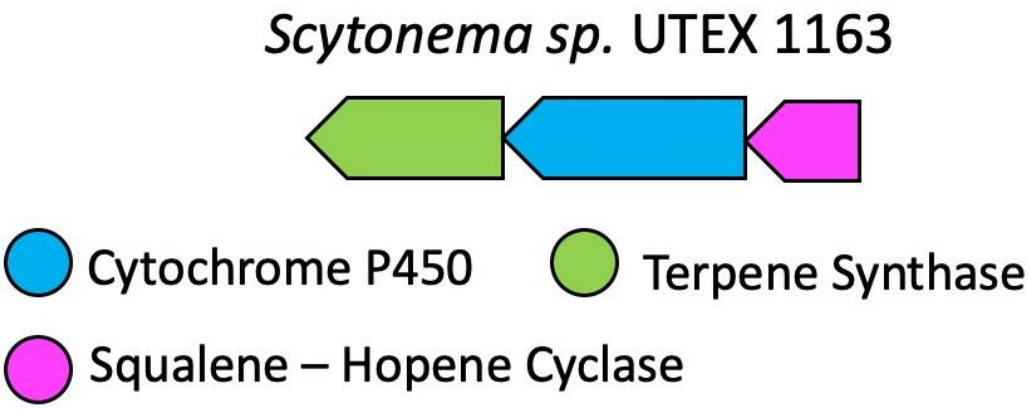

Figure 4.4 A gene cluster from *Scytonema* sp. UTEX 1163 that antiSMASH predicts will produce a terpene. This was determined to be the most likely cluster to produce scytoscalarol. The hypothetical protein is predicted to contain a terpene synthase domain by antiSMASH. The cluster also contains a cytochrome P450 gene which can catalyse hydroxylations.

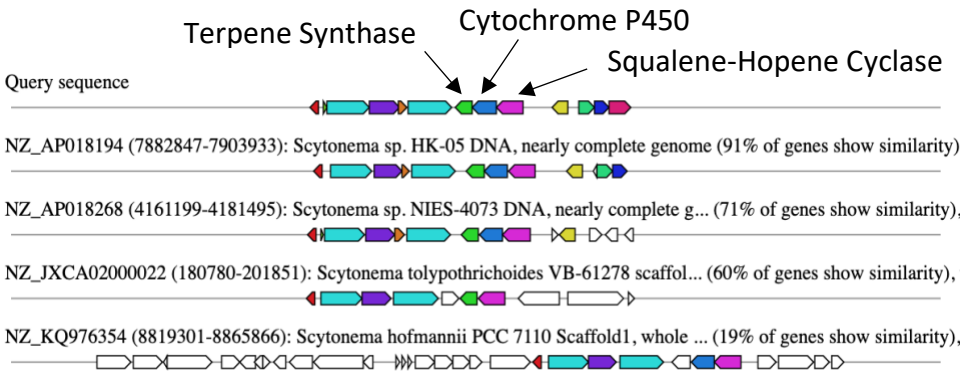

Figure 4.5: Multigeneblast result for possible scytoscalarol gene cluster found in *Scytonema* sp. UTEX 1163, shown as query sequence. The closest hits are shown below it in descending order of similarity. Homologous are seen in multiple *Scytonema* strains.

4.2.3 Cloning and heterologous expression of predicted scytoscalarol gene cluster

In order to confirm whether this candidate BGC is responsible for producing scytoscalarol, we decided to clone it and heterologously express it in *E. coli*. *E. coli* was chosen as the heterologous host due to the existence of plasmids that contain genes to boost the precursor supply of sesterterpenes. Heterologous expression of P450s is known to be difficult in *E. coli* due to their requirement of heme[144] as a cofactor. Heme supply in *E. coli* is limited, often resulting in incomplete holo-heme protein46. Therefore, only the putative terpene synthase and the squalene – hopene cyclase from the BGC were cloned and expressed meaning that the final scytoscalarol molecule could not be produced in these expression experiments. As well as this, the substrate for the installation of the guanidino group is unknown so it was unclear whether the guanidino group would be installed in this experiment. The main goal of this experiment was to see the processing of the GFPP precursor into the cyclised scytoscalarol terpene skeleton. Cloning methodology can be found in detail in section 7.4.

Prof. Dickschat supplied a plasmid named pYE_WP_004941318 (henceforth referred to as pYE) which contains a GFPP synthase gene from *Streptomyces mobaraensis*, accession number WP_004941318. Based on its characterisation in the biosynthesis of the sestermobaraenes[148], this gene should provide a supply of the GFPP precursor, however it will require a supply of the five carbon isoprene subunits. The plasmid, pMBIS, would be used to boost supply of the five carbon isoprenoid subunits, DMAPP and IPP by including genes from the mevalonate-dependant pathway[149]. The genes included in the pMBIS vector are a mevalonate kinase, phosphomevalonate kinase, pyrophosphate decarboxylatase, FPP synthase and an IPP isomerase which allows the conversion of IPP into DMAPP. The pMBIS plasmid requires mevalonate as a precursor which would be supplemented in the growth medium.

If the predicted gene cluster is responsible for producing scytoscalarol then processing of the C_{25} GFPP precursor into the scytoscalarol skeleton would be seen, which could be detected by LC-MS. These two precursor systems were therefore co-expressed alongside

each of the terpene synthases from the predicted scytoscalarol gene cluster in order to boost terpene biosynthesis.

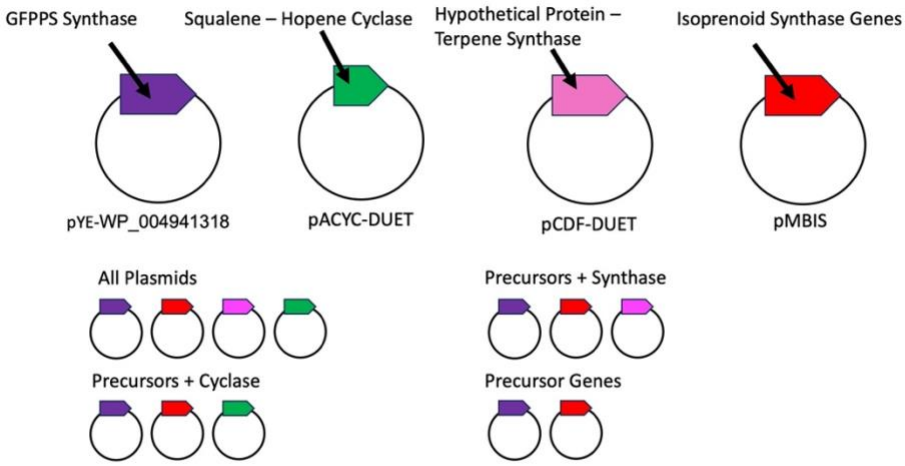

Figure 4.6: Top: The four plasmids used to try and validate the gene cluster responsible for producing scytoscalarol. Bottom: The four combinations of plasmids that were expressed and analysed. Each individual plasmid was also expressed on its own. All combinations were complimented with the empty vectors of the omitted plasmids so that all expressions contained all four antibiotics.

The squalene – hopene cyclase gene (1071 bp) was amplified from genomic DNA. The cyclase gene was cloned into the Pst1 and Nco1 sites of a pCDF-DUET vector using Gibson Assembly to produce the pCDF-DUET_cyclase plasmid with a total size of 4790 bp. The terpene synthase domain containing hypothetical protein (1730 bp) was amplified from genomic DNA. This synthase domain containing gene was cloned into the Pst1 and Nco1 sites of pACYC-DUET using Gibson Assembly to produce the pACYC-DUET_synthase plasmid with a total size of 5676 bp. Successful cloning was confirmed by whole plasmid Oxford Nanopore sequencing, carried out by Plasmidsaurus. These vectors were chosen because they have compatible resistance markers and origins of replication with the pMBIS and pYE vectors that contain the precursor synthesis genes. Also, pCDF-DUET and pAYC-DUET are protein expression vectors with strong inducible promoters to maximise expression of the insert genes. These constructs were then introduced via electroporation into *E. coli* BL21 DE3. Two vectors were transformed in at a time, followed by the making of new electrocompetent cells containing the two vectors. The remaining two vectors were then transformed into these fresh cells. This two-step approach was used in order to improve the efficiency of transformation. Control strains were generated by omitting one or several of the four constructs, and using the corresponding empty vector(s) instead. For the cyclase and synthase domain containing genes standard pACYC-DUET and pCDF-DUET were used, respectively. We were unable to source the empty vectors for the pMBIS and pYE vectors. As a result, vectors with a matching resistance marker and a compatible origin of replication were used. To replace the kanamycin resistant pYE vector pET-28a was used. To replace the tetracycline resistant pMBIS plasmid, pME6032[150] was used. As the goal was to determine which of the two genes was responsible for the first step cyclisation to form the scytoscalarol backbone four variations, as shown in Figure 4.6, were produced. In addition to these four variations each construct was also expressed on its own, alongside the three other empty vector controls.

Table 4.1: Summary of the constructs used to try and validate the scytoscalarol BGC. A vector with a compatible origin of replication and equivalent resistance marker was chosen for the pMBIS and pYE vectors as their empty vectors were not available.

Plasmid	**Insert**	**Resistance Marker**	**Empty Vector Equivalent**
pCDF-DUET	Terpene Cyclase	Spectinomycin	pCDF-DUET
pACYC-DUET	Terpene Synthase	Chloramphenicol	pACYC-DUET
pMBIS	Mevalonate Pathway Genes	Tetracycline	pME6032
pYE	GFPP Synthase	Kanamycin	pET-28a(+)

4.2.4 Expression of terpene synthase and cyclase genes for the validation of the scytoscalarol gene cluster

E. coli BL21 DE3 strains containing the four combinations of constructs (Figure 4.5), alongside strains containing each individual construct were fermented and assessed for terpene production. Initially expression was trialled in both LB and M9 minimal medium. M9, as a minimal medium gives a very clean background when analysing extracts by LC-MS, and has proven successful for the production of other specialised metabolites in *E. coli*. Both media were supplemented with 10 mM mevalonic acid[149] which is the substrate required for the increase of precursor supplied by the genes encoded on the pMBIS plasmid. All expression work was done in duplicate. For the first round of expression, cultures were grown at 37 ^{o}C until they reached an OD of 0.6 at which point all cultures were induced with 0.1 mM IPTG. Once induced cultures were grown at 18 ^{o}C for sixteen hours. After this the cultures were extracted using methanol and whole culture methanol extracts were analysed by LC-MS. Both positive and negative mode analysis were carried out as well as MS/MS analysis. Co-expression of the pYE and pMBIS vectors on their own should lead to increased production of the GFPP precursor . Expression of these plasmids in combination with either the synthase or the cyclase (as well as both) could then lead to processing of the GFPP precursor into a cyclised terpene. On a first instance, LC-MS data

were analysed manually to see if any obvious differential peaks were present in some samples but not in others. A deeper sift through the data was carried out using untargeted metabolomics data with the software Compound Discoverer (Thermo). No significant differential production was seen in this first experiment, it was also not possible to detect any production of GFPP in either positive or negative mode.

After the failed first expression the constructs were re-transformed into two *E. coli* strains, BL21 DE3 and NICO[151]. The expression conditions were also altered, after reaching an OD of 0.6 the cultures were induced with 0.3 mM of IPTG and they were then incubated at 30 ^{o}C for three hours before samples were taken. Samples were processed and analysed as in the first experiment. Once again no differential production of molecules was seen, and it was again not possible to see production of the GFPP precursor, or any associated molecules such as adducts or GFPP with the diphosphate hydrolysed.

The constructs containing the terpene synthase domain containing protein and squalene – hopene cyclase were sent to Prof. Dickschat's lab for in vitro testing. Currently, no notable activity has been detected.

4.3 Discussion of results

Unfortunately, this experiment was not successful and it was not possible to validate the gene cluster responsible for producing scytoscalarol. The expression of the pYE and pMBIS vectors which should produce the GFPP precursor may be responsible. The GFPP precursor was never detected by LC-MS, which could indicate that there was an issue with the expression of these genes. The expression of these genes could have been checked by proteomics to confirm if they were successfully expressed or not. However, these plasmids were expressed in accordance with existing literature[148,149]. Another reason for not being able to detect GFPP could be that the molecule quickly degrades or is used in another pathway. Another issue could be that LC-MS is not suitable to detect the molecule, GC-MS is commonly used to detect terpenes[152] and could have been used in this work. It is also possible that the gene cluster chosen is not the scytoscalarol gene cluster. The chosen gene

cluster has a homologue in *Scytonema* sp. NIES-2130 which does not produce scytoscalarol, which could indicate that it's the wrong gene cluster. If this experiment was to be revisited it would be prudent to test other BGCs. Given the similarity of cybastacine A and B to scytoscalarol it could be worthwhile to try and obtain *Nostoc* sp. BEA-0956, sequence it and compare the BGCs to *Scytonema* sp. UTEX 1163 to see if any terpene gene clusters are similar. Lastly, it is possible that the cytochrome P450 acts prior to the cyclisation of the terpene or is required for the terpene to be cyclised. If this experiment was to be re-started it would be prudent to first validate that the pYE and pMBIS plasmids are successfully expressed and the terpene precursors are produced, which could be done by proteomics to look for expression of the genes. Based on the lack of promising preliminary data in the co-expression experiments, it was decided to focus efforts on the identification of novel natural products.

