th

Ireland's Hidden Pharmacy: Deep Sea Sponges

Science Foundation Ireland funded project

Conor McKeon

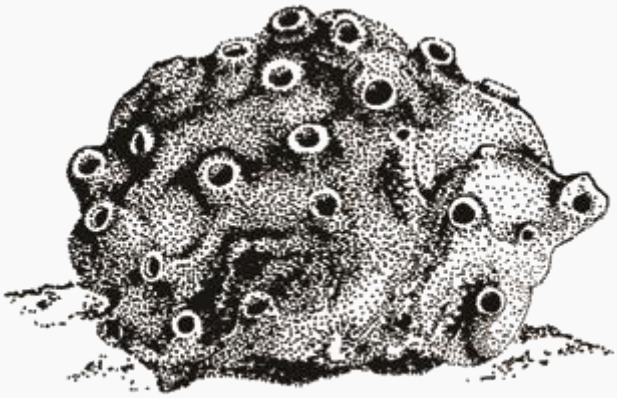
March 18, 2019

Under the supervision of Prof. Olivier P. Thomas,

Guidance and review of Dr. Patrick O'Leary

&

Co-supervision of Sam Afoullouss



Marine Biodiscovery, School of Chemistry and Ryan Institute,
National University of Ireland Galway (NUI Galway), University Road,
H91TK33 Galway, Ireland

AFFIDAVIT

I declare that this material, which I now submit for assessment, is my own work and that any assistance I received in its preparation is fully acknowledged and disclosed in the document. To the best of my knowledge and belief, all sources have been properly acknowledged, and the assessment task contains no plagiarism. I understand that plagiarism, collusion, and/or copying are grave and serious offences and am aware that penalties could include a zero mark for this assessment, suspension or expulsion from NUI Galway. I have read the NUI Galway code of practice regarding plagiarism at www.nuigalway.ie/plagiarism. I acknowledge that this assessment submission may be transferred and stored in a database for the purposes of data-matching to help detect plagiarism. I declare that this document was prepared by me for the purpose of partial fulfilment of requirements for the programme for which I am registered with the AUA. I also declare that this assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study or another college.

Student Name _____Conor McKeon_____

Conor McKeon March 18th 2019

Student Signature Date

Acknowledgements

I wish to express my special thanks to the whole Marine Biodiscovery group for their humour and sincerity which together made this research a pleasure,

To Dr. Kevin Calabro
&
Dr. Laurence Jennings,

For offering their expertise at key stages.

To Sam Afoullouss,

For his enthusiasm and dedicated assistance throughout.

Table of Contents

List of Figures

List of Tables

Abbreviations

TLC - Thin-Layer Chromatography
GC - Gas Chromatography
HPLC - High Performance Liquid Chromatography
UV - Ultra-Violet
TFA - Trifluoroacetic Acid
NP - Natural Product
ATP - Adenosine Tri-Phosphate
MNP - Marine Natural Product
TNP - Terrestrial Natural Product
MAA - Mycosporine-like Amino Acids
SAM - S-Aadenosyl Methionine
SFI - Science Foundation Ireland
ILV - Irish Lights Vessel
RV - Research Vessel
NMR - Nuclear Magnetic Resonance
ELSD - Evaporative Light Scattering Detector
MS - Mass Spectrometry
UPLC - Ultra Performance Liquid Chromatography
GNPS - Global Natural Products Social molecular networking
DAD - Diode Array Detector
HRESIMS - High Resolution Electro-Spray Ionisation Mass Spectrometry
SOP - Standard Operating Procedure
FID - Free Induction Decay
TMS - TriMethylSilyl
COSY - Correlation Spectroscopy
HMBC - Heteronuclear Multiple Bond Coherence
HSQC - Heteronuclear Single Quantum Coherence
NOESY - Nuclear Overhauser Effect Spectroscopy
VLC - Vacuum Liquid Chromatography
ROV - Remotely Operated Vehicle
QTOF - Quadrupole Time Of Flight
SPE - Solid Phase Extraction
PTFE - PolyTetraFluoroEthylene
BEH - Bridged Ethylene Hybrid

Abstract

Evolutionary survival and defensive mechanisms of sessile marine organisms have guided the production of a huge variety of evolutionarily optimised bioactive specialized metabolites. Sponges in particular represent an abundance of underexplored biomolecules, providing almost one third of marine natural products discovered to date. The extraordinary structural complexity of these metabolites renders them immensely valuable as potential lead compounds in modern drug development. This investigation employs chemical screening and molecular netwoking to evaluate the potential of a range of deep-sea sponge samples based on chemical diversity and novelty. The aim to further the search for potentially bioactive metabolites was achieved with the isolation and characterisation of the unprecedented histidine derivative 6-methyl-hercynine from the North Atlantic deep-sea sponge *Characella pachastrelloides*. Reported also are the isolations of three known compounds: the nucleosides Adenosine and Thymidine from the *Characella* sponge & the bromotyrosine alkaloid Aplysamine 7 from a *Pseudoceratina* sp. The isolated compounds were fully characterised by 1D and 2D nuclear magnetic resonance and MS.

1. Introduction

1.1 Natural products

The molecules of life, that underpin every biological phenomenon, are as extraordinary as the animals, plants, bacteria, fungi and protists in with they are created. These biological substances are sources of endless academic and industrial interest as the intricate production of these metabolites gives us a seemingly infinite supply of novel chemical species, discovery of which would be unimaginable by synthetic routes. Natural selection is the driver of chemodiversity and this is exemplified in the variety of chemical species exhibited by an organism, and their associated ecological roles. Investigating the rationale for a chemical's biosynthesis offers clues to the ecological interplay between species and extrapolates the blueprints of evolution.

Natural Products (NPs) are, in the broadest sense, any low molecular weight substance produced by life (usually less than 1500Da). These vast arrays of compounds are produced by multiple enzyme-controlled metabolic pathways. Primary metabolites are vital for a organism to sustain life and reproduce (e.g ATP for energy transfer). The genes responsible for primary metabolite production are heavily conserved and lead to identical or similar compounds occurring in species of different phyla.[5] Specialized (secondary) metabolites on the other hand are compounds that provide a specific ecological function and allow a species to excel and thrive in its particular niche. These metabolites are integral in communication, symbiosis, competition and defence, and overcoming stressors such as parasitism, predation, pollution, radiation and climatic effects. [6]

Due to the universality of primary metabolites and their established biological capacity, they represent effectual building blocks to which specific modifications can be made to actuate specialized functions resulting in 'secondary' metabolites.[6] As a result of this process, many NPs exhibit potent bioactivities and display strong potential as new pharmaceutical agents. [7] These modified metabolites which prove beneficial to the organism are retained, and the enzymes required for the secondary biosynthesis are usually encoded as a gene cluster.[8] The distinction between primary and secondary is not indisputable, given that all metabolites have roles presumably essential for the survival of the organism, and although convenient, this distinction casts uncertainty on the importance of these specialized metabolites for those unfamiliar with natural products. The importance of specialized metabolites is starkly seen in sessile organisms which can not run to escape predation or rely on an immune system to ward off pathogens. To survive in such a dynamic environment, the sessile organism must be able to adapt its chemistry relatively quickly or face extinction.[9]

Industrial efforts in NP biodiscovery has dwindled, due in part to inefficiency and high rates of rediscovery, but academic research remains persistent. By examining biosynthesis, physico-chemical patterns and structure-activity relationships we can direct our investigations towards biologically active chemicals.[10] The ever-growing library of information coupled with increasingly advanced analytical technology is accelerating progress in the perpetual task of exploring the natural world for it's chemistry and with the search for interesting compounds unrelenting, the future of NP biodiscovery appears optimistic.

1.2 Marine natural products

The earth is undeniably an ocean planet with ocean covering 71% of the Earth's surface and with life found in every level of the water column, from the surface to the deepest abyss, it contains the majority of the available living space on the planet.[11] Exploration of this vast environnment has only recently become possible with the arrival of SCUBA and submersibles since the 1950's which enabled deep-sea sampling for the first time. The inaccessibility, coupled with the available terrestrial biodiversity meant that the marine environment went largely ignored until the race to isolate bioactive marine compounds began in the early 1960's, spurred by the realization of the pharmaceutical potential these marine derived compounds possess.[12]

It is reasonable to suppose that due to the massively different conditions in which marine natural product (MNP) and terrestrial natural product (TNP) biosynthesis occurs, there may be fundamental differences in the chemistry of marine and terrestrial metabolic products. It should therefore be asked whether metabolites which have been modified and refined for a particular biological role in the aquatic environment are likely to possess drug-like physicochemical characteristics for use in terrestrial mammals. This question has been the focus of several cheminformatic studies on a wide range of new molecular entities (NMEs) identified from marine and terrestrial sources, concluding that MNPs exhibit equal potential as drug candidates despite appreciable variances in architecture, elemental makeup and pharmacokinetic properties. [13]

The characteristics detailed in **Fig 1** were chosen from a study of 60 physicochemical properties of 151,609 TNPs and 35,883 MNPs. Based on the data, MNPs are more likely to have charge, more mass and larger rings, which allow more flexibility under pressure [14]. TNPs appear to dominate in oxygen prevalence, solubility and aromaticity. Although many surface marine organisms have developed unique UV protection with non-aromatic mycosporine-like amino acids (MAAs), UV protection and pigmentation become less important with depth, partially explaining the relative scarcity of aromatic moieties among MNPs.[15]

Despite the difficulty in sub-aqua sampling, marine biodiscovery remains promising as the aquatic domain contains vastly more biodiversity, and bizarre marine biosynthetic products such as unnatural amino acids continue to enamor researchers.

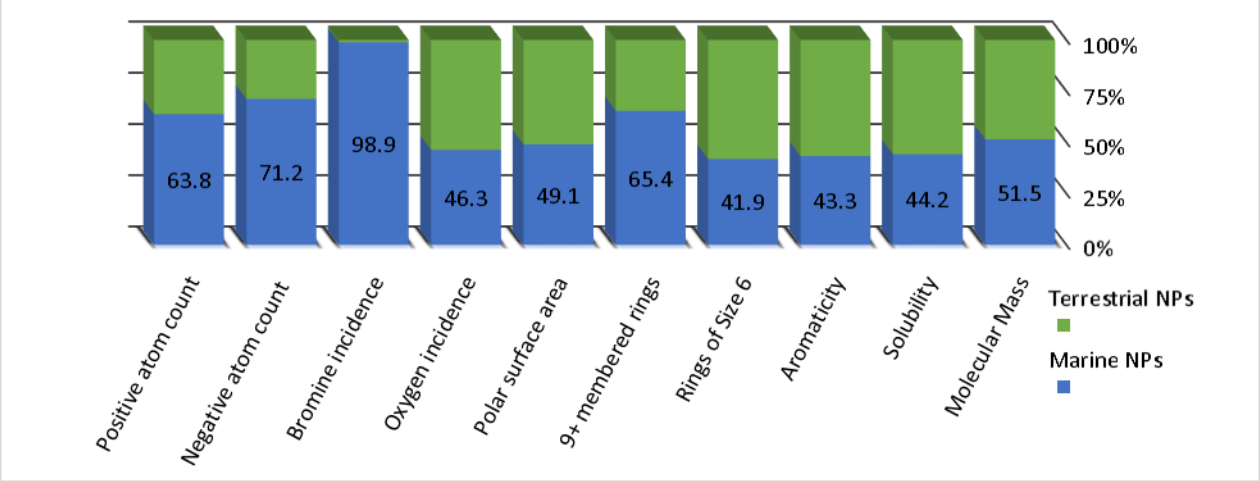


Fig 1 – Terrestrial NPs vs Marine NPs. Each bar represents the relative ratios (i.e. roughly one Br atom is present in TNPs for every 99 found in MNPs) [13]

1.3 Sponges

1.3.1 Introductory sponge biology

Sponges, the members of the phylum porifera, are the oldest metazoans (multicellular animals) still in existence. They were first to branch off the evolutionary tree from the common ancestor of all animals. As of 2019, there are 8873 living, recognized species of sponge with current trends predicting this will reach over 12000 within the century.[16] Porifera are divided into four distinct classes based predominantly on the characteristics of microgranular structural components called spicules (**Fig 2 (C)**). The shapes and chemical composition of these inorganic architectures are emblematic of poriferean class: Demospongiae (82% of all species - siliceous monaxone or tetraxone spicules), Hexactinellida (siliceous triaxone spicules) and Calcarea (calcareous spicules). The fourth class, Homoscleromorpha containing just two of the twenty three sponge families, has only recently been classified as distinct from Demospongiae based on genomic and morphological evidence. [16]

Sponges are sessile, invertebrate filter feeders with no nervous, digestive or circulatory systems, found in almost every marine environment and from epipelagic to hadalpelagic depths (0 – 6000+ m). The sponge exterior has tiny incurrent pores, called ostia, which lead to a system of internal canals lined with flagellated water pumping cells, eventually leading out to the central 'oscula' through which water and waste exit the sponge (**Fig 2 (B)**).[17] The spicule 'skeleton' supports the canal system and the connecting tissue – the 'mesohyl' in which the internal cells are suspended. Despite having a simple body plan and simplistic functions, the sponge displays enormous complexity in many areas with relatively little known about biosynthesis, cellular communication and intimate symbiotic relationships. Taxonomical identification may be carried out by analysis of spicule structure or morphological evidence as sponges vary widely in size and shape, from small cushion shapes to elaborate branching structures, from millimetre thick encrustations to sponges several metres long (**Fig 2 (A)**).

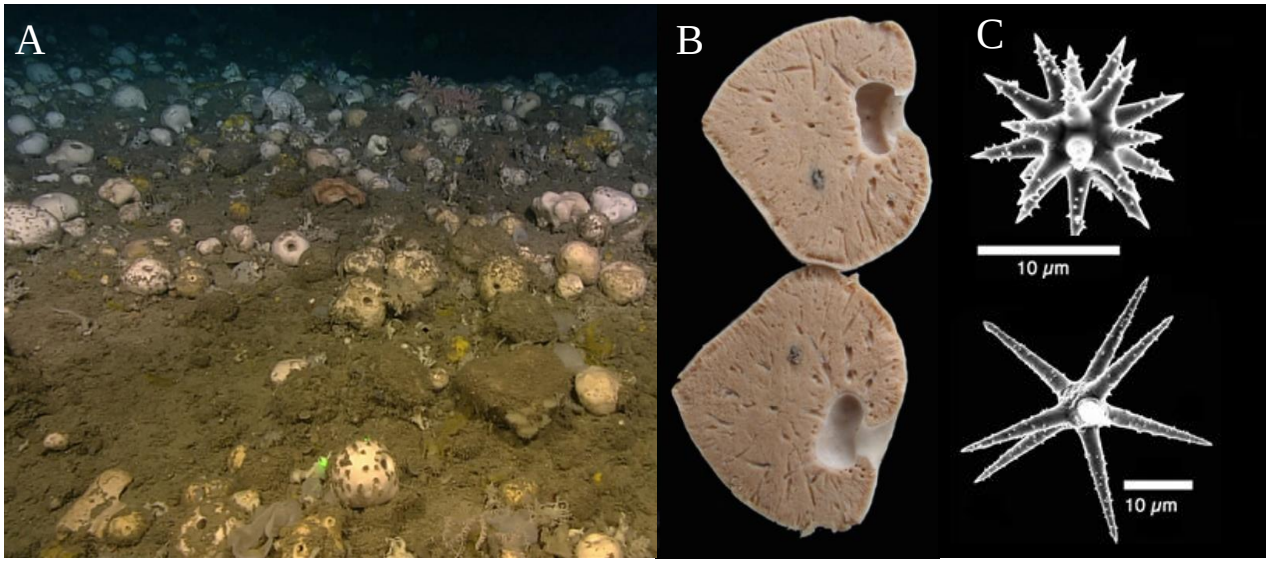


Fig 2 **(A) – *Geodia* sponge ground**. Flemish Cap 1581m.[2]
(B) –Internal morphology of cut *Geodia atlantica* specimen.[2]
(C) – Spicules of *Geodia atlantica*.[2]

1.3.2 Chemicals of combat and defence

The production of chemicals as a means to deter predation and maintain ecological eminence is paramount to the survival and development of sessile organisms such as sponges. Allelochemicals is the broad term for these metabolites which affect the health, growth or behavior of other organisms, and allelopathy may help to explain the extraordinary chemical diversity of marine sponges. [18] Competition for resources in marine benthic habitats has been a driver of evolution and diversification, and any means by which an organism can take advantage will likely lead to the detriment of its opponent.

Fouling (accumulation of microorganisms, plants or animals on the host surface) is one such way in which an organism can become impaired. Unwelcome epibionts may compete with the host for nutrients and resources, and obstruct functions such as gas exchange and waste excretion. Biosynthesized antifouling metabolites are of economical importance as organotin biocides, which have seen widespread use since the 1960s as antifoulant coatings for boats, have been linked to defomities in shellfish and have since been banned by the International Maritime Organization.[19] Some of the low-molecular weight bioactive antifoulants that have been isolated from sponges include bromopyrrole alkaloids such as sceptrin from the Carribean *Agelas conifera*, (**Fig 3 (A)**) and kalihipyran-7, the diterpene isonitrile isolated from the marine sponge *Acanthella cavernosa* collected off Yakushima Island in Japan (**Fig 3 (B)**).[20,21]

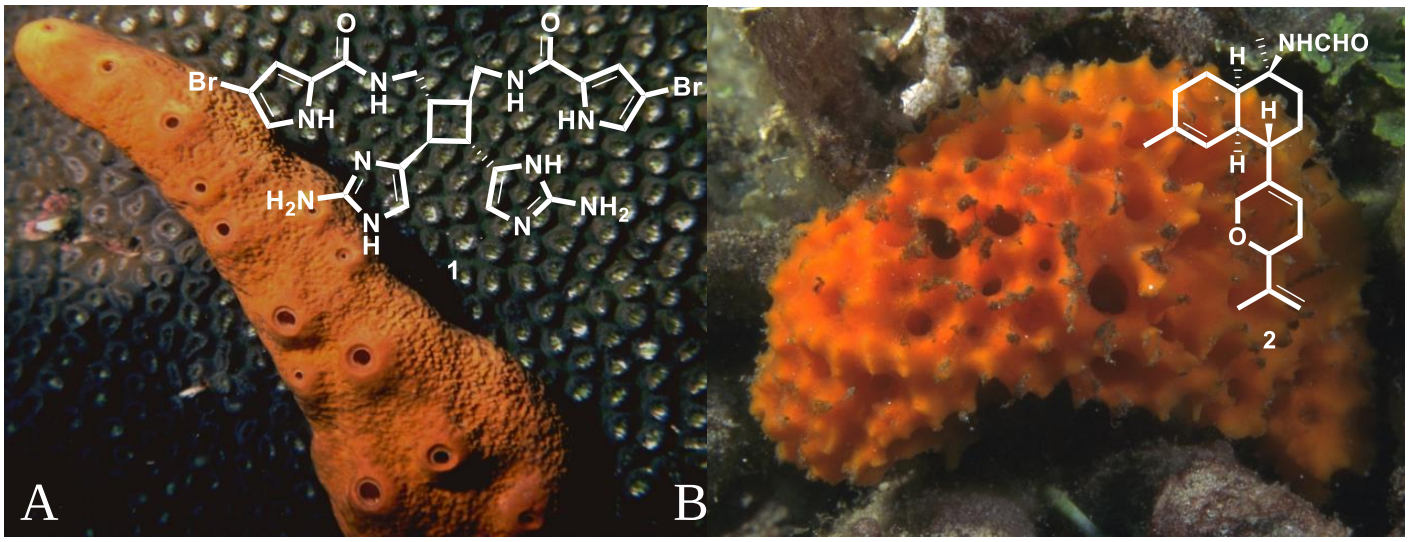


Fig 3 **(A) - *Agelas conifera* and antifoulant sceptrin (1)**
(B) - *Acanthella cavernosa* and antifoulant kalihipyran-7 (2)

An interesting observation stemming from contention for space is the aggressive overgrowth of one species relative to another. It was demonstrated in 1975 that sponges had adverse effects on neighboring coral and bryozoa. This enabled the sponge to be more effective in competitive interactions, making it more likely to overgrow.[22] A 1998 study, spurred by the observation that *Dysedea* sponges commonly overgrew *Cacospongia* spp. on Guam, isolated the sesquiterpene-ester 7-deacetoxyolepupuane as both a tactic to increase spatial occupation and a deterrent to spongiverous fish, highlighting the multiple defensive roles secondary metabolites can play.[23]

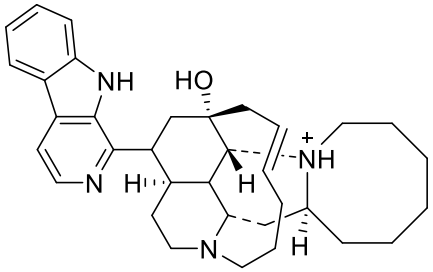
Natural immunity against pathogenic microbes is perhaps the most affluent source of bioactive

substances from the natural world.[9] The abundance and diversity of microbial symbionts, coupled with the enormous seawater throughput ensures that sponges encounter numerous infectious microbes. The success and proliferation of marine sponges despite the apparent dangers proves their proficiency in disease control, partially accounting for the wealth of specialized bioactive metabolites produced.

1.3.3 Sponge symbionts and the metabolome

Sponges are host to an enormously diverse fauna of microorganisms, many of which are reportedly specific to the sponge microbiota, with microbial assemblages in some cases comprising 40% of sponge bodymass.[24] Intimate bacterial partnerships between bacteria and porifera are thought to be evolutionary and, potentially predating sponge speciation.[25] Evidence for this is the unlikely co-existence of bacteria and choanocytes (bacteria and food digesters) in the sponge mesohyl, implying recognition and suggesting the utilization of antimicrobial NPs by the sponge to regulate the relationship.

The understanding of sponge-symbiont relationships is important to sponge NP investigation as several novel bioactive compounds, thought to be sponge metabolites, have been shown to be products of specific microbial symbionts. This finding was anticipated due to the recognition of highly structurally similar compounds, across different sponge species and in geographically disparate regions. The potent, antimalarial alkaloid Manzamine A, for example, was first extracted from *Haliclona* sp. in Okinawa and subsequently in other species, elsewhere in the world (**Fig 4**). Elucidation of the microbial colony of two Indonesian Manzamine A-producing sponges yielded actinobacteria, subsequent cultivation and analysis of which yielded the compound, verifying its bacterial origin.[26]



3

Fig 4 – Actinobacteria-produced manzamine A (3), found in *Haliclona* sponges.

Much is still unknown about sponge-symbiont relationships such as how identical symbiotic partnerships are recurrent in a species of different regions, while specific bacterial inhabitants of these species have never been found in seawater or elsewhere. Culturing of sponge bacteria *in vitro* has had limited accomplishments due to their complex requirements for growth (pH, temperature, nutrients *etc*.), and the extraction of NPs from the sponge hologenome remains the best way to examine these organisms for bioactive metabolites.[27] Although genomic sequencing efforts continue to advance our understanding of metabolite biosynthesis and symbiosis regulation, this area remains as a currently untapped reserve of information which will likely lead to revelations in future pharmaceutical development.

1.3.4 Sponge NP chemistry

The wide range of chemical entities biosynthesized by sponges and their associated microbiota can be classified according to different characteristics such as physiological effects (antimicrobial, antineoplastic), biosynthetic precursors and mechanisms (alkaloid, phenylpropanoid *etc*), or structural similarities (polysulfated glycolipid, prenylated hydroquinone). Many such structural descriptors exist for the enormous complexity of chemicals although these metabolites can be attributed to a small number of precursors and biosynthetic pathways.

Fatty acids, prostaglandins, aromatic polyketides and macrolide antibiotics are some of the products of the acetate pathway's use of the C_1 donor S-adenosyl methionine (SAM), C_2 unit donor Acetyl-CoA, and C_3 unit propionyl Co-A *e.g* Sponge-derived macrolide tedanolide (**4**)(**Fig 5**).

The terpenoids, a chemical class widespread in nature, are produced by successive additions of the C_5 mevalonate unit which is itself composed of of three Acetyl-CoA elements, but channelled by the mevalonate pathway into a different series of compounds.[28] Some members of this class are the sesterterpenoids (C_{25}) to which Ansellone A (**5**) (**Fig 5**) from the *Phorbas* sponge genus belongs, the modified triterpenoids (C_{30}) steroids such as cholesterol, and higher order polyisoprenes $(C_5)_n$ such as natural rubbers.

Alkaloids, which make up the largest and most diverse chemical class in sponges, providing the greatest number of novel compounds (20% between 2001 & 2010), contain a nitrogen heteroatom which confers basicity to the molecule.[29] The biosynthesis of alkaloids is based on amino acid precursors such as proline & histidene which give pyrrole and imidazole structural components respectively. On the basis of radiolabeling experiments, the aminoimidazole moiety of Oroidin (**6**)(**Fig 5**) of *Agelas* sponges, has been found to be a modified lysine rather than a histidine or arginine as was predicted, demonstrating the diversity in alkaloid biosynthetic pathways.[30]

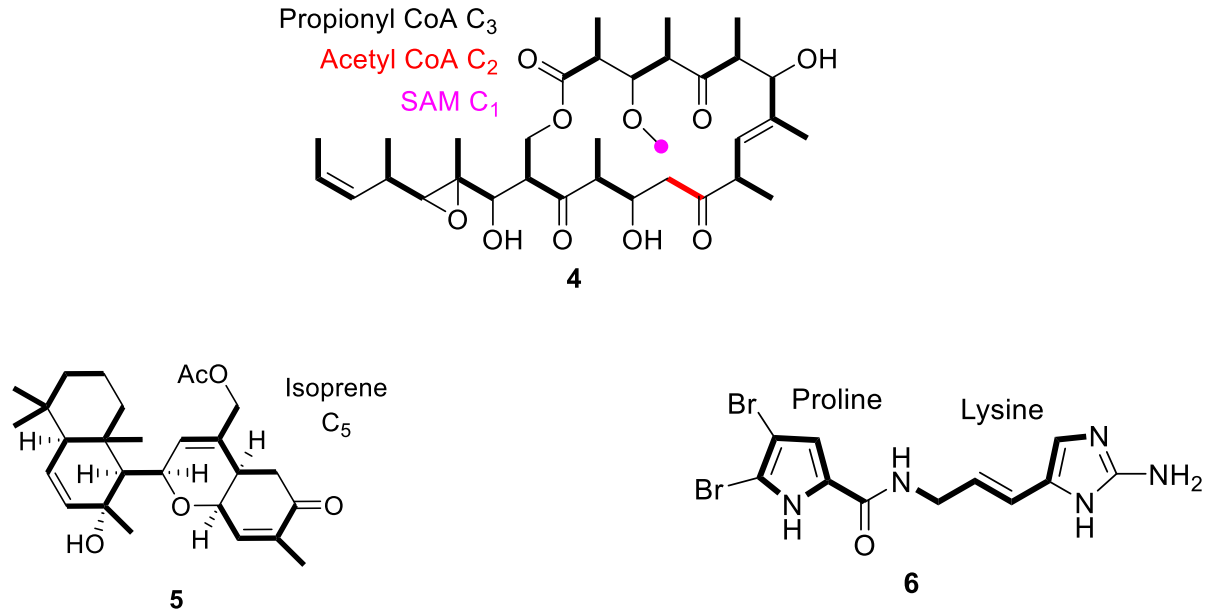


Fig 5 – Tedanolide (4), Ansellone A (5) & Oroidin (6). Biosynthetic precursors indicated in bold

1.4 Drug discovery & pharmacy

With the rise of modern medicine in the 19th century owing to advances in chemistry and biology, analytical investigation of traditional medicines began. In-depth chemical analyses of living entities were carried out, yielding novel compounds such as salicin in Willow bark, *Salix alba*, which had been chewed historically for its analgesic effects.[31] Synthetic chemistry served to mass produce these chemicals in more efficient ways, and in 1897 scientists at Bayer laboratories achieved manufacture of one of the first commercial drugs, the salicin analogue acetylsalicylic acid, paving the way to aspirin's global renown.[31]

Among of the oldest recognized medicinal NP is the anti-inflammatory resinous exudate of *Boswellia serrata* species, Frankincence.[32] Modern analysis has identified the pentacyclic triterpene boswellic acid as the major component.[33] Although the anti-inflammatory mode-of-action has been established, boswellic acid has not been exploited by pharmaceutical companies and crude *Boswellia* extracts remain only as an alternative medicine or 'nutraceutical'. Many other anti-inflammatory neutraceuticals such as cod liver oil exist as alternatives to the industry-regulated blockbuster drugs, however, the isolation of the active ingredient, complete synthesis, and development towards conventional drug status prove unprofitable for pharmaceutical companies. As a result, NP based drug discovery saw a decline in interest towards the end of the 20th century with discontinuation of industrial biodiscovery programs in favor of synthetic methods.[34] To tackle ailments that current medicine can not, while satisfying the criteria of pharmaceutical companies, drug discovery must be heavily guided towards chemical novelty.