Chapter 5: Heterologous expression of cyanobacterial gene clusters

5.1 Introduction and chapter aims

Heterologous expression is a common way to explore biosynthetic gene clusters of particular interest[153,154]. One advantage is you can express a single cluster in a clean background allowing for easier screening of production. Heterologous expression also allows the ability to work with a host organism that is genetically tractable and well-studied. Many heterologous hosts are available, allowing for a large choice when selecting the best host for a given expression. *E. coli* is a common heterologous host due to its fast growth and wide amount of tools existing for it[155,7.] There also exists a vast amount of literature to optimise the cloning and expression of BGCs in *E. coli.* Alternate heterologous hosts include *Streptomyces*[154]*,* which is a high GC organism and due to codon usage can be a good choice as a host for high GC organisms. Multiple cyanobacterial heterologous hosts also exist. One of the best studied cyanobacterial heterologous hosts is *Anabaena* PCC 7120[157]. Inducible riboswitches have been developed for PCC 7120 which allows for controllable heterologous expression of pathways[158]. One drawback of using a cyanobacterial heterologous host is that even the fast-growing strains grow very slowly when compared to *E. coli.* For this reason, we preferred *E. coli* as a heterologous host for the expression of BGCs from Gram-negative bacteria.

One class of BGC we were particularly interested in were type V lanthipeptides, which are a class of RiPP that contain a HopA1- domain containing protein. HopA1 is an effector protein found in *Pseudomonas syringae* and is an avirulence gene[159]. These HopA1 like proteins are widespread, 41 % of these proteins are found in Cyanobacteria and the majority of remaining clusters come from Actinobacteria7. It is thought that this HopA1, phosphotransferase protein pair represents an understudied RiPP modification which could be used to identify novel RiPPs. Recently a novel glycosylated lantibiotic, cacaoidin, was isolated from *Streptomyces cacoi*[160]*.* Cacaoidin contains a N,N-dimethyl lanthionine system which combines lanthionine rings, found in lanthipeptides, and N-terminus demethylation, found in linaridins[160]. Due to this combination of features cacaoidin has

been described as the first member of the new lanthidin class of RiPP[161]. The cacaoidin biosynthetic pathway contains a HopA1-like protein and a phosphotransferase, which are homologous to TsaD and TsaC[56], genes from the thiostreptamide biosynthetic pathway. Exploring these HopA1 proteins has led to the discovery of novel RiPPs as shown from the *Streptomyces* strains described above. However, nothing is currently known about the products of HopA1 biosynthetic gene clusters found in Cyanobacteria.

A HopA1-like protein, TsaD, is part of the thiostreptamide biosynthetic pathway. Thiostreptamide is a thioamitide, which is a structurally complex family of RiPPs that shows promising anti-tumour activity[56,162,163]. It has been proposed[56] that TsaD in conjunction with a phosphotransferase, TsaC, performs a key dehydration step in thiostreptamide biosynthesis. A search for TsaD like proteins using a 95% identity cut off identified 742 proteins, almost all of which are found next to a phosphotransferase7. These proposed functions for the HopA1-like protein and phosphotransferase were later confirmed through reconstitution of the cacaoidin biosynthetic pathway[164].

5.2 Heterologous expression of a HopA1-like protein containing gene cluster from *Scytonema hofmannii* UTEX 1834 using *E. coli.*

Previous bioinformatic analysis in our lab[56], later confirmed by AntiSMASH analysis of *Scytonema* strains revealed many HopA1-like gene containing clusters. *Scytonema hofmannii* PCC 7110 is predicted to contain four and all of our sequenced strains contained at least one HopA1-like protein containing BGC.

Table 5.1: The number of HopA1 gene clusters in each of our sequenced strains as predicted by antiSMASH.

Strain	Number of HopA1-like Containing Clusters
Scytonema bohnerii SAG 255.80	2
Scytonema crispum UTEX 1556	1
Scytonema hofmannii UTEX 1834	4
Scytonema javanicum SAG 39.90	1
Scytonema mirabile SAG 83.79	1
Scytonema myochrous SAG 46.87	2
Scytonema sp. PCC 7814	2
Scytonema sp. UTEX 1163	2
Scytonema sp. UTEX EE33	1
Scytonema spec. SAG 67.81	5

Since no natural products produced by a *Scytonema* HopA1-like gene containing cluster had been characterised so far, we decided to try and express one. A gene cluster from *Scytonema hofmanni* UTEX 1834 was chosen due to the cluster being well conserved amongst cyanobacteria, and this cluster is shown in Figure 5.1. This conservation meant that it was easy to propose the boundaries of the gene cluster based on multigene blast comparisons. No product is known for this gene cluster or any homologous gene clusters. Interestingly this cluster does not contain a putative cyclase like some previously characterised examples[161,7.] It is worth noting that thioamitides do not have a well-conserved cyclase. Generally type V lanthipeptides require a cyclase as the formation of some rings in lanthipeptides is energetically unfavoured[164]. Thus, it is unknown what structural motifs these BGCs products could possess and whether the product would be linear or cyclic. The refactoring of the BGC was designed to split and clone the cluster into two separate vectors, pET-28a(+) and pCDF-DUET, as shown in Figure 5.1. The gene cluster contains two operons, which were cloned into individual vectors. The precursor peptides, HopA1-domain containing protein and the phosphotransferase were cloned into pET-28a(+) and the ATP-binding cassette domain containing protein, HylD family secretion

protein and the peptidoprolyl isomerase were cloned into pCDF-DUET. Full details on cloning methodology can be found in section 7.4.

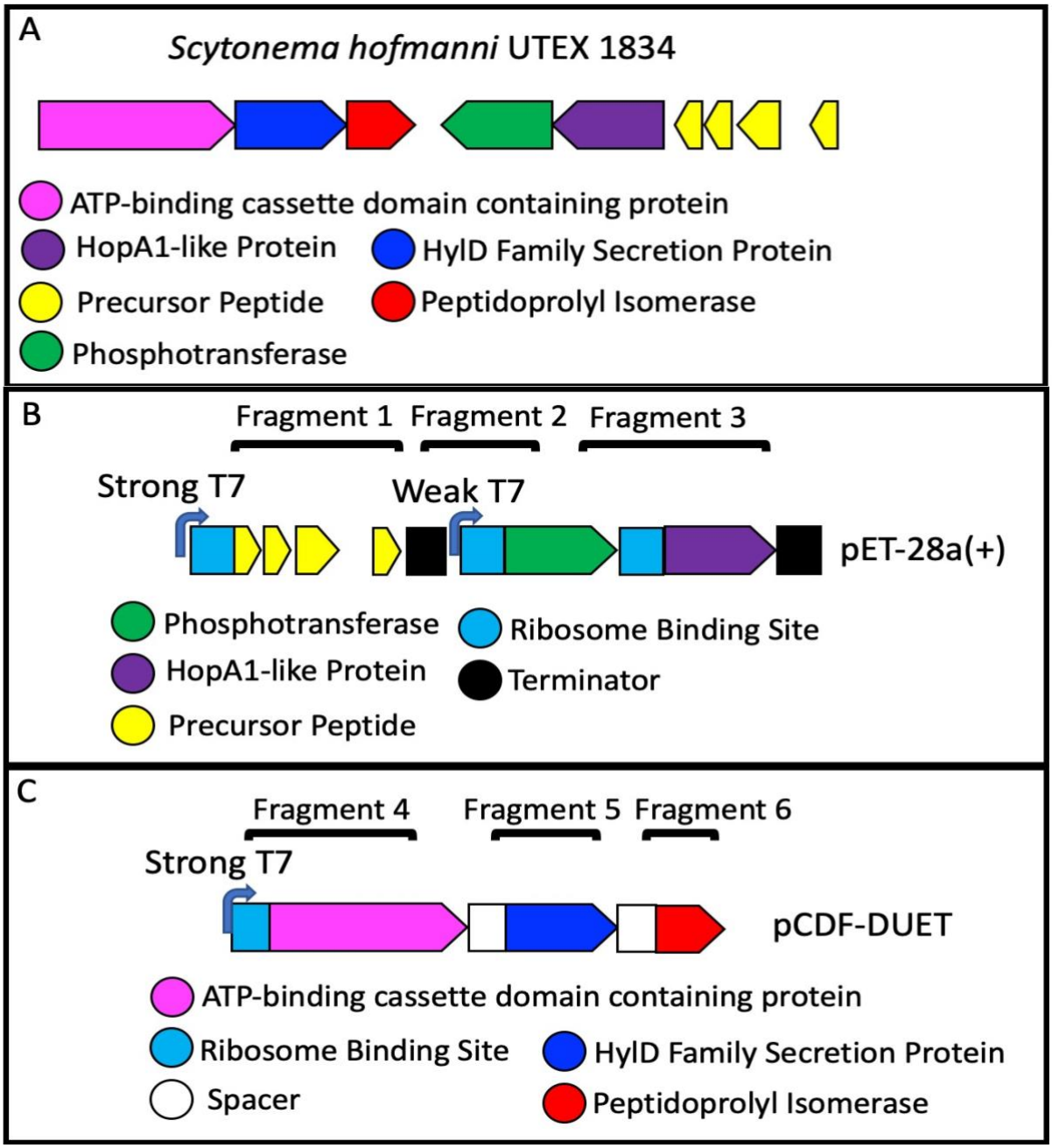

Figure 5.1: A – Schematic of the HopA1-protein containing gene cluster from *Scytonema hofmannii* UTEX 1834 selected for cloning and heterologous expression. Gene annotations were assigned by using both Prokka and antiSMASH. Gene cluster boundaries were determined based on MultigeneBlast comparisons between homologous gene clusters found in other strains of *Scytonema*. B & C – Schematic of the planned refactoring of the gene cluster in panel. The gene cluster was spilt across two plasmids as the gene cluster contains multiple operons that are presumably under the control of different promoters. Splitting the gene cluster also reduces the size of the expression cassette, which is a positive for successful cloning and expression. The refactored cluster contains canonical RBSs, shown in blue, terminators – shown in black, spacers – which are unique 80 bp sequences and the pET-28a(+) construct in panel B includes a weak version of the T7 promoter to ensure correct transcription levels

The cluster was refactored to include multiple synthetic DNA elements, promoters, spacers, RBSs and terminators. This was done so that the expression level of each gene would hopefully be sufficient for an active pathway. For example, it has been reported that in RiPP clusters the precursor peptide needs to be expressed in much higher quantities than the tailoring enzymes and transporters[165], which is because each precursor peptide molecule produced is processed into one molecule of the final product. Whereas, each discrete tailoring enzyme can process many substrates. To achieve this, the precursor peptides were placed directly under the control of the strong T7 promoter located on the pET-28a(+) plasmid. A terminator immediately follows on from the precursor peptide sequence to avoid read through from the strong T7 promoter onto the HopA1 and phosphotransferase genes. Another T7 promoter, but with weaker transcription levels, was placed after the terminator to drive transcription of the phosphotransferase and the HopA1-like genes. Each of these genes are preceded by a canonical ribosome binding site, to ensure efficient translation.

Both plasmids, pET-28a(+) and pCDF-DUET were digested with the restriction enzymes NcoI and XhoI. The phosphotransferase gene was amplified by PCR, the forward primer contained the sequence for the RBS. These fragments were then combined in a fusion PCR with the two fragments being used as the DNA template.. The last fragment for the pET-28a(+) construct contained the weak T7 promoter. This was amplified by PCR from an existing construct in the lab (Yoonia_BGC_pET29b), a primer contained the sequence for a terminator, to terminate the transcription from the strong promoter used to express the precursor peptides. The reverse primer contained the sequence for a canonical ribosome binding site. The fragments were then combined using Gibson Assembly with the pET-28a(+) linearised vector to produce the final construct. The whole sequence of this final construct was confirmed to be correct using Oxford Nanopore long read sequencing. The HylD and peptidoprolyl isomerase fragments were combined in a fusion PCR. The next step was to combine the HylD-Peptidoprolyl isomerase fragments with the ATP-binding fragment and the pCDF-DUET backbone using Gibson Assembly. This was attempted multiple times. After the first attempt, following the manufacturers recommendations was not successful, a second attempt was made testing multiple molar ratios of insert to backbone: 1:1, 2:1 and 3:1. Unfortunately none of these attempts worked. This experiment

was repeated once more, after re-amplifying all of the DNA segments by PCR, with the same result. As a result, it was decided that simply re-attempting the Gibson Assembly was no longer the best course of action and that the cloning would need to be re-designed. Rather than re-do the more complex cloning, including lots of refactoring, it was designed to instead attempt to clone a simpler gene cluster.

5.3 Heterologous expression of a HopA1-like gene containing cluster from *Scytonema* sp. SAG 67.81 using *E.coli.*

Due to previous attempts at cloning a HopA1-like protein containing gene cluster being unsuccessful, a simpler gene cluster, with fewer genes, was chosen to be expressed. A gene cluster from *Scytonema* sp. SAG 67.81 was chosen and the genes are shown in Figure 5.2. This gene cluster is the simplest example of a HopA1-like protein containing gene cluster among the clusters in our dataset. It has an annotated precursor peptide gene encoding four consecutive glycines. Double glycine is a classical cleavage motif[166] which separates the core and leader regions of the precursor peptide in lanthipeptides, thus we can speculate about the structure of the core peptide. The core peptide has a repeating section of nine lysines and eight serines forming a highly polar section. The core region of the precursor also contains terminal cysteines. The serines could potentially be dehydrated. No cyclase is found in this cluster so any further reactions would have to be spontaneous or catalysed by an enzyme not in the cluster. There are no characterised examples of such a highly charged cyanobacterial peptide natural product. Lysine-rich peptides have been isolated from other bacterial species[167,168], no lysine rich RiPPs have so far been found to be antimicrobial[169]. Cationic peptides, rich in lysine and arginine are known, and are known to be profoundly antimicrobial[170]. An example is the cathelicidin family of peptides, which forms part of the immune system in many verterbrates[171].

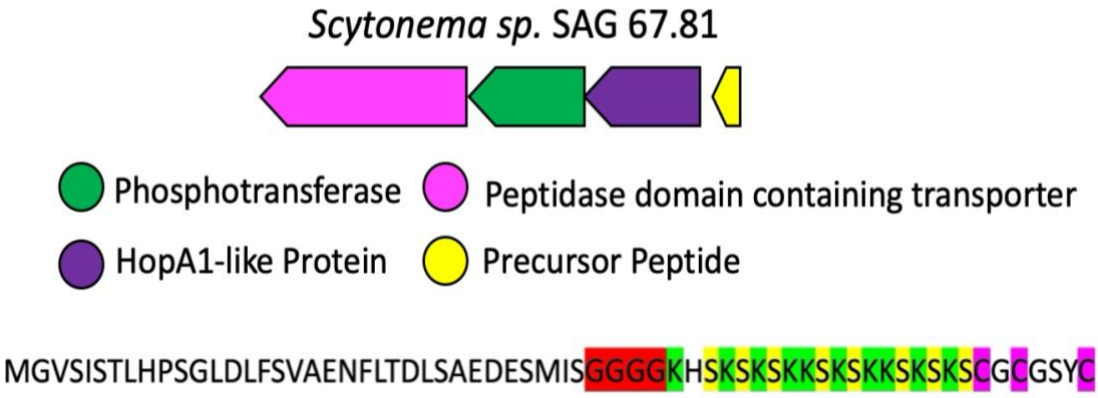

Figure 5.2: BGC from *Scytonema* sp. SAG 67.81 that was chosen to be cloned and expressed. This BGC is a simple example of type V lanthipeptide as it contains only the necessary genes, a putative precursor peptide, a HopA1-like protein and a phosphotransferase, alongside a transporter which contains a peptidase domain that we can hypothesise cleaves the core region of the precursor peptide from the leader region. The sequence of the putative precursor peptide contains a double glycine motive, a canonical cleavage site for lanthipeptides. The sequence of the predicted core region of the precursor peptide contains repeated lysines and serines as well as C-terminal cysteines.

This gene cluster is well conserved within the *Scytonema* genus. In this case the gene cluster was split into two parts for cloning into separate vectors. The putative precursor peptide gene was amplified by PCR using the 67_PP_Fwd and 67_PP_Rev primers to produce a 248 bp fragment. This fragment was cloned into pET-28a that had been made linear using the restriction enzymes NcoI and XhoI using Gibson Assembly to produce a 5425 bp final construct. The final sequence of the construct was confirmed by Oxford Nanopore sequencing. This way, the precursor peptide gene is under the control of the strong T7 promoter found on the protein expression vector, pET-28a(+).

The three remaining genes, likely part of the same operon, were amplified as one single fragment. pCDF-DUET was digested using the Nco1 and Xho1 restriction enzymes. The genes were cloned into the pCDF-DUET backbone using Gibson assembly. The final sequence was confirmed by whole plasmid Oxford Nanopore sequencing. The two vectors were then introduced via electroporation into BL21 DE3 and NICO *E. coli* strains. The vectors were transformed either together, individually alongside a partner empty vector and a dual empty vector control. Expression was carried out in both Lysogeny broth (LB) and M9 minimal media. Initially expression was carried out in M9 and LB, cultures were grown at 37 ^{o}C and induced once an optical density (OD) of 0.6 was reached using 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Once induced cultures were grown at eighteen degrees overnight. After expression 1 mL of culture was extracted with methanol analysed using a Q-Exactive LC-MS.

Comparative metabolomics was carried out using Compound Discoverer, but no production was initially detected. Therefore, multiple expression conditions were tested. For the second expression trial, cultures were induced using 0.3 mM IPTG, but once again no differential expression of molecules was detected by LC-MS. Lastly, in an attempt to force expression, cultures were induced using 1 mM IPTG and after induction were grown at thirty-seven degrees for three hours. However, no new metabolites were detected.

5.4 Discussion

Despite heterologous expression being a well-used method that has been used to discover at least 63 new families of bacterial natural products[153], the overall success rate is low[153]. One reason for the lack of successful expression could be that only two gene clusters were attempted to be expressed. In this project, only two HopA1-containing BGCs were selected for expression in *E. coli*, whereas a higher throughput approach of expressing more BGCs could have increased the likelihood of success. More variations in the culture conditions could have been tried, such as further tweaks to the length of fermentation, the temperature the cultures were incubated at, different production media. Codon optimisation was not used, the BGCs were checked for codons that are rare in *E. coli*, but codon optimisation has been shown to boost heterologous expression success[172].

Cyanobacterial BGCs have been previously heterologously expressed in a range of hosts including *E. coli*[173]*, Streptomyces venezuelae*[174] and *Anabaena* PCC 7120[157]. Of these strains, expression was only attempted in *E. coli, and* perhaps this could have been expanded to *Streptomyces.* Though issues with codon usage[175] could arise when trying to express cyanobacterial genes in *Streptomyces,* which is because *Scytonema* have low GC content and *Streptomyces* have high GC content.

For the expression of the HopA1 gene cluster from *Scytonema* sp. SAG 67.81, the cluster was successfully cloned however no candidate product could be detected. An SDS page gel did not show evidence that any of the cluster enzymes were being expressed, which could have been further explored using a proteomics approach,. This approach would be particularly relevant as the precursor peptide, a crucial substrate for RiPP biosynthesis requiring a high production level, can be difficult to see on an SDS page gel due to its small size (21 kDa). An alternate to this would be to tag the precursor peptide so it could be purified and then to reconstitute the pathway in vitro7. This approach would simplify expression and avoid possible degradation of the precursor peptide. This method of tagging and purifying the proteins has been successful previously, it was used to discover the polytheonamide family of cytotoxic peptides[177], and the selidamide family of lipopetides[178].

It may also be possible to synthetically produce the precursor peptide and then modify it with purified tailoring enzymes.

If this experiment was to be picked-up by someone in the future, the logical first step would be to choose multiple gene clusters and work on them concurrently. Initially only one gene cluster was expressed at a time due to this work happening concurrently with bio-informatics and screening attempts. It would also be pertinent to design the clusters for expression in other heterologous hosts, such as *Streptomyces*.

Chapter 6: Discussion and Future Work

At the start of this PhD the goals were set out as follows:

- Research international cyanobacterial strain collections and purchase all available strains of *Scytonema.*
- Attempt to culture all obtained strains of *Scytonema*, and establish a protocol for their propagation in our lab.
- Obtain high-quality whole genome sequences for all obtain strains of *Scytonema*.
- Through the use of computational genomic tools, analyse the *Scytonema* genomes to assess their capabilities to produce specialised metabolites.
- Screen the *Scytonema* strains to analyse their ability to produce bioactive metabolites. The culture conditions of the strains could would be altered to try and trigger the production of specialised metabolites
- From the genomic analysis of the *Scytonema* strains, identify particular BGCs of interest and then attempt to express them heterologously.

We were able to obtain 16 strains of *Scytonema,* though four of these strains were not culturable leaving twelve workable strains. A protocol was established for their continuous subculture in the lab. High quality genome sequences were obtained for ten of our *Scytonema* strains in collaboration with the Quadram Institute. I assembled these genomes using Unicycler and was able to analyse the genomes using a suite of computational tools. This analysis supported our theory that *Scytonema* strains have large genomes which encode for many potentially novel natural products.

To try and identify and characterise some of these novel natural products, multiple BGCs were chosen to be heterologously expressed. HopA1-like gene containing clusters were chosen for this due to general interest in these clusters across our lab. Expression of multiple clusters was attempted however none were successfully expressed.

More success was had through the classical screening approach. Initially, no bio-active metabolites were seen; however, this was attributed to the concentration of extracted

metabolites being too low. For the second screen the final extracts were one-hundred fold more concentrated with respect to the first screen. This screen identified an already known molecule, scytoscalarol. This showed that our approach was sufficient in being able to extract and identify bio-active metabolites, and also led to some work into the biosynthesis of scytoscalarol. This work led to a hypothesised biosynthetic pathway for the production of scytoscalarol. A putative biosynthetic gene cluster was identified and the genes were successfully cloned into *E.coli* protein expression vectors. Unfortunately, no evidence of biotransformation was seen. One major issue with this work was that no evidence of production of the terpene pre-cursor molecules was seen. This work will be continued by a future lab member and optimisation of the expression of the precursors will be required. Despite this, not much work is required to finish this project and determine if the putative gene cluster is indeed responsible for the production of scytoscalarol.

Despite the second screen successfully seeing bioactivity, it was not novel and thus was not a stunning success. The third screen further limited nutrients to impart a greater stress on the *Scytonema*. This approach was deemed a success as bioactivity was found in the screen that was not seen in the previous screens. This molecule, isolated from *Scytonema* sp. UTEX 1834, was purified and isolated through HPLC purification. LC-MS analysis of the purified molecule determined a putative molecular mass, which suggested that the molecule was novel. Initially, the molecule was isolated in very small amounts. Through the trialling of extraction methods, the efficiency of extraction was increased ten-fold. The improvement in extraction was due to switching the extraction solvent to methanol and only extracting from the cellular mass. Extracting with methanol from the cells being ten-fold more effective suggests that the molecule is retained within the cell, and not exported. Extracting the cells with methanol allowed 2.5 mg of the purified molecule to be obtained, which was enough for NMR analysis. There were issues with the NMR analysis; the initial analysis was carried out in deuterated chloroform however the molecule was not dissolved into solution. As a result, the analysis needs to be repeated and different NMR solvents need to be trialled in order to obtain full dissolution and more intense spectra. Deuterated dimethyl-sulfoxide (DMSO) would be a starting point. Despite issues with solubility, good quality spectra were obtained; however, not enough

time remained to fully analyse these spectra. Very little future work remains for this project. The NMR needs to be repeated in a different solvent, then structural characterisation needs to be carried out. This combined with the LC-MS data (importantly the MS/MS data) should allow for a putative structure to be determined. From a putative structure a putative BGC could be found. This analysis would lead to enough data for a high quality publication.

Ultimately, fewer bioactive molecules were seen than one would expect. From ten *Scytonema* strains, with over three-hundred predicted, novel BGCs, one novel compound was detected. This doesn't seem like very many. Some reasons for this could be: firstly, we did a poor job at turning on BGCs. We could have tested different light intensities or temperatures. Secondly, novel natural products could have been produced, but they our screens were not sufficient in picking them up. We could have screened against plants or for cytotoxic activity.

Chapter 7: Materials and methods

7.1 Cyanobacterial methods

7.1.1 Chemicals and reagents

Unless otherwise specified, all chemicals and reagents were purchased from Merck (previously Sigma Aldrich). All solvents used for extractions and chromatography were purchased from Fisher Scientific. Unless otherwise specified all enzymes and competent cells were purchased from New England Biolabs (NEB).

7.1.2 Growth of cyanobacteria

Initially Cyanobacterial Strains were grown in the medium recommended by the strain collections that they came from. All strains were grown in liquid media. The media recipes are shown below. *Scytonema hofmannii* UTEX 1834, Scytonema sp. UTEX EE33, *Scytonema lyngbyoides* SAG 40.90, *Scytonema* sp. NIES 2130, *Scytonema* PCC 7110, *Scytonema* PCC 7814 and *Scytonema* sp. UTEX LB 2588 were all recommended to be grown in BG-11 media (Table 7.1). *Scytonema* sp. UTEX 1163 was originally grown in Soilwater GR+ (Table 7.3). *Scytonema crispum* UTEX LB 1566 was originally grown in modified Bold 3N (Table 7.4). *Scytonema hofmannii* UTEX B 1834 was originally grown in soil extract media (Table 7.3). *Scytonema bohnerii* SAG 255.80, *Scytonema mirabile* SAG 83.79, *Scytonema myochrous* SAG 46.87 and *Scytonema* sp. SAG 67.81 were all originally grown in ES media (Table 7.6). *Scytonema javanicum* SAG 39.90 was originally grown in Z 45/4 (Table 7.8). Scytonema spirulinoides SAG 41.90 was originally grown in MiEB12 (Section 7.1.8). As well as sub culturing in their recommended media the growth of all acquired *Scytonema* strains in liquid BG-11 was tested. All strains that were able to be subcultured grew equally as well in liquid BG-11 as in their recommended medium. As a result all further growth was carried out in liquid BG-11.

7.1.3 BG-11 medium recipe

Table 7.1: Components used to make BG-11 media. All components are added to 1 L of ddH_2O and then autoclaved at 121 ^oC for one hour. To make solid media 15 g/L of agar is added. Stock solutions were made up prior and kept at 4 ^oC for storage. Recipe for trace metal solution shown in table 3.

Component	**Amount**	**Stock Solution Concentration**
$NaNO_3$	10 mL/L	15 g/100 mL
K_2HPO_4	10 mL/L	0.4 g/100 mL
$MgSO_4 \bullet 7H_2O$	10 mL/L	0.75 g/100 mL
$CaCl_2 \bullet 2H_2O$	10 mL/L	0.36 g/100 mL
Citric Acid $\bullet H_2O$	10 mL/L	0.06 g/100 mL
Ferric ammonium citrate	10 mL/L	0.06 g/100 mL
$Na_2EDTA \bullet 2H_2O$	10 mL/L	0.01 g/100 mL
Na_2CO_3	10 mL/L	0.2 g/ 100 mL
BG-11 trace metals solution	1 mL/L	-

Table 7.2: Recipe for the trace metals solution for BG-11 media. All components are added to 900 mL of ddH_2O then once all components have been added and are dissolved the flask is topped up to 1 L total volume using ddH_2O. Trace metal solution kept at 4 ^oC for storage.

Component	**Amount**	**Final Concentration**
H_3BO_3	2.86 g/L	46 mM
$MnCl_2 \bullet 4H_2O$	1.81 g/L	9 mM
$ZnSO_4 \bullet 7H_2O$	0.22 g/L	0.77 mM
$Na_2MoO_4 \bullet 2H_2O$	0.39 g/L	1.6 mM
$CuSO_4 \bullet 5H_2O$	0.79 g/L	0.3 mM
$Co(NO_{3)2} \bullet 6H_2O$	49.4 mg/L	0.17 mM

7.1.4 Soilwater GR+ medium recipe

Once the components are combined the container is covered and steamed for two consecutive days for three hour each day. The media should reach a temperature of 98 ^{o}C and then, after three hours, be left to cool at room temperature. Following two days of steaming the media is left at 4 ^{o}C for twenty-four hours then is ready to use. Soil was provided by JIC Horticultural services.