1.4.1 Pharmaceutical potential of sponges

Several novel NPs have historically revolutionized cancer therapy, such as the nucleoside spongothymidine (**7**)(**Fig 6**) from the sponge *Tectitethya crypta*, the analog of which – arabinocytosine (**9**)(**Table** 1) – was marketed following isolation in 1951.[35,36] This sparked enormous interest in biodiscovery, but many further efforts in anti-cancer drug discovery were not as immediately successful. Spongistatin (**8**)(**Fig 6**) for example, isolated in 1992 from the *Spongia* genus, though one of the most active anti-tumor compounds in history, could not be obtained in the amount required for drug development.[37] An effort to process 400kg of sponge yielded only (13.8 mg of the macrolide (3.4×10^{-7} % yield). The first full stereoselective synthesis was only achieved ten years later with 29 steps and 0.5% overall yield.[38]

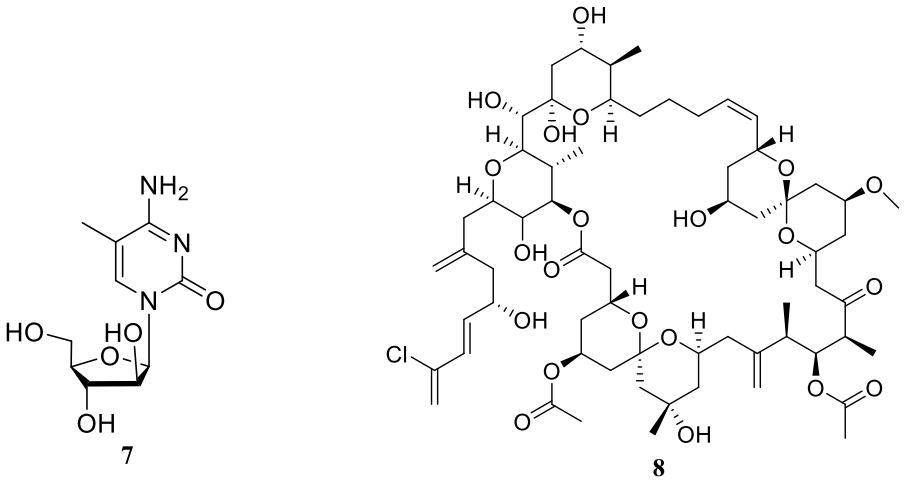
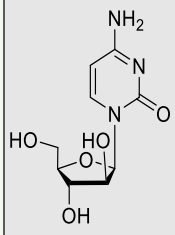
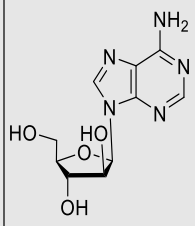
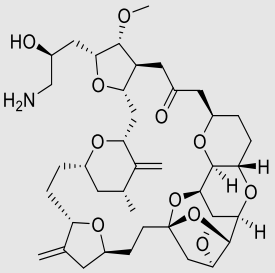
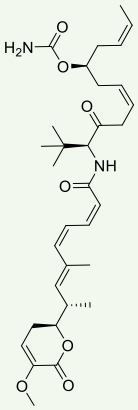
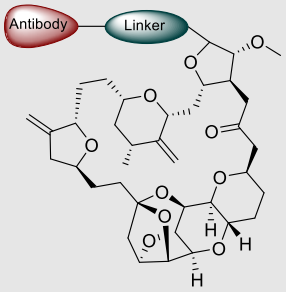


Fig 6 – Spongothymidine (7) and spongistatin (8)

Marine sponges represent the biggest producers of bioactive MNPs, with 30.5 % of discovered compounds exhibiting bioactivity.[39] Sponge derived compounds are famed for their activity against various cancer cell lines with over half of the 1615 bioactive sponge compounds isolated from 2001 – 2010 possessing cytotoxicity and/or anticancer properties.[29] Clinical approval and pharmaceutical development of a drug lead compound can often take decades and as a result, only three sponge-derived compounds have gained FDA approval to date (**Table 1**).[40,41] Despite this seemingly low proportion, identification of bioactive structural moieties (pharmacophores) from NPs can contribute to semi-synthetic drug development routes. Given the success of the three approved sponge-derived anticancer and antiviral compounds it is hoped that MNPs will help combat some of the most problematic diseases of the modern era.

Table 1 – Sponge-derived compounds approved or in clinical trials.[41]

Clinical status	**Approved**			**Phase II**	**Phase I**
Compound name	Cytarabine	Vidarabine	Eribulin Mesylate	Plocabulin	MORAb-202
Trademark	Cytosar-U®	Vira-A®	Halaven®	NA	NA
FDA Approved	1969	1976	2010	NA	NA
Chemical class	Nuceoside	Nuceoside	Macrolide	Polyketide	Macrolide Antibody-drug conjugate
Disease Area	Anticancer	Antiviral	Anticancer	Anticancer	Anticancer
Source sponge species	*Tectitethya crypta*		*Halichondria okadai*	*Lithoplocamia lithistoides*	*Halichondria okadai*
Structure	 **9**	 **10**	 **11**	 **12**	 **13**

2. Project overview

2.1 Background

The potential of biological systems to produce chemicals of enormous diversity and complexity has been instrumental in the development of traditional medicines and modern pharmaceuticals alike. This project aims to further the search for novel chemistry with a view toward biological activity, by investigating some of the least-well known organisms on the planet; deep-sea sponges. This project is an affiliate of the 5-year Science Foundation Ireland (SFI) project "***exploiting and conserving deep-sea genetic resources***" led by PI Prof Louise Allcock from NUI Galway. The multi-directive marine biodiscovery project seeks to isolate new bioactive NPs from deep-sea corals and sponges in Irish offshore waters. Of significant relevance to this research is the project's aim to develop bioprospecting approaches for isolating such compounds with maximum efficiency, given the difficulty in deep-sea sampling.

The seven samples investigated were collected roughly 200km west of Ireland at a depth of 900-1200m on the Porcupine Bank of the Irish continental shelf using the deep-water remotely operated vehicle *Holland I* during a National Parks & Wildlife cruise of the *ILV Granuaile* and the CE16006 cruise of the *RV Celtic Explorer*. The Porcupine Bank is known for its excellently preserved coral reefs which have been found to promote sponge abundance and species richness.[42]

Five of the seven samples used in this research are not yet formally identified. The specimen BDV-1541 used in screening was previously identified as a *Geodia pachydermata (Sollas, 1886)* through light-microscope spicule analysis. The specimen BDV-1826 was previously extracted and fractionated within the NUI Galway Marine Biodiscovery group and identified as a *Characella pachastrelloides (Carter, 1876)* by Dr. Christine Morrow through light-microscope spicule analysis. A voucher sample of each specimen is retained in 96% ethanol in the Ryan Institute, National University of Ireland, Galway, Ireland.

Approximately 1 mg of a pure compound is needed for NMR characterization using the available instrumentation. This may not be possible to obtain for each compound, depending on the abundance in the organism and available biomass but every effort must be made to minimize sample waste by retaining glassware washings after use. Investigation will continue as long as time constraints will allow in an effort to characterize as many metabolites as possible in the hope of identifying novel, bioactive chemical species.

2.2 Project Aims

This research seeks to isolate and characterize novel bioactive metabolites of deep-sea sponges from the Irish continental shelf.

Six sponge samples are first chemically screening to dereplicate known molecules, guide isolation and prioritize one specimen for investigation based on chemical diversity and novelty (**Fig 7**).

Based on the screening results, one of the six screened samples may be selected for full-scale extraction. The *Characella pachastrelloides* sample is not subject to screening but will be purified and analysed in the hope of isolating novel bioactive metabolites if the six screened samples are deemed unsuitable.

2.3 Workflow Overview

<table>
<tr><th></th><th colspan="3">Chemical Screening</th><th colspan="5">Purification and molecular analysis</th></tr>
<tr><th></th><th>Solid-Phase Extraction</th><th>HPLC Profiling</th><th>MS/MS</th><th>Chemical Extraction</th><th>Purification</th><th>1D NMR</th><th>Repurification</th><th>2D NMR</th></tr>
<tr><td>SFI-378</td><td>✓</td><td>✓</td><td>✓</td><td colspan="5" rowspan="4">Discontinued</td></tr>
<tr><td>SFI-387</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>SFI-390</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>SFI-391</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>SFI-394</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td colspan="2">F4 15 Peaks
C-18 column</td><td>F4-P3
Aplysamine-7</td><td></td></tr>
<tr><td>BDV-1541</td><td>✓</td><td>✓</td><td>✓</td><td colspan="5"></td></tr>
<tr><td>BDV-1824</td><td colspan="3">Previously Done by Dr. Kevin Calabro</td><td colspan="2">F4 12 Peaks
C-18 column</td><td>F2- P7 & P11
Adenosine &
Thymidine</td><td>F4-P2 6 Peaks
Amide column</td><td>F2-P2-P10
6-methyl-hercynine</td></tr>
</table>

Fig 7 - Workflow Overview and results from key stages

3. Screening and prioritization of 6 sponge samples

3.1 Overview of screening methodology

3.1.1 Literature review

Geodia sponges which are massively abundant along the Irish continental shelf, are well-known producers of the antifouling, dipeptide Barettin. This extensively studied 6-bromindole alkaloid is produced from the coupling of amino acid residues; tryptophan and arginine.[43] High-stability 'cysteine knot' polypeptides –barrettides A & B– have also been identified as antifoulants employed by *Geodia* sponges.[44] A study of North Atlantic Geodia sponges identified 24-methylenecholesterol as the major component of the sponge's sterol assemblage, reaching up to 77% in some samples.[45]

Chemical analyses of *Characella* sponges are far fewer in number but recent publications have identified the presence of complex bioactive metabolites, such as the highly light-sensitive, cytotoxic macrolide Poecillastrin H from a segment of sponge obtained from dredging at a depth of 191 m, in Southern Japan.[46] An investigation by the NUI Galway Marine Biodiscovery group of the methanolic fraction of North-Atlantic *Characella pachastrelloides* yielded the anti-inflammatory lipoglycotripeptides – characellides A−D.[47]

A community composition analysis of 2,585 sponge samples collected on the Porcupine Bank and the southeastern slopes of Rockall Bank (500-900m) reported a diversity of 105-122 sponge species with predominance of the Demospongiae orders; Verongiida and Poecilosclerida.[42]

3.1.2 Profiling – DAD, ELSD & HR-MS

Profiling involves a basic forensic analysis and documentation of the chemical composition of a sample. Chemical profiling of an organism is carried out through separations and mass spectrometry analysis on sample fractions. This indispensable stage yields information on component masses, elemental nature of components, peptide proportion, component amounts and separability. Profiling can be undertaken on a small portion (2g) of biomass and if deemed favourable, full scale extraction may follow. The data obtained are tabulated in a database and the novelty of the organism's metabolome is evaluated.

HPLC is an invaluable technique to the screening process. The HPLC chromatogram displays the separation profile of the mixture by pumping it through a stationary phase chromatographic column in a polarity-varied mobile phase The sample fractions are subject to the same optimised standard column and solvent gradient conditions so as to produce standardized, comparable results. The goal of this process is to identify sample fractions with many moderate-polarity compounds, which elute with good spectral resolution. Combining chromatographic and spectroscopic methods into 'hyphenated techniques' such as HPLC-ELSD (evaporative light scattering detector), is an extremely powerful tool for analysts as an extra dimension of data are gained accompanying separation. Application of ELSD following HPLC informs the analyst of the amount of solute molecules being detected at each point in time. This additional information

eliminates the risk of pursuing components with high absorptivity which may only be present in minor quantities and reveals the presence of compounds which are not UV-active.

Mass spectrometry (MS) and the tandem application (MS/MS) provide a highly accurate and informative map of component molecular masses and their abundances. To prepare molecules for MS they must be ionised and vapourised, the most common method being electrospray ionisation (ESI) in which the high density of protonated sample molecules in a droplet initiates coulombic explosion producing gaseous ions. These gaseous ions are accelerated to a specified velocity by a charged plate and are separated according to mass by their deflection a magnetic field. When combined with Ultra Performance Liquid Chromatography, (UPLC-MS/MS), a temporal dimension is acquired and the fragmentation mass spectrum can be examined at each instant along the retention profile. With analytical competence, and through deductive reasoning, the analyst may be able to infer elemental composition, chemical formula and structural motifs from the wealth of information gained from a UPLC-MS/MS experiment.

3.1.3 Molecular networking for dereplication and chemical diversity analysis

Dereplication is the process of identifying known compounds in a sample extract or fraction. With this knowledge purification can be guided towards the isolated of novel compounds. There is an enormous time and resource cost associated with isolation and structural elucidation of a metabolite and any means to help dereplicate in the early stages of investigation helps improve the efficiency of biodiscovery programs. Given the abundance of information amassed throughout investigation, it should be possible to cross reference data with that of previous investigations, building on research already done. Conventional dereplication methods attempt to identify similar or identical compounds by cross referencing discrete physical properties such as molecular mass, maximal UV-absorption wavelength, NMR chemical shifts or HPLC retention time against databases in the hope of finding matches. Scifinder, NP Atlas, MarinLit, AntiMarin and Derep-NP are just some of the databases that can be utilized to gain information on the chemical diversity in a crude extract prior to purification.

A new method of dereplication is emerging through ‘molecular networking’, based on the fact that each molecule’s MS/MS spectrum acts as a chemical fingerprint, containing information about the structural makeup of the molecule.[48] Each compound in the network is represented by a *node*. Families of compounds will have distinct similarities in their fragmentation fingerprints due to the common structural motifs they share. When two compounds have a minimum of 6 common fragments, their nodes are connected by an *edge*. This is how molecular networking exploits similar fragments to build up *clusters* of related compounds making it easier to visualise the chemical diversity in the sample (**Fig 8**).

Molecular networking based dereplication is being led by the specialized, open-access online platform Global Natural Products Social Molecular Networking (GNPS) with the goal of making mass spectrometry data of NPs more accessible and valuable to researchers.[49] MS/MS spectra are uploaded by users into a GNPS database which generates the molecular network. Uploaded spectra can be cross

referenced against multiple databases within GNPS, containing pure NPs. This helps identifying compounds within a molecular network. With identification of any one node in a cluster, the mass difference between connected nodes can indicate the structural differences and identify analogues within the cluster. When a bioactive standard (e.g. halichondrin B) is incorporated in the crude extract, it is possible to target the isolation of analogs within a cluster containing the standard by using the UPLC retention times.[3]

Molecular Networking has several major advantages over other dereplication strategies. The molecular network is generated from all of the MS/MS data simultaneously, increasing efficiency. Database cross-referencing, annotation and network visualisation can be easily carried out, displaying the output in a concise diagram. However, as GNPS is a relatively new dereplication tool, databases are not comprehensive and rely heavily on user-generated data. As a result, molecular networking is an evolving tool for dereplication, observing chemodiversity and targeted NP isolation, that will likely have extensive application in the future of NP research.