Table 7.3: Recipe for Soilwater GR+ Media. It is important to make sure that the soil contains no herbicides or pesticides. Recipe is for a total volume of 200 mL.

Component	**Amount**
Soil	15 g
$CaCO_3$	1 mg
ddH_2O	200 mL

7.1.5 Modified bold 3N medium

All components apart from the vitamins are added to 1 L of ddH_2O and then autoclaved at 121 ^oC for one hour. The pH was brought to 7.2 using 1 M HCl. Vitamin stocks were filter sterilised and added once the media cooled. The medium was stored at 4 ^oC.

Table 7.4: Composition of Modified Bold 3N medium. To make solid media 15 g/L of agar is added. Stock solutions were made up prior and kept at 4 ^oC for storage. Recipe for P-IV metal solution shown in table 5.

Component	**Amount**	**Stock Solution Concentration**
$NaNO_3$	30 mL/L	2.5 g/ 100mL
K_2HPO_4	10 mL/L	0.75 g/100 mL
$MgSO_4 \bullet 7H_2O$	10 mL/L	0.75 g/100 mL
$CaCl_2 \bullet 2H_2O$	10 mL/L	0.25 g/100 mL
KH_2HPO_4	10 mL/L	1.75 g/100 mL
NaCl	10 mL/L	0.25 g/100 mL
Soilwater GR+ Media	40 mL	-
Vitamin B_{12}	1 mL/L	0.1 mM
Biotin	1 mL/L	0.1 mM
Thiamine	1 mL/L	0.1 mM
P-IV Metal Solution	6 mL/L	-

Table 7.5: Composition of P-IV Metal Solution for Modified Bold 3N medium. Components were added to 950 mL of ddH_2O and dissolved whilst continuously stirring. Total volume is then brought to 1 L and stored at 4 ^oC.

Component	Amount	Final Concentration
$Na_2EDTA• 2H_2O$	0.75 g/L	2 mM
$FeCl_3•6H_2O$	0.097 g/L	0.36 mM
$MnCl_2•4H_2O$	0.041 g/L	0.21 mM
$ZnCl_2$	0.005 g/L	0.037 mM
$CoCl_2•6H_2O$	0.002 g/L	0.0084 mM
$Na_2MoO_4•2H_2O$	0.004 g/L	0.017 mM

7.1.6 ES medium

All components are added to 1 L of ddH_2O and autoclaved at 121 ^oC for one hour. To make soil extract, an amount of soil, obtained from JIC horticultural services, is added to a flask and water poured in until it's above the soil line. The flask is boiled for one hour in a steamer twice in a twenty-four hour period. The supernatant is then filtered out and autoclaved at 121 ^oC.

Table 7.6: Recipe for ES medium. For solid medium 15 g/L of agar is added. All components are stored at 4 ^oC for storage. The recipe for the micronutrient solution is found in table 7.7.

Component	Amount	Stock Solution Concentration
KNO_3	20 ml/L	1 g/100 mL
K_2HPO_4	20 ml/L	0.1 g/100 mL
$MgSO_4 • 7H_2O$	20 ml/L	0.1 g/100 mL
Soil extract	30 ml/L	-
Micronutrient solution	5 mL	-

Table 7.7: Recipe for the micronutrient solution used to make ES medium. Stocks are made up in 100 mL of ddH_2O and then the corresponding amount of stock added to 900 mL of ddH_2O. Once all components have been added the solution is made up to a volume of 1 L with ddH_2O and autoclaved at 121 ^oC for one hour. Solution is kept at 4 ^oC for storage.

Component	**Amount**	**Stock Solution Concentration**
$ZnSO_4 \bullet 7H_2O$	1 ml/L	0.1 g/100 mL
$MnSO_4 \bullet 4H_2O$	2 ml/L	0.1 g/100 mL
H_3BO_3	5 ml/L	0.2 g/100 mL
$Co(NO_{3)2} \bullet 6H_2O$	5 ml/L	0.02 g/100 mL
$Na_2MoO_4 \bullet 2H_2O$	5 ml/L	0.02 g/100 mL
$CuSO_4 \bullet 5H_2O$	1 ml/L	0.0005 g/100 mL
$FeSO_4 \bullet 7H_2O$	0.7 g	-
EDTA	0.8 g	-

7.1.7 Z 45/4 media

All components are added to 1 L of ddH_2O and autoclaved at 121 °C for one hour. The medium was stored at 4 °C

Table 7.8: Composition of Z 45/4 medium. Composition of the micronutrient solution is shown in table 7.9. For solid media 15 g/L of agar is added.

Component	**Amount**
K_2HPO_4	41.0 mg
KH_2HPO_4	17.0 mg
$CaCl_2 \bullet 2H_2O$	37.0 mg
$MgSO_4 \bullet 7H_2O$	25.0 mg
FeEDTA$\bullet 2H_2O$	10.0 mg
Micronutrient Solution	0.06 mL

Table 7.9: Components for the micronutrient solution for Z 45/4 medium. All components are added to 100 mL of ddH_2O.

Component	Amount
H_3BO_3	310.0 mg
$MnSO_4 \bullet 4H_2O$	223.0 mg
$Na_2WO_4 \bullet 2H_2O$	3.3 mg
$(NH_4)_6Mo_7O_{24} \bullet 4H_2O$	8.8 mg
KBr	11.9 mg
KI	8.3 mg
$ZnSO_4 \bullet 7H_2O$	28.7 mg
$Cd(NO_3)_2 \bullet 4H_2O$	15.4 mg
$Co(NO_3)_2 \bullet 6H_2O$	14.6 mg
$CuSO_4 \bullet 5H_2O$	12.5 mg
$NiSO_4(NH_4)_2SO_4 \bullet 6H_2O$	19.8 mg
$Cr(NO_3)_3 \bullet 7H_2O$	3.7 mg
$VOSO_4 \bullet 2H_2O$	2.0 mg
$Al_2(SO_4)_3K_2SO_4 \bullet 2H_2O$	46.4 mg

7.1.8 $MiEB_{12}$ medium

$MiEB_{12}$ medium was prepared in the same way as ES medium, however 5 x 10^{-6} g/L of vitamin B12 was also added before the media is autoclaved.

7.1.9 Cyanobacterial growth conditions

Unless otherwise specified, liquid cultures of cyanobacteria were grown in 175 mL tissue culture flasks containing either 50 or 100 mL of media. These flasks were laid flat and stacked to maximise the surface area exposed to light. Strains were initially grown in the plant growth room on a 16:8 day night cycle, during the day portion the light intensity was 1000 lux. The temperature throughout the cycle was maintained at 25 ^{o}C. When strains were grown in the plant growth cabinet the same conditions were maintained.

7.1.10 Sub-culturing of cyanobacteria

For general maintenance cyanobacterial strains were sub-cultured every four to six weeks. As long-term freezer stocks of the cyanobacterial strains were not viable continuous sub-culturing was a necessity. To subculture 1 mL of a cyanobacterial culture that had been grown for four to six weeks was added to 50 mL of media and then incubated as described above.

7.2 General methods

7.2.1 *E. coli* methods

E. coli DH5α was grown in liquid LB media at 37 ^{o}C for sixteen hours whilst shaking at 200 rpm. *E. coli* DH5α was grown on LB agar at 37 ^{o}C until colonies were visible and large enough to be picked. Having been grown, plates were stored at 4 ^{o}C. Stocks of *E.coli* were stored in twenty-five percent glycerol at -80 ^{o}C.

7.2.2 Genomic DNA extraction

Genomic DNA was extracted from cyanobacterial strains using the Fast DNA^{TM} SPIN Kit for Soil (MP Biomedicals, Loughborough, UK).

7.2.3 Transformation of chemically competent *E. coli* (DH5α)

One 50 µL aliquot of *E. coli* per transformation was taken from storage at – 80 °C and left on ice to thaw. Once thawed 2 µL of each plasmid (up to three plasmids) being transformed was added. One the plasmid was added the *E. coli* was left on ice for thirty minutes. After thirty minutes the mixture is heat shocked by being placed in a 42 °C water bath for forty-five seconds before being placed on ice for two minutes. After heat shocking 1 mL of LB is added and then incubated at 37 °C for one hour whilst shaking at 200 rpm. The cells are then centrifuged at 8000 rpm and 900 µL of the supernatant was removed. The cells pellet was then resuspended in the remaining supernatant and spread on an LB agar plate containing the appropriate antibiotics. The plate was then incubated at 37 °C until colonies are seen (approximately sixteen hours).

7.2.4 Transformation of electrocompetent *E.coli* (BL21)

One 50 µL aliquot of *E. coli* per transformation was taken from storage at – 80 °C and left on ice to thaw. Once thawed 2 µL of each plasmid (up to three plasmids) being transformed was added. This mixture was then transferred to an electroporation cuvette with a 2mM path length (purchased from Cell Projects). It is important to tap the cuvette to ensure the liquid covers the whole of the cuvette and to make sure no bubbles remain. Electroporation was carried out using an Eppendorf Eporator at a setting of 2500 V. 900 µL of LB is then added and transferred from the cuvette to a 1.5 µL Eppendorf tube and incubated at 37 °C for one hour whilst shaking at 200 rpm. The cells are then centrifuged at 8000 rpm and 900 µL of the supernatant was removed. The cells pellet was then resuspended in the remaining supernatant and spread on an LB agar plate containing the appropriate antibiotics. The plate was then incubated at 37 °C until colonies are seen (approximately sixteen hours).

7.2.5 Preparation of electrocompetent *E. coli* cells

An aliquot of the desired competent cells were grown up overnight in liquid LB containing the appropriate antibiotics. 1 mL of the overnight culture was added to 10 mL of LB containing the appropriate antibiotics and grown at 37 ^{o}C and shaking at 200 rpm until the optical density (OD) reached 0.4. Once the OD is reached the flask was split into two 50 mL falcon tubes and centrifuged at 4 ^{o}C and 7000 rpm for five minutes. After centrifuging the supernatant was discarded the pellet was resuspended in 50 mL of sterile ice cold ddH_2O. Centrifugation was then repeated at 4 ^{o}C and 7000 rpm for five minutes and the supernatant discarded. The two tubes were then pooled together by adding 50 mL of ice cold ddH_2O to one tube, resuspending the pellet and transferring it to the second and resuspending the second pellet. Centrifugation was then repeated at 4 ^{o}C and 7000 rpm for five minutes and the supernatant discarded. The cells were then resuspended in 50 mL of sterile, ice cold 10 percent glycerol, centrifugation was then repeated at 4 ^{o}C and 7000 rpm for five minutes and the supernatant discarded. Cells were then resuspended in 1 mL of ten-percent sterile, ice-cold glycerol and centrifuged at 4 ^{o}C and 7000 rpm for five minutes. The supernatant was then discarded and the pellet resuspended in the desired volume of ten-percent glycerol. This suspension was then dispensed into 50 µL aliquots which were stored at -80 ^{o}C until use.

7.3 Expression of predicted scytoscalarol gene cluster

All combinations of the four plasmids shown in Table 4.1 as well as the four empty vector equivalents were transformed into *E. coli* BL21 DE3 to produce 16 different variations. These sixteen variations were grown, in duplicate, in both LB and M9 media. The media were supplemented with mevalonic acid to produce a final concentration of 10 mM. Cultures were grown at 37 ^{o}C until they reached an OD of 0.6 at which point the cultures were induced using 0.1 mM IPTG. After inducing cultures were incubated at 18 ^{o}C for sixteen hours. Cultures were extracted with an equal volume of methanol and these extracts were analysed by LC-MS.

7.4 Cloning and sequencing methods

7.4.1 Polymerase Chain Reaction (PCR)

PCR reactions were carried out using the components and timings shown in the tables below, following manufacturers recommendations. Annealing temperature was calculated using Geneious Prime (Dotmatics). High fidelity PCR reactions, intended for cloning, was carried out using Q5 DNA Polymerase (NEB) and colony PCR reactions were carried out using GoTaq GRN 2z Master Mix (Promega).

Table 7.10: Reaction components for PCR reactions carried out using Q5 High-Fidelity Polymerase (NEB). Ratios were doubled or halved depending on how much product was required and the efficiency of the PCR.

Reaction Components	
Reagent	**Volume (µL)**
5x Q5 Reaction Buffer	10
10 mM dNTPS	1
Each Primer (10 mM)	2.5
Template DNA	1
Q5 High-Fidelity DNA Polymerase	0.5
Milli-Q H_2O	32.5

Table 7.11: Thermocycler conditions for PCR reactions carried out using Q5 High-Fidelity Polymerase (NEB). Number of cycles could be tweaked to improve yield.

Thermocycler Conditions		
Temperature (^{O}C)	**Time (s)**	
98	120	
98	15	35x
Annealing Temperature	30	
72	30 / kbp	
72	300	
4	Indefinitely	

Table 7.12: Reaction components for colony PCR reactions carried out using GoTaq GRN 2x Master Mix (Promega)

Reaction Components	
Reagent	**Volume (µL)**
GoTaq GRN 2x Master Mix	10
Each Primer (10 mM)	0.5
Template DNA	Colony
Milli-Q H_2O	9

Table 7.13: Thermocycler conditions for colony PCR reactions carried out using GoTaq GRN 2x Master Mix (Promega)

Thermocycler Conditions		
Temperature (^{O}C)	**Time (s)**	
95	480	
95	30	35x
Annealing Temperature	30	
72	60 / kbp	
72	300	
4	Indefinitely	

7.4.2 Agarose gel electrophoresis

Generally, 0.8 percent agarose gels with 3 µL ethidium bromide per 100 mL were used for electrophoresis. If the DNA strand of interest was less than 500 bp an agarose gel of up to 1.5 % agarose was used. All gels were run for forty-five minutes at 110 V in 1x TBE (Tris Borate EDTA) buffer. 1 µL of 6x loading dye (NEB) per 5 µL of sample was added. In all cases the 1KB+ ladder from NEB was used. Gels were visualised using UV light.

7.4.3 Purification of DNA from agarose gels and PCR reactions

PCR reaction mixtures were ran on an agarose gel in accordance with the specifications above. Once ran, DNA fragments were excised from the gel using a scalpel and placed into an Eppendorf tube. Purification was carried out using the GFX™ PCR DNA and Gel Band Purification from Cytiva. Purified DNA was eluted in 24 µL of ddH_2O.

7.4.4 DNA digestion of plasmids

All restriction enzymes were purchased from New England Biolabs (NEB) and used with the accompanying CutSmart™ buffer. Digestions were carried out for between one and sixteen hours. After digestion enzymes were inactivated either by being immediately placed at -80 ^oC or by being immediately gel purified.

7.4.5 Gibson assembly

Gibson Assembly was carried out using NEBuilder™ HiFi DNA Assembly Mix in accordance with manufacturer's instructions. 100 ng of vector DNA was added to 200 ng of insert DNA and the equivalent volume of master mix was added. Reactions were incubated in a thermocycler at fifty degrees for sixty minutes. After this 2 µL of the Gibson Assembly mix was transformed as shown above.

7.4.6 Plasmid sequencing

All sequencing of plasmids was carried out by Plasmidsaurus in accordance with their specifications.

7.5 Metabolomics and Chromatography Methods

7.5.1 LC-MS sample preparation of *E. coli*

E. coli bacterial cultures that were screened using LC-MS were prepared in the following way. 1 mL of bacterial culture was spun down at 16,000 rpm and 700 µL of the supernatant is removed and placed into a 1.5 mL glass vial which can be stored at -20 ^{o}C.

7.5.2 LC-MS sample preparation of *Scytonema* cultures

Mature *Scytonema* cultures were placed in a -80 ^{o}C freezer until fully frozen. Once frozen they were freeze dried until all H_2O was removed (approximately twenty-four hours). The remaining crude extract was resuspended in 5 mL of ethyl acetate and centrifuged at 16,000 rpm for ten minutes. The supernatant was then taken to remove anything insoluble and dried completely using the High BP setting on a Genevac EZ-2 Elite Personal Evaporator from SP Scientific. This dried sample was stored at -80 ^{o}C until it was analysed. For analysis the sample was dissolved in 100 µL of fifty-percent acetonitrile and placed in a 1.5 mL glass vial containing a glass insert.

7.5.3 Analytical LC-MS conditions

All LC-MS was carried out using QExactive LC-MS system (Thermo Fisher Scientific). Unless otherwise specified a Kinetex 2.6 µM C18 100 Å column (Phenomenex). The solvents used were 0.1% formic acid in H_2O and acetonitrile (MeCN). The gradient was a linear increase from 0-95% organic solvent over six minutes followed by 100% organic solvent up until nine minutes to fully wash off material still stuck to the column. Generally results were acquired in positive mode with full MS/MS. The mass range was 50 – 3300 m/z. Results were analysed using Freestyle and Compound Discoverer both provided by Thermo Fisher Scientific.

7.5.4 Semi-preparative HPLC for the isolation of bioactive compound from *Scytonema* sp. UTEX 1834

This section refers to the work carried out in section 3.5. Semi-Prep HPLC was carried out using an Agilent 1290 HPLC-MS-ELSD instrument and a Kinetex 5 µM XB-C18 250 x 10 mm 100 Å column. The solvents used were 0.1% formic acid in H_2O and acetonitrile. A gradient of 10-98% organic solvent was used, ramping up linearly over sixty minutes. A flow rate of 3 mL /min was used. Fractions were collected from minutes thirty to forty, each fraction being 1.5 mL (thirty seconds of flow). This time frame was deduced from a previous run where fractions were taken across the whole run and screened for bioactivity. These fractions were then analysed on the LC-MS and all samples with scytoscalarol in were pooled. These pooled samples were then run on the semi-prep HPLC with the same conditions above. However, the collection settings were altered to collect six second slices (300 µL) from thirty-five to thirty-seven minutes. Once again, these fractions were analysed on the LC-MS, in accordance with the conditions above, to confirm which fractions contained pure scytoscalarol.

7.5.5 Biological screening assays for antimicrobial activity

The following indicator strains were used to screen *Scytonema* extracts for anti-bacterial activity. B17.06226 *Enterococcus faecium* (VRE), *Staphylococcus aureus* (MRSA), *Pseudomonas aeruginosa* PA01 and *E. coli* NR698. *E. coli* NR698 has increased permeability to antibiotics due to a compromised outer-membrane lipopolysaccharide assembly. These strains were acquired from a lab collection and were originally obtained as isolated from the Norfolk and Norwich Hospital. A stock of each desired indicator strain was taken from a -80 ^{o}C freezer and a small amount of frozen cells was transferred to a universal containing 10 mL of LB using a toothpick. This culture was grown for sixteen hours at 37 ^{o}C and 200 rpm. After sixteen hours 1 mL of this solution was added to 10 mL of LB in a universal and grown until an OD of 0.4-0.6 was reached. Whilst this culture was growing 100 mL of solid soft nutrient agar (SNA) was melted in a steamer. Once melted the SNA was placed in a water bath set to 50 ^{o}C until ready to be used. Once the indicator strain reached an OD between 0.4 and 0.6, 5 mL of the indicator strain was added to 100 mL of 50 ^{o}C SNA then poured into 10 cm square plates (50 mL of SNA per plate). This plates were then left to dry. Once dried, extracts and controls were spotted onto the plate (5 µL per spot) and once all spots added the plate was placed at 4 ^{o}C for an hour. Controls included a positive control of 50 µM apramycin and a negative control of fifty-percent MeCN. Plates were then grown for sixteen hours at 30 ^{o}C after which they could be analysed and photographed.

7.6 Bioinformatic analysis

7.6.1 Assembly of cyanobacterial genomes

Combined Illumina-Nanopore data was obtained from The Quadram Institute of Bioscience in collaboration with Prof. Alison Mather and Dr Dave Baker. All work was carried out myself. All data is archived on the Truman Lab shared drive with plans for it to be uploaded and made publicly available. Raw Nanopore reads wee filtered using Nanofilt[179] with a length cut off of 1kbp and a quality cut off of >7. Illumina reads were trimmed and filtered using FastP[180] using the default parameters. Hybrid assemblies were carried out using Unicycler[104] version 0.5.0. This version of Unicycler was installed and run on the John Innes Centres high performance computing cluster. Unicycler was run using the default parameters. The assembled contigs were separated into strains (binned) using Metabat2[106] using default parameters.

7.6.2 CheckM analysis of binned contigs

CheckM[107] v1 was ran on the JIC high performance cluster using the Metabat2 bins as input. The following commands were used, qa – to asses bins for contamination and completeness, analyse – to identify marker gens in bins, unique – to ensure no sequences were assigned to multiple bins, and merge – to identify and merge bins with complementary sets of marker genes.

7.6.3 Annotation of assembled genomes

The assembled genomes were annotated using Prokka[181], default settings were used however the locus tag was set as the strain name.

7.6.4 Pan-genome analysis using Roary

Roary[182] was ran locally using annotated *Scytonema* genomes outputted by Prokka as GFF3 files as input, as well as *Scytonema* files downloaded from NCBI as GFF3 files. Default parameters were used apart from the -I variable, the identity cut off, which was varied between analyses.

7.6.5 BiG-SCAPE analysis of *Scytonema* genomes

BiG-SCAPE was run on the JIC high-performance cluster using both our assembled *Scytonema* genomes as well as high-quality *Scytonema* genomes from NCBI. These genomes were ran through antiSMASH and the resulting gene clusters regions were used as input. The following parameters were added to the default parameters. –hybrids-off, which splits hybrid clusters into their constituent gene clusters. The –cutoffs which is the raw distance cut-off value, was varied between analysis.

7.6.6 AntiSMASH

All Scytonema genomes were analysed using antiSMASH 7.0[16] to determine the number of gene cluster regions. *Scytonema* genomes were inputted as .gbk files using the antiSMASH website. The following options were used for the antiSMASH analysis: detection strictness – relaxed, KnownClusterBlast, ClusterBlast, SubClusterBlast, MIBiG cluster comparison, ActiveSiteFinder, RREFinder, Cluster Pfam analysis and Pfam-based GO term annotation.

7.6.7 Construction of phylogenetic trees

Phylogenetic trees were constructed using autoMLST[111]. Genomes were uploaded as .gbk files, publicly available genomes were inputted using their NCBI accession number. AutoMLST maintains a database of the ten highest quality genomes for each species. The average nucleotide identity (ANI) of input strains is compared to this database and the

nearest reference organisms from NCBI RefSeq are selected. Single copy genes with dN/dS values less than one are then identified. These single copy genes are then aligned and trimmed before the tree is produced.

7.7 Primers

Table 7.14: Table of primers used for the experiments within this thesis.

Primers / Oligonucleotide	Sequence (5'-3')
Cyclase_Fwd_pCDF	GTTTAACTTTAATAAGGAGATATACATGGCGAGTGATTTAGAACAACA
Cyclase_Rev_pCDF	GCGGCCGCAAGCTTGTCGACCTGCACTACTGCAGAGCCGCCAG
Synthase_Fwd_pACYC	GTTTAACTTTAATAAGGAGATATACATGGTTACTTCTTTAACCAAGCAAA
Synthase_Rev_pACYC	GCGGCCGCAAGCTTGTCGACCTGCACTATCCAGCAACAGCTTGCT
1834_Phospo_Fwd	TCAAATCAATCATAGGAGGGCTAAAATGACATTTCTATTAAGTTCTCAAAATG
1834_Phospho_Rec	GTAATTGCATTGGCTCCTTTGTATCTTACTTCAGCAAGCGCAAAACTTT
1834_HopA1_Fwd	GATACAAAGGAGCCAATGCAATTACTAGATTCGCCTACA
1834_HopA1_Rev	TGTCGACGGAGCTCGAATTCGGATCTCACTGTGAATCGGATAAATCTAAA
1834_PP_Fwd	CGGCCTGGTGCCGCGCGGCAGCCATATGTCTATTTATGTCGATTACATCA
1834_PP_Rev	TTAGAGGCCCCAAGGGGTTATGCTAAACAAGCGGAAAAAGGAGATACCAA
T7_Weak_Fwd	CGGCCTGGTGCCGCGCGGCAGCCATATGTCTATTTATGTCGATTACATCA
T7_Weak_Rev	TTAGAGGCCCCAAGGGGTTATGCTAAACAAGCGGAAAAAGGAGATACCAA
1834_ATP_Fwd	CCTCAACCAACAACAACTGCCTCAAGTTGCTTGACCTGACATTGTAATAGCCACCAAAA
1834_ATP_Rev	AGAACAGACAATGAATTAAAAGATGGATTTGCCATTAGGAACCTCCTTCGCTGGTTTA
1834_PPI_Fwd	ATGGCAAAAACTTTTAATTTTTCCGC
1834_PPI_Rev	GCAGCGGTTTCTTTACCAGACTCGATTAAGTAGGTTCTAGCAAGAGCC
1834_HylD_Fwd	TTAAGAAATTGCAAAAAGGCGGTTTAGATCTTTAGCCTGACATTGTAATAGCCACCAA
1834_HylD_Rec	ATGTCTTCAGCGGAAAAATTAAAAGTTTTTGCCATGGAAAACCTCCTTCGATTTCAGAA
67_PP_Fwd	ATTTTGTTTAACTTTAAGAAGGAGATATACATGGGTGTTTCCATTTCTACCCT
67_PP_Rev	GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTTAGCAGTAGGAGCCACAACC
67_Tailoring_Fwd	ATTTTGTTTAACTTTAATAAGGAGATATACATGGTATTTATATTAAGTTATCACAATGCT
67_Tailoring_Rev	CGACCTGCAGGCGCGCCGAGCTCGAATTCGTCACGGATTTAAAAAGTCTAAATGCTCA

7.8 Plasmids

Table 7.15: Table of plasmids used and plasmids generated for the experiments within this thesis.

Plasmids	**Use**
pCDF-DUET_cyclase	Validation of scytoscalarol BGC
pACYC-DUET_synthase	Validation of scytoscalarol BGC
pCDF-DUET	Protein expression vector
pACYC-DUET	Protein expression vector
pET-28a(+)	Protein expression vector
pME6032	Tetracycline resistant vector
pYE_WP_004941318	Terpene precursor supply gene
67_tailoring_pCDF	Attempted heterologous expression
67_PP_pET	Attempted heterologous expression
1834_pET_express	Attempted heterologous expression

References

1 M. I. Hutchings, A. W. Truman and B. Wilkinson, *Curr. Opin. Microbiol.*, 2019, **51**, 72–80.

2 E. P. Abraham, E. Chain, C. M. Fletcher, A. D. Gardner, N. G. Heatley, M. A. Jennings and H. W. Florey, *The Lancet*, 1941, **238**, 177–189.

3 A. Fleming, *The Lancet*, 1943, **242**, 434–438.

4 D. C. Hodgkin, *Adv. Sci.*, 1949, **6**, 85–89.

5 E. Prohaska, *Arzneimittelforschung.*, 1961, **11**, 400–402.

6 C. E. Roberts, J. D. Allen and W. M. Kirby, *Clin. Pharmacol. Ther.*, 1961, **2**, 70–79.

7 H. Eagle, *J. Bacteriol.*, 1946, **52**, 81–88.

8 S. A. Waksman, A. Schatz and D. M. Reynolds, *Ann. N. Y. Acad. Sci.*, 2010, **1213**, 112–124.

9 J. Davies, *Can. J. Infect. Dis. Med. Microbiol.*, 2006, **17**, 287–290.

10 E. M. Darby, E. Trampari, P. Siasat, M. S. Gaya, I. Alav, M. A. Webber and J. M. A. Blair, *Nat. Rev. Microbiol.*, 2023, **21**, 280–295.

11 T. M. Embley and E. Stackebrandt, *Annu. Rev. Microbiol.*, 1994, **48**, 257–289.

12 G. M. Eliopoulos, S. Willey, E. Reiszner, P. G. Spitzer, G. Caputo and R. C. Moellering, *Antimicrob. Agents Chemother.*, 1986, **30**, 532–535.