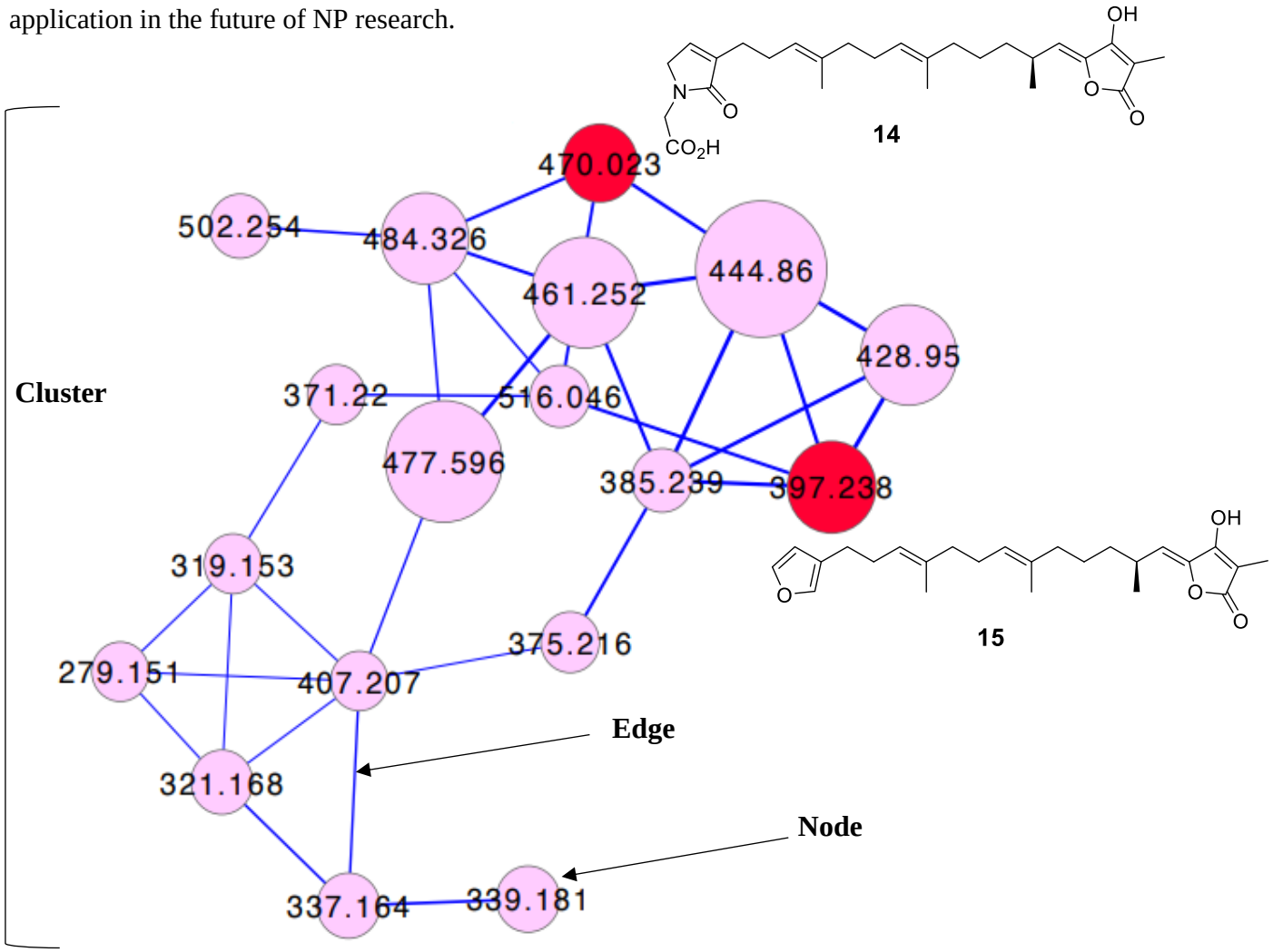


Fig 8 – Sponge sesterterpene tetronic acid cluster. The red nodes describe incorporated known bioactive standards ircinialactam A (**14**) and variabilin (**15**). The pink nodes represent unknown structural analogues.[3]

3.2 Screening results and discussion

3.2.1 HPLC-DAD-ELSD profiles

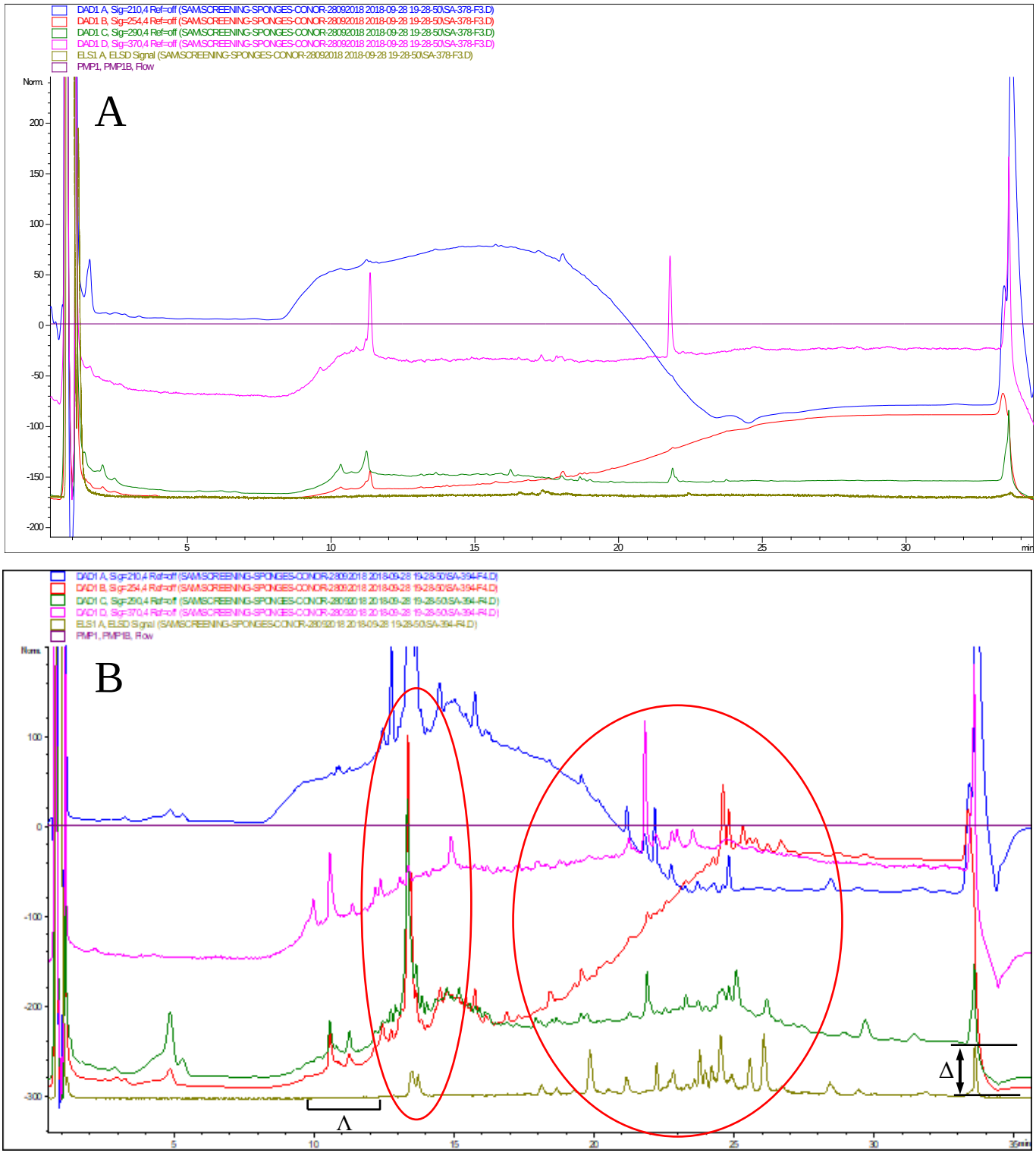


Fig 9 **(A) – BDV #378 - F3 HPLC-ELSD profile.** Example of poor separation chromatogram.
(B) – BDV #394 - F4 HPLC-ELSD profile. Example of favourable separation chromatogram. Areas of interest are circled in red. This fraction was prioritized for purification.
Blue – 210nm, Red – 254nm, Pink – 370nm, Green – 390nm, Brown – ELSD

Fig 9 (B) displays the excellent UV-Vis spectral resolution at several wavelengths of SFI-394, with favourable ELSD peaks indicating the abundance of material comprising each peak. The ELSD profile validates the choice of column and concentration gradient as one can see the majority of material was eluted in the timeframe, with relatively little material being detected upon cleansing of the column, quantified by the height of the ELSD peak (marked Δ). The ELSD spectrum is beneficial for later avoiding lengthy purification of highly-chromophoric compounds which are only present in minute quantities such as those being detected at 12 minutes (marked Λ). As a result of this clear separation of metabolites in SFI-394 - F4, this sample and fraction were prioritized for full scale extraction.

3.2.2 UPLC-DAD-HRESIMS

A moderate-polarity polypeptide was resolved in the 394 separation window, exhibiting a characteristic bell-shaped peak with monoisotopic mass 1075.0842g.mol^{-1} with peak separation of 0.33 indicating detection of the triply-ionised – $[M + 3H]^{3+}$ species. As expected, the doubly-ionised – $[M + 2H]^{2+}$ species with peak separation of 0.5 was observed at 1612.1208g.mol^{-1} revealing a peptide of 3222.2310g.mol^{-1}. The characteristic shape of the peptide peak is due to the 1.109% abundance ^{13}C. The distribution of isotopomers follows a binomial distribution pattern which was modelled suggesting an approximate 140 carbon atoms implying roughly 28 residues.

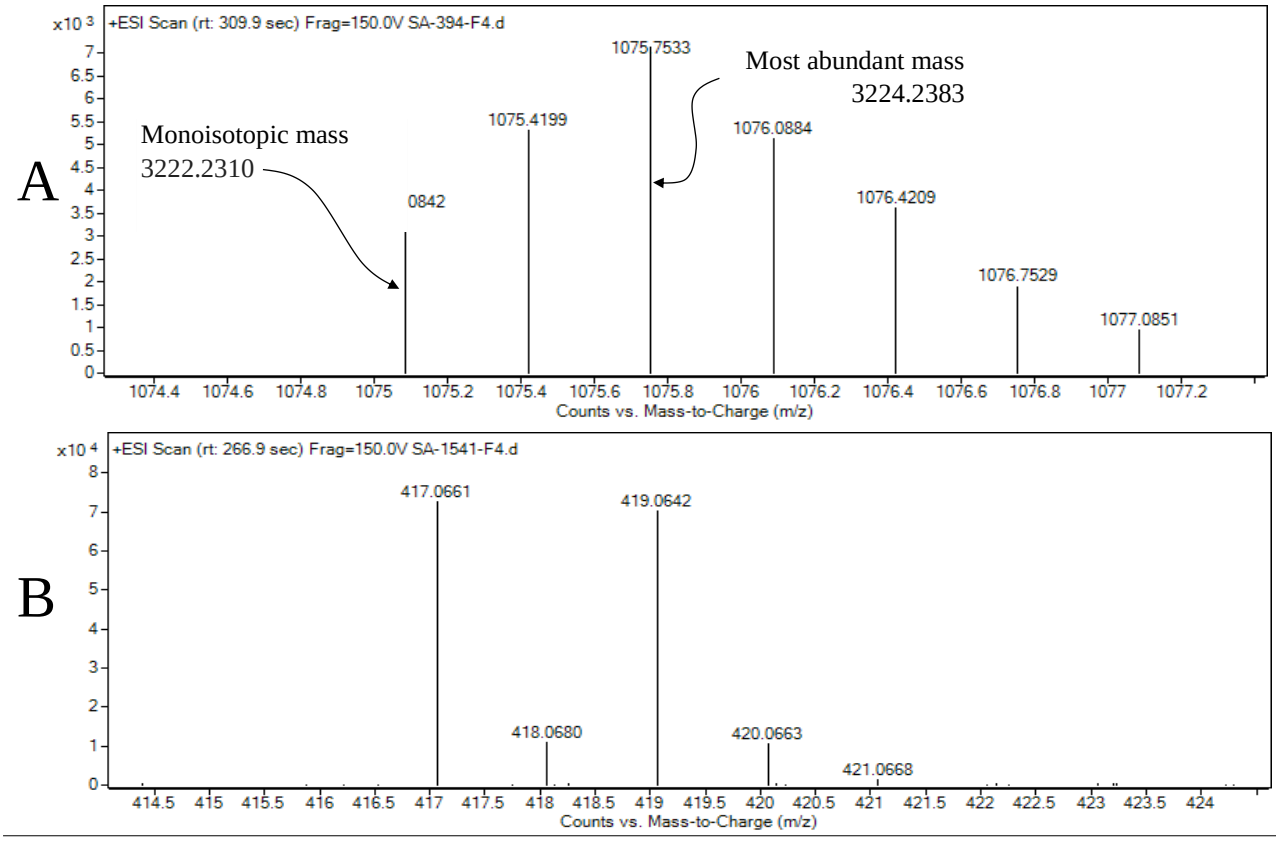
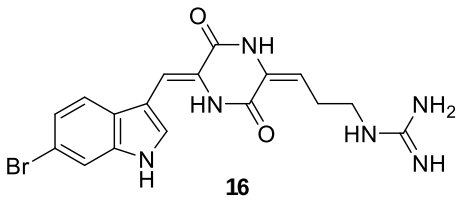


Fig 10 **(A)** – **394 -F4 –** **MS Peak at 310s.** Characteristic peptide isotopic distribution.
(B) – **1541 -F4 –** **MS Peak at 267s.** Characteristic mono-Br isotopic distribution.



16

Fig 11 – Geobarrettin B.