13 L. Katz and R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 155–177.

14 J. W.-H. Li and J. C. Vederas, *Science*, 2009, **325**, 161–165.

15 S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill and D. A. Hopwood, *Nature*, 2002, **417**, 141–146.

16 K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W. W. Metcalf, E. J. N. Helfrich, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2023, **51**, W46–W50.

17 J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn and A. Bateman, *Nucleic Acids Res.*, 2021, **49**, D412–D419.

18 D. H. Haft, J. D. Selengut, R. A. Richter, D. Harkins, M. K. Basu and E. Beck, *Nucleic Acids Res.*, 2013, **41**, D387–D395.

19 I. Letunic, S. Khedkar and P. Bork, *Nucleic Acids Res.*, 2021, **49**, D458–D460.

20 A. J. van Heel, A. de Jong, C. Song, J. H. Viel, J. Kok and O. P. Kuipers, *Nucleic Acids Res.*, 2018, **46**, W278–W281.

21 B. R. Terlouw, K. Blin, J. C. Navarro-Muñoz, N. E. Avalon, M. G. Chevrette, S. Egbert, S. Lee, D. Meijer, M. J. J. Recchia, Z. L. Reitz, J. A. van Santen, N. Selem-Mojica, T. Tørring, L. Zaroubi, M. Alanjary, G. Aleti, C. Aguilar, S. A. A. Al-Salihi, H. E. Augustijn, J. A. Avelar-Rivas, L. A. Avitia-Domínguez, F. Barona-Gómez, J. Bernaldo-Agüero, V. A. Bielinski, F. Biermann, T. J. Booth, V. J. Carrion Bravo, R. Castelo-Branco, F. O. Chagas, P. Cruz-Morales, C. Du, K. R. Duncan, A. Gavriilidou, D. Gayrard, K. Gutiérrez-García, K. Haslinger, E. J. N. Helfrich, J. J. J. van der Hooft, A. P. Jati, E. Kalkreuter, N. Kalyvas, K. B. Kang, S. Kautsar, W. Kim, A. M. Kunjapur, Y.-X. Li, G.-M. Lin, C. Loureiro, J. J. R. Louwen, N. L. L. Louwen, G. Lund, J. Parra, B. Philmus, B. Pourmohsenin, L. J. U. Pronk, A. Rego, D. A. B. Rex, S. Robinson, L. R. Rosas-Becerra, E. T. Roxborough, M. A. Schorn, D. J. Scobie, K. S. Singh, N. Sokolova, X. Tang, D. Udwary, A. Vigneshwari, K. Vind, S. P. J. M. Vromans, V. Waschulin, S. E. Williams, J. M. Winter, T. E. Witte, H. Xie, D. Yang, J. Yu, M. Zdouc, Z. Zhong, J. Collemare, R. G. Linington, T. Weber and M. H. Medema, *Nucleic Acids Res.*, 2023, **51**, D603–D610.

22 J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L. T. D. Cappelini, A. W. Goering, R. J. Thomson, W. W. Metcalf, N. L. Kelleher, F. Barona-Gomez and M. H. Medema, *Nat. Chem. Biol.*, 2020, **16**, 60–68.

23 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J.

Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828–837.

24 J. A. van Santen, E. F. Poynton, D. Iskakova, E. McMann, T. A. Alsup, T. N. Clark, C. H. Fergusson, D. P. Fewer, A. H. Hughes, C. A. McCadden, J. Parra, S. Soldatou, J. D. Rudolf, E. M.-L. Janssen, K. R. Duncan and R. G. Linington, *Nucleic Acids Res.*, 2022, **50**, D1317–D1323.

25 M. R. Jones, E. Pinto, M. A. Torres, F. Dörr, H. Mazur-Marzec, K. Szubert, L. Tartaglione, C. Dell'Aversano, C. O. Miles, D. G. Beach, P. McCarron, K. Sivonen, D. P. Fewer, J. Jokela and E. M.-L. Janssen, *Water Res.*, 2021, **196**, 117017.

26 V. Vannucchi, *Riv. Crit. Clin. Med.*, 1952, **52**, 128–137.

27 O. Tacar, P. Sriamornsak and C. R. Dass, *J. Pharm. Pharmacol.*, 2013, **65**, 157–170.

28 A. Nivina, K. P. Yuet, J. Hsu and C. Khosla, *Chem. Rev.*, 2019, **119**, 12524–12546.

29 J. Wang, R. Zhang, X. Chen, X. Sun, Y. Yan, X. Shen and Q. Yuan, *Microb. Cell Factories*, 2020, **19**, 110.

30 H. L. Robertsen and E. M. Musiol-Kroll, *Antibiotics*, 2019, **8**, 157.

31 T. F. Paine, H. S. Collins and M. Finland, *J. Bacteriol.*, 1948, **56**, 489–497.

32 B. R. Miller and A. M. Gulick, *Methods Mol. Biol. Clifton NJ*, 2016, **1401**, 3–29.

33 R. D. Süssmuth and A. Mainz, *Angew. Chem. Int. Ed.*, 2017, **56**, 3770–3821.

34 M. H. Mccormick, J. M. Mcguire, G. E. Pittenger, R. C. Pittenger and W. M. Stark, *Antibiot. Annu.*, 1955, **3**, 606–611.

35 D. H. Moore, J. A. Blessing, C. Dunton, R. E. Buller and G. C. Reid, *Gynecol. Oncol.*, 1999, **75**, 473–475.

36 J. S. Dickschat, *Angew. Chem. Int. Ed Engl.*, 2019, **58**, 15964–15977.

37 J. D. Rudolf, T. A. Alsup, B. Xu and Z. Li, *Nat. Prod. Rep.*, 2021, **38**, 905–980.

38 A. Masyita, R. M. Sari, A. D. Astuti, B. Yasir, N. R. Rumata, T. B. Emran, F. Nainu and J. Simal-Gandara, *Food Chem. X*, , DOI:10.1016/j.fochx.2022.100217.

39 W. Yang, X. Chen, Y. Li, S. Guo, Z. Wang and X. Yu, *Nat. Prod. Commun.*, 2020, **15**, 1934578X20903555.

40 P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K.-D. Entian, M. A. Fischbach, J. S. Garavelli, U. Göransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Müller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. T. Reaney, S. Rebuffat, R. P. Ross, H.-G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Süssmuth, J. R. Tagg, G.-L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, *Nat. Prod. Rep.*, 2013, **30**, 108–160.

41 A. D. Moffat, J. Santos-Aberturas, G. Chandra and A. W. Truman, in *Antimicrobial Therapies: Methods and Protocols*, eds. C. Barreiro and J.-L. Barredo, Springer US, New York, NY, 2021, pp. 227–246.

42 S. Luo and S. H. Dong, *Molecules*, 2019, 24.

43 C. J. Schwalen, G. A. Hudson, S. Kosol, N. Mahanta, G. L. Challis and D. A. Mitchell, *J. Am. Chem. Soc.*, 2017, **139**, 18154–18157.

44 P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S.

Garavelli, U. Göransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Müller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. T. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Süssmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. Van Der Donk, *Nat. Prod. Rep.*, 2013, 30, 108–160.

45 H. J. Sofia, G. Chen, B. G. Hetzler, J. F. Reyes-Spindola and N. E. Miller, *Nucleic Acids Res.*, 2001, **29**, 1097–1107.

46 A. M. Kloosterman, K. E. Shelton, G. P. van Wezel, M. H. Medema and D. A. Mitchell, *mSystems*, , DOI:10.1128/msystems.00267-20.

47 K. E. Shelton and D. A. Mitchell, *Methods Enzymol.*, 2023, **679**, 191–233.

48 M. Montalbán-López, T. A. Scott, S. Ramesh, I. R. Rahman, A. J. van Heel, J. H. Viel, V. Bandarian, E. Dittmann, O. Genilloud, Y. Goto, M. J. Grande Burgos, C. Hill, S. Kim, J. Koehnke, J. A. Latham, A. J. Link, B. Martínez, S. K. Nair, Y. Nicolet, S. Rebuffat, H.-G. Sahl, D. Sareen, E. W. Schmidt, L. Schmitt, K. Severinov, R. D. Süssmuth, A. W. Truman, H. Wang, J.-K. Weng, G. P. van Wezel, Q. Zhang, J. Zhong, J. Piel, D. A. Mitchell, O. P. Kuipers and W. A. van der Donk, *Nat. Prod. Rep.*, 2021, **38**, 130–239.

49 N. Leikoski, L. Liu, J. Jokela, M. Wahlsten, M. Gugger, A. Calteau, P. Permi, C. A. Kerfeld, K. Sivonen and D. P. Fewer, *Chem. Biol.*, 2013, **20**, 1033–1043.

50 M. A. Ortega and W. A. Van Der Donk, *Cell Chem. Biol.*, 2016, 23, 31–44.

51 J. A. McIntosh, Z. Lin, M. D. B. Tianero and E. W. Schmidt, *ACS Chem. Biol.*, 2013, **8**, 877–883.

52 M. Y. Phyo, C. Y. G. Ding, H. C. Goh, J. X. Goh, J. F. M. Ong, S. H. Chan, P. Y. M. Yung, H. Candra and L. T. Tan, *J. Nat. Prod.*, 2019, **82**, 3482–3488.

53 L. M. Repka, J. R. Chekan, S. K. Nair and W. A. van der Donk, *Chem. Rev.*, 2017, **117**, 5457–5520.

54 E. Gross and J. L. Morell, *J. Am. Chem. Soc.*, 1971, 93, 4634–4635.

55 N. Schnell, K. D. Entian, U. Schneider, F. Götz, H. Zähner, R. Kellner and G. Jung, *Nature*, 1988, **333**, 276–278.

56 T. H. Eyles, N. M. Vior, R. Lacret and A. W. Truman, *Chem. Sci.*, **12**, 7138–7150.

57 K. P. Yan, Y. Li, S. Zirah, C. Goulard, T. A. Knappe, M. A. Marahiel and S. Rebuffat, *ChemBioChem*, 2012, **13**, 1046–1052.

58 A. A. Bowers, C. T. Walsh and M. G. Acker, *J. Am. Chem. Soc.*, 2010, **132**, 12182–12184.

59 Y. Yu, L. Duan, Q. Zhang, R. Liao, Y. Ding, H. Pan, E. Wendt-Pienkowski, G. Tang, B. Shen and W. Liu, *ACS Chem. Biol.*, 2009, **4**, 855–864.

60 Y. Fan, H. Chen, N. Mu, W. Wang, K. Zhu, Z. Ruan and S. Wang, *Bioorg. Med. Chem.*, , DOI:10.1016/j.bmc.2020.115970.

61 S. Mazard, A. Penesyan, M. Ostrowski, I. T. Paulsen and S. Egan, *Mar. Drugs*, , DOI:10.3390/md14050097.

62 P. N. Leão, N. Engene, A. Antunes, W. H. Gerwick and V. Vasconcelos, *Nat. Prod. Rep.*, 2012, **29**, 372–391.

63 J. Rikkinen, *Biodivers. Conserv.*, 2015, **24**, 973–993.

64 A. M. Thawabteh, H. A. Naseef, D. Karaman, S. A. Bufo, L. Scrano and R. Karaman, *Toxins*, 2023, **15**, 582.

65 N. J. Smucker, J. J. Beaulieu, C. T. Nietch and J. L. Young, *Glob. Change Biol.*, 2021, **27**, 2507–2519.

66 V. I. Kapkov, S. G. Vasilieva and E. S. Lobakova, *Mosc. Univ. Biol. Sci. Bull.*, 2019, **74**, 15–20.

67 G. Gao, Y. Wang, H. Hua, D. Li and C. Tang, *Mar. Drugs*, 2021, **19**, 363.

68 D. G. Nagle and V. J. Paul, *J. Exp. Mar. Biol. Ecol.*, 1998, **225**, 29–38.

69 S. C, *Curr Pharm. Des.*, 2001, **13**, 1259–1277.

70 S. Vijayakumar and M. Menakha, *J. Acute Med.*, 2015, **5**, 15–23.

71 M. Kobayashi, T. Natsume, S. Tamaoki, J. Watanabe, H. Asano, T. Mikami, K. Miyasaka, K. Miyazaki, M. Gondo, K. Sakakibara and S. Tsukagoshi, *Jpn. J. Cancer Res.*, 1997, **88**, 316–327.

72 Q. Yan, Y. Wang, W. Zhang and Y. Li, *Mar. Drugs*, 2016, **14**, 85.

73 M. Akaiwa, J. Dugal-Tessier and B. A. Mendelsohn, *Chem. Pharm. Bull. (Tokyo)*, 2020, **68**, 201–211.

74 Z. Fu, C. Gao, T. Wu, L. Wang, S. Li, Y. Zhang and C. Shi, *iScience*, 2023, **26**, 107778.

75 R. Carpine and S. Sieber, *Curr. Res. Biotechnol.*, 2021, **3**, 65–81.

76 M. Preisitsch, K. Harmrolfs, H. T. Pham, S. E. Heiden, A. Füssel, C. Wiesner, A. Pretsch, M. Swiatecka-Hagenbruch, T. H. Niedermeyer, R. Müller and S. Mundt, *J. Antibiot. (Tokyo)*, 2015, **68**, 165–177.

77 R. Montaser, V. J. Paul and H. Luesch, *Phytochemistry*, 2011, **72**, 2068–2074.

78 A. Allsopp, *Nature*, 1968, **220**, 810.

79 C. P. Wolk, A. Ernst and J. Elhai, in *The Molecular Biology of Cyanobacteria*, ed. D. A. Bryant, Springer Netherlands, Dordrecht, 1994, pp. 769–823.

80 S. Halary, C. Duval, B. Marie, C. Bernard, B. Piquet, O. Gros, M.-L. Bourguet-Kondracki and S. Duperron, *FEMS Microbes*, 2024, **5**, xtad024.

81 D. Sen, M. M. Chandrababunaidu, D. Singh, N. Sanghi, A. Ghorai, G. P. Mishra, M. Madduluri, S. P. Adhikary and S. Tripathy, *Genome Announc.*, 2015, **3**, e00009-15.

82 R. Lücking, J. D. Lawrey, M. Sikaroodi, P. M. Gillevet, J. L. Chaves, H. J. M. Sipman and F. Bungartz, *Am. J. Bot.*, 2009, **96**, 1409–1418.

83 A. Giraldo-Silva, V. M. C. Fernandes, J. Bethany and F. Garcia-Pichel, *Microorganisms*, 2020, **8**, 397.

84 I. Becerra-Absalón, M. Á. Muñoz-Martín, G. Montejano and P. Mateo, *Front. Microbiol.*, 2019, **10**, 937.

85 T. Dagan, M. Roettger, K. Stucken, G. Landan, R. Koch, P. Major, S. B. Gould, V. V. Goremykin, R. Rippka, N. Tandeau de Marsac, M. Gugger, P. J. Lockhart, J. F. Allen, I. Brune, I. Maus, A. Pühler and W. F. Martin, *Genome Biol. Evol.*, 2013, **5**, 31–44.

86 C. M. Crnkovic, J. Braesel, A. Krunic, A. S. Eustáquio and J. Orjala, *Chembiochem Eur. J. Chem. Biol.*, 2020, **21**, 845–852.

87 C. P. Mason, K. R. Edwards, R. E. Carlson, J. Pignatello, F. K. Gleason and J. M. Wood, *Science*, 1982, **215**, 400–402.

88 S. Carmeli, R. E. Moore and G. M. Patterson, *J. Nat. Prod.*, 1990, **53**, 1533–1542.

89 R. Nakamura, K. Tanino and M. Miyashita, *Org. Lett.*, 2003, **5**, 3579–3582.

90 K. Delawská, P. Divoká, D. Sedlák, M. Kuzma, K. Saurav, M. Macho, G. Steinbach and P. Hrouzek, *ChemBioChem*, 2022, **23**, e202100489.

91 V. Reshef, E. Mizrachi, T. Maretzki, C. Silberstein, S. Loya, A. Hizi and S. Carmeli, *J. Nat. Prod.*, 1997, **60**, 1251–1260.

92 U. Matern, L. Oberer, R. A. Falchetto, M. Erhard, W. A. König, M. Herdman and J. Weckesser, *Phytochemistry*, 2001, **58**, 1087–1095.

93 U. Matern, L. Oberer, M. Erhard, M. Herdman and J. Weckesser, *Phytochemistry*, 2003, **64**, 1061–1067.

94 H. R. Bokesch, B. R. O'Keefe, T. C. McKee, L. K. Pannell, G. M. L. Patterson, R. S. Gardella, R. C. Sowder, J. Turpin, K. Watson, R. W. Buckheit and M. R. Boyd, *Biochemistry*, 2003, **42**, 2578–2584.

95 S. H. Shim, G. Chlipala and J. Orjala, *J. Microbiol. Biotechnol.*, 2008, **18**, 1655–1658.

96 R. P. Rastogi, R. R. Sonani and D. Madamwar, *Appl. Biochem. Biotechnol.*, 2015, **176**, 1551–1563.

97 An Antimicrobial Guanidine-Bearing Sesterterpene from the Cultured Cyanobacterium Scytonema sp. | Journal of Natural Products, https://pubs.acs.org/doi/10.1021/np900288x, (accessed 15 May 2024).

98 A. H. Cabanillas, V. Tena Pérez, S. Maderuelo Corral, D. F. Rosero Valencia, A. Martel Quintana, M. Ortega Doménech and Á. Rumbero Sánchez, *J. Nat. Prod.*, 2018, **81**, 410–413.

99 A. Krunic, A. Vallat, S. Mo, D. D. Lantvit, S. M. Swanson and J. Orjala, *J. Nat. Prod.*, 2010, **73**, 1927–1932.

100 L. M. P. Heinilä, D. P. Fewer, J. K. Jokela, M. Wahlsten, A. Jortikka and K. Sivonen, *Front. Microbiol.*, 2020, **11**, 578878.

101 K. Heck, G. S. Machineski, D. O. Alvarenga, M. G. M. V. Vaz, A. de M. Varani and M. F. Fiore, *J. Microbiol. Methods*, 2016, **129**, 55–60.

102 H. F, W. Sa, B. P, W. W and G. S, *Toxicon Off. J. Int. Soc. Toxinology*, , DOI:10.1016/j.toxicon.2015.07.014.

103 N. De Maio, L. P. Shaw, A. Hubbard, S. George, N. D. Sanderson, J. Swann, R. Wick, M. AbuOun, E. Stubberfield, S. J. Hoosdally, D. W. Crook, T. E. A. Peto, A. E. Sheppard, M. J. Bailey, D. S. Read, M. F. Anjum, A. S. Walker and N. Stoesser, *Microb. Genomics*, 2019, **5**, e000294.

104 R. R. Wick, L. M. Judd, C. L. Gorrie and K. E. Holt, *PLoS Comput. Biol.*, 2017, **13**, e1005595.

105 S. Gao, W.-K. Sung and N. Nagarajan, *J. Comput. Biol.*, 2011, **18**, 1681–1691.

106 D. D. Kang, F. Li, E. Kirton, A. Thomas, R. Egan, H. An and Z. Wang, *PeerJ*, 2019, **7**, e7359.

107 D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz and G. W. Tyson, *Genome Res.*, 2015, **25**, 1043–1055.

108 A. Erlacher, T. Cernava, M. Cardinale, J. Soh, C. W. Sensen, M. Grube and G. Berg, *Front. Microbiol.*, , DOI:10.3389/fmicb.2015.00053.

109 Q. Wu, X. Zhang, S. Jia, J. Li and P. Li, *J. Freshw. Ecol.*, 2019, **34**, 663–673.

110 K. R. Moore, C. Magnabosco, L. Momper, D. A. Gold, T. Bosak and G. P. Fournier, *Front. Microbiol.*, , DOI:10.3389/fmicb.2019.01612.

111 M. Alanjary, K. Steinke and N. Ziemert, *Nucleic Acids Res.*, 2019, **47**, W276–W282.

112 K. D. Pruitt, G. R. Brown, S. M. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, C. M. Farrell, J. Hart, M. J. Landrum, K. M. McGarvey, M. R. Murphy, N. A. O'Leary, S. Pujar, B. Rajput, S. H. Rangwala, L. D. Riddick, A. Shkeda, H. Sun, P. Tamez, R. E. Tully, C. Wallin, D. Webb, J. Weber, W. Wu, M. DiCuccio, P. Kitts, D. R. Maglott, T. D. Murphy and J. M. Ostell, *Nucleic Acids Res.*, 2014, **42**, D756–D763.

113 A. Undabarrena, R. Valencia, A. Cumsille, L. Zamora-Leiva, E. Castro-Nallar, F. Barona-Gomez and B. Cámara, *Microb. Genomics*, 2021, **7**, 000621.

114 A. J. Page, C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden, M. Fookes, D. Falush, J. A. Keane and J. Parkhill, *Bioinformatics*, 2015, **31**, 3691–3693.

115 B. R. McDonald and C. R. Currie, *mBio*, 2017, **8**, 10.1128/mbio.00644-17.

116 A. C. E. Darling, B. Mau, F. R. Blattner and N. T. Perna, *Genome Res.*, 2004, **14**, 1394–1403.

117 H. B. Bode, B. Bethe, R. Höfs and A. Zeeck, *ChemBioChem*, 2002, **3**, 619–627.

118 C. Pinedo-Rivilla, J. Aleu and R. Durán-Patrón, *Mar. Drugs*, 2022, **20**, 84.

119 R. Pan, X. Bai, J. Chen, H. Zhang and H. Wang, *Front. Microbiol.*, 2019, **10**, 294.

120 I. Ra, H. E, A. Jh and E. Hc, *J. Appl. Phycol.*, , DOI:10.1007/s10811-015-0631-4.

121 S. Sánchez, A. Chávez, A. Forero, Y. García-Huante, A. Romero, M. Sánchez, D. Rocha, B. Sánchez, M. Ávalos, S. Guzmán-Trampe, R. Rodríguez-Sanoja, E. Langley and B. Ruiz, *J. Antibiot. (Tokyo)*, 2010, **63**, 442–459.

122 S. Romano, S. A. Jackson, S. Patry and A. D. W. Dobson, *Mar. Drugs*, 2018, **16**, 244.

123 K. Kurosawa, I. Ghiviriga, T. G. Sambandan, P. A. Lessard, J. E. Barbara, C. Rha and A. J. Sinskey, *J. Am. Chem. Soc.*, 2008, **130**, 1126–1127.

124 M. Frank, F. C. Özkaya, W. E. G. Müller, A. Hamacher, M. U. Kassack, W. Lin, Z. Liu and P. Proksch, *Mar. Drugs*, 2019, **17**, 99.

125 N. M. Tran-Cong, A. Mándi, T. Kurtán, W. E. G. Müller, R. Kalscheuer, W. Lin, Z. Liu and P. Proksch, *RSC Adv.*, 2019, **9**, 27279–27288.

126 B. Rp, M. Mc, C. A, S. S and de C. Jc, *Biotechnol. Bioeng.*, , DOI:10.1002/bit.21771.

127 W. Dg, R. Be, D. S, H. Rd, S. Df and G. Ss, *Trends Microbiol.*, , DOI:10.1016/j.tim.2018.11.002.

128 P. Casanova-Ferrer, J. Muñoz-García and S. Ares, *Front. Cell Dev. Biol.*, 2022, **10**, 959468.

129 T. Ott, J. T. van Dongen, C. Gu¨nther, L. Krusell, G. Desbrosses, H. Vigeolas, V. Bock, T. Czechowski, P. Geigenberger and M. K. Udvardi, *Curr. Biol.*, 2005, **15**, 531–535.

130 A. U. Igamberdiev and R. D. Hill, in *Methods in Enzymology*, ed. R. K. Poole, Academic Press, 2008, vol. 436, pp. 379–391.

131 M. V. Thorsteinsson, D. R. Bevan, R. E. Ebel, R. E. Weber and M. Potts, *Biochim. Biophys. Acta*, 1996, **1292**, 133–139.

132 A. Parsaeimehr, I. I. Ahmed, M. L. K. Deumaga, B. Hankoua and G. Ozbay, *Algal Res.*, 2022, **66**, 102805.

133 A. Y. Alwali and E. I. Parkinson, *J. Ind. Microbiol. Biotechnol.*, 2023, **50**, kuad019.

134 A. P. Tyurin, V. A. Alferova and V. A. Korshun, *Microorganisms*, 2018, **6**, 52.

135 N. Ruiz, B. Falcone, D. Kahne and T. J. Silhavy, *Cell*, 2005, **121**, 307–317.

136 C.-I. Cheigh and Y.-R. Pyun, *Biotechnol. Lett.*, 2005, **27**, 1641–1648.

137 D. Kallifidas, D. Dhakal, M. Chen, Q.-Y. Chen, S. Kokkaliari, N. A. Colon Rosa, R. Ratnayake, S. D. Bruner, V. J. Paul, Y. Ding and H. Luesch, *Org. Lett.*, 2024, **26**, 1321–1325.

138 S. Colas, B. Marie, E. Lance, C. Quiblier, H. Tricoire-Leignel and C. Mattei, *Environ. Res.*, 2021, **193**, 110590.

139 M. T. Runnegar, C. Xie, B. B. Snider, G. A. Wallace, S. M. Weinreb and J. Kuhlenkamp, *Toxicol. Sci.*, 2002, **67**, 81–87.

140 C. Z. To and A. K. Bhunia, *Front. Microbiol.*, , DOI:10.3389/fmicb.2019.00949.

141 B. Rgs, B. Af, T. M and R. Jpg, *Nat. Prod. Rep.*, , DOI:10.1039/c7np00037e.

142 I. Ohtani, R. E. Moore and M. T. C. Runnegar, *J. Am. Chem. Soc.*, 1992, **114**, 7941–7942.

143 G. Shalev-Alon, A. Sukenik, O. Livnah, R. Schwarz and A. Kaplan, *FEMS Microbiol. Lett.*, 2002, **209**, 87–91.

144 B. Hu, H. Yu, J. Zhou, J. Li, J. Chen, G. Du, S. Y. Lee and X. Zhao, *Adv. Sci.*, 2023, **10**, 2205580.

145 P. Moosmann, F. Ecker, S. Leopold-Messer, J. K. B. Cahn, C. L. Dieterich, M. Groll and J. Piel, *Nat. Chem.*, 2020, **12**, 968.

146 P. R. Ortiz de Montellano, *Chem. Rev.*, 2010, **110**, 932.

147 K. Fiege, C. J. Querebillo, P. Hildebrandt and N. Frankenberg-Dinkel, *Biochemistry*, 2018, **57**, 2747–2755.

148 A. Hou and J. S. Dickschat, *Angew. Chem. Int. Ed.*, 2020, **59**, 19961–19965.

149 V. J. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman and J. D. Keasling, *Nat. Biotechnol.*, 2003, **21**, 796–802.

150 R. Wilton, A. J. Ahrendt, S. Shinde, D. J. Sholto-Douglas, J. L. Johnson, M. B. Brennan and K. M. Kemner, *Front. Plant Sci.*, 2018, **8**, 2242.

151 C. Robichon, J. Luo, T. B. Causey, J. S. Benner and J. C. Samuelson, *Appl. Environ. Microbiol.*, 2011, **77**, 4634–4646.