Fig 10 (B) shows the MS peak separation of a small monobrominated species with peaks differing by 2 Da due to the ^{79}Br and ^{81}Br isotopes, with 51% & 49% abundances respectively. The mass of the compound which was examined against the online MNP literature database MarinLit yielding solely the monobrominated indole alkaloid Geobarrettin B (**Fig 11**), confirming the taxonomical classification of sample BDV-1541 as *Geodia pachydermata*. As a result of this finding, BDV-1541 was deprioritized to avoid repeating isolation of this metabolite.

3.2.3 Molecular networking

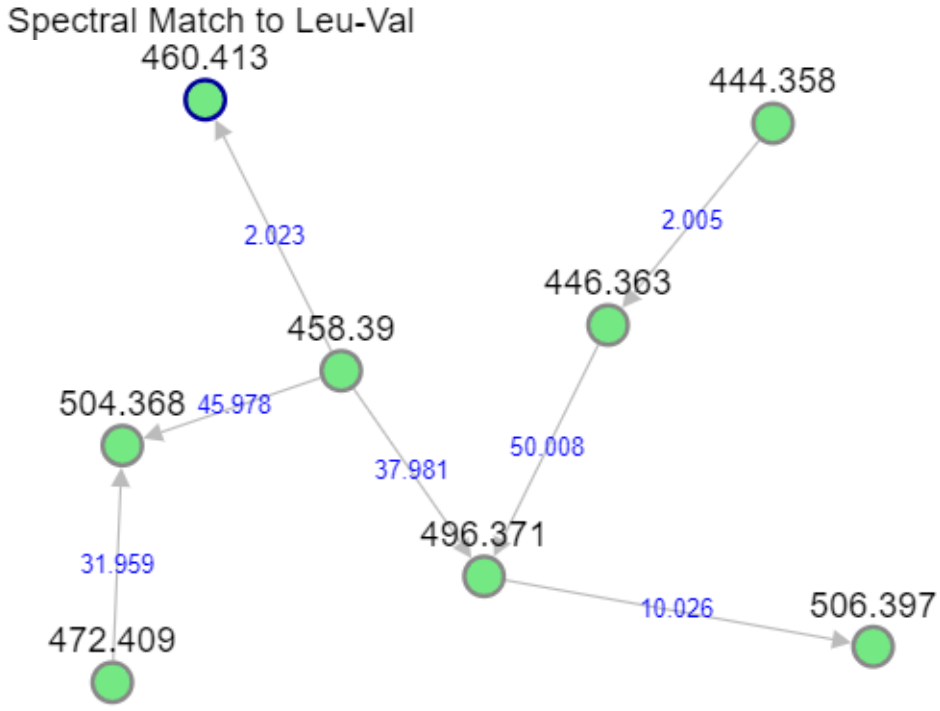


Fig 12 – SFI-394-F4 Molecular network cluster. Node at 460.413g.mol^{-1} circled in blue due to GNPS spectral match

The molecular network of SFI-394-F4 which was prioritized due to the HPLC-ELSD separation profile yielded a molecular cluster containing 8 nodes (**Fig 12**). This cluster was of interest as none of the molecular masses had matches to known compounds in MarinLit, however, the compound with mass 460.413 g.mol^{-1} was found by GNPS to have a spectral match to a leucine-valine moiety. As a result, this molecular cluster supported prioritization of SFI-394 and this cluster was targeted during purification.

4. Purification and chemical analysis of two sponge extracts

4.1 Overview of purification and identification methodology

4.1.1 Characterisation

In order to achieve the project aim, it is necessary to build up a picture of the sample composition by considering all the data obtained from previous analysis and rationalising how best to proceed with the purification and structural determination. Each sample will have unique characteristics to which the analysis must be tailored so as to ensure definitive, reproducible findings with minimal sample and resource waste. It is important to maintain a broad viewpoint, taking into account that the sample may be light or heat sensitive and may degrade over the course of purification misleading characterisation or producing artefacts. With each new piece of information obtained, literature should be consulted to improve understanding and continually aid correct characterisation.

Based on the results of HPLC-DAD, MS and molecular networking of 394-F4, it was expected the major components in the analysis window were small (300-600 Da) moderate-polarity brominated alkaloids. Preliminary characterisation such as this is essential for effective purification and massively aids later structural determination, especially for large compounds with recognizable structural motifs.

4.1.2 Purification: Theory & Application

The ability to isolate just milligrams of a rare NP from kilograms of sample biomass is paramount to chemical investigation, and improvements to the modus operandi continuously evolve in parallel with instrumentation development. Exploitation of metabolite physicochemical properties is the basis for purification methodology, and separation is largely on account of differences in molecular polarity, size and charge. The main four chromatographic techniques: ion exchange, surface adsorption, partition, and size exclusion are based on the movement of molecules across a sorbent (stationary phase) in a liquid or gas solvent (mobile phase). Many different types of chromatography exist, but for purification and analysis of NPs, column chromatography and HPLC are most widely used, with gel permeation having especial application to macromolecules such as proteins. Depending on the choice of solvent and sorbent, Polarity based chromatography can be normal phase (aqueous eluent with hydrophobic, non-polar stationary phase) or reversed phase (organic eluent with hydrophillic, polar stationary phase).

According to a 2012 review of NP purification, the most prevalent isolation methodology consists of normal-phase silica gel, repeated gravity-driven column chromatography, and HPLC.[50] This procedure is likely oriented towards the more prevalent chemical investigation of plants in which hydrophobic compounds such as terpenoids are more often of interest. For pharmacologically oriented MNP investigation, the focus lies more towards hydrophilic compounds such as alkaloids and nucleotide derivatives which are more likely to occupy biologically relevant chemical space[51]. As a result, reverse phase C-18 column chromatography is employed herein so as to remove the abundant ionic and very polar compounds first, collect the moderate-polarity eluates, and eliminate the risk of contamination by large, non-polar compounds such as fatty acids which are often over-retained on the stationary phase. As per the SOP, The analyst performing HPLC will decide which one of three methods to employ depending on the

amount and the required degree of separation; preparative, semi-preparative or analytical, each with their own demands and limitations. The objective of preparative and semi-preparative chromatography is to purify significant quantities of a sample and collect the components for further application. Analytical separation is done with a view towards investigating a property of the analyte and is commonly coupled with instrumentation for quantitative analysis.

Purification of a crude mixture follows a branching network guided by analytical and logical intuition of where the metabolites of interest reside. As detailed in **Fig 13**, following preparative-HPLC purification, an isolated peak may contain several compounds which again have to be repurified while ensuring to obtain the requisite 0.5-1mg needed for structural elucidation. Considering the degree of sample loss through handling and removal of impurities, repurification may not be possible if too little sample is present.

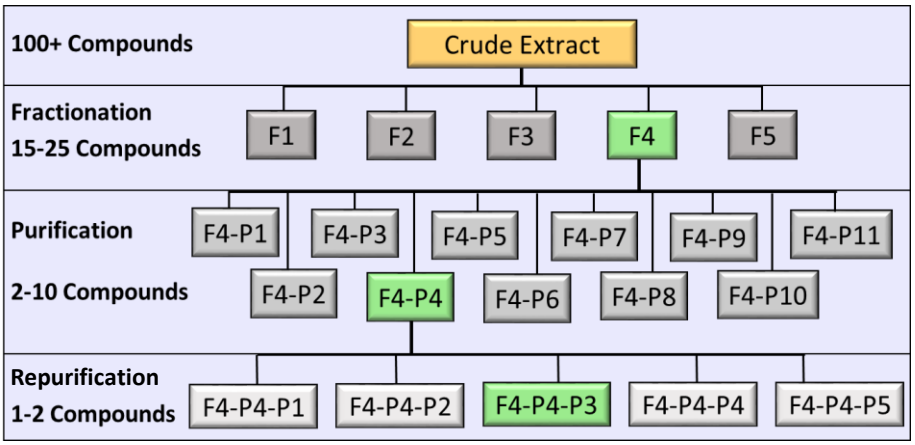


Fig 13 – Example purification flowchart.

4.1.3 Nuclear Magnetic Resonance: Theory & Application

NMR is currently the most useful tool in the natural product researchers arsenal. It enables preliminary purity and composition analysis of unrefined samples, and allows for unambiguous, absolute structural elucidation of purified compounds. The result of a scan is a series of signals, each with a chemical shift characteristic of a nucleus' relaxation response to a pulsed external magnetic field. The chemical shift reflects the extent to which the intrinsic Larmor frequency of a nucleus is altered by its surrounding electronic environment. The frequency of the free-induction decay (FID) electromagnetic wave produced by nuclear spin relaxation is monitored and a Fourier transform is applied to convert the information to a discrete readable signal with units of ppm (as it is the 10^{6}-fold ratio of signal to the spectrometer's field strength). As the Larmor frequency is proportional to the applied magnetic field strength, oscillation frequency is expressed on a relative scale, with a known signal as a reference point. This reference signal may be a known inert additive such as trimethylsilyl (TMS) or based on the residual signal of the solvent used. However NMR can only be applied to 'NMR-active' nuclei with non-zero nuclear spin and hence, a

magnetic dipole moment. As a result, spin-paired nuclei such as deuterium (^{2}H) and ^{12}C are not oriented by the external magnetic field and do not reveal structural information in an NMR experiment.

Two dimensional (2D) NMR is an extension of these basic principles whereby the acquisition of two correlated frequency data sets is separated by a specified delay known as the 'evolution time'. Radiofrequency pulses before and after the interval register correlations between neighbouring nuclei. By incrementing the evolution time and varying the pulse sequence and intensity, many different types of couplings can be detected, each leading to a specialised experiment; homonuclear coupling through bonds (COSY), heteronuclear coupling through bonds (HSQC & HMBC), or through-space coupling of nuclei in proximity up to 4Å (NOESY). The result of deconvoluting the nuclear FID is a 2D 'map' with a 1D NMR spectrum on each axis and an intensity value for each pair of 1D signals depicting atom connectivity in a molecule and facilitating unambiguous structural determination.

In the course of this project, samples of varying purity were subject to 1D and 2D NMR to aid characterisation, purity determination and structural elucidation. Due to the non-destructive nature of NMR analysis, samples could be analysed for ^{1}H-NMR at the 1D level quickly and without risk of sample loss. Longer scans may be required for ^{13}C-NMR due to the small relative abundance (1.1%) of the NMR-active ^{13}C isotope or less concentrated samples to amplify the output and distinguish it from the noise signal. Due to longer run-times, 2D analyses were carried out only when a full data set was required for structural elucidation.

4.2 Results and discussion – Sample SFI-394

Screening presented SFI-394 as the optimal sample for investigation. The remaining biomass (34.6 g) was extracted routinely (**Fig 14**).

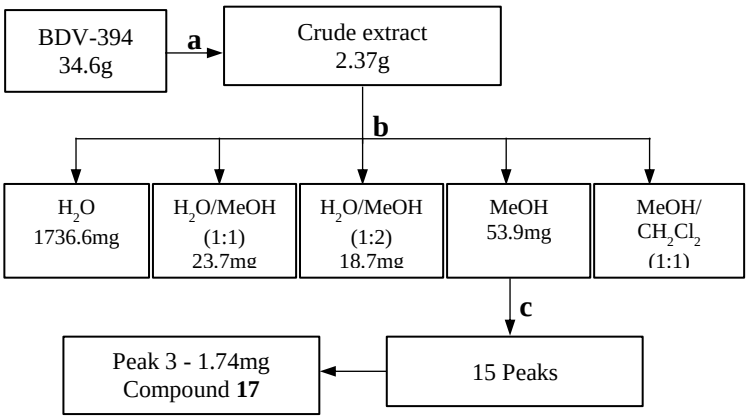


Fig 14 – SFI-394 purification. **a -** Extraction (3 x 200mL solvent aliquots of MeOH/CH_2Cl_2 (1:1))
b - RP C-18 VLC fractionation - 500mL solvent elution volume - decreasing polarity
c – RP HPLC purification C-18 column, gradient mobile phase of H_2O & MeOH

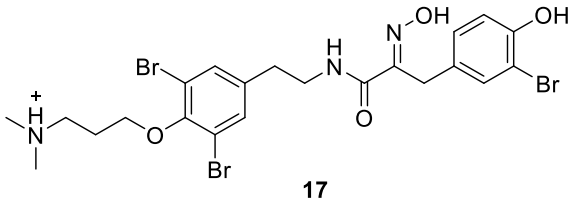


Fig 15 – Aplysamine 7 (17).

4.2.1 Structure elucidation - Aplysamine-7 (17)

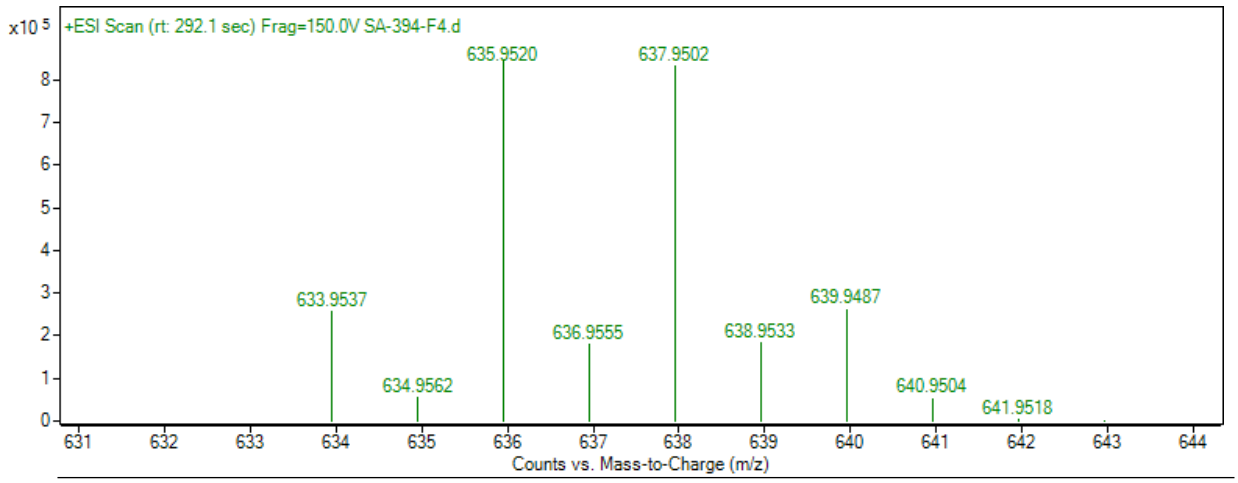


Fig 16 – Aplysamine 7 mass spectrum. 1:3:3:1 Br isotopic pattern

Table 2 – Agilent Masshunter formula predictions.