152 T.-D. Nguyen, S. Riordan-Short, T.-T. T. Dang, R. O'Brien and M. Noestheden, *ACS Omega*, 2020, **5**, 5565–5573.

153 A. E. Kadjo and A. S. Eustáquio, *J. Ind. Microbiol. Biotechnol.*, 2023, **50**, kuad044.

154 H.-S. Kang and E.-S. Kim, *Curr. Opin. Biotechnol.*, 2021, **69**, 118–127.

155 T. J. Harris and J. S. Emtage, *Microbiol. Sci.*, 1986, **3**, 28–31.

156 R. Jiang, S. Yuan, Y. Zhou, Y. Wei, F. Li, M. Wang, B. Chen and H. Yu, *Biotechnol. Adv.*, 2024, **75**, 108417.

157 A. Taton, A. Ecker, B. Diaz, N. A. Moss, B. Anderson, R. Reher, T. F. Leão, R. Simkovsky, P. C. Dorrestein, L. Gerwick, W. H. Gerwick and J. W. Golden, *ACS Synth. Biol.*, 2020, **9**, 3364–3377.

158 J. Svoboda, B. Cisneros and B. Philmus, *Synth. Biol. Oxf. Engl.*, 2021, **6**, ysab019.

159 S. H. Kim, S. I. Kwon, D. Saha, N. C. Anyanwu and W. Gassmann, *Plant Physiol.*, 2009, **150**, 1723–1732.

160 J. P. Deisinger, M. Arts, I. Kotsogianni, J.-S. Puls, F. Grein, F. J. Ortiz-López, N. I. Martin, A. Müller, O. Genilloud and T. Schneider, *iScience*, 2023, **26**, 106394.

161 F. J. Ortiz-López, D. Carretero-Molina, M. Sánchez-Hidalgo, J. Martín, I. González, F. Román-Hurtado, M. de la Cruz, S. García-Fernández, F. Reyes, J. P. Deisinger, A. Müller, T. Schneider and O. Genilloud, *Angew. Chem. Int. Ed.*, 2020, **59**, 12654–12658.

162 L. Frattaruolo, R. Lacret, A. R. Cappello and A. W. Truman, *ACS Chem. Biol.*, 2017, **12**, 2815–2822.

163 L. Frattaruolo, M. Fiorillo, M. Brindisi, R. Curcio, V. Dolce, R. Lacret, A. W. Truman, F. Sotgia, M. P. Lisanti and A. R. Cappello, *Cells*, 2019, **8**, 1408.

164 Z.-F. Pei, L. Zhu, R. Sarksian, W. A. van der Donk and S. K. Nair, *J. Am. Chem. Soc.*, 2022, **144**, 17549–17557.

165 N. Miguel-Vior, E. Cea, T. Eyles, G. Chandra and A. Truman, *Front. Microbiol.*, , DOI:10.3389/fmicb.2020.00495.

166 M. J. van Belkum, R. W. Worobo and M. E. Stiles, *Mol. Microbiol.*, 1997, **23**, 1293–1301.

167 S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas and H. B. Bode, *Org. Biomol. Chem.*, 2011, **9**, 3130–3132.

168 T. Nakashima and S. W. Fox, *J. Mol. Evol.*, 1980, **15**, 161–168.

169 A. P. Decker, A. F. Mechesso and G. Wang, *Int. J. Mol. Sci.*, , DOI:10.3390/ijms232112874.

170 J. Lei, L. Sun, S. Huang, C. Zhu, P. Li, J. He, V. Mackey, D. H. Coy and Q. He, *Am. J. Transl. Res.*, 2019, **11**, 3919–3931.

171 E. M. Kościuczuk, P. Lisowski, J. Jarczak, N. Strzałkowska, A. Jóźwik, J. Horbańczuk, J. Krzyżewski, L. Zwierzchowski and E. Bagnicka, *Mol. Biol. Rep.*, 2012, **39**, 10957–10970.

172 A. B. Al-Hawash, X. Zhang and F. Ma, *Gene Rep.*, 2017, **9**, 46–53.

173 S. E. Ongley, X. Bian, Y. Zhang, R. Chau, W. H. Gerwick, R. Müller and B. A. Neilan, *ACS Chem. Biol.*, 2013, **8**, 10.1021/cb400189j.

174 E. J. Kim, J. H. Lee, H. Choi, A. R. Pereira, Y. H. Ban, Y. J. Yoo, E. Kim, J. W. Park, D. H. Sherman, W. H. Gerwick and Y. J. Yoon, *Org. Lett.*, 2012, **14**, 5824–5827.

175 F. Wright and M. J. Bibb, *Gene*, 1992, **113**, 55–65.

176 S. Mordhorst, B. I. Morinaka, A. L. Vagstad and J. Piel, *Angew. Chem. Int. Ed.*, 2020, **59**, 21442–21446.

177 A. Bhushan, P. J. Egli, E. E. Peters, M. F. Freeman and J. Piel, *Nat. Chem.*, 2019, **11**, 931–939.

178 F. Hubrich, N. M. Bösch, C. Chepkirui, B. I. Morinaka, M. Rust, M. Gugger, S. L. Robinson, A. L. Vagstad and J. Piel, *Proc. Natl. Acad. Sci.*, 2022, **119**, e2113120119.

179 W. De Coster, S. D'Hert, D. T. Schultz, M. Cruts and C. Van Broeckhoven, *Bioinformatics*, 2018, **34**, 2666–2669.

180 S. Chen, Y. Zhou, Y. Chen and J. Gu, *Bioinformatics*, 2018, **34**, i884–i890.

181 T. Seemann, *Bioinformatics*, 2014, **30**, 2068–2069.

182 Roary: rapid large-scale prokaryote pan genome analysis | Bioinformatics | Oxford Academic, https://academic.oup.com/bioinformatics/article/31/22/3691/240757, (accessed 22 May 2024).

Appendices

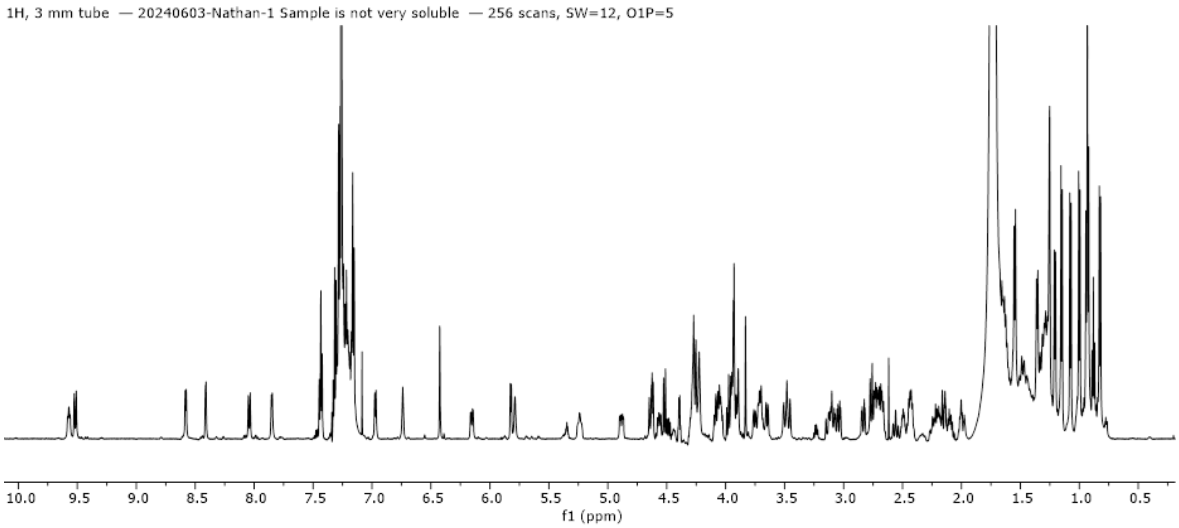

Figure 8.1: 1H NMR of antimicrobial molecule isolated from *Scytonema* sp. UTEX 1834.

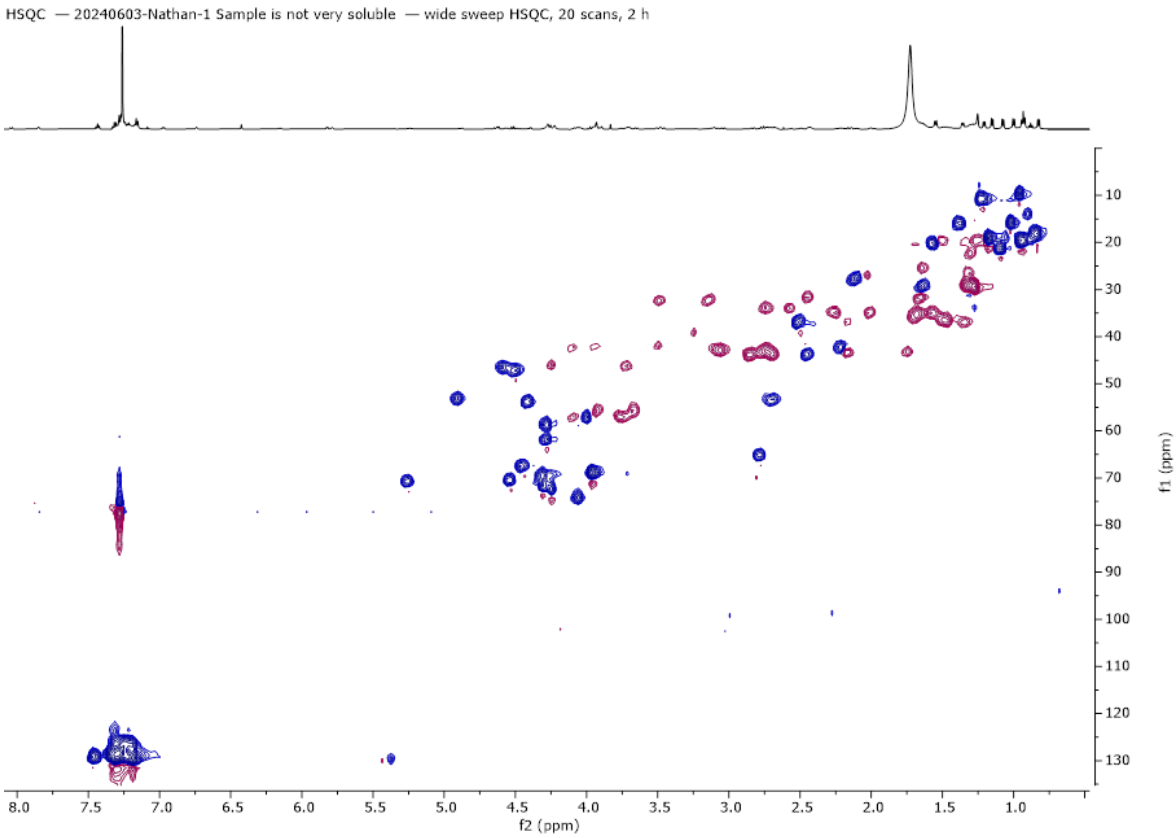

Figure 8.2: HSQC NMR of unknown molecule isolated from *Scytonema* sp. UTEX 1834.

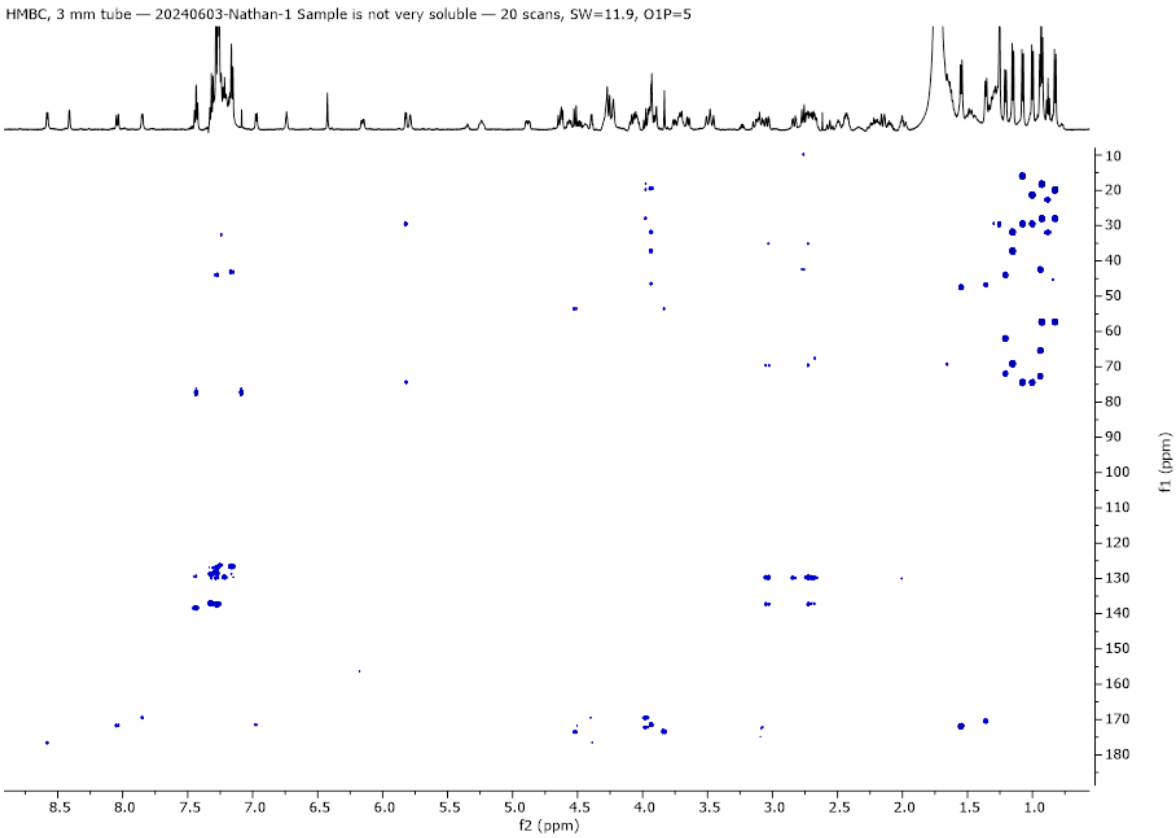

Figure 8.3: HMBC NMR for unknown antimicrobial isolated from *Scytonema* sp. UTEX 1834.