	FORMULA	SPECIES	SCORE
1	C20 H24 Br3 N6 O3	(M+H+)	97.84
2	C22 H26 Br3 N3 O4,	(M+H+)	97.11
3	C13 H24 Br3 N12 O	(M+H+)	96.96

Table 3 – Aplysamine 7 NMR tables. Experimental & original values**.**

Current Research			Aplysamine 7 - Original report [52]		
NMR solvent: MeOH-d4			NMR solvent: DMSO-d6		
position	$\int_H$	δ_H, mult. (*J* in Hz)	position	$\int_H$	δ_H, mult. (*J* in Hz)
2	1H	7.35, d (2.0)	2	1H	7.35, d (1.8)
5	1H	6.76, d (8.2)	5	1H	6.76, d (8.4)
6	1H	7.04, dd (8.2, 2.0)	6	1H	7.04, dd (8.3, 1.8)
7	2H	3.76, s	7	2H	3.75, s
11	2H	3.43, t (7.0)	11	2H	3.43, t (7.1)
12	2H	2.75, t (7.0)	12	2H	2.74, t (7.1)
14, 18	2H	7.44, s	14, 18	2H	7.43, s
20	2H	4.07, t (5.5)	20	2H	4.06, t (5.6)
21	2H	2.27, tt (7.7, 5.5)	21	2H	2.26, tt (7.7, 5.6)
22	2H	3.52, t (7.7)	22	2H	3.52, t (7.7)
$N(CH_3)_2$	6H	2.97, s	$N(CH_3)_2$	6H	2.96, s

(+)-HRESIMS analysis of 17 revealed a monoisotopic molecular ion peak for $[M+H]^+$ of *m/z* 633.9537 , which led to the chemical formula predictions in **Table 2**. Chemical formula prediction 2 matched Aplysamine 7, identified previously in the screening stage. The ^{1}H NMR of the isolated compound was analysed and compared against the ^{1}H NMR values obtained in the original report of Aplysamine 7.[52] Inspection of the ^{1}H NMR spectrum revealed two methyl groups at δ_H 2.97 (s, 6H, $N(CH_3)_2$), one equivalent methylene singlet with at δ_H 3.76 (s, 2H, H-7), several equivalent methylenes arranged in two distinct spin-coupled systems; one triplet at δ_H 2.75 (t, *J = 7.00 Hz*, 2H, H-12) coupled with δ_H 3.43 (t, *J* = 7.00 Hz, 2H, H-11), and a triplet of triplets at δ_H 2.27 (tt, *J = 7.7, 5.5* Hz 2H, H-21) coupled to two triplets at δ_H 3.52 (t, *J* = 7.7Hz, 2H, H-22) & 4.07 (t, *J* = 5.5 Hz, 2H, H-20). An aromatic system was identified by the 1-2-4 aromatic coupling arrangement of signals at δ_H 6.76 (d, *J = 8.2* Hz, 1H, H-5), 7.04 (dd, *J = 8.2, 2.0* Hz, 1H, H-6) & 7.35 (d, *J = 2.0 Hz,* 1H, H-2). The second aromatic system was identified by the equivalent aromatic singlet at δ_H 7.44 (s, 2H, H-14/18). These data were found to match the values in the original report of Aplysamine 7. To confirm the structure, the MS/MS fragmentation pattern was analysed (**Fig 19**).

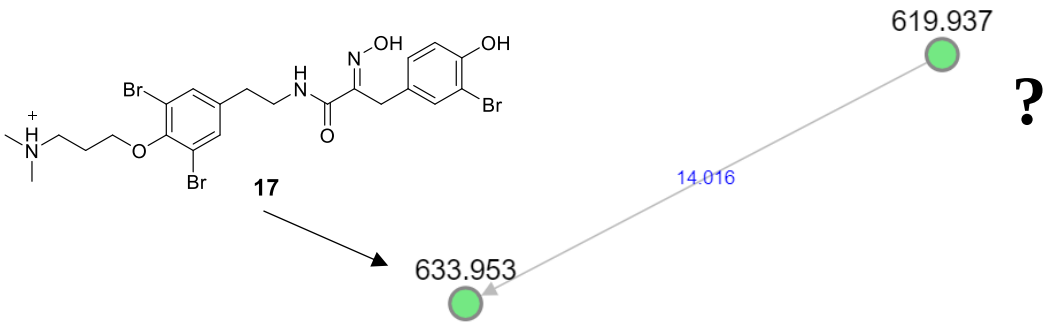


Fig 17 – Molecular cluster containing Aplysamine 7.

The molecular cluster containing Aplysamine 7 was consulted, exhibiting just two nodes (**Fig 17**). The two compounds had a mass difference of 14.016 g.mol^{-1} equating to a CH_2 unit. The smaller compound was not isolated in sufficient purity for characterisation but is likely to be produced by de-methylation of the terminal amine position. This proposed compound has not yet been discovered in nature though may be an artefact produced as a result of the purification process.

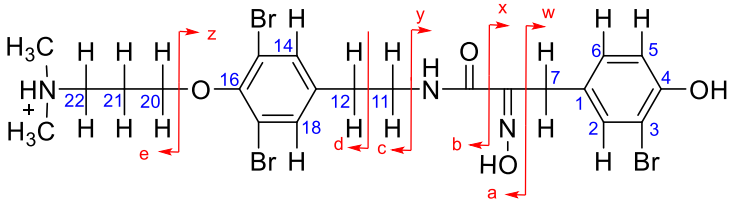


Fig 18 – Aplysamine 7 linear structure. Fragmentation cleavage sites & atom numbering shown**.**

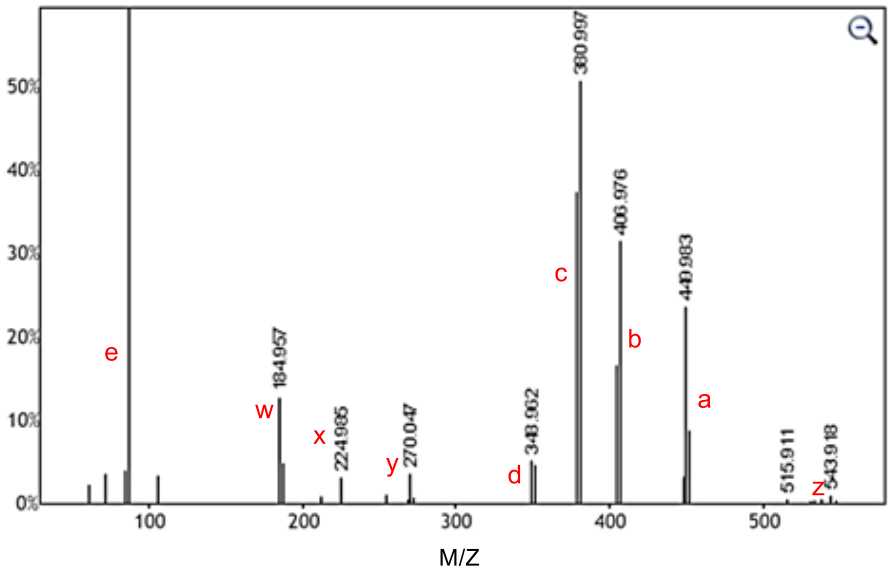


Fig 19 – Aplysamine 7 MS/MS fragmentation spectrum.

4.3 Results and discussion – BDV-1826

The isolation and structural elucidation of three metabolites from the fraction F2 of BDV-1826 is reported. The novel compound (**20**) was isolated in sufficient purity for analysis after a second HPLC purification.

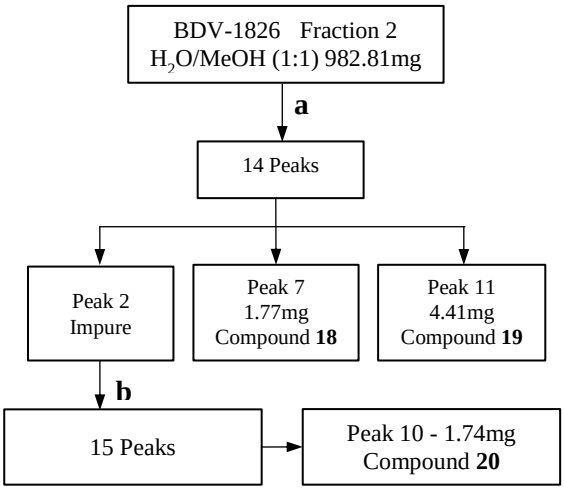


Fig 20 – SFI-394 purification.

a – RP HPLC purification C-18 column, gradient mobile phase of H_2O & MeOH

b – RP HPLC repurification, BEH amide column, gradient mobile phase of ACN & H_2O

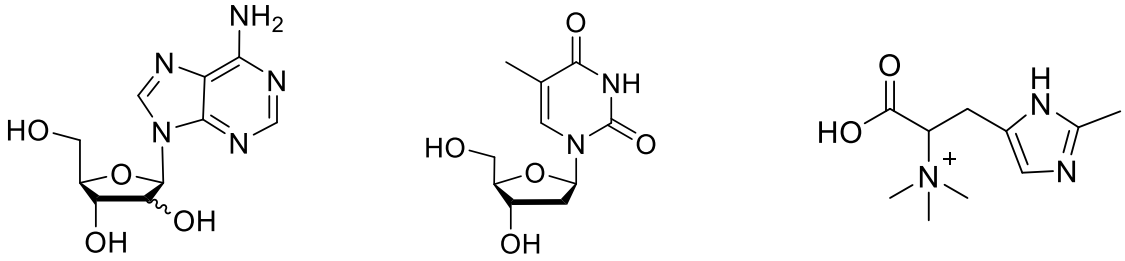


Fig 21 – BDV-1826 isolated compounds.

18 – Adenosine
19 – Thymidine
20 – 6-methyl-hercynine

4.3.1 Structure elucidation - Adenosine

Table 4 – Adenosine NMR table.

	NMR solvent: MeOH-d4		
position	δ_H, mult. (*J* in Hz)	δ_C	COSY
2	8.64s	143.23	
8	8.40s	145.21	
1'	6.11d (5.0)	89.6	2'
2'	4.62t (5.0)	74.9	1', 3'
3'	4.31dd (4.0, 4.9)	70.6	2', 4'
4'	4.15dt (4.0, 3.0)	86.2	3', 5', 5'
5'	3.79dd (12.3, 2.9)	61.3	4', 5'
	3.87dd (12.5, 3.1)	61.3	4', 5’

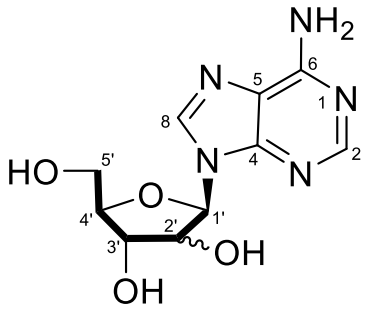


Fig 22 – Adenosine atom numbering and key COSY correlations.

(+)-HRESIMS analysis of **18** revealed a molecular ion peak for $[M+H]^+$ at *m/z* 268.2485. Structural elucidation followed intuitively from the signal at δ_H 6.1 (d, *J = 5.1* Hz, 1H, H-1’), indicative of an anomeric proton on a sugar unit. The COSY spectrum revealed a 5 member spin coupled system. The integrations, chemical shifts and splitting patterns of which indicated the presence of a ribose sugar. The signals obtained were matched against literature values for ribose.[53] δ_H 6.1 (d, *J = 5.1* Hz, 1H, H-1’), 4.62 (dd, *J = 5.1, 4.9* Hz, 1H, H-2’), 4.31 (dd, *J* = 4.0, 4.6 Hz, H-3’) and 4.15 (dt, *J* = 3.9, 3.2 Hz, 1H, H-4’) all had singular integration and splitting patterns with chemical shifts indicative of a C-O bond at each position. The two signals at δ_H 3.79 (dd, *J = 12.3, 3.3* Hz, 1H, H-5’a) & 3.87 (dd, *J* = 12.3, 2.5 Hz, 1H, H-5’b) had slightly different couplings to H-4’ as expected due to the chiral centre, confirming the presence of a ribose unit. The terminal carbon’s chemical shift at δ_C 61.3 (CH_2, C-5’) indicated the presence of a OH. The two aromatic signals at δ_H 8.64 (s, 1H, H-8) & δ_H 8.40 (s, 1H, H-2) were characteristic of the commonly found purine base adenine. Based on literature values, the more downfield signal at δ_H 8.64 (s, 1H, H-8) was assigned to the 8 position.[53]

4.3.2 Structure elucidation – Thymidine

Table 5 – Thymidine NMR table.

	NMR solvent: MeOH-d4
position	δ_H, mult. (*J* in Hz)
6	7.81s
7	1.88s
1'	6.28, t (6.7)
2’	2.21, dt (6.7, 11.7)
	2.25, dt (6.7, 11.7)
3'	4.40, td (6.7, 5.8)
4'	3.91, dt (5.8, 3.1)
5'	3.74, d (12.0, 3.1)
	3.79, d (12.0, 2.9)

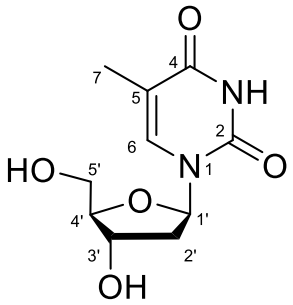


Fig 23 – Thymidine atom numbering.

(+)-HRESIMS analysis of **19** revealed a molecular ion peak for $[M+H]^+$ at *m/z* 243.2358. Structural elucidation followed intuitively by comparison with the Adenosine H-NMR and from the triplet signal at δ_H 6.28 (t , *J = 6.7* Hz, 1H, H-1’), indicative of an anomeric proton on a sugar unit. The pyrimidine group of thymine was matched to literature findings by the sole aromatic signal at δ_H 7.81s (s, 1H, H-6) and the methyl group identified by the singlet at δ_H 1.9 (s, 3H, H-5). The deoxyribose sugar unit was identified by the splitting patterns and integration of the signals at δ_H 2.21 (dt, *J = 6.7, 11.7* Hz, 1H, H-2’), 2.25 (dt, *J = 6.7, 11.7* Hz, 1H, H-2’), 4.40 (td, *J = 6.7, 5.8* 1H, H-3’), 3.91 (dt, *J = 5.8, 3.1* Hz, 1H, H-4’), 3.74 (d, *J = 12.0, 3.1* Hz, 1H, H-5’a) & 3.79 (d, *J = 12.0, 2.9 Hz*, 1H, H-5’b). The identity of the compound was confirmed by comparison to literature values for nucleosides.[53]

4.3.3 Structure elucidation – 6-methyl-hercynine

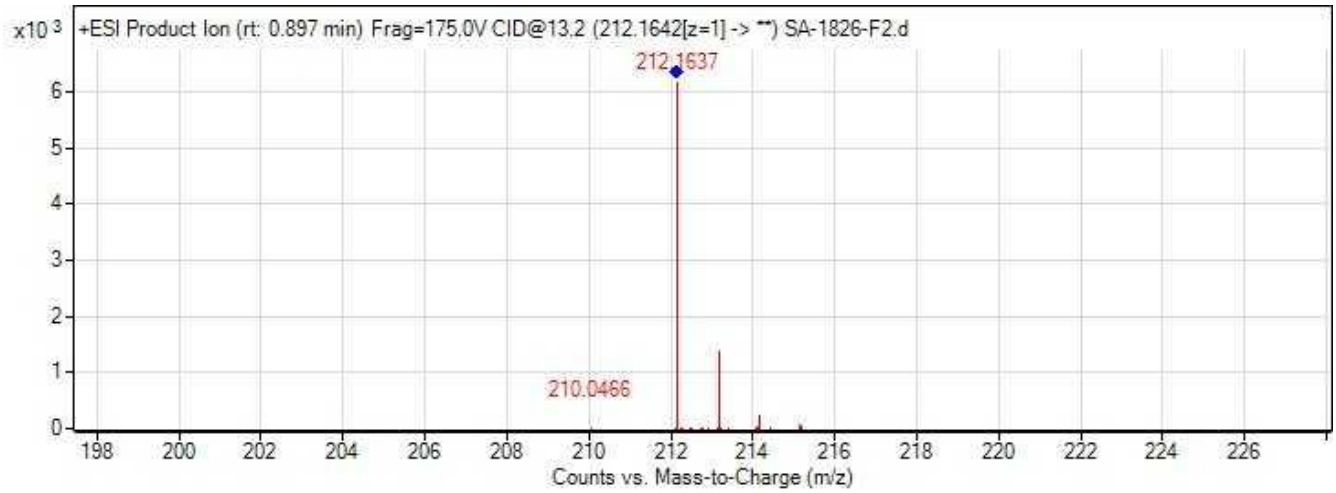


Fig 24 - 6-methyl-hercynine $[M]^+$ HRESIMS peak

Table 6 – 6-methyl-hercynine NMR table.

	NMR solvent: MeOH-d4	
position	δ_H, mult. (*J* in Hz)	δ_C
1		169.8
2	4.00, dd (12.1, 3.7)	76.38
3	3.45, dd (14.0, 3.7) 3.35, dd (14.0, 12.1)	22.61
4		127.8
6		144.1
8	7.29, s	119.1
9	3.29, s	52.2
10	2.73, s	15.5

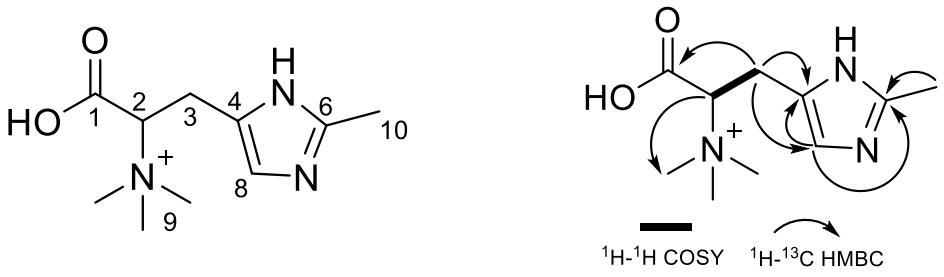


Fig 25 – 6-methyl-hercynine atom numbering and key 2D NMR correlations.

Compound **20** was isolated as a colourless crystalline solid. The HRESIMS had a molecular ion peak for $[M]^+$ of *m/z* 212.1637 consistent with a molecular formula of $C_{10}H_{18}N_3O_2$. This indicated the presence of 4 unsaturated carbons. The ^{1}H NMR data revealed four methyl singlets at δ_H 3.32 (s, 9H, H-9), 2.73 (s, 3H, H-10), one ABX system at δ_H 3.38 (dd, 1H, *J = 13.93, 3.59 Hz*), 3.45 (dd, 1H, *J = 14.05, 3.59 Hz*), one aromatic methine singlet δ_H 7.29 (s, 1H), one methine doublet of doublets at δ_H 4.00 (dd, 1H, *J = 12.12, 3.59 Hz*). The HSQC and HMBC indicated resonances associated with three deprotonated carbons, a carboxylic acid at δ_C 169.6 (C-1) and two aromatic carbons at δ_c 144.1 (C-6) and 127.8 (C-4). A COSY correlation between H-2 and H-3 allowed a spin coupled system (SCS) between C-2/C-3 to be assigned. The HMBC correlations from H-3 to C-1 and three *N*-methyls at δ_c 52.2 established the presence of a methylated amino acid derivative. The connection between the aromatic ring and the SCS was evident by a key HMBC correlation between H-3/C-4 and H-3/C-8. The chemical shifts of C-4, C-6 and C-8 indicated the presence of an imidazole ring. The aromatic carbons were connected with key HMBC signals H-8/C-4, H-8/C-6. Further examinations revealed a methylation of the imidazole at C-6 concluded from the key H-10/C-6 HMCB correlation. H-10 only had one correlation in HMBC experiment confirming the presence of an imidazole ring.

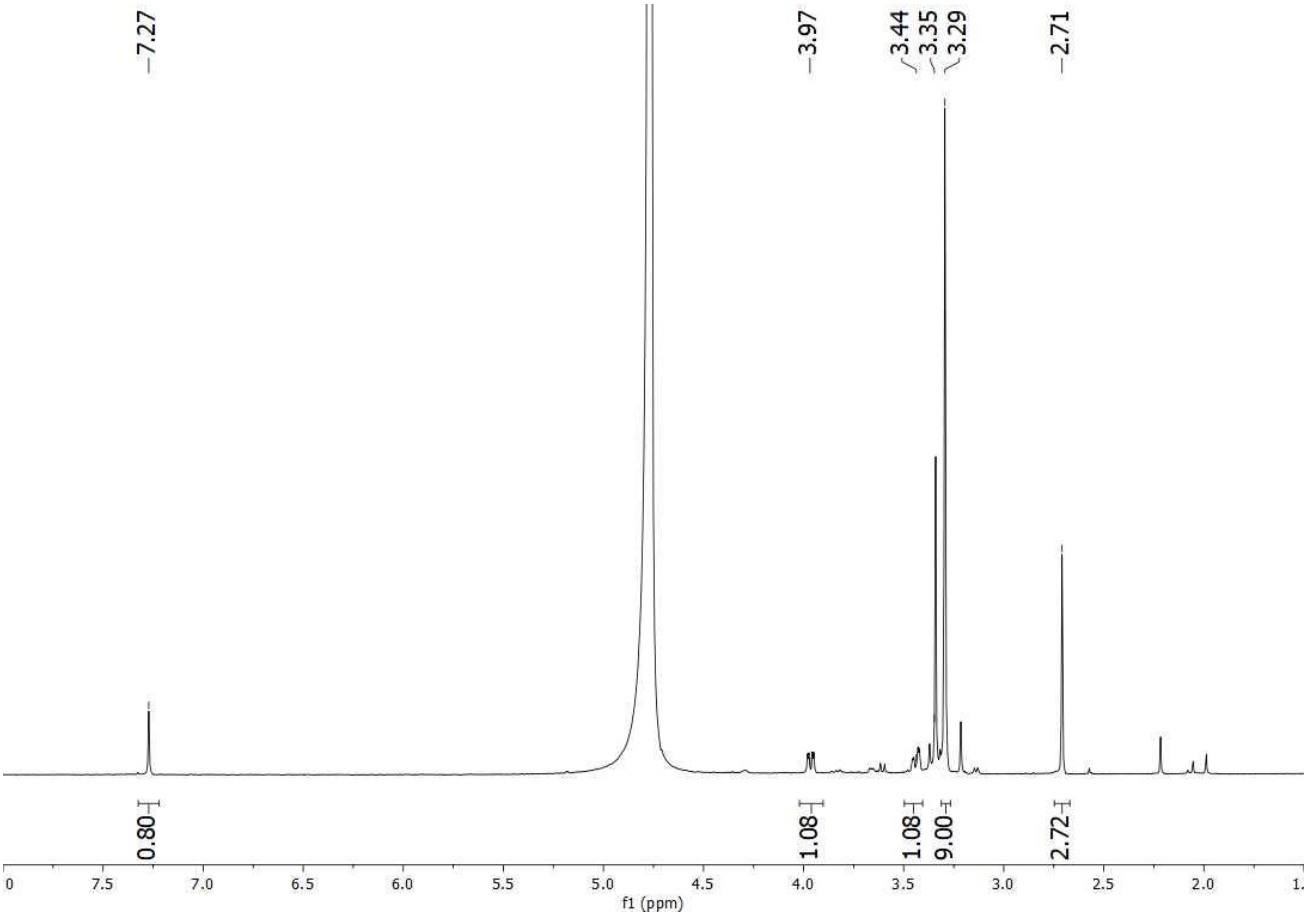


Fig 26 – 6-methyl-hercynine ^{1}H NMR spectrum.

5. Experimental methods

5.1 Sampling and instrumentation

5.1.1 Sampling and preparation

The samples SFI-378, SFI-387, SFI-390, SFI-391, SFI-394 and BDV-1541were collected at a depth of 900-1200m on the Porcupine bank off the Irish continental shelf using the deepwater remotely operated vehicle (ROV) *Holland I* during a National Parks & Wildlife cruise of the *ILV Granuaile.* Sample BDV-1824 was collected on the Goban Spur at 809 m depth (48.6509° N, 10.4846° W) *on the CE16006* crusie of the *RV Celtic Explorer*. A small portion of each specimen was retained in 96% ethanol as a voucher sample for DNA analysis and identification in the Ryan Institute, National University of Ireland, Galway, Ireland. The rest of each sample was lyophilized and stored at -20 °C. A ball mill was used to powder the sample prior to extraction.

5.1.2 General experimental procedure

HPLC-DAD-ELSD separation profiles were obtained on an Agilent 1260 analytical HPLC series equipped with a DAD detector & coupled to an Agilent-380-ELSD detector. UPLC separation for mass analysis was done on an Agilent 1290. High Resolution Electrospray Ionisation Mass Spectrometry (HRESIMS) data was obtained from a QTOF Agilent 6540 in ESI(+). NMR experiments were measured on a 600 MHz equipped with a cryoprobe (Varian). Chemical shifts (δ in ppm) are referenced to carbon (δ_C 49.0) and residual proton (δ_H 3.31) signals of MeOH-d4.

5.2 Screening and prioritization of 6 sponge samples

5.2.1 Extraction & Fractionation

A 1g subsample of freezedried biomass from each sample was extracted (3 x 50mL solvent aliquots) in MeOH/CH_2Cl_2 (1:1 v/v), aided by ultrasonication followed by gravity filtration of the supernatant. 0.5g polygoprep 60-50 µm C-18 (Macherey-Nagel) bulk powder was added to each extract and the mixture were dried. Each crude extract was fractionated in a Solid Phase Extraction (SPE) (1g, RP-C-18) cartridge prepared by washing with two column volumes of MeOH/DCM (1:1) and conditioning with two volumes of distilled H_2O. The samples were loaded on the cartridges. Fractions (F1-F5) were obtained by eluting two column volumes of each solvent according to the decreasing solvent polarity gradient: **1).** H_2O, **2).** H_2O/MeOH (1:1), **3.)** H_2O/MeOH (1:2), **4.)** MeOH, **5).** MeOH/ CH_2Cl_2 (1:1). Each fraction was dried and weighed.

5.2.2 Analytical HPLC

Fractions F2-F5 for each sample were made up to a 10mg.mL^{-1} concentration with MeOH & 0.2µm polytetrafluoroethylene PTFE syringe filtered. The standard separation profiles were obtained on an Agilent 1260 analytical HPLC using a reversed-phase-C-18 Column (Waters analytical C-18 column 4.6mm x 250mm) with 1.5 mL.min^{-1} flow rate and 10µL injection over a 31 min run time with mobile phase solvent gradient of H_2O (A) & MeOH (B) both acidified with 0.1% v/v TFA. The separation procedure used was standard for all samples with specifications: 5 min isocratic at 10% B, then a linear

gradient for 15 min to 100% B, isocratic at 100% B for 10 min, returning linearly to 10 % B over 2 min at which it was held for 4 min until completion.

5.2.3 UPLC-HRESIMS

The 10mg.mL^{-1} solutions previously used for analytical HPLC, were used for UPLC-HRESIMS analysis. UPLC separation was achieved with an Agilent 1290 analytical reversed-phase (RP)-UPLC using a C-18 Column (Waters analytical BEH C-18 column 4.6mm x 250mm). An Agilent 6540 in positive mode - ESI(+) with a QTOF detection method was used to record mass spectra in both MS & automatic MS/MS analysis.

5.2.4 Molecular networking

The data were converted to mzXML format and uploaded to the GNPS server. A molecular network was created using the online workflow at GNPS. The data was filtered by removing all MS/MS peaks within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 peaks in the +/- 50Da window throughout the spectrum. The data was then clustered with MS-Cluster with a parent mass tolerance of 2.0 Da and a MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra . Further, concensus spectra that contained less than 2 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.

5.3 Purification and molecular analysis of 2 sponge extracts

5.3.1 Extraction & Fractionation 394

The freeze-dried, powdered sponge sample (34.6g) was extracted three times (3 x 200mL solvent aliquots) of MeOH/CH_2Cl_2 (1:1) at ambient temperature aided by ultrasonication and gravity filtration, yielding 2.37g of extract after evaporation of the solvent (6.85% w/w of dry biomass). The extract was redried onto 5g polygoprep 60-50 µm C-18 (Macherey-Nagel) bulk powder in preparation for RP-C-18 vacuum liquid chromatography (VLC). For VLC, a 250mL capacity sintered disc Büchner funnel was filled halfway with polygoprep 60-50 µm C-18 bulk powder and the column washed with 500mL MeOH/CH_2Cl_2 (1:1) before conditioning with 500mL H_2O. The extract was loaded onto a filter paper on the column and separated yielding fractions (F1-F5) under vacuum without allowing the column to run dry. Each fraction was eluted with 500mL of the corresponding solvent mixture according to the decreasing polarity gradient **1).** H_2O, **2).** H_2O/MeOH (1:1), **3.)** H_2O/MeOH (1:2), **4.)** MeOH, **5).** MeOH/ CH_2Cl_2 (1:1). Each fraction was dried and weighed.

5.3.2 Purification 394-F4 & 1826-F2

Purification of the fraction 394-F4 (53.9mg) was performed on a Waters 2695 Semi-prep HPLC using a reversed-phase T3 Column (Waters semi-preparative T3 column 10mm x 250mm) with 5 $mL.min^{-1}$ flow rate and injections ranging from 10µL to 60 µL with a gradient mobile phase of H_2O (A) & MeOH (B) both acidified with 0.1% v/v TFA. The gradient method was developed with a 36 min run time and gradient specifications: 4 min isocratic at 70% B, then a linear gradient for 19 min to 90% B, isocratic at 90% B for 8 min, returning linearly to 70 % B over 2 min at which it was held for 5 min until completion. Compound **17** (RT 13.0 min; 1.74 mg) was isolated in sufficient purity for NMR identification

Purification of the fraction 1826-F2 (982.81mg) was performed on a JASCO (Tokyo, Japan) Semi-prep HPLC using a reversed-phase-C18 Column (Waters semi-preparative XSelect T3 10mm x 250mm) with 5 $mL.min^{-1}$ flow rate and injections ranging from 10µL to 80 µL with a gradient mobile phase of H_2O (A) & MeOH (B) both acidified with 0.1% v/v TFA. The gradient method was developed with a 32 min run time and gradient specifications: 5 min isocratic at 8% B, then a linear gradient for 18 min to 65% B, isocratic at 65% B for 5 min, returning linearly to 8 % B over 1 min at which it was held for 2 min until completion. Compounds **18** (RT 13.0 min; 3.11 mg) and **19** (RT 15.5 min; 4.41 mg) were obtained in sufficient purity for NMR identification. Compound **20** was isolated but not in sufficient purity for definitive characterisation and was subject to repurification.

Repurification of the peak 1826-F2-P2 (52mg) was performed on a Waters 2695 HPLC using an semi-preparative reversed-phase amide column (Waters analytical BEH column 5µm 10mm x 250mm) with 5 $mL.min^{-1}$ flow rate and injections ranging from 10µL to 60 µL with a gradient mobile phase of H_2O (A) & ACN (B) both acidified with 0.1% v/v TFA. The gradient method was developed with a 19 min run time and gradient specifications: 2 min isocratic at 90% B, then a linear gradient for 13 min to 80% B, followed by an instant return to 90% B at which it was held for 4 min until completion. Aas expected due to the strong absorption of the histidine aromaticity at 254nm, compound **20** (RT 8.5 min; 2.10 mg) was obtained in sufficient purity for NMR identification.

6. Conclusions

The continued search for potentially bioactive specialized metabolites in the North Atlantic deep-sea has yielded the novel histidine derivative 6-methyl-hercynine from the sponge *Characella pachastrelloides*. This unprecedented metabolite, isolated from the aqueous-methanolic fraction, is likely to exist in biologically relevant chemical space due to its small size and amino-acid origin. Further studies will require bioassay investigation to assess the bioactivity of this compound and evaluate its pharmaceutical potential.

The *Pseudoceratina* specimen prioritized in screening was extracted, purified and analysed yielding the known compound Aplysamine 7. The molecular network was reviewed and the node representing Aplysamine 7 was identified in a cluster containing two nodes. The revealed analogue was found to differ from Aplysamine 7 by 14.016g.mol^{-1} representing a CH_2 unit. Unfortunately this analogue was not obtained sufficient purity for characterisation. Aplysamine 7 was the sole metabolite identified in this specimen as the other metabolites were present in too low quantity for characterisation from the available biomass, demonstrating the 'supply problem' often encountered in MNP investigation.

This research focused on applying various analytical techniques to investigate the pharmaceutical potential of deep-sea sponges, and using the data obtained to guide isolation of potentially bioactive NPs. Mature analytical methods such as UPLC and HRESIMS were combined with the emerging informatics tool molecular networking to help dereplicate known molecules and evaluate the chemical diversity in six sponge samples. Molecular networking of NPs through GNPS is in infant stages as much of the benefit associated with its use is reliant on community-sourced data, which may not be comprehensive or accurate. Despite these limitations, the molecular networks generated as part of the screening process proved very helpful in chemical diversity visualisation and sample prioritization.

Molecular networking is just one emerging technology which will undoubtedly have a major role to play in the future of biodiscovery. As demonstrated throughout this research, technological advances are the driving force behind natural products research and by combining mature technologies with modern developments, researchers will accelerate the process of finding novel bioactive metabolites wherever life is found.

7. References

1. J. Levinton1. J. Levinton. Stony Brook University Department of Ecology & Evolution. (2005). Available at: http://life.bio.sunysb.edu/marinebio/westmeadow/sponge.html. (Accessed: 17th March 2019). Stony Brook University Department of Ecology & Evolution. (2005). Available at: http://life.bio.sunysb.edu/marinebio/westmeadow/sponge.html. (Accessed: 17th March 2019)
2. Cárdenas, P. *et al.* Taxonomy, biogeography and DNA barcodes of Geodia species (Porifera, Demospongiae, Tetractinellida) in the Atlantic boreo-arctic region. *Zool. J. Linn. Soc.* **169**, 251–311 (2013).
3. Khushi, S., Nahar, L., Salim, A. A. & Capon, J. R. Cacolides: Sesterterpene Butenolides from a Southern Australian Marine Sponge, Cacospongia sp. *Marine Drugs* **16**, (2018).
4. Engelhardt, H. One century of liquid chromatography: From Tswett's columns to modern high speed and high performance separations. *J. Chromatogr. B* **800**, 3–6 (2004).
5. Moghe, G. D. & Last, R. L. Something Old, Something New: Conserved Enzymes and the Evolution of Novelty in Plant Specialized Metabolism. *Plant Physiol.* **169**, 1512–1523 (2015).
6. Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3–19 (2003).
7. Dias, D. A., Urban, S. & Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**, (2012).
8. Theobald, S. *et al.* Uncovering secondary metabolite evolution and biosynthesis using gene cluster networks and genetic dereplication. *Sci. Rep.* **8**, 17957 (2018).
9. Rohde, S., Nietzer, S. & Schupp, P. J. Prevalence and Mechanisms of Dynamic Chemical Defenses in Tropical Sponges. *PLoS One* **10**, e0132236 (2015).
10. Khan, R. A. Natural products chemistry: The emerging trends and prospective goals. *Saudi Pharm. J.* **26**, 739–753 (2018).
11. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl. Acad. Sci.* **115**, 6506 LP-6511 (2018).
12. John Faulkner, D. Highlights of marine natural products chemistry (1972–1999). *Nat. Prod. Rep.* **17**, 1–6 (2000).
13. Shang, J. *et al.* Cheminformatic Insight into the Differences between Terrestrial and Marine Originated Natural Products. *J. Chem. Inf. Model.* **58**, 1182–1193 (2018).
14. Usami, Y. Recent Synthetic Studies Leading to Structural Revisions of Marine Natural Products. *Marine Drugs* **7**, (2009).
15. Depth-dependent photoacclimatization to solar ultraviolet radiation in the Caribbean coral Montastraea faveolata . *Mar. Ecol. Prog. Ser.* **192**, 137–151 (2000).
16. Van Soest, R. W. M. *et al.* Global Diversity of Sponges (Porifera). *PLoS One* **7**, e35105 (2012).
17. Werner E.G. Müller(*), I. M. M. Porifera: an enigmatic taxon disclosed by molecular biology/cell biology. in *Sponges (Porifera)* 89–106 (2007).
18. Helber, S. B., Hoeijmakers, D. J. J., Muhando, C. A., Rohde, S. & Schupp, P. J. Sponge chemical defenses are a possible mechanism for increasing sponge abundance on reefs in Zanzibar. *PLoS One* **13**, e0197617–e0197617 (2018).
19. Omae, I. Organotin antifouling paints and their alternatives. *Appl. Organomet. Chem.* **17**, 81–105 (2003).
20. Assmann, M., Lichte, E., Pawlik, J. R. & Köck, M. Chemical defenses of the Caribbean sponges Agelas wiedenmayeri and Agelas conifera. *Mar. Ecol. Prog. Ser.* **207**, 255–262 (2000).
21. Okino, T., Yoshimura, E., Hirota, H. & Fusetani, N. New Antifouling Kalihipyrans from the Marine Sponge Acanthella cavernosa. *J. Nat. Prod.* **59**, 1081–1083 (1996).
22. Jackson, J. & Buss, L. W. *Jackson JBC, Buss L.. Allelopathy and spatial competition among coral-reef invertebrates. Proc Natl Acad Sci USA 72: 5160-5163. Proceedings of the National Academy of Sciences of the United States of America* **72**, (1976).
23. Thacker, R. W., Becerro, M. A., Lumbang, W. A. & Paul, V. J. Allelopathic Interactions between Sponges on a Tropical Reef. *Ecology* **79**, 1740–1750 (1998).
24. Taylor, M. W., Radax, R., Steger, D. & Wagner, M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.* **71**, 295–347 (2007).
25. Wilkinson, C. R. Immunological Evidence for the Precambrian Origin of Bacterial Symbioses in Marine Sponges. *Proc. R. Soc. London. Ser. B, Biol. Sci.* **220**, 509–518 (1984).
26. Hamann, M. T. The manzamines as an example of the unique structural classes available for the discovery and optimization of infectious disease controls based on marine natural products. *Curr. Pharm. Des.* **13**, 653–660 (2007).
27. Schippers, K. J. *et al.* Cultivation of Sponges, Sponge Cells and Symbionts: Achievements and Future Prospects. *Adv. Mar. Biol.* **62**, 273–337 (2012).
28. Miziorko, H. M. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* **505**, 131–143 (2011).
29. Mehbub, M. F., Lei, J., Franco, C. & Zhang, W. Marine sponge derived natural products between 2001 and 2010:

trends and opportunities for discovery of bioactives. *Mar. Drugs* **12**, 4539–4577 (2014).
30. Genta-Jouve, G. *et al.* Corrigendum: New Insight into Marine Alkaloid Metabolic Pathways: Revisiting Oroidin Biosynthesis. *ChemBioChem* **12**, 2238 (2011).
31. Desborough, M. J. R. & Keeling, D. M. The aspirin story – from willow to wonder drug. *Br. J. Haematol.* **177**, 674–683 (2017).
32. Siddiqui, M. Z. Boswellia serrata, a potential antiinflammatory agent: an overview. *Indian J. Pharm. Sci.* **73**, 255–261 (2011).
33. Al-Harrasi, A. *et al.* Chemical, molecular and structural studies of Boswellia species: β-Boswellic Aldehyde and 3-epi-11β-Dihydroxy BA as precursors in biosynthesis of boswellic acids. *PLoS One* **13**, e0198666–e0198666 (2018).
34. Chin, Y.-W., Balunas, M. J., Chai, H. B. & Kinghorn, A. D. Drug discovery from natural sources. *AAPS J.* **8**, E239–E253 (2006).
35. Bergmann, W. & Feeney, R. J. THE ISOLATION OF A NEW THYMINE PENTOSIDE FROM SPONGES1. *J. Am. Chem. Soc.* **72**, 2809–2810 (1950).
36. Stonik, V. A. Marine natural products: a way to new drugs. *Acta Naturae* **1**, 15–25 (2009).
37. Gomes, N. G. M., Dasari, R., Chandra, S., Kiss, R. & Kornienko, A. Marine Invertebrate Metabolites with Anticancer Activities: Solutions to the 'Supply Problem'. *Mar. Drugs* **14**, 98 (2016).
38. Pettit, G. R. *et al.* Antineoplastic agents. 257. Isolation and structure of spongistatin 1. *J. Org. Chem.* **58**, 1302–1304 (1993).
39. Hu, Y. *et al.* Statistical research on the bioactivity of new marine natural products discovered during the 28 years from 1985 to 2012. *Mar. Drugs* **13**, 202–221 (2015).
40. Mayer, A. M. S. *et al.* The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol. Sci.* **31**, 255–265 (2010).
41. Marine Pharamacology: Clinicial Pipeline | Midwestern University. Available at: http://marinepharmacology.midwestern.edu/clinical_pipeline.html. (Accessed: 17th March 2019)
42. van Soest, R. *et al. Sponge diversity and community composition in Irish bathyal coral reefs*. *Contributions to Zoology* **76**, (2007).
43. Sölter, S., Dieckmann, R., Blumenberg, M. & Francke, W. Barettin, revisited? *Tetrahedron Lett.* **43**, 3385–3386 (2002).
44. Carstens, B. B. *et al.* Isolation, Characterization, and Synthesis of the Barrettides: Disulfide-Containing Peptides from the Marine Sponge Geodia barretti. *J. Nat. Prod.* **78**, 1886–1893 (2015).
45. Hougaard, L., Christophersen, C., Nielsen, P. H., Klitgaard, A. & Tendal, O. The chemical composition of species of Geodia, Isops and Stryphnus (Choristida: Demospongia: Porifera)—A comparative study with some taxonomical implications. *Biochem. Syst. Ecol.* **19**, 223–235 (1991).
46. Suo, R. *et al.* Poecillastrin H, a Chondropsin-Type Macrolide with a Conjugated Pentaene Moiety, from a Characella sp. Marine Sponge. *J. Nat. Prod.* **81**, 1295–1299 (2018).
47. Afoullouss, S. *et al.* Treasures from the Deep: Characellides as Anti-Inflammatory Lipoglycotripeptides from the Sponge Characella pachastrelloides. *Org. Lett.* **21**, 246–251 (2019).
48. Yang, J. Y. *et al.* Molecular Networking as a Dereplication Strategy. *J. Nat. Prod.* **76**, 1686–1699 (2013).
49. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
50. Pauli, G. F., Chen, S.-N., Friesen, J. B., McAlpine, J. B. & Jaki, B. U. Analysis and Purification of Bioactive Natural Products: The AnaPurNa Study. *J. Nat. Prod.* **75**, 1243–1255 (2012).
51. Muigg, P., Rosén, J., Bohlin, L. & Backlund, A. In silico comparison of marine, terrestrial and synthetic compounds using ChemGPS-NP for navigating chemical space. *Phytochem. Rev.* **12**, 449–457 (2013).
52. Salim, A. A., Khalil, Z. G. & Capon, R. J. Structural and stereochemical investigations into bromotyrosine-derived metabolites from southern Australian marine sponges, Pseudoceratina spp. *Tetrahedron* **68**, 9802–9807 (2012).
53. ·, E. P. · P. B. & Badertscher, M. Structure Determination of Organic Compounds. doi:10.1007/978-3-540-93810-